



Antinomies

ALGORITHMS, ARTIFICIAL INTELLIGENCE AND BEYOND

THEORISING SOCIETY AND CULTURE
OF THE 21ST CENTURY

Edited by Dariusz Brzeziński, Kamil Filipek,
Kuba Piwowar and Malgorzata Winiarska-Brodowska



“A journey across the blurry landscapes of a society reshaped by algorithms and AI. This book brings clarity and depth through a wide set of theoretical lenses that cover the technological, political, economic and cultural aspects of our current algorithmic condition.”

Massimo Airoidi, *Assistant Professor of Sociology,
University of Milan, Italy*

“Algorithms and artificial intelligence are calling into question many established conceptions and theories of the social world. By integrating a range of theories, this comprehensive new book provides the conceptual tools needed to gain insight and perspective on these shifts.”

David Beer, *Professor of Sociology, University of York, UK*

“This is a wonderfully timely book on algorithmic culture and artificial intelligence. I am pleased to see this theme historicized from a modernization perspective as a contrast to the presentism of so much of contemporary work in this area. This is essential reading for anyone who wishes to understand the present algorithmic condition in its wider historical context.”

Göran Bolin, *Professor of Media and Communication
Studies, Södertörn University, Sweden*

“This is an important new collection on the ‘algorithmic condition.’ The pervasiveness of algorithms and AI in everyday life demands not only empirical work but new (and old) theoretical frameworks. This volume offers those frameworks from a rich variety of disciplinary and interdisciplinary perspectives with excellent contributions from both established and early career scholars and from the Global South as well as the North. The focus on theorising means that the significance of the collection will endure long beyond the latest version of an AI chatbot or instalment of the algorithmic turn.”

John Downey, *Professor of Comparative Media Analysis,
Loughborough University, UK, and former President of European
Communication Research & Education Association (ECREA)*

“A rich and diverse collection of essays using social theory to understand the profound impacts of artificial intelligence and ‘algorithms’ on society and culture. Researchers and students hungry for fresh ways of examining the vital issues raised by these technologies should devour this volume.”

David Hesmondhalgh, *Professor of Media,
Music and Culture, University of Leeds, UK*

“This theoretically and empirically rich edited volume by Brzeziński and others offers novel insights in the adoption of AI and algorithms in the private sector. In exploring various dimensions of algorithmic conditions, practical cases and ethical implications, the book provides a comprehensive, multidimensional, and multi-perspective study on the algorithmic revolution. The book will be of great interest to scholars in sociology, media and communication, politics, computer sciences, and governance.”

Carolijn van Noort, *Associate Professor of Blue Governance,
Aalborg University, Denmark*

“Is AI a new ‘social actor?’ The new ‘engine’ of social change? Or are these 19th-century concepts not adequate to the challenges posed by the revolution that seemingly everyone – in industry, state, science and the arts – wants AI to be? Offering a much-needed antidote to the uncritical use of outdated concepts in today’s ‘AI debate,’ this book opens up a broad range of promising paths towards a new theory of technology and society in the face of AI.”

Noortje Marres, *Professor in Science, Technology and Society, Centre for
Interdisciplinary Methodologies, University of Warwick, UK*

“The book is a timely and important contribution to understanding how AI and data driven technologies interact with society from the perspective of social sciences and humanities. The latter branch of science is all too often left out in the current AI research and funding. The different chapters significantly fill the gap by taking a conceptual, interdisciplinary, international, and critical perspective on key issues co-shaping the algorithmic condition.”

Jo Pierson, *Professor of Responsible Digitalisation, Hasselt University,
Belgium, and Professor of Media and Communication Studies,
Vrije Universiteit Brussel, Belgium*

“This book explores the profound impact of AI and algorithms on society and culture. It examines the dual nature of digital technologies as tools of innovation and instruments of power, addressing their applications, ethical challenges, and influence on social inequalities. A must-read for understanding the intersection of technology and society in the digital age.”

Aleksandra Przeglasińska, *Vice Rector of ESR and International Affairs,
Kozminski University, Poland, and Senior Research Associate,
Center for Labour and Just Economy, Harvard University, USA*

Algorithms, Artificial Intelligence and Beyond

This volume brings together eminent scholars from various parts of the world, representing different fields of knowledge in order to explore the social, cultural, political and economic effects of the development of new technologies.

On the one hand, the book contextualises the discussion of algorithms and artificial intelligence (AI) within the broader framework of the digital revolution, on the other it also examines individual experiences and practices. Moreover, in light of the speed at which algorithms and AI are being incorporated into various aspects of life, contributors also question the ethical implications of their development. The widespread development of AI and algorithmic solutions is one of the most important contemporary phenomena. It has an overwhelming impact on the social and cultural life of the 21st century. In this context, one can point to both exciting examples of the application of algorithms and AI in business and popular culture, as well as the challenges of widening social inequality or the expanding scope of surveillance.

The scope of the impact of algorithms and AI makes the formation of new theoretical frameworks vital. This is the aim of this book, which will be of interest to academics within the humanities and social sciences with an interest in technology and the impact of algorithms and AI on society and culture.

Dariusz Brzeziński is Associate Professor at the Institute of Philosophy and Sociology at the Polish Academy of Sciences, Poland and a Visiting Research Fellow at the School of Sociology and Social Policy at the University of Leeds, UK.

Kamil Filipek is Assistant Professor in the Department of Economic Sociology and Social Research Methods at Maria Curie-Skłodowska University in Lublin (UMCS), Poland.

Kuba Piwowar is Assistant Professor at the Department of Culture and Media at the SWPS University in Warsaw, Poland.

Malgorzata Winiarska-Brodowska is Assistant Professor in the Department of Journalism and International Communication at the Institute of Journalism, Media and Social Communication at Jagiellonian University in Krakow, Poland.

Antinomies

Innovations in the Humanities, Social Sciences and Creative Arts

Series Editor: Anthony Elliott

Distinguished Professor, University of South Australia

This Series addresses the importance of innovative contemporary, comparative and conceptual research on the cultural and institutional contradictions of our times and our lives in these times. *Antinomies* publishes theoretically innovative work that critically examines the ways in which social, cultural, political and aesthetic change is rendered visible in the global age, and that is attentive to novel contradictions arising from global transformations. Books in this series are from authors both well-established and early careers researchers. Authors will be recruited from many, diverse countries – but a particular feature of the series will be its strong focus on research from Asia and Australasia. The series addresses the diverse signatures of contemporary global contradictions and as such seeks to promote novel transdisciplinary understandings in the humanities, social sciences and creative arts.

The series editor is especially interested in publishing books in the following areas that fit with the broad remit of the series:

- New architectures of subjectivity
- Cultural sociology
- Reinvention of cities and urban transformations
- Digital life and the post-human
- Emerging forms of global creative practice
- Culture and the aesthetic

Computational Power

The impact of ICT on law, society, and knowledge

Massimo Durante

The Algorithmic Unconscious

How Psychoanalysis Helps in Understanding AI

Luca M. Possati

Beyond Capitalist Dystopia

Reclaiming Freedom and Democracy in the Age of Global Crises

Davor Džalto

Algorithms, Artificial Intelligence and Beyond

Theorising Society and Culture of the 21st Century

*Edited by Dariusz Brzeziński, Kamil Filipek, Kuba Piwowar
and Malgorzata Winiarska-Brodowska*

For a full list of titles in this series, please visit www.routledge.com/Antinomies/book-series/ANTIMN.

Algorithms, Artificial Intelligence and Beyond

Theorising Society and Culture
of the 21st Century

**Edited by Dariusz Brzeziński,
Kamil Filipek, Kuba Piwowar and
Malgorzata Winiarska-Brodowska**

First published 2025
by Routledge
4 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2025 selection and editorial matter, Dariusz Brzeziński, Kamil Filipek,
Kuba Piwowar and Malgorzata Winiarska-Brodowska; individual chapters,
the contributors

The right of Dariusz Brzeziński, Kamil Filipek, Kuba Piwowar and Malgorzata
Winiarska-Brodowska to be identified as the authors of the editorial material,
and of the authors for their individual chapters, has been asserted in accordance
with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or
utilised in any form or by any electronic, mechanical, or other means, now
known or hereafter invented, including photocopying and recording, or in
any information storage or retrieval system, without permission in writing
from the publishers.

Trademark notice: Product or corporate names may be trademarks or
registered trademarks, and are used only for identification and explanation
without intent to infringe.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-1-032-64691-6 (hbk)

ISBN: 978-1-032-64692-3 (pbk)

ISBN: 978-1-032-64693-0 (ebk)

DOI: [10.4324/9781032646930](https://doi.org/10.4324/9781032646930)

Typeset in Times New Roman
by KnowledgeWorks Global Ltd.

Contents

<i>List of Figures and Tables</i>	<i>ix</i>
<i>Notes on Contributors</i>	<i>x</i>
<i>Acknowledgements</i>	<i>xv</i>
 Theorising the Algorithmic Condition: Challenges, Strategies, and Perspectives	 1
DARIUSZ BRZEZIŃSKI, KAMIL FILIPEK, KUBA PIWOWAR AND MALGORZATA WINIARSKA-BRODOWSKA	
 PART I	
The Algorithmic Revolution: New Theoretical Frameworks	15
 1 A Theory of Algorithmic Sociality (TaS)	17
ROBERT SEYFERT	
 2 Theorising Artificial Intelligence as Practice	39
KAMIL FILIPEK	
 PART II	
Algorithms, Artificial Intelligence, and the Theories of Modernity	55
 3 Images of Algorithmic Modernity: Agency and Autonomy in the Age of Predictive Analytics	57
ANTHONY ELLIOTT	
 4 Zuckerberg’s Cave: Smartness and Discipline in Digital Modernity	72
KIERON O’HARA	

PART III

Theory, Culture, and Society in the Algorithmic Condition 89

5 Algorithmic Nostalgia: Longing for the Past in the Age of Artificial Intelligence 91

DARIUSZ BRZEZIŃSKI

6 Socio-cultural Artificial Intelligence (SCAI) and Journalism: A Transdisciplinary Perspective in Media and Communication Studies 105

MALGORZATA WINIARSKA-BRODOWSKA AND KATALIN FEHER

7 Narrating AI Stories: Selected Strategies 121

MIROSLAW FILICIAK AND KUBA PIWOWAR

PART IV

Theorising AI-Driven Power and Politics 137

8 If (world2vec) Then vec2politics: On Machine Learning and the Performativity of Recursive Power 139

NICOLAS CHARTIER-EDWARDS, ETIENNE GRENIER
AND JONATHAN ROBERGE

9 Generative AI, Big Tech and the Value of Political Economy Analysis for the Sociology of Artificial Intelligence 155

ZHENG LIU

10 Four Models of Algorithmic Bureaucracy in China 173

BAOGANG HE AND JINJIN WU

PART V

Critical Theory of Artificial Intelligence 191

11 Social Inequalities and Artificial Intelligence: How Digital Inequality Scholarship Enhances Our Understanding 193

CHRISTOPH LUTZ

12 AI Racial Bias: How Text-to-Image Artificial Intelligence Generators Construct Prestigious Professions 211

ANNA M. GÓRSKA AND DARIUSZ JEMIELNIAK

Index 227

Figures and tables

Figures

1.1	Elements of algorithmic sociality.	25
1.2	Relations of algorithmic sociality.	29
1.3	Processes of algorithmic sociality.	32
4.1	The institutional dimensions of digital modernity.	75
6.1	Co-occurrence network of terms related to “journalism” and “social media” within the same sentence (2009–2019).	109
6.2	A word cloud – disinformation in context of SCAI (2020–2023).	111
6.3	Increasing interest in journalism theories related to artificial intelligence – number of scientific articles on this topic, 2009–2023.	111
9.1	Private investment in AI by geographic area, 2013–22.	159
10.1	Algorithmic functions (the automation systems and the predictive decision), algorithmic application system (departmental algorithmic applications and cross-departmental algorithmic applications), and algorithmic bureaucracy.	176
10.2	Impact of the departmentalised algorithmic system on bureaucracy.	180
10.3	Impact of cross-departmental algorithms on bureaucracy.	181
12.1	Results of respondents’ evaluation of 97 AI-generated images with distinction to 4 professions.	219
12.2	Results of respondents’ evaluation of 97 AI-generated images with distinction to 4 professions, excluding “unknown” race and “inhuman” answers.	219
12.3	Results of respondent’s evaluation of 97 AI-generated images with distinction to 9 text-to-image generators.	220

Tables

9.1	Selected large language models released in China	162
10.1	Four forms of algorithmic bureaucracy	182
11.1	Digital inequalities within and between organisations framework	196
12.1	Text-to-image generators with a number of generated images	217

Notes on Contributors

Dariusz Brzeziński is Associate Professor at the Institute of Philosophy and Sociology of the Polish Academy of Sciences (Department of Theoretical Sociology) and a Visiting Research Fellow at the School of Sociology and Social Policy at the University of Leeds. He teaches sociology and methodology of social sciences at the Jagiellonian University in Cracow as well. His research focuses on social thought, sociology of culture, and contemporary social and cultural transformations. He is the author of *Zygmunt Bauman and the Theory of Culture* (McGill-Queen's University Press, 2022), a co-editor of *Bronisław Malinowski and His Legacy in Contemporary Social Sciences and Humanities* (Routledge, 2024) and *Revisiting Modernity and the Holocaust: Heritage, Dilemmas, Extensions* (Routledge, 2022), and also a co-editor of a three-volume series of Zygmunt Bauman's *Selected Writings* (Polity Press, 2021–2024). Dariusz Brzeziński is a co-founder and vice president of the Digital Sociology Group within the Polish Sociological Association.

Nicolas Chartier-Edwards is a PhD student at Institut national de la recherche scientifique (INRS) in Québec City, Canada, where he pursues a custom program in Politics, Science and Technology. His thesis focuses the transformation of the Canadian Federal state and the Québec provincial's state's statecraft process by artificial intelligence. His broader work in the context of the *Shaping 21st Century AI – Controversies and Closure in Media, Policy, and Research* project, published in journals such as *AI & Society*, *Journal of Digital Social Research* and *Session Sociologique*, investigates the political economy of AI innovation clusters and the modalities of the exercise of power by artificial intelligence.

Anthony Elliott is Executive Director of the Jean Monnet Centre of Excellence at the University of South Australia, where he is Research Professor of Sociology and Dean of External Engagement. He is also Super-Global Professor of Sociology (Visiting) at Keio University, Japan. He is Fellow of the Academy of the Social Sciences in the United Kingdom, Fellow of the Academy of the Social Sciences in Australia, Fellow of the Cambridge Commonwealth Society and a Senior Member of King's College, Cambridge. His recent books include *The Culture of AI* (2019), *Concepts of the Self* (2020, 4E), *Reinvention* (2021, 2E), *Contemporary Social Theory: An Introduction* (2022, 3E), *Making Sense*

of *AI: Our Algorithmic World* (2022) and *Algorithmic Intimacy: The Digital Revolution in Personal Life* (2023). He is editor of *The Routledge Social Science Handbook of AI* (2022).

Katalin Feher is Associate Professor at Ludovika University of Public Service in Hungary, a Visiting Professor for the European Union's Transatlantic Horizon Europe NGI programme at the Applied AI Institute of Concordia University in Canada and Vice Chair of the European Communication Research and Education Association's Mediatization Section. Additionally, she serves as an Expert for the European Innovation Council and SMEs Accelerator Program and is a Member of the editorial board for KOME. She is a Fulbright and Bolyai HAS Alumna. Her research interests focus on the impact of sociotechnical systems, AI media, generative AI and responsible AI.

Mirosław Filiciak is Associate Professor, Director of the Institute of Humanities and Head of the Center for Cultural Research of Technologies at SWPS University in Warsaw, Poland. His research interests include the Internet, computer games, transformation of television, contemporary culture and informal distribution of information. He collaborates with numerous public cultural institutions, businesses and NGOs. He is also the co-creator of *Kultura 2.0* [Culture 2.0], a project devoted to cultural transformations in the digital era as well as *Medialab*, the first Polish initiative combining social activism, art and technology.

Kamil Filipek is Assistant Professor in the Department of Economic Sociology and Social Research Methods, Maria Curie-Skłodowska University (UMCS) in Lublin, Poland. He is Director of the Centre for Artificial Intelligence and Computational Modelling at UMCS. Additionally, he works in the business field as a data scientist and data analyst. His research interests include: social media, social practices, quantitative research and algorithmic bias. His works have been published in prestigious journals e.g. *Journal of Consumer Affairs*, *Communications: The European Journal of Communication Research*, *PLOS One*.

Anna M. Górska is Assistant Professor at Kozminski University, and Director of Women and Diversity in Organizations Research Center. Anna Górska specialises in gender and diversity in organisations. Her works have been published in international academic journals such as *Gender, Work and Organization*, *Gender in Management, Information & Management, Accounting, Auditing & Accountability Journal*, *Journal of Global Information Technology Management*. She has participated as a principal investigator and researcher in multiple research grants funded by Polish National Center of Science, EU funds and Norwegian Funds. She has also been awarded scholarships and awards, including Fulbright Junior Research Award. She has gained her international experience as a visiting scholar at Columbia Business School and ESCP Business School.

Étienne Grenier is artist and researcher working in the field of digital culture. Currently a Ph.D. candidate at the Institut National de la Recherche Scientifique (INRS) in Montreal, Canada. He studies the impact of datafication on cultural

production with different projects, among which Shaping AI. His artistic installations and performances have been presented in institutions and major festivals in Europe and the Americas.

Baogang He (Ph.D., ANU 1994) is Alfred Deakin Professor at Deakin University and a Fellow of the Academy of Social Sciences in Australia. He is widely known for his work in Chinese politics, in particular deliberative politics in China as well as in regionalism, international relations, federalism and multiculturalism in Asia. His publications are found in top journals including *Science*, *British Journal of Political Science*, *Journal of Peace Research*, *Political Theory*, *Political Studies*, and *Perspectives on Politics*. Two of his papers (“Authoritarian Deliberation: The Deliberative Turn in Chinese Political Development,” co-authored with Mark Warren and “The Domestic Politics of the Belt and Road Initiative and its Implications”) were in the top 1% of highly cited research papers in the Web of Sciences as of 11 Feb 2021. His published books include *The Democratisation of China*, *The Democratic Implications of Civil Society in China*, *Rural Democracy in China* and *Governing Taiwan and Tibet*. His recent co-edited books include *Deliberative Democracy in Asia*, (with Michael Breen and James Fishkin), Routledge in 2021, *China and North Korea’s Human Rights* (with David Hundt, Chengxin Pan), Routledge, 2021 and *Comparative Federalism in Asia* (with Michael Breen and Laura Reumann), Routledge, June 2023.

Dariusz Jemielniak is Full Professor and head of Management in Networked and Digital Environments (MINDS) department, Kozminski University, and faculty associate at Berkman-Klein Center for Internet and Society, Harvard University. He is a corresponding member of the Polish Academy of Sciences (elected as the youngest in history in social sciences and humanities). His recent books include *Collaborative Society* (2020, MIT Press, with A. Przegalinska), *Thick Big Data* (2020, Oxford University Press), *Common Knowledge? An Ethnography of Wikipedia* (2014, Stanford University Press). His current research projects include disinformation and bot detection. He currently serves on the Wikimedia Foundation Board of Trustees, as well as a vice-chair for the Polish Academy of Sciences.

Zheng Liu is Lecturer in Innovation Management at the University of Bristol. She has a PhD in Sociology from the University of Cambridge, where she is affiliated as a Senior Research Associate in Jesus College, a Research Affiliate in the Department of Sociology and an Associate Member with Cambridge Digital Humanities. Her most recent research has been about developing a systematic sociology of AI, having published a widely cited paper on this topic in *Sociology Compass* in 2021. She is currently PI on a British Academy and Sino-British Fellowship Trust jointly funded project on analysing the social structures of China’s AI industry. She has served as an Expert Reviewer for prestigious international funding bodies such as the European Research Council, UK Research and Innovation, the German Research Foundation, and Social Sciences and Humanities Research Council of Canada. Her other research interests include industrial emergence and transformation in the digital age, innovation

theory, digitalisation, social history of digital technologies, political economy, and social theory. Previously, she held research fellowships at the University of Macau and studied at Tsinghua University.

Christoph Lutz is Full Professor in the Department of Communication and Culture, BI Norwegian Business School (Oslo, Norway). He is also the co-director of the Nordic of Internet and Society, a leading Internet and Society research centre in Europe. His research interests include digital inequality, social media, privacy, the sharing economy and social robots. Christoph has published widely on these topics in top-tier journals such as *New Media & Society*, *Information, Communication & Society*, *Journal of Business Research*, *Journal of Management Information Systems* and *Science and Engineering Ethics*.

Kieron O'Hara is Emeritus fellow in electronics and computing at the University of Southampton. His research has focused on technology, particularly the World Wide Web and political philosophy. His recent book, *Four Internets* (with Dame Wendy Hall, 2021), describes the geopolitics of the internet and data governance under digital modernity.

Kuba Piowar is a sociologist and cultural scientist, and holds a Ph.D. in the humanities, specialises in the socio-cultural aspects of technology, data activism, new media and market research. Between 2008 and 2024, he worked at Google, where he was an analyst and advisor to key business partners, working with the company's clients in Poland in the processes of digital transformation. He is Assistant Professor at the Department of Culture and Media at the SWPS University in Warsaw, he is also a member of the Polish Sociological Association. He holds an MA diploma in sociology at the SWPS University, he also graduated from postgraduate Gender Studies at IBL PAN, art history at Collegium Civitas and Marketing Academy at The Wharton School at the University of Pennsylvania.

Jonathan Roberge Jonathan Roberge is Full Professor at the Institut National de la Recherche Scientifique (INRS) in Montreal, Canada. He funded the Nenic Lab as part of the Canada Research Chair in Digital Culture and he has held since 2012. He also co-directed the *Shaping 21st Century AI – Controversies and Closure in Media, Policy, and Research* project and is currently co-chair of the Chaire de recherche du Québec sur l'intelligence artificielle et le numérique francophones (chaire IANF). His most recent edited volumes include *Algorithmic Culture* (Routledge, 2016) and *The Cultural Life of Machine Learning* (Palgrave, 2020).

Robert Seyfert is Professor of Sociology at Kiel University. He studied philosophy, sociology and political science in Dresden and New York. In 2010, he completed his doctorate in sociology at the University of Konstanz (*Das Leben der Institutionen: Zu einer Allgemeinen Theorie der Institutionalisierung*, Velbrück 2011). The habilitation (postdoctoral lecture qualification) in sociology took place in 2019 at the Faculty of Social Sciences at the University of Duisburg-Essen (*Beziehungsweisen: Elemente einer relationalen Soziologie*,

Velbrück 2019). His main areas of research include the digital transformation and algorithmic sociality. Relevant publications can be found in the collected volume (ed. with Jonathan Roberge) *Algorithmic Cultures. Essays on Meaning, Performance and New Technologies*. London/New York: Routledge, 2016; and “Algorithms as regulatory objects,” *Information, Communication & Society* 25 (2022), 1542–1558.

Malgorzata Winiarska-Brodowska is Vice-President of the European Communication Research and Education Association (ECREA) and member of ECREA Executive Board (Governing Body). She is Assistant Professor at the Department of Journalism and International Communication, Institute of Journalism, Media and Social Communication, Jagiellonian University in Krakow, Poland. She holds a Ph.D. in social sciences (political science). She graduated in journalism, cultural studies and European studies. Prior to working in academia she gained professional experience in governmental administration and media. She has been involved in management of different entities: e.g. as a co-founder of the Digital Sociology Section within the Polish Sociological Association, chair of the Central and East European Network/CEECON, and vice-chair of the ECREA Communication and the European Public Sphere Temporary Working Group. Her research and teaching interests lie at the intersection of media, politics and international relations. She authored several dozen scientific publications in the field of political communication, media and communication in international relations, organisational and strategic communication, as well as AI use in media and communication field.

Jinjin Wu is Associate Professor in the School of Government at Shenzhen University and has been a visiting fellow at Nanyang Technological University, Singapore. He has published widely on algorithms in public service, local budget process and fiscal capacity, and public deliberation. He has published more than 30 papers in Chinese and international peer-reviewed journals including *Contemporary Politics*, *American Review of Public Administration* and *Chinese Public Administration* (Chinese).

Acknowledgements

The idea for this book came from the symposium *Algorithms, Artificial Intelligence & Beyond: Theorising Society and Culture of the 21st Century*, organised by the later editors as part of the 18th Polish Sociological Congress, which took place in Warsaw on 14–17 September, 2022. The idea was endorsed by Professor Anthony Elliott, one of the participants of this symposium, who decided to consider the volume for publication in *Antinomies*, a Routledge book series he edits. We are very grateful to him for his help, support and commitment to this project. We also sincerely thank Helen Pritt from Routledge for her guidance and cooperation and two anonymous reviewers for their insightful comments. We would like to express our gratitude to the authors who accepted the invitation to contribute to this book. It was a real honour and a great pleasure for us to work with them. We hope that this collaboration will result in further research and publishing projects in the future.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Theorising the Algorithmic Condition

Challenges, Strategies, and Perspectives

*Dariusz Brzeziński, Kamil Filipek, Kuba Piwowar
and Malgorzata Winiarska-Brodowska*

Introduction

“It is reasonable to suppose that the proliferation of information-processing machines is having, and will continue to have, as much of an effect on the circulation of learning as did advancements in human circulation (transportation systems) and later, in the circulation of sounds and visual images (the media),” Jean-François Lyotard (1983 [1979]: 4) concluded in *The Postmodern Condition: A Report on Knowledge*, which originally appeared in French in 1979. This by now classic study introduced Lyotard’s famous thesis that a momentous transformation was taking place in knowledge production and distribution in the second half of the twentieth century, and that this transformation was intimately related to changes in the condition of the times. In short, Lyotard argued that postmodernity was characterised by “incredulity towards metanarratives,” that is, the unifying stories that had once been the ultimate reference point for all knowledge. Lyotard’s analysis of “computerised societies” formed a very important part of his reflections on the genesis, course, and consequences of the decline of metanarratives. He stated that the knowledge produced in such societies was being increasingly commoditised, becoming a good for which industry, commerce, the military, and politics were competing. In Lyotard’s view, new technologies were decisively changing the balance of power in the contemporary condition, which also necessitated a significant transformation in the way this condition should be theorised.

Lyotard’s recognition of the relationship between the development of technological knowledge and social change has lost none of its relevance today, almost half a century after the publication of *The Postmodern Condition*. If anything, the current processes of algorithmisation and the development of artificial intelligence (AI)¹ have elevated this relationship to a higher level. The contemporary world, which tends to be referred to as the “algorithmic condition,” echoing Lyotard’s work among others (Coleman et al. 2018; Pedwell 2024), is characterised by the pervasiveness of technology in all spheres of social life. For example, economic processes, modes of governance, and security systems are increasingly based on quantification carried out by algorithms (e.g. Crawford 2021; Elliott 2022; Beer 2024). The causality of algorithms is also very much present in culture since, as Elliott (2019) has observed, “[t]his explosion in AI, machine learning and big data

goes to the core of the daily routine interactions in which people are embedded – from personalised social media to facial recognition software to gain access to offices” (xviii). What is more, technology is becoming an ever more conspicuous presence in the realm of individual development, profoundly transforming the ways in which individuals make both everyday and long-term decisions (e.g. [Airoidi 2022](#); [Elliott 2023](#); [Novotny 2021](#)).

In the context of all these transformations co-shaping the “algorithmic condition,” it is a matter of some urgency to develop new theoretical frameworks that will help us analyse it adequately. Significant efforts have already been made in this regard (e.g. [Elliott 2019](#); [Hristova et al. 2021](#); [Marres 2017](#); [Roberge and Castelle 2021](#)); however, in view of the dynamics of contemporary reality, the existing theorisations must constantly be revised, and new social and cultural phenomena must be thematised as they arise. This is precisely the purpose of this book. The authors of the papers it includes explore various dimensions of the algorithmic condition, seeking novel ways to theorise it. They come from different fields of knowledge (sociology, cultural studies, media and communication studies, philosophy, political science, management and administrative sciences, and computer sciences) and from various countries (Australia, Canada, China, Germany, Norway, Hungary, Poland, and the United Kingdom). All this makes the volume a comprehensive, multidimensional, and multi-perspective attempt to reflect on the genesis, course, and implications of the algorithmic revolution currently underway.

This chapter introduces the reader to the issues addressed by the contributors and presents the lines of argumentation they advance in their chapters. It is divided into three sections. The first section offers an overview of the development of the algorithmic condition up to the present day. Corresponding to the multidisciplinary nature of the volume, the next section outlines the various perspectives from which algorithms and AI are researched in the social sciences and humanities today. The third section sketches out the book’s structure and the objectives pursued by the authors in each of its five parts. The conclusion draws on Lyotard’s *The Postmodern Condition* to reflect on both the “heteromorphous nature” of and the utopian position² adopted by the theorisations of the algorithmic condition in the volume.

“Where Do We Come from? What Are We? Where Are We Going?”

Technologies have never been and are not independent of the context in which they operate ([Ede 2019](#)). The multiple ways in which they are used are shaped by the prevailing culture, the political system, and the economic conditions of their time ([MacKenzie and Wajcman 1999](#)). Similarly, algorithms and AI are produced by specific norms, principles, rules, and geopolitical contexts. Thus, a look back at the milestones of AI development helps better understand the challenges faced by social scientists and humanities scholars who investigate broadly understood technological development. It is also an excellent starting point to reflect on the future of the algorithmic condition.

The genesis and evolution of the algorithmic condition is a saga filled with relentless experimentation, unexpected developments, and intense competition, propelled

by commercial and political motives (e.g. [Wooldridge 2021](#)). Numerous accounts and narratives have chronicled this complex history, as examined by Mirosław Filiciak and Kuba Piwowar in this volume. The key turning points usually cited in this context are: Alan Turing’s research on “machine intelligence” in the early 1950s ([Turing 1950](#)); the Dartmouth Summer Research Project on AI in 1956 ([McCarthy et al. 1955](#)); the development of programming languages and machine learning over the following decades, in which the “summers” of rapid growth alternated with the “winters” of stalled progress ([Gonsalves 2019](#)); the further advancement of the natural language processing (NLP) technology ([Indurkha and Damareu 2010](#)); and, finally, the still continuing surge of interest in deep learning, which uses multi-layer neural networks to process and analyse data ([Prince 2023](#)).

From the point of view of the investigations undertaken in this volume, one very important moment in the latest history of AI was the introduction in 2017 of the AlphaGo Zero model, demonstrating that, although training data are crucial, they are not an ultimate prerequisite for intelligent systems. To explain, AlphaGo Zero is a more advanced version of the AlphaGo model, which defeated Korean Go³ grandmaster Lee Sedol in 2015 ([Silver et al. 2016](#)). Massimo Airolti (2022) captures the significance of this technology by elucidating that: “Far from mechanically executing tasks, current AI technologies can learn from (datafied) experience, a bit like human babies. And as with human babies, once thrown into the world, these machine learning systems are no less than social agents, who shape society and are shaped by it in turn” (14). The multiple and widely varying implications of the fact that algorithms are now “social agents” are among the pivotal topics addressed in this book.

The advent of Keras, an open-source Python library, represents yet another milestone in the latest history of AI. Conceived by François Chollet (2021), Keras heralded a “democratic” phase in AI development by providing an unparalleled opportunity for individuals with basic programming skills to leverage the potential of AI across various fields, such as business, medicine, and the military. Since its inception in 2015, the Keras library has fuelled a remarkable upswing in designing AI models. Many of these models still continue to underpin a wide array of solutions leveraging AI. However, the democratic phase in the development of AI is coming to an end. The arrival of large models, initially represented by Large Language Models (LLMs), was a harbinger of this shift ([Amaratunga 2023](#)). ChatGPT, the most well-known example of the business success of LLMs, is in fact the fastest-growing consumer application in history. Currently, Large Visual Models are joining LLMs, along with multimodal models that easily switch between various data sources.

A critical aspect in this process is that building large models is not something that everyone can afford. Creating such a model requires access to immense computational power, capacity to process astronomical amounts of data, and a host of outstanding AI professionals to devise, test, and refine the developed solution. Consequently, the centres capable of building AI models are few and far between, and fees for using these large models have been imposed. Besides, given that the data processed by these centres are now fundamental for politics, the economy, and social life, the centrality of AI hubs in contemporary power networks is ever

more pronounced. This has made the current phase essentially undemocratic and is, moreover, having a vital part in the processes taking place in contemporary democracies. This issue is investigated in this volume mostly in the chapters by Kieron O'Hara, and Jonathan Roberge, Nicolas Chartier-Edwards, and Étienne Grenier.

What will be the next step in the development of the algorithmic condition? Many researchers envisage the emergence of artificial general intelligence (AGI), which will surpass everything that has been achieved so far in this very dynamic field of science. Whether there will be one AGI or rather multiple versions of AGI is difficult to predict today. What is less challenging to foresee is that if AGI comes into being, it will change our lives even more profoundly, introducing new dimensions of risk, new lines of social division, and new conflicts that we are not yet able to imagine (Sigalos and Browne 2024).⁴ For the time being, we can only make conjectures about the consequences of the development of AGI, and the analyses in this volume can usefully support such speculations. The optimistic vision of AI-using communities of practice presented by Kamil Filipek and the algorithmic bureaucracy depicted by Baogang He and Jinjin Wu are worth exploring in this context. However, they need to be juxtaposed with the insights from chapters by Christoph Lutz, who discusses AI in conjunction with social inequality, and by Anna Górska and Dariusz Jemielniak, who deal with the problem of racial bias in AI.

It is evident that AI has not reached the pinnacle of AGI yet. It continues to be embodied in models meticulously coded by programmer teams, each designed to perform narrowly and explicitly defined assignments. As AI models grow in size and enhance their capability across a spectrum of tasks, including text analysis, image processing, and sound recognition, their performance hinges on several critical factors. These include their creators' expertise, the choice of training data, the methodologies employed to evaluate their effectiveness, and the criteria adopted for their continual assessment. In its ongoing evolutionary phase, AI has seamlessly woven itself into the fabric of our existence, bringing about permanent shifts in various domains of our lives. This phenomenon is brilliantly rendered in articles by Robert Seyfert and Anthony Elliott. The horizon of the algorithmic condition teems with further transformations, though the mechanism to steer this extensive metamorphosis, which affects every aspect of human endeavour, remains elusive. This situation calls for a heightened awareness and understanding of AI's trajectory in order to make sure that its future advancements align with ethical standards and constructively contribute to social progress.

Multiple Research Perspectives for Exploring Algorithms and Artificial Intelligence

In *The Postmodern Condition: A Report on Knowledge*, Jean-François Lyotard (1983 [1979]) noted that the classic dividing lines between the various disciplines were being called into question: "[D]isciplines disappear, overlappings occur at the borders between sciences, and from these new territories are born" (39). In the algorithmic condition, sciences not only overlap and promote new, often interdisciplinary, research, but also go beyond their disciplinary boundaries in these explorations

to integrate non-scientific entities, including non-academic stakeholders in knowledge production. AI research is increasingly being undertaken as part of cross-disciplinary studies. It is becoming necessary not only to combine the STEM fields (science, technology, engineering, and mathematics) with the social and humanistic sciences (sociology, law, philosophy, media and social communication sciences, economics, political sciences, and psychology) but also to seek partners for collaboration outside of academia to launch transdisciplinary investigations.

The trans-disciplinarity of today's inquiries is well reflected in this book, since its chapters have been written by scholars who come from very different disciplines but are uniformly committed to transdisciplinary research. This transdisciplinary mindset is explicitly communicated, for example, in the very title of the article authored by Malgorzata Winiarska-Brodowska and Katalin Feher. Furthermore, in parallel to their academic research, some of the authors of the chapters either work with technology companies or are artists using new media and technologies. The aggregation of all these perspectives in this volume is in line with the "algorithmic turn," currently unfolding in the social sciences and humanities⁵ (Edwards and Veale, 2017; Gurumurthy and Bharthur 2018; Kiberg and Spilker 2023). Dariusz Brzeziński discusses this issue in more detail in his chapter.

Three major thematic concerns can be identified among the multiple approaches that researchers of the algorithmic condition encounter in expert, scientific, and media discussions. They are the social and cultural embeddedness of algorithms and AI, the political and economic applications of the new technologies, and their legal and ethical aspects. Concerning the former, scholars investigate how algorithms and AI influence and are influenced by societal norms and cultural values. Thereby, their focus is often on understanding and designing AI systems that are culturally sensitive and socially beneficial (e.g. Elliott 2022; Hristova et al. 2021). The second stream examines how algorithms and AI are used within political and economic contexts, and how they impact public policy, governance, and power dynamics. This includes the use of digital technologies in surveillance, national security, and public administration. Attention is also paid to geopolitical aspects, such as how nations leverage AI for strategic advantages, and how global technological companies are shaping the current algorithmic condition (e.g. Tinnirello 2022; Zuboff 2019). The third thematic preoccupation ties in with the fact that the more AI technologies become more integrated into our everyday life, the more significant legal and ethical questions they raise. This research area addresses, for example, the liability of AI systems, intellectual property rights, and the ethical design of autonomous systems. It also tackles the formulation of regulations and guidelines to ensure that AI development and deployment are safe, transparent, and fair (e.g. Brevini 2022; Crawford 2021). All these points of interest are covered in the papers in this volume.

Notably, the respective chapters theorise the algorithmic condition from various geographical perspectives, from the US and Canada to the European Union and the Far East. This is of special importance as there are some significant differences between approaches to digital technologies embraced in various cultures (e.g. Kai-Fu 2018; Wang et al. 2023; Zeng 2022). Andrés Ortega Klein (2020) underlines that "technology ushers in a clash of cultures to which insufficient attention

has been paid” (3). In his analysis of the narrative and reality around the nexus of new technologies, defence of shared values, and regulation, Ortega Klein stresses that values are an ever more salient element in discussions on technology, in particular regarding AI in the context of surveillance systems and other aspects of exportable techno-authoritarianism. Human rights approaches for the digital era vary across the globe, and these divergencies are important for understanding research and theory-building around AI since values that are cultivated within a socio-political context are integrated into technology. This technology subsequently drives soft power and hard, military power alike. AI applied, for example, in surveillance systems may be used for both political (e.g. [Qiang 2019](#)) and economic purposes (e.g. [Zuboff 2019](#)). Embedding research on AI in different scientific perspectives and in different cultures may contribute to a better understanding of current processes, helping researchers and practitioners define social values of the future. This context is particularly highlighted in the chapter of Zheng Liu.

All these differences notwithstanding, there is ample room for collaboration. For example, Huw [Roberts et al. \(2021\)](#) identify the areas where the EU and especially the US need improvement in order to achieve ethical outcomes and deepen their cooperation. Likewise, comparing the AI policies of the US and China and constructing a framework for capturing the origin and trajectory of policy differences, [Emmie Hine and Luciano Floridi \(2024\)](#) argue that while there are impediments to cooperation, there is also room for dialogue and mutual growth. Similarly, on the global level (for instance, as expressed in the UN document *The Right to Privacy in the Digital Age: Report [2021]*), many countries work to ensure the ethical, human-centred, and responsible development and implementation of AI, act towards an enabling environment for AI policy and insist that this setting should foster a variety of approaches in the creation of AI technologies. These issues are also discussed in this book.

Five Theoretical Approaches to the Algorithmic Condition

The volume consists of five parts. Each of them looks into the relationship between theory and the algorithmic condition from a different angle. The chapters in [Parts I](#) and [II](#) are focused on finding new theoretical frameworks or revisiting the existing ones to adapt them to the social and cultural challenges that result from the development of algorithms and AI. [Parts III](#) through [V](#) are devoted to investigating the planes of analysis related to the social, cultural, political, and economic changes and linked to the ethical and legal challenges, as outlined in the foregoing.

[Part I](#), titled **The Algorithmic Revolution: New Theoretical Frameworks**, includes two chapters. In his opening piece, “A Theory of Algorithmic Sociality (TaS),” sociologist Robert Seyfert proposes the eponymous framework as a way to conceptualise the new (post-) social relations spawned by the development of algorithms and AI. According to Seyfert, the current phase of digitalisation (which he calls Digitalisation 2.0), “can be understood (...) as the comprehensive digital transformation of society in which conventional face-to-face *relations* are replaced by digitally and algorithmically mediated relations as the dominant modes of

relating in society” (p. 18, *italics original*). Seyfert does not consider traditional approaches to theorising human relations obsolete but claims that we need to expand the field of analysis to problematise the increasingly complex and multiple forms of these relations. His very approach is relational, meaning that he aims to theorise the reciprocal relationships of algorithms and humans.

In the following chapter, titled “Theorising Artificial Intelligence as Practice,” sociologist and data scientist Kamil Filipek offers an innovative perspective on the algorithmic condition, inspired by the theories of practice (see, e.g. Knorr Cetina, Schatzki, and von Seigny 2000). Filipek argues that this theoretical approach, which has been used and honed for some three decades now, provides an excellent platform for theorising new social and cultural phenomena related to the development of algorithms and AI. As he argues, “in this view, what really matters is not just what we formally call AI, but rather how people actually use and shape it, how they practise AI” (p. 39). Like Seyfert, Filipek focuses on theorising the relations that arise at the intersection of society and technology. He does this by applying the long-established concept of a “community of practice” (Wenger 1998) to his case study.

Part II of the volume is titled **Algorithms, Artificial Intelligence, and the Theories of Modernity** and contains two chapters that revise the theories of modernity from the vantage point of the algorithmic condition. In “Images of Algorithmic Modernity: Agency and Autonomy in the Age of Predictive Analytics,” sociologist Anthony Elliott explores the profound effects of digital technology on contemporary society, addressing the promise and perils of our increasing reliance on algorithmic systems. Elliott builds on his concept of “algorithmic modernity” (Elliott 2022), but he primarily looks at the work of Cornelius Castoriadis as an inspiration for critical reflection on the algorithmic condition. His text critically examines the impact of AI on various aspects of daily life, from social interactions to the economy, and highlights the dual nature of such technologies as both enablers of convenience and facilitators of social disparities. Elliott’s discussion extends onto the complexities of navigating our increasingly digital world, where algorithmic decisions determine human experiences and agency. His narrative urges a balanced view as he recognises the transformative power of AI while calling for a cautious approach to the integration of AI into societal structures. In doing that, Elliott emphasises the imperative of addressing the resulting inequalities and ensuring that technology serves the broader public good.

The next chapter, titled “Zuckerberg’s Cave: Smartness and Discipline in Digital Modernity,” is authored by philosopher, computer scientist, and political writer Kieron O’Hara. Drawing on Anthony Giddens’ (1990) concept of “late” or “reflexive” modernity, O’Hara (2020) develops his theory of “digital modernity,” emphasising that its institutions and axiology are increasingly shaped by technology. While O’Hara refers to very optimistic narratives of the algorithmic condition, he also points to its not-always-recognised dystopian potential. In this context, he revisits Plato’s metaphor of the cave and concludes that the inmates of the modern Zuckerberg Cave usually do not want to leave it at all. Rather, he argues, “dwellers in the metaverse are fully aware, preferring the discipline supplied by the virtual

and in the ideal accepting its priority over the physical. Digital modernity hits a sweet spot for the prisoners” (p. 81). O’Hara’s chapter represents a critical and engaged approach to the development of the algorithmic condition.

Part III of the book, titled **Theory, Culture, and Society in the Algorithmic Condition**, includes three chapters. Its opening text, “Algorithmic Nostalgia: Longing for the Past in the Age of Artificial Intelligence,” written by sociologist and culture scholar Dariusz Brzeziński, navigates the intersection of digital advancements and longing for the past. Brzeziński identifies four dimensions of algorithmic nostalgia: algorithmic generation of nostalgic content on social media, promotion of nostalgic longing on streaming platforms, digital manipulation of content for nostalgic effects, and digital reconstruction of the past. These developments, ranging from Facebook’s “On This Day” to Netflix’s revival of classic shows and the technologically resurrected performances of bands such as ABBA, illustrate the transformation of nostalgia in the digital age. Brzeziński contends that these phenomena not only cater to but also fuel nostalgic yearnings, blurring the line between reality and hyperreality in the process, and challenging the traditional notions of nostalgia. Driven by commercial interests and technological capacities, the proliferation of digitally induced nostalgia calls for a re-evaluation of nostalgia’s conceptual framework and brings into relief the need for empirical research to understand its evolving role in culture and society.

Penned by media scholars Malgorzata Winiarska-Brodowska and Katalin Feher, the next chapter is titled “Socio-Cultural Artificial Intelligence (SCAI) and Journalism: A Transdisciplinary Perspective in Media and Communication Studies” and theorises the intersection between the algorithmic condition and the mediasphere, with a particular emphasis on journalism. To this end, the authors develop a socio-cultural AI (SCAI) perspective, which, as they argue, differs from techno-AI in viewing “technology as a tool and not as an ultimate goal” (p. 107). They discuss the results of an empirical study of articles published in high-ranked academic journals between 2009 and 2023. Its findings reveal that the advantages and challenges of the algorithmic condition are ever more relevant to the media (as epitomised by “filter bubbles,” “fake news,” or “deep fake”). The authors also analyse transformations in theories of journalism related to the development of algorithms and AI, emphasising their transdisciplinary character.

In the following chapter, titled “Narrating AI Stories: Selected Strategies,” culture and media scholar Mirosław Filiciak and sociologist and culture researcher Kuba Piwowar scrutinise the narratives surrounding AI, particularly those disseminated by Big Tech companies, and their impact on public perceptions and policy. They outline the dichotomy of optimistic and pessimistic portrayals of AI and examine the sources and evolution of these narratives. Their discussion spells out how Big Tech’s dominance in the AI discourse often marginalises critical voices, emphasising the need for diverse perspectives in the conversation around AI’s societal impact. Filiciak and Piwowar explore the role of governments, critical collectives, and popular culture in constructing alternative narratives that challenge the status quo. By underscoring the importance of critical engagement with

AI narratives, they aim to foster a more inclusive and critical discourse on AI and encourage a holistic approach to its development and implementation.

Part IV of the book, titled **Theorising AI-Driven Power and Politics**, focuses on searching for new theoretical frameworks related to politics, economy, and management in the algorithmic condition. In their opening piece, “If (world2vec) Then vec2politics: On Machine Learning and the Performativity of Recursive Power,” sociologist Nicolas Chartier-Edwards, artist and researcher in digital cultures Étienne Grenier, and sociologist and cultural scholar Jonathan Roberge examine the profound influence of AI on contemporary societies. As AI transcends mere prediction to shape the societal fabric through smart monitoring, the manipulation of information flow, and automated financial systems, its entrenchment raises concerns about the complexity it adds to social and symbolic realms. The authors introduce the concept of “vectorial politics,”⁶ which is underpinned by AI’s recursive impact on society and driven by algorithmic techniques, such as backpropagation and vectorisation. By simplifying complex data into manageable forms, these mechanisms inadvertently simplify social complexities as well and, as a result, undermine traditional decision-making paradigms. The contributors further examine the societal and political theories emerging from AI’s influence; they rely on Yoshua Bengio’s work as a case study to unravel the entanglement of AI in social structuring and political discourse. Through the lens of backpropagation and vectorisation, they critique the circular logic that pervades AI’s approach to societal issues and call for a more inclusive and critically informed engagement with technology’s role in shaping future societal norms.

In the following chapter, titled “Generative AI, Big Tech, and the Value of Political Economy Analysis for the Sociology of Artificial Intelligence,” cultural and economic sociologist Zhung Liu makes a case for a political economy framework as indispensable in developing a comprehensive sociology of AI. Liu spotlights the shift in AI innovation from academic institutions to industry giants and startups, prompted by commercial interests. This market-driven “AI summer” is characterised by considerable investments from both the public and the private sectors, notably in the US and China, with tech giants leading breakthrough innovation in areas such as LLMs and generative AI. However, Liu cautions against oversimplifying the narrative to dominance by these tech behemoths and foregrounds the critical role of startups, academia, and the intricate web of investments, which together shape the AI landscape. The political economy approach championed by Liu aims to dissect these complex interactions, revealing the nuanced power dynamics, ownership structures, and market forces that boost AI research and innovation. By applying political economy analysis, Liu seeks to unearth the sociopolitical underpinnings of AI development and embraces a sociology of AI that fully integrates economic and political considerations to better understand AI’s societal implications and potential future trajectories.

The next chapter, *Four Models of Algorithmic Bureaucracy in China*, written by political economist Baogang He and public administration scholar Jinjin Wu, explores the ways in which big data, innovative technology, and AI have transformed

the traditional bureaucracy into an algorithmic bureaucracy. They develop an analytical framework to demonstrate how different models of algorithmic bureaucracy emerge from the interaction of algorithmic functions and application systems. Their framework refines the understanding of the algorithmisation of bureaucratic organisations by analysing the role of algorithmic functions and their application systems in transitioning bureaucracy. They identify four categories of algorithmic bureaucracy models, which stem from the interplay of the automation and prediction functions and the departmental versus cross-departmental applications. These models offer insights into how algorithms impact bureaucratic structures, processes, and power relations, providing a nuanced view of AI applications in public administration beyond the dominant narrative of digital authoritarianism in China. The four case studies reported by the authors illustrate the diverse impacts of algorithmic bureaucracy and illuminate its potential to address inefficiencies and improve public service performance.

Part V, the book's last, is titled **Critical Theory of Artificial Intelligence**. It consists of two chapters that critically examine the adverse social consequences of the algorithmic condition. The opening chapter, "Social Inequalities and Artificial Intelligence: How Digital Inequality Scholarship Enhances Our Understanding," written by communication and media scholar Christoph Lutz, focuses on how inequality is exacerbated by the ascendancy of digital technology. Lutz delineates a "digital divide" across three levels: access, skills, and uses, and inspects the outcomes of technology use, particularly within organisational contexts where AI systems are both developed and applied. His discussion of intra-organisational and inter-organisational dynamics reveals that these digital divides comprehensively affect the workplace, from access to AI technologies and the skills to use them to the benefits from their use. If these divides are to be effectively counteracted, AI inequality must be recognised as a significant issue, and socio-technical approaches, legal interventions, and measures that promote empowerment, collaboration, and equitable access to AI technologies must be put in place. According to Lutz, future research should critically evaluate interventions, explore the "inequality footprint" of organisations in magnifying or mitigating AI inequality, and aim to build an infrastructure that facilitates comprehensive data sharing on AI-related organisational practices.

In the last chapter, "AI Racial Bias: How Text-to-Image Artificial Intelligence Generators Construct Prestigious Professions," management scholars Malgorzata Górska and Dariusz Jemielniak address a critical gap in scholarship by exploring the intersection of AI with race and gender. To explain how biases underlying AI technologies perpetuate social inequities, they survey the historical and institutional factors that have contributed and still contribute to the underrepresentation of marginalised groups in technological discourses. Their discussion draws on examples from algorithmic processes and content generation to shed light on the role of AI in reinforcing stereotypes and aggravating discrimination, and to underscore the importance of incorporating diverse perspectives into AI research and development. Their plea for a more inclusive approach to AI dovetails with and adds to the broader conversation about the imperative to critically examine and address the social implications of technological advancements. They call for a balanced and comprehensive

engagement with these issues so as to unveil and challenge the racial and gender biases within AI systems and to promote a more equitable technological future.

Conclusion

“The heteromorphous nature of language games” is one of the most important aspects of Lyotard’s analyses in *The Postmodern Condition* (1983 [1979]: 66). Meticulously investigated by other intellectuals who have since studied social and cultural transformations, this phenomenon concerns all forms of expression, including scientific discourse. It involves a shift away from a search for a unified scientific worldview, based on solid foundations, to the proliferation of diverse theories, concepts, and orientations. In this context, Lyotard’s suppositions on the consequences of the computerisation of society are still highly relevant. In one of his scenarios, this process could lead to advances in the regulation and control of social life, which would *de facto* entail the formation of a disciplinary society. His other scenario assumed a contrary outcome, with informatisation facilitating and intensifying social dialogue. Lyotard (1983 [1979]: 67) did not prophesy which direction things would take, but he stated that the realisation of the latter option was premised on citizens’ free access to memory and data banks.

The texts collected in this volume reveal that both scenarios outlined by Lyotard frame the discourse on the theorisation of the algorithmic condition today. Some of the chapters show that algorithms and AI create innovative disciplinary mechanisms or breed new dimensions of social inequality and bias. Others point out that, in the algorithmic condition, both information transfer and social practices are remarkably improved, as a result of which further advances in democratisation are also possible. Both ways of theorising and the numerous attempts to combine them, also represented in this volume, contribute to the contemporary “heteromorphous nature of language games” involved in the development of algorithms and AI. It is this heteromorphous nature, rather than the search for a single, all-encompassing theory of the algorithmic condition, that corresponds to and helps capture its complexity, multidimensionality, and ambivalence.

Crucial to deliberations in this book is also a reflection on and advocacy for conditions necessary for the algorithmic condition to increase individual autonomy and reflexivity and to buttress the construction of a more equitable future. In this respect, the contributors’ analyses are utopian, and in line with today’s discourse on utopia (e.g. Jacoby 2005; Levitas 1990, 2013; Sargent 2005), their utopian investment is intended as a reference point for further social and cultural development. This type of utopian thinking is consistent with the “heteromorphous nature of language games,” because, instead of proposing specific blueprints, this thinking sketches the general contours of the transformations needed and encourages dialogue on their directions. This approach is also fully in harmony with the crowning conclusion of *The Postmodern Condition*, in which Lyotard (1983 [1979]) lauds open dialogue underpinned by free access to information and proclaims: “This sketches the outline of a politics that would respect both the desire for justice and the desire for the unknown” (67).

Notes

- 1 The term “artificial intelligence” has a very broad semantic field. Anthony Elliott (2022) proposes seven ways of defining AI, as: (i) computer programmes capable of intelligent activity; (ii) autonomous systems that rival or exceed human capabilities; (iii) technologically driven forms of thought; (iv) a project of the automated production of meanings, signs, and values; (v) the study and design of “intelligent agents”; (vi) systems imitating intelligent human behaviour; and (vii) an imitation of biological intelligence.
- 2 In line with recent trends in the social sciences and humanities (e.g. Jacoby 2005; Levitas 2013), utopian thinking is understood here not as an aspiration to design a perfect and unchanging society (which, of course, would altogether contravene the assumptions of the postmodern condition), but as relativising the *status quo*, combined with exploring alternatives to it. We explain the idea in more detail further in the text.
- 3 Go is a board game invented in ancient China that has been gaining popularity around the world in recent years.
- 4 OpenAI CEO Sam Altman contends that AGI will not have as much impact on the world as it seems (Sigalos and Browne 2024). It is hard to take his words at face value.
- 5 “The algorithmic turn” is one of the cultural turns deriving from the linguistic turn (Bachmann-Medick 2016 [2006]).
- 6 The coinage “vectorial politics” was previously used by Paul Virilio in *Negative Horizon: An Essay in Dromoscopy* (2005 [1984]), where he wrote: “Hyperconnected in these terminal points and other control and surveillance points, the anational State prepares to cast off its moorings with urban populations, its milieu is henceforth the non-lieu, the non-place, of speed, the non-territory of an essentially vectorial politics where the pre-eminence of Time replaces that of Space” (95).

References

- Airoidi, M. *Machine Habitus: Towards a Sociology of Algorithms*. Cambridge: Polity Press, 2022.
- Amaratunga, T. *Understanding Large Language Models: Learning Their Underlying Concepts and Technologies*. Berkeley, CA: Apress, 2023.
- Bachmann-Medick, D. *Cultural Turns: New Orientations in the Study of Culture*, translated by A. Blahut. Berlin: De Gruyter, 2016.
- Beer, D. *The Tensions of Algorithmic Thinking: Automation, Intelligence and the Politics of Knowing*. Bristol: Bristol University Press, 2024.
- Brevini, B. *Is AI Good for the Planet*. Cambridge: Polity Press, 2022.
- Cetina, Knorr, T.R Schatzski, and E von Seivigny. *The Practice Turn in Contemporary Sociology*. New York: Routledge, 2000.
- Chollet, F. *Deep Learning with Python*. New York: Simon & Schuster, 2021.
- Coleman, F., V Bühlmann, A O'Donnell, and I van der Tuin. 2018. “Ethics of Coding: A Report on the Algorithmic Condition.” https://ualresearchonline.arts.ac.uk/id/eprint/14236/1/Ethics_of_Coding_A_Report_on_the_Algorit.pdf Access 11 April 2024.
- Crawford, K. *The Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press, 2021.
- Ede, A. *Technology and Society: A World History*. Cambridge: Cambridge University Press, 2019.
- Edwards, L., and M Veale. “Slave to the Algorithm? Why a ‘Right to an Explanation’ Is Probably Not the Remedy You Are Looking For.” *Duke Law & Technology Review* 16 (2017): 18–84.
- Elliott, A. *The Culture of AI: Everyday Life and the Digital Revolution*. London: Routledge, 2019.
- Elliott, A. *Making Sense of AI: Our Algorithmic World*. Cambridge: Polity Press, 2022.

- Elliott, A. *Algorithmic Intimacy: The Digital Revolution in Personal Relationships*. Cambridge: Polity Press, 2023.
- Giddens, A. *The Consequences of Modernity*. Cambridge: Polity Press, 1990.
- Gonsalves, T. "The Summers and Winters of Artificial Intelligence." In *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction*. Vol. 1, edited by M. Khosrow-Pour, 168–179. Hershey, PA: IGI Global, Engineering Science Reference, 2019.
- Gurumurthy, A. and D. Bharthur. "Democracy and the Algorithmic Turn: Issues, Challenges and the Way Forward." *Sur-International Journal of Human Rights* 15, no. 27 (2018): 39–50.
- Hine, E. and L. Floridi. "Artificial Intelligence with American Values and Chinese Characteristics: A Comparative Analysis of American and Chinese Governmental AI Policies." *AI & Society* 39 (2024): 257–278.
- Hristova, S., S. Hong, and J. Daryl Slack. *Algorithmic Culture: How Big Data and Artificial Intelligence Are Transforming Everyday Life*. Lanham, MD: Lexington Books, 2021.
- Indurkha, N. and F. J. Damareu. *Handbook of Natural Language Processing*. Boca Raton, FL: Taylor and Francis Group, 2010.
- Jacoby, R. *Picture Imperfect: Utopian Thought for an Anti-Utopian Age*. New York: Columbia University Press, 2005.
- Kai-Fu, L. *AI Superpowers: China, Silicon Valley, and the New World Order*. New York: Houghton Mifflin Harcourt, 2018.
- Kiberg, H. and H. Spilker. "One More Turn After the Algorithmic Turn? Spotify's Colonization of the Online Audio Space." *Popular Music and Society* 46, no. 2 (2023): 151–171.
- Levitas, R. *The Concept of Utopia*. London and New York: Philip Allan, 1990.
- Levitas, R. *Utopia as Method: The Imaginary Reconstitution of Society*. Basingstoke: Palgrave Macmillan, 2013.
- Liotard, J. F. *The Postmodern Condition: A Report on Knowledge*, translated by G. Bennington and B. Massumi. Minneapolis: University of Minnesota Press, 1983 [1979].
- MacKenzie, D. and J. Wajcman. *The Social Shaping of Technology*. Buckingham, UK: Open University Press, 1999.
- Marres, N. *Digital Sociology: The Reinvention of Social Research*. Cambridge, UK: Polity Press, 2017.
- McCarthy, J., M. L. Minsky, N. Rochester, and C. E. Shannon. 1955. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> Access 11 April 2024.
- Novotny, H. *In AI We Trust: Power, Illusion and Control of Predictive Algorithms*. Cambridge, UK: Polity Press, 2021.
- O'Hara, K. "The Contradictions of Digital Modernity." *AI and Society* 35, no. 1 (2020): 197–208.
- Ortega Klein, A. 2020. "The U.S.-China Race and the Fate of Transatlantic Relations. Part 1: Tech, Values, and Competition." *Center for Strategic and International Studies*. <https://www.csis.org/analysis/us-china-race-and-fate-transatlantic-relations> Access April 11, 2024.
- Pedwell, C. "Speculative Machines and Us: More-than-Human Intuition and the Algorithmic Condition." *Cultural Studies* 38, no. 2 (2024): 188–218.
- Prince, S. J. D. *Understanding Deep Learning*. Cambridge: MIT Press, 2023.
- Qiang, X. "President Xi's Surveillance State." *Journal of Democracy* 30, no. 1 (2019): 53–69.
- Roberge, J. and M. Castelle, eds. *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. London: Palgrave Macmillan, 2021.
- Roberts, H., J. Cowls, E. Hine, F. Mazzi, A. Tsamados, M. Taddeo, and L. Floridi. "Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US." *Science and Engineering Ethics* 27 (2021): 1–25.

- Sargent, L. T. "A Necessity of Utopian Thinking: A Cross-National Perspective." In *Thinking Utopia: Steps into Other Worlds*, edited by J. Rüsen, M. Fehr and Th. W. Rieger, 1–14. Oxford and New York: Berghahn Books, 2005.
- Sigalos, M. and R Browne. 2024. "OpenAI's Sam Altman Says Human-level AI Is Coming but Will Change World Much Less Than We Think." <https://www.cnn.com/2024/01/16/openais-sam-altman-agi-coming-but-is-less-impactful-than-we-think.html> Access 11 April 2024.
- Silver, D., A. Huang, and C. Maddison, et al. "Mastering the Game of Go With Deep Neural Networks and Tree Search." *Nature* 529 (2016): 484–489. <https://www.nature.com/articles/nature16961>
- The Right to Privacy in the Digital Age: Report*. 2021. "Issued by Office of the High Commissioner for Human Rights, A/HRC/48/31." <https://www.ohchr.org/en/calls-for-input/2021/right-privacy-digital-age-report-2021> Access 11 April 2024.
- Tinnirello, M. *The Global Politics of Artificial Intelligence*. Boca Raton, FL: Taylor and Francis Group, 2022.
- Turing, A. M. "Computer Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433–460.
- Virilio, P. *Negative Horizon. An Essay in Dromoscopy*, translated by M. Degener. London: Continuum, 2005 [1984].
- Wang, W., J. Downey, and F. Yang. 2023. "AI Anxiety? Comparing the Sociotechnical Imaginaries of Artificial Intelligence in UK, Chinese and Indian newspapers". *Global Media and China* (online first).
- Wenger, E. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press, 1998.
- Wooldridge, M. *A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going*. New York: Flatiron Books, 2021.
- Zeng, J. *Artificial Intelligence with Chinese Characteristics*. Singapore: Palgrave Macmillan, 2022.
- Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.

Part I

The Algorithmic Revolution

New Theoretical Frameworks



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

1 A Theory of Algorithmic Sociality (TaS)

Robert Seyfert

Introduction

Several attempts have been made in recent years towards theorising digital transformations ([Daniels et al. 2017](#); [Lupton 2015](#); [Marres 2017](#); [Selwyn 2019](#); [Maasen and Passoth 2020](#)). These approaches are concerned with examining the digital transformation in society as a whole (cf. [Baecker 2018](#); [Nassehi 2024](#); Seyfert and Roberge 2016) or in its social domains, e.g., the infrastructure ([Healy 2015](#); [Latour et al. 2012](#)), the economy ([Pfeiffer 2021](#)), the financial market ([MacKenzie 2021](#)), forms of subjectification ([Goriunova 2019](#); [Reckwitz 2020](#)) and everyday life ([Amoore and Piotukh 2016](#); [Elliott 2022](#)).

In the following, I propose a theory of algorithmic sociality (TaS) following on from and in continuation of existing conceptual approaches. This theory aims to (1) conceptualise new (post-)social relations that have emerged in the process of digital transformation and add them to other types of social relations prevalent in sociology, such as intersubjective and interobjective relations. In doing so, digitalisation will not be understood as a unifying (or perhaps even simplifying) movement, but (2) as a process in which both the forms of relationships and the individuals in these relationships become more diverse and heterogeneous. (3) The individuals or elements (human subjects, algorithmic objects, etc.) that play a role in these relations and processes are typologically mapped. Gilbert Simondon's concept of disparity (*disparation*) is used to conceptualise the relations and processes between the individuals and such elements. The notion of *algorithmic sociality* is proposed to describe these new (post-)social relations.

A Relational Theory of Algorithmic Sociality (TaS)

The fundamental need for such a theory arises from changes to our societal situation. The digital transformation in recent decades has changed social coexistence, the institutions and organisations of society as a whole to an extent that also necessitates a transformation of sociological theory. In order to gain conceptual access to the digital transformation of society, the first step could be to distinguish the current (algorithmic) age as Digitalisation 2.0 from the age of Digitalisation 1.0.

In a very abstract manner, *digitalisation* can be defined as the “transformation of analogue values into digital formats [within a] binary distinction” (Baecker 2018: 60). The focus is on creating digital representations of physical entities (digital images, music, etc.) (Hedrick et al. 2020).

The central problems of *Digitalisation 1.0* in the late 1990s and early 2000s revolved around questions of transmitting these digital representations, interconnectedness (network society) and information exchange (information society), i.e., how information and data “are produced, circulated, and remediated via networked digital technologies” (Grusin 2010: 6; cf. Payne 2013). Initially, however, this was not a phenomenon that encompassed society as a whole but rather related to specific industrial sectors, expert cultures and niche milieus.

Digitalisation 2.0 can be understood as the generalisation of these phenomena, as the comprehensive digital transformation of society in which conventional face-to-face *relations* are replaced by digitally and algorithmically mediated relations as the dominant modes of relating in society. However, this dominance of relations to and with digital systems by no means renders the intersubjective and bodily-physical presence relationships traditionally studied in sociology obsolete. Rather, they expand the diversity and heterogeneity of already existing social relations, whereas these can even experience special appreciation and cultural status through the hegemony of digital relations.

Digitalisation 2.0¹ can be defined by two central criteria which at the same time make it distinguishable from Digitalisation 1.0. On the one hand, it is defined by digital collaboration and participation (Jenkins et al. 2006). Here, the separation of producers and consumers is overcome: everyone becomes a producer as well as a consumer of digital phenomena (Ritzer and Jurgenson 2010). Examples for this include Web 2.0 (e.g., social media in contrast to static websites as phenomena of Digitalisation 1.0), in which everyone can (potentially) both consume and produce. A second criterion is the spatio-temporal mobilisation and flexibilisation that actually makes the digitalisation of all areas of everyday life possible: “Technology itself—in terms of both applications and operating software—moves from desktop to WebTop” (Lash 2006: 580).

In the context of Digitalisation 2.0, there is often talk of increasing democratisation. In contrast to participation and spatio-temporal mobilisation, however, this is not an empirical finding, but rather rhetoric (Beer and Burrows 2007). In view of the private-sector status of the digital platforms in which the central social relations take place today, there can hardly be talk of democratisation (Zuboff 2019). We can, however, sociologically translate the rhetoric of democratisation, namely as generalised participation: while participation in central social media is not democratically organised, it is widespread. This generalisation refers to the difference mentioned between Digitalisation 1.0 and Digitalisation 2.0: to the transition from a niche culture to a phenomenon that affects society as a whole, to the spread of algorithms into all areas of social life (Roberge and Seyfert 2016).

With TaS, this chapter aims to contribute to a theoretical understanding of this general proliferation of digital relationships. In the process, extensive references can be made to current debates in the sociology of technology, science and

technology studies, the sociology of the digital (or digital sociology) and others. In principle, these theoretical approaches are not interested in sociological analysis with and through algorithms (such as Data Science), but in the sociological elucidation of a society that is increasingly determined by algorithmically mediated relations (individual, institutional, organisational, etc.). The concept of relationship is quite broad, encompassing interactions (Rammert 2016; Rammert and Schulz-Schaeffer 2016) as well as interpassions (Seyfert 2019: Chapter 10). Such an approach points to a methodological relationalism that sees the relationships between individuals and elements (e.g., human individuals and algorithmic objects) as constitutive and primary vis-à-vis these entities. If one understands algorithms as relational mechanisms, one can assume that both the algorithm as well as the individuals emerge from this relationship or are changed by it and (in the case of machine-learning algorithms) even learn from each other (Brooker et al. 2017; Gillespie 2016). Such theories are thus directed against two dominant traditions. They are directed (1) against an *algorithmic determinism* that considers algorithms primarily an instrument for the configuration of users (Woolgar 1991), or even for the manipulation and control of human individuals (cf. Harari 2017). And they are directed (2) against an *algorithmic instrumentalism* that sees algorithms merely as a passive tool of humans (cf. Steiner 2012). Algorithms are neither adequately described as objective representations of the social world nor as subjectifying technologies. The “truth” of algorithms is no more in their code than the truth of individuals is in their brain structures. Both emerge from interrelationships in line with Simmel’s *reciprocal relationships* (Simmel, 1910: 386), i.e., people co-produce algorithms just as algorithms conversely co-produce people (Kalpokas 2019: 70). In this context, algorithms take on a special role within the digital transformation of society because they represent one of the crucial types of reciprocity in digital societies.

The relationality of the theory, however, does not assert a dissolution of the elements and individuals interwoven in them. Through their interrelationships, they find themselves in a state of permanent change, yet still remain distinct elements. With Gilbert Simondon, I will describe their relationship as a *disparity*. This is a metastable equilibrium, i.e., (1) a quasi-coherent entity in which (2) a problematic and tension-laden relationship between the elements persists. This problematic relationship is neither dialectically eliminated nor do the elements merge into a socio-technical symbiont. One consequence of such a theoretical constellation is to distance oneself from post-digital approaches, i.e., those that seek to level the distinction between digital and analogue. I argue for a theory of social modes of relating (Seyfert 2019) that does not overcome the distinction between digital and analogue, but rather assumes the multiplication and heterogenisation of social relations: the new forms of digital (or algorithmic) social relations are joining, supplementing, and multiplying the classical analogue (or intersubjective and bodily-physical) relationships.

Accordingly, the theory (TaS) unfolds in the following order: (1) the individuals or *elements* (e.g., human individuals and algorithmic technologies), (2) the *relations* that these elements enter into and (3) the *processes* associated with these relations.

In order to adequately present the theoretical context, I will first trace the environment and the elements in which these relations take place (and which in turn shape the elements). This is in no way intended to indicate a primacy of a (natural or technical) environment or of the elements (people, algorithms). A good theory has no core or hierarchical structure, it has no beginning and no end (Deleuze 1994: 129). If I begin by presenting the environment and the elements of algorithmic sociality, it is simply because of the convention that presenting a theory has to start somewhere.

Elements of Algorithmic Sociality

For the sake of simplicity, I am developing the TaS using the heuristic distinction between two systems and environments²: digital and analogue environments or digital and analogue systems.

Before I go into this in more detail, I would like to explain the distinction between analogue and digital. Representatives of post-digital approaches in particular will assume that this is an adoption of an outdated conceptual differentiation. These approaches argue that it is precisely the fact that digital relationships have become ubiquitous and self-evident that makes distinguishing between digital and analogue obsolete. For example, Nicholas Negroponte (1998) observes – during Digitalisation 1.0 – in the context of computers that we only notice them when they are not there or when they are not working. In everyday life, we do not encounter the digital as something particular or specific. The difference between the digital and the analogue seems to have disappeared. The same seems to be true for algorithms today. Just like in the past, we are now – in the second wave of (algorithmic) digitalisation – dealing with a radical structural transformation of society and all social relations. Analogous to the thesis of post-digitality, one could therefore speak of a post-algorithmic age today, which would refer to the fact that precisely the ubiquity and self-evidence of algorithms make the distinction between algorithmic and non-algorithmic phenomena superfluous. Without being able to trace the discussion about the concept of the post-digital constellation in detail, one would first have to point out the ideology of progress implicit in this theory, which becomes apparent, among other things, when the distinction between digital and analogue or online and offline is described as “anachronistic” due to our permanent digital networking (Berry 2015: 50). From this perspective, Digitalisation 2.0 is the age of the post-digital. I consider the notion of a post-digital (or post-algorithmic) age to be implausible in empirical and theoretical terms, especially with regards to the transformation of social relations. The COVID-19 pandemic of the past years has allowed individuals to experience first-hand the differences and values of various modes of relation, including digital-analogue, digital-digital and analogue-analogue.

Within the pandemic situation, relations between physically present bodies gained special appreciation precisely because of the dominance of digital relationships. However, this valorisation of intersubjective relations in pandemic times presupposes a distinction between digital and analogue relations. The concept of post-digitality, however, systematically levels such empirical differences. New forms of social

relations – such as between algorithms and subjects – are invisibilised, and this is because the comparison to other forms (e.g., to analogue face-to-face interactions), which could make the specifics of such forms of relationship visible at all, is lost.

The transformation of social relations and the emergence of new forms (including subject-algorithmic relations) is also not accompanied by the dissolution of the elements involved (human subjects, algorithmic objects, etc.) and their merging into socio-technical symbionts. Rather, the individuals and elements are subjected to their own transformation processes, e.g., in which new forms of subjectivity are developed. Everywhere, the distinction between algorithmic object and human subject is performatively produced and stabilised (e.g., through the valorisation of intersubjective relations during the pandemic or through regulatory demarcations of algorithmic objects [Seyfert 2022]). At the same time, the relationship between the two is context-dependent: some human subjects claim complete control over “their” algorithmic tools, while others are much more willing to “surrender” to the algorithms and be guided by them. Strictly speaking, one can no longer talk of algorithmic objects and human subjects, but (following Michel Serres) of algorithmic and human quasi-subjects and quasi-objects (Lange et al. 2019: 604 ff.).

If the representatives of post-digital approaches assert an increasing indistinguishability of digital and analogue phenomena or living environments – as a dialectical abolition of distinction or even as a transition into symbiotic relations – the TaS conceptualises the relationship between analogue and digital elements (in section 2.2) with Gilbert Simondon as a *disparity* and a metastable equilibrium.

The realm of the analogue includes *analogue systems* (animals, plants, people, etc.) and the *analogue environment* (the natural environment of material things and artefacts, etc.). The analogue environment also includes technical artefacts that are not digital. For reasons of space, the analogue environment will only be mentioned in passing in the following, e.g., in connection with the object recognition of autonomous vehicles. I mention them here for the sake of completeness of theory. In contrast, analogue systems will play a more important role, especially human subjects.

The digital realm includes the *digital environment*: these are technical infrastructures, networks, servers, “integrated circuits, printed circuit boards, cables, power supplies, memory and printers,” etc. (Tanenbaum and Austin 2013: 8). In this context, infrastructures must be thought of as relational and procedural, not as static and fixed. Star and Ruhleder propose the following definition: “Infrastructure is a fundamentally relational concept. It becomes infrastructure in relation to organized practices” (Star and Ruhleder 1996: 113). Thus, an algorithm itself can also become the infrastructure and digital environment of another algorithm, e.g., (in the case of machine learning) when an algorithm recognises patterns in training data and develops a model from them. Such a model can in turn form the digital environment or infrastructure for an algorithm to identify inferences in live data.

The digital environment also includes *data* (big and small). Data, too, should not be thought of as singular entities, but rather as relational components. They only gain their independent value based on their specific position and the relationships they can enter into within the data structure: “Digital data are thus designed to be linked with other data for specific calculations in order to algorithmically produce

result data” (Häußling 2020: 141). Data are not simply raw data, but specifically structured information. Each phenomenon that an algorithm is to process must be modelled in an algorithm-specific way (Manovich 1999: 84). In other words, data must be transferred into a concrete data form and formalised so that algorithms can operate with them automatically. One can then also analyse algorithms by looking at how information has to be tailored to the algorithms, how information is made *algorithm-ready* (Gillespie 2014: 171). In the context of a relational TaS, the development of relational object-oriented database systems is important. While older systems organised data in inflexible hierarchies, in these systems “information can be organized in more flexible ways, where bits of data can have multiple associations with other bits of data, categories can change over time, and data can be explored without having to navigate or even understand the hierarchical structure by which it is archived” (Gillespie 2014: 171).

Furthermore, the digital environment is also a *codified space*, a measured and digitally mapped space. A digital environment is, e.g., an environment for which a digital map exists that can be downloaded and used for orientation. If such a map does not exist, it is an analogue environment. This example shows that terms such as digital environment, analogue environment, digital systems and analogue systems are heuristic terms whose reference to reality can only be clarified in empirical studies. A digital space (for which a digital map exists in principle) becomes an analogue space for those who lack access (who do not have this map, do not have online access or access rights, etc.).

Digital systems now include *algorithms*, e.g., in digital platforms, automated systems, social media, etc. Data are fundamentally part of the digital environment, but they only become part of digital systems at the moment of their algorithmic relation, i.e., when they pass from the status of virtuality (in the database) to the status of actualisation. Data can only operate as interfaces between social and algorithmic processes (Häußling 2019: 144) if algorithms mediate the connection between data and social processes when they actualise them.

What applies to the distinction between digital environments and digital systems also applies to the distinction between analogue and digital environments, namely that it is a largely heuristic distinction. The well-known insight testifies to this: “Hardware and software are logically equivalent” (Tanenbaum and Austin 2013: 8). This refers to the “softening of hardware,” i.e., the fact that the processing systems are increasingly embedded directly in the hardware (Vahid 2003, 27 ff.). The reference to embedding software in hardware simultaneously helps to define the digital environment more precisely. This does not necessarily have to be imagined as something external, as a zone “out there.” Instead, the environment can be very close. Digital environments are not only “big” infrastructures like cable lines, radio towers and satellites, but all kinds of digital shells and infrastructures in which digital systems are embedded. I already pointed out that a model or virtual data can also form an environment.

The crucial aspect is that algorithms are the central mechanisms of digital systems in digital environments. For the TaS, they thus represent the main elements of a digitally transformed society. This approach thus goes beyond the idea of

an information society (Beck et al. 2014) or a network society (Castells 2000) and does justice to the algorithmisation of contemporary society. Algorithmic sociality does not so much refer to the fact that access to the social world is mediated via data or networks (cf., e.g., Häußling 2020); rather, algorithms are the relations and processes that allow data to become part of a data network in the first place. Without relationising data (its actualisation) by algorithms, data are *objects* in Latour's sense, i.e., they are disconnected elements without social existence. What Latour said about archaeological excavation objects and archaeologists applies to data and algorithms: "as soon as they grasp in their hands these poor fossilised or dusty objects, these relics immediately cease to be objects" and return to the social (relationising) world (Latour 2000: 11).

Below, I would like to go into more detail on the definition of *algorithms* specified here. In social and cultural studies, algorithms are conceptualised in various ways, such as rules, formulae, functions or mathematical abstractions and situated practices (Mackenzie 2017: 8).

Relational and operational (actualising) aspects are central to the TaS: "Relationships between real-world surveillance data and machines capable of making statistically relevant inferences about what that data can mean" (Cheney-Lippold 2011: 178). In this context, an algorithm is not just a formal instruction, not just a "set of rules proceeding from inputs to output" executed by a computer (Soare 1997: 222). It is not just a table of symbols or a formula (cf. Cleland 2001). The notion of an algorithm as a formal instruction separates instructions from their implementation and execution. Algorithms are not only formal procedural descriptions, however, but above all *processes*. If data are the "products of relations" (Häußling 2020: 139), then algorithms are the *processes* that set this relationising in motion. They not only establish relationships between different environments and systems, but also inscribe specific temporalities and temporalities in the social world; "they need unfolding, and thus they embody time" (Miyazaki 2016: 129). In this respect, I understand algorithms not only as formal schemata but also as real processes and operations. In fact, I follow the definition of an algorithm by Gurevich and Blass: "By definition, a computable function is a function computable by an algorithm" (Gurevich 2000: 80). In this respect, in terms of the TaS, "an algorithm is a machine" (Blass and Gurevich 2004a: 216). This also overcomes the distinction between algorithms as formal rules and their technical implementation and execution. Those wishing to separate algorithms and implementation use the word "implementation for the things that we call algorithms" (Blass and Gurevich 2004a: 234).

In the context of establishing relationships between different environments and systems, another common specification is to define algorithms as "recursive functions" (Moschovakis 2001). Thus, Totaro and Ninno see recursion as the central defining feature of algorithms and at the same time the interpretative key to understanding modern rationality (2014, 2020). A recursive function is a process in which the output is processed as input in a potentially infinite loop (Wirth 1996: 140). It is a "nesting, and variations of nesting" (Hofstadter 1979: 127). To illustrate this, reference is often made to an object positioned between two mirrors directly facing each other, producing an infinite chain of reflections of that object

(Totaro and Ninno 2014: 30). However, defining an algorithm as a recursive function does not do justice to the diversity of algorithms. It is virtually useless for a general TaS. Such a definition obviously does not include those types of algorithms that are interrelated with an unknown (digital) environment and can also compute non-recursive functions (Blass and Gurevich 2004b: 296 and 289). The TaS therefore follows such definitions that understand algorithms as “interactive” (without attaching much importance to the activist slant of interactionist terminology at this point). Interactive algorithms interact with an unknown (digital) environment that provides random bits (Blass and Gurevich 2004b: 289f.). Only a definition like this does justice to a relational approach in which algorithms are not defined ontologically by the content of an instruction or a code, but as machines that are structured in relation to the environment. Algorithms are relational mechanisms that relate elements (e.g., data) to each other (i.e., process them) and enact³ or co-produce them while doing so⁴ (Seyfert 2022). This process is followed by another algorithm that places (processes) the data produced in this way in new relationships, and so on. If data are structurally relational, i.e., they are determined by their constellation to other data, algorithms are procedurally relational, i.e., they put data into relationships, actualise them and thus enact them. The procedural-relational character of algorithms is also the reason why data alone are not part of algorithms; rather, they form the environment of algorithmic systems: data are only related to each other (i.e., processed) through algorithms; they are actualised and produced by algorithms.

This rather general definition of algorithms can now be used to specify further forms, which I can only hint at due to the limited space available. In discourse on algorithms, a distinction is usually made between learning and non-learning (based on human rules) algorithms. However, this is not an entirely trivial distinction, which makes the differences seem greater than they actually are. On the one hand, we can make this distinction appear as a qualitative difference: while rule-based algorithms produce an output by processing an input based on explicit rules, *machine learners* (Mackenzie 2017) achieve this goal based on automatically learned knowledge. The resulting outputs are ambivalent (for the human) participants (which I will discuss in more detail in section 2.3). However, we can also understand this distinction as an internal increase in complexity, whereby the prominent distinction of learning/non-learning loses significance. Then, while rule-based algorithms produce an output by processing an input according to a fixed rule, machine learners produce algorithms themselves starting from given inputs and outputs: “Learning algorithms [...] are algorithms that make other algorithms. With machine learning, computers write their own programs” (Domingos 2015: 6).

In addition, there are hybrid forms (Chiticariu et al. 2013: 827), and sub-forms (supervised, unsupervised, reinforcing, etc.), which I cannot go into here due to lack of space.

The key significance of relations and processes already becomes clear in this basic representation (Figure 1.1) of the system-environment distinction.

The next section will deal with the new forms of social relations that become visible through the TaS.

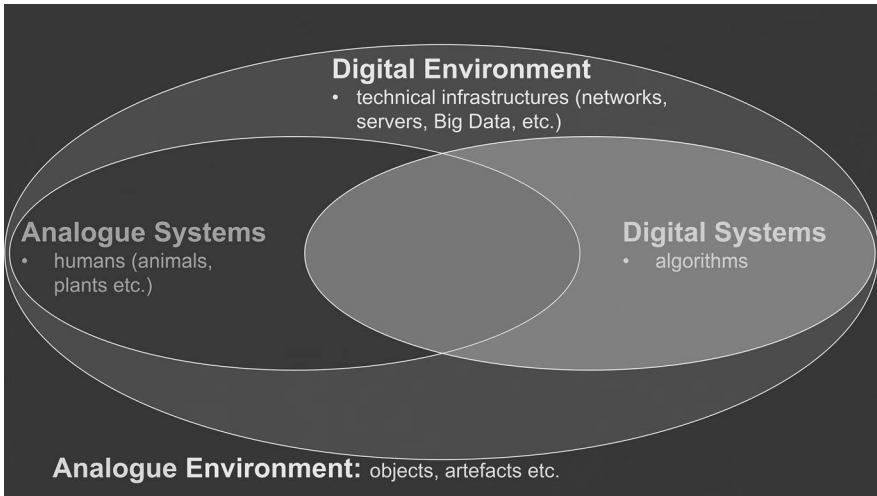


Figure 1.1 Elements of algorithmic sociality.

Relations of Algorithmic Sociality

Sociology usually understands social relations as *intersubjective* relations, i.e., the interactions and interpassions of humans. In addition, approaches interested in *subject-objective* relations, i.e., the relationship of humans to non-human and more-than-human entities and things, can be found – although not in similar distribution – e.g., in the theories of post-social (Knorr Cetina 1997) and interobjective relations (Latour 1996), New Materialism (Lemke 2015), heterological ecologies (Kwek and Seyfert 2018) or the sociology of technology (Rammert 2016; Rammert and Schulz-Schaeffer 2016; Häußling 2019). The TaS is now also taking a look at these social relations that are emerging with the digital transformation. This already gives a first indication of the heterogeneity of social relations. The intersubjective and subject-objective relations are supplemented by *inter-algorithmic*, *subject-algorithmic* and *object-algorithmic* relations. An object-algorithmic relation, e.g., is the relationship of automated systems to a natural (analogue) environment. A classic example is in the automated object recognition of autonomous vehicles (cf. Hind 2019). Autonomous vehicles should not only recognise and react to other vehicles or pedestrians but also to things in the natural environment (such as trees, paths, but also weather and light conditions). In this context, it is important to point out that digital systems only have mediated access to the analogue environment via the digital environment. Strictly speaking, an object-algorithmic relation is a heuristic idea rather than an empirical object. No algorithmic system has full access to the natural environment. Each of these systems only recognises objects in the natural (analogue) environment when mediated via the digital environment, meaning that they are highly digitally mediated. Autonomous vehicles access their environment with the help of a variety of sensors (lidar, 360-degree cameras, GPS, etc.). The algorithm uses these data to create an environment model in the digital environment, i.e., a representation of the

vehicle environment in real time. The acquired data and the model are used by the vehicle to recognise the environment (object recognition, traffic light signals, etc.), to determine the position, to plan the movement (e.g., evasive manoeuvres) and to control (e.g., braking and accelerating), etc. (Devi et al. 2020: 2123).

Although not at the same level, a similar digitally mediated image can also be found in subject-objective relations, e.g., for our relationship to the analogue natural environment. This relationship is also increasingly only possible through and via the digital environment, e.g. when we move in a coded environment with the help of digital maps (determining our position), in which an algorithm suggests possible routes (planning) or informs us about dangerous situations through warning apps (control). Algorithms are constantly present. This also means, e.g., that the natural environment is increasingly perceived on the basis of its algorithmic representability: movement in space is co-determined by the affordances of navigation devices. For example, holiday destinations might be chosen based on whether they can be sufficiently represented on social media (Sari and Trinanda 2020).

Algorithms put data (elements in the digital environment) into a relationship and thus make them flow (or process them), e.g., when they sort data. For the algorithms relevant here (interactive algorithms), it is true that data themselves are (structurally) relational objects, i.e., they can in turn influence the activities of algorithms (Häußling 2020: 142). The relational character of algorithmic sociality can be illustrated by the various forms of algorithmic discrimination (cf. Hajian et al. 2016). This is the relational entanglement of specific elements in the digital environment (e.g., data sets), digital systems (e.g., algorithms) and analogue systems (e.g., people) and the analogue environment. Thus, the analogue (social) environment exhibits forms of discrimination that machine-learning algorithms can pick up and reproduce. This is similar to other processes of socialisation: if I socialise or train a learning algorithm, e.g., exclusively on people with a similar physical characteristic (e.g., skin colour), this will result in models that become blind to characteristics that deviate from this. If an algorithm now uses this model with live data, this consequently leads to biased and discriminated outputs. This is even more true when it is used on data sets that take into account more diverse physical characteristics. In this case, training on biased data leads to a discriminatory model or algorithm. It is also conceivable that the data (of the digital environment) show discriminations that are not found in the analogue (social) environment, i.e., that algorithms even intensify discrimination. This raises the question, e.g., which data should be taken into account in order to overcome the discrimination (Williams et al. 2018). One proposed solution is to balance data sets so that they better represent the social world (Pessach and Shmueli 2020: 17). Even more complex forms of discrimination are conceivable, e.g., by applying models to data that have little to do with the training data (cf. Tal et al. 2019). Even this brief overview points to the deep relational interconnectedness of analogue environment (social discrimination), analogue systems (subjective bias), digital environment (data bias) and digital systems (algorithmic processing bias).

Despite this fundamental interconnection, it is necessary, as explained in the beginning, to distinguish algorithmic from non-algorithmic relationships,

relationships between humans and algorithms and relationships between algorithms. Karin Knorr Cetina and Donald MacKenzie have shown how the operations of algorithms can be studied using the notion of *interaction* traditionally reserved for human relationships in sociology (Knorr Cetina 2013; MacKenzie 2019). Beyond the notion of an interaction of algorithms, other approaches conceptualise inter-algorithmic relations through the concept of the algorithmic *mass* or *avalanche* (Borch 2020) and the algorithmic *series* or *queue* (Pardo-Guerra 2019), e.g. In such approaches, ideas of the production of technologies are developed that extend beyond the idea of a purely technical evolution or economic efficiency. In his studies of algorithmic stock market trading, Donald MacKenzie is interested in how algorithms communicate or *interact* with each other. Their interaction is also part of a complex political economy. MacKenzie is also concerned with the question of how inter-algorithmic communication is enabled and co-produced by political and social processes, since the emergence of algorithmic stock market trading cannot be explained without the very specific regulation of financial markets and the market microstructure that this regulation helped to produce (MacKenzie 2021).

In addition to the conceptualisation of inter-algorithmic relations, I would like to highlight another form of social relation, which, according to my thesis, is the central relational form of algorithmic sociality. This concerns *subject-algorithmic relations*. It is not an algorithmically mediated form of intersubjectivity (as in social media) or a kind of inter-algorithmic sociality (as in the case of financial market algorithm communication); nor is it a form of sociality based on algorithmic processes (a kind of mechanical or machinic sociality). Rather, the concept of *subject-algorithmic relations* refers to the relations between analogue systems (plants, animals, humans, etc.) and digital systems. Endless examples can be found, ranging from wearables for humans (Nosthoff and Maschewski 2019) and animals (Lindzon 2015); interaction with assistance systems and robots (Karafillidis and Weidner 2018); the use of search engines, sales platforms, computer and video games; to smart farming in agriculture (Wolfert et al. 2017). It is the subject-algorithmic relations that form the core of algorithmic sociality. This is overlooked by studies that approach the digital transformation exclusively from the technical side, privileging the role of technology over non-technical entities and processes (Suchman 2006: 269f.), e.g., by exclusively studying the possibilities, capabilities and limitations of new technologies such as machine learning and Big Data. However, it is also overlooked by those that approach these phenomena exclusively phenomenologically and subject-theoretically, i.e., focusing exclusively on the influence, experiences and impressions of individual subjects. Both the technical focus and the phenomenological focus presuppose one of the two sides of the relation as constitutive. Thus, algorithms are often understood exclusively as instruments of control that exert influence on people as strangely passive elements and do not experience any change in these relationships themselves (cf. Zuboff 2019). Or, conversely, algorithms are passive tools created by humans, which in turn leave them completely unchanged.

The TaS shifts the primacy from both ends of the relation (algorithmic object-human subject) to the relation *between* these two sides and from there analyses how both sides co-produce and change in their reciprocity. Such an analysis must

(1) focus on the different facets of the relation. It is not empirically sufficient to simply speak of the relationship between people and algorithms, people and machines, people and robots, etc. Rather, it is important to provide a conceptual toolkit that is complex enough to make the details of these relationships visible. Elsewhere, I have suggested that these *modes of relating* should be explored through an *analysis of affect*, i.e., the ways in which each side of the relation is affected by the other and the ways in which each affects the other side of the relation (Seyfert 2019: Chapter 8). Thus, such a relationship goes far beyond cognitive, intentional and reflexive processes. For example, an ethnography of high-frequency algorithmic trading has brought to light, among other things, the socio-technical relations between human subjects and their technical devices, including relations to algorithmic technologies. This is not simply a form of screen-based work and information processing but rather a multi-sensory experience. Traders engage deeply in trading activities through a combination of visual and auditory cues, and occasionally, chemical stimuli (Seyfert 2018). Furthermore, one can investigate the relationships of human subjects and algorithmic objects alongside the distribution of attention, cognition, action, control and responsibility. These are quite dynamic and changeable constellations that can be present in very different forms of relation and participatory relationship depending on the context and situation.

Beyond the diverse and heterogeneous forms of relation, an analysis of algorithmic sociality (2) can make the transformative effects visible. This can analyse which processes are set in motion by these relations. Here, a central argument of a relational theory is that the two sides (e.g., human individuals and algorithmic technologies) are not determined by themselves, but that these are only enacted and transformed in the interrelationship. Studying processes of algorithmic sociality can make this change visible. As already mentioned above, this is also accompanied by processes of subjectification, which can refer to both human individuals as well as algorithmic technologies (e.g., anthropomorphised robots).

Despite this reciprocity, the TaS assumes that the elements remain fundamentally distinguishable. They are neither annulled in a dialectical process nor do they merge into a socio-technical symbiont. I prefer to describe their relationship (3) as *disparate* and metastable. The concept of *disparity* was introduced by Gilbert Simondon in his theory of individuation, among others (Simondon 1995). He thus solves the problem of dialectical methods, in which the production of a systematic unity always takes place at the expense of the heterogeneity of its elements (cf. Deleuze 1994). A disparate relationship is the mediation of (at least) two positions, in which the original contradictoriness continues to exist after the mediation. In contrast, dialectics mediates the contradiction and eases the tension: “The essential moment thereby vanishes from the fluctuating interplay, namely, that of disintegrating into extremes of opposed determinatenesses” (Hegel 2018: 112). Disparity is now a process in which “two disparate realities can become a system” without the opposition itself being cancelled out (Simondon 1995: 29). Gilbert Simondon uses the example of human vision to illustrate how incompatible elements are held together under tension in a disparate relationship: each eye perceives reality as a two-dimensional image. As the left and right eyes are in different positions, the two images can never overlap.

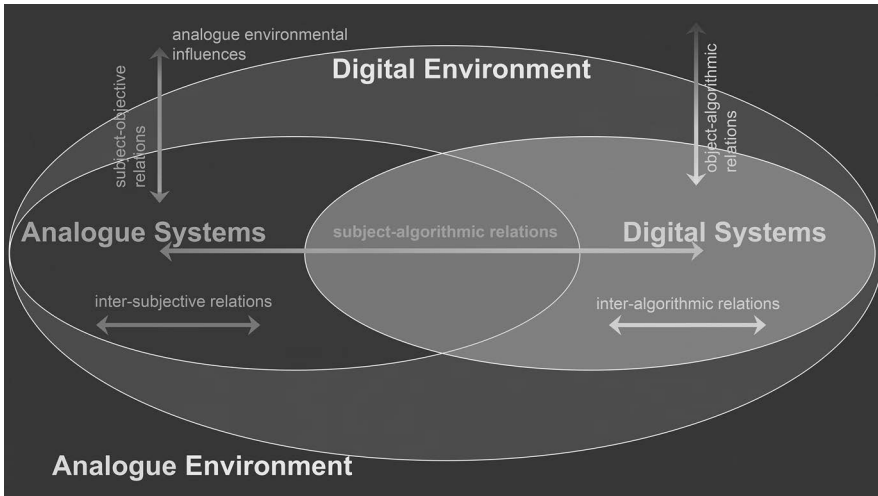


Figure 1.2 Relations of algorithmic sociality.

There is no third image that can be formed from this, i.e., the images are disparate (parallax). They are in contradiction, in tension with one another. The solution lies in the integration of these two two-dimensional images into an “integrated system of the two images, a system that exists according to a higher level than that of each of the images, but does not contradict them” (Simondon 1995: 206): a three-dimensional image. In doing so, the contradictoriness of the two two-dimensional images is not (dialectically) eliminated; rather it remains (disparate). Not only is the unity and integrity of the system taken into account (as in dialectic), but so are the difference and contradiction of the elements. The tension and the contradiction between the elements do not disappear; they are not dialectically abolished, but remain in a “metastable equilibrium” throughout our lives (Simondon 1995: 33).

I will now transfer this conception of disparate and metastable relationships to the (post-)social relations in digital constellations (Figure 1.2).

Processes of Algorithmic Sociality

In my schema, I will limit myself in the following exclusively to the *subject-algorithmic relations* of algorithmic sociality, so to the interrelationships (interactions and interpassions) of human subjects with algorithms or the interrelationships mediated via algorithms. These relations are interrelationships in which every action on one side leads to a reaction on the other side, which in turn is an action or passion. The reciprocity is to be taken seriously, as it is not a one-sided relationship or influence, but a transformative interrelationship in which both sides undergo processes of transformation, e.g., de-subjectification and re-subjectification. Informed by Spinoza’s affect theory, interrelationship means that one cannot affect without being affected at the same time. All receptions and passions (forms of

being affected), as well as all actions and reactions (forms of affecting), must be understood as transformative events according to Spinoza (Spinoza 1994).

To describe this relationship, I will again choose an entry point that is contingent and does not make any normative statements about the first relationship, i.e., does not grant any primacy to the first elements mentioned. I shall start with the receptive side of the algorithmic system. This can be identified heuristically as the attempt to recognise patterns (Nassehi 2024), e.g., in the form of *behaviour recognition*. For example, an algorithmic system can attempt to “understand” human behaviour by *recognising intent*. Behaviourist models usually come into play in these experiments. This is important because pattern recognition in no way represents or maps “society”: algorithmic pattern recognition is not digital society. Rather, they are observations of the digital environment by specific digital systems with their own epistemic regimes (Seyfert 2016), so specific perspectives of observation based on specific assumptions which come about by using specific technical devices. Observing a digital system reduces complexity, e.g., by assuming that analogue systems can be understood or – more precisely – modelled with certain (implicitly embedded) behaviourist approaches: it is often a matter of trying to convince and capture analogue systems (Lischka and Stöcker 2017). These are not simply harmless correlation analyses, but models with behaviourist claims: *habit-forming products* (Eyal 2014). Attempts to control behaviour follow on from behavioural recognition processes, e.g., to encourage the purchase of product X. The behaviourist nature of digital control means that such algorithms are successful in some areas, but not in others. They are successful precisely where people’s actions match the probabilistic predictions of behaviour by algorithmic systems, but not in other areas. There are also other reasons for the frequent lack of fit: the personal relevance of certain products is not determined on the basis of the preferences of specific individuals, but rather from a statistical assemblage of different components, a “matrix of probabilities of associations between people and things that shifts in time” (Mackenzie 2018: 11).

These ambivalent results also explain the mystification of algorithms: they make precise predictions and yet are completely off the mark at the same time. In some areas of social life, they are impressively accurate, yet this is now an even stronger contrast to the equally regular fulminant failure. Just as frequently as we are amazed at the precision of the prediction, we are stunned at how wrong they are. This also brings the other side of the *subject-algorithmic* relation into play. Empirically, the (e.g., human) subject in this relationship would be described neither as a mirror of the behaviourist predictions (and thus as a plaything of the algorithmic system) nor as a technologically enlightened subject (who would be able to see through the algorithmic operations completely). Rather, it is an individual for whom algorithms appear as largely non-transparent objects. This has to do with (1) that the majority of people are simply not technologically literate; or (2) that they are proprietary algorithms meant to be kept secret; or (3) that algorithms are often not designed to be interpretable by human subjects at all (Burrell 2016). However, this does not make human subjects externally controlled machines, as both behaviourist models (implicit) and critical sociology (explicit) assume.

Rather, analogue systems (predominantly humans in our examples) operate largely with *vernacular knowledge*, i.e., they develop a pragmatic knowledge of the operations of algorithms. I call it vernacular knowledge and not implicit knowledge because vernacular knowledge can, by all means, be explicit knowledge. The opposite of vernacular knowledge is not explicit knowledge, but expert knowledge. Think of search engine optimisation (SEO), e.g., search engine operators have a multiple interest in keeping their algorithms secret. On the one hand, they are a source of enormous economic profit. A search engine also depends on its search results not being easy to interpret and thus easy to manipulate. The culture of secrecy of search engine operators leads to users developing their own, indirect ways of knowing. They try to deduce indirectly how the algorithm works in a process of reverse engineering. In the case of SEO, an entire industry has emerged from this that deals with nothing else. This form of vernacular knowledge often even exists as explicit knowledge. We can also choose more everyday examples, such as the way we practically use digital route planners (Pentenrieder 2020), tap into the specific logics of social media and deal with recommendations from sales platforms. In all these examples we are working with vernacular knowledge, with “folk theories” (Eslami et al. 2016), i.e., with “intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems, which guide reactions to and behaviour towards said systems” (DeVito et al. 2017, 3165). This knowledge is vernacular insofar as it emerges from use, from which one makes sense of the workings of the algorithms. And it is vernacular insofar as it spreads in semantic form (e.g., in the form of DIY videos or life hacks on internet forums and social media). It is possible that this knowledge presents itself as “false” knowledge from the perspective of those who designed these systems, from the perspective of the experts. This could even be associated with appeals for the technical literacy of the population (computer science as a compulsory subject!). However, this techno-enlightenment perspective underestimates the performative effects of the *digital practices* that emerge from vernacular knowledge. For human subjects do not simply become manipulable objects for two reasons. One is that the assumptions and predictions of algorithmic systems (implicit or explicit) are largely socially unrealistic. The other is that “false” predictions can also produce real effects, make themselves true, as it were. This has been shown, among other things, by research on financial market models. Thus, the Black-Scholes model for option pricing was considered a rather inefficient calculation model in the financial market. However, due to its clarity and ease of use, it has become the most popular form of option price calculation in the financial market. The widespread dissemination in turn led to the influencing of market prices in a way that made the previously inefficient model the most efficient calculation model (MacKenzie 2006; MacKenzie and Millo 2003⁵). This example illustrates the unpredictability of digital practices. The unpredictability – which algorithmic systems will be successful and how they will be used specifically in practice – means that the effects of behaviourist experiments are not predictable, but contingent.

With the reference to the performative effects of digital practices, the processes of algorithmic sociality are fully represented (Figure 1.3). Digital practices are not

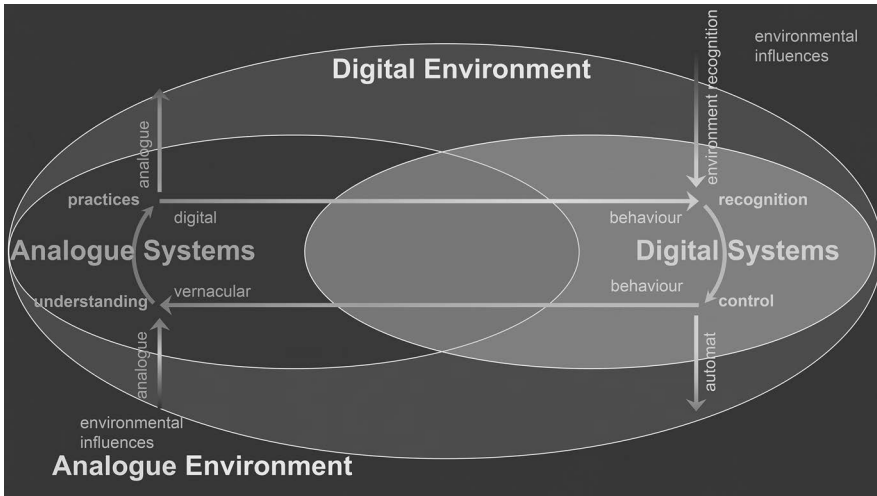


Figure 1.3 Processes of algorithmic sociality.

directly linked to algorithmic attempts to control behaviour, but deviate from their expectations. The digital practices thus, in turn, become the source data for new behaviourist predictions. These processes are interrelationships of (1) algorithmic pattern recognition, (2) behaviourist behaviour control attempts, (3) vernacular recognition and (4) digital practices, which in turn form the material for algorithmic pattern recognition.

Conclusion

The TaS is intended to provide an opportunity to clarify and empirically examine the diversity of forms of relationships that accompany the digital transformation, which here is understood as a multiplication rather than a transition (into a post-digital age) of social relations. A transformation understood like this requires a theory that is able to take into account all modes of relating, not only the emerging algorithmic ones but also the (seemingly) anachronistic ones, e.g., face-to-face interactions and interpassions. Beyond the intersubjective relations that sociology classically examines, it is now a matter of conceptually looking at relations to or with digital systems, or the intersubjective relations mediated by them. The reciprocity of algorithmic objects and human subjects in particular plays an important role.

The relationship of the elements (algorithmic objects, human subjects, etc.) is understood as a disparity that brings these elements into a system of relationships, yet does not make the elements disappear, instead understands their relationship as contradictory and fraught with tension. At the same time, the elements in this relation do not remain unchanged, but are in a relationship of mutual affectation, i.e., change. Newly emerging social relations (e.g., between human subjects and algorithmic objects) do not form an emergent system in which the elements would

merge, and thus do not lead to the end of subjectivity, but rather go hand in hand with new forms of subjectification and objectification.

As a conceptualisation of the new (digital) landscape of relations, the TaS can also form a basis for empirical studies. The formulation of correct questions is the start of every empirical investigation. In the theoretical framework of the TaS, these could be as follows: which elements (humans, animals, technologies, algorithms, etc.) are relevant in the relations to be empirically observed? Which forms of interrelationships (interactions and interpassions) between human subjects and algorithmic objects exist and what exactly do they look like? Which changes can be observed through these new modes of relating? What does embedding the elements in the digital environment look like (e.g., through sensors, cameras, digital communication, etc.)? What is the relationship between data and algorithms, or how can this be explained in terms of their mutual relationship?

The new conceptualisation of social relations and social relationships proposed by the TaS and the refocusing of methodological procedures and empirical investigations that this has made possible thus make a contribution to setting sociology on the path for researching society in times of the digital transformation.

Acknowledgements

This chapter is the English translation of Seyfert, Robert. “Die Theorie algorithmischer Sozialität (TaS),” *Österreichische Zeitschrift für Soziologie*, 5 June 2023. <https://doi.org/10.1007/s11614-023-00535-1>. All quotes that originated from non-English works have been freely translated.

Notes

- 1 Beer and Burrows (2007) point out that 2.0 terms (Web 2.0, Digitalisation 2.0, etc.) are contaminated concepts that often stem from strategic efforts by the advertising industry. I follow their example and use Digitalisation 1.0 and 2.0 exclusively as a “sensitising concept” (Swedberg 2014: 64).
- 2 The system and environment distinction does not indicate a reference to Luhmann’s systems theory at this point, as one might perhaps assume.
- 3 For the concept of enactment see Mol (2002: 33).
- 4 For the concept of co-production see Jasanoff (2004).
- 5 Donald MacKenzie also points out here the crucial difference between performativity theory and Merton’s theory of self-fulfilling prophecy: the Black-Scholes model for option pricing prevailed and has proven itself, even though the majority didn’t believe in its efficiency.

References

- Amoore, Louise and Volha Piotukh, eds. *Algorithmic Life: Calculative Devices in the Age of Big Data*. London, New York: Routledge, Taylor & Francis, 2016.
- Baecker, Dirk. *4.0 oder die Lücke die der Rechner lässt*. Leipzig: Merve, 2018.
- Beck, Ulrich, Anthony Giddens, and Scott Lash, eds. *Reflexive Modernisierung: eine Kontroverse*. Frankfurt a. M.: Suhrkamp, 2014.
- Beer, David and Roger Burrows. “Sociology and, of and in Web 2.0: Some Initial Considerations.” *Sociological Research Online* 12 (2007): 67–79.

- Berry, David M. "The Postdigital Constellation." In *Postdigital Aesthetics*, edited by David M. Berry and Michael Dieter, 44–57. London: Palgrave Macmillan, 2015.
- Blass, Andreas and Yuri Gurevich. "Algorithms vs. Machines." In *Current Trends in Theoretical Computer Science: The Challenge of the New Century*, vol. 2: Formal Methods and Semantics, edited by Gheorghe Păun, Grzegorz Rozenberg, and Arto Salomaa, 215–236. Singapore: World Scientific, 2004a.
- Blass, Andreas and Yuri Gurevich. "Algorithms: A Quest for Absolute Definitions." In *Current Trends in Theoretical Computer Science: The Challenge of the New Century*, vol. 2: Formal Methods and Semantics, edited by Gheorghe Păun, Grzegorz Rozenberg, and Arto Salomaa, 283–311. Singapore: World Scientific, 2004b.
- Borch, Christian. *Social Avalanche: Crowds, Cities and Financial Markets*. Cambridge: Cambridge University Press, 2020.
- Brooker, Phillip, William Dutton, and Christian Greiffenhagen. "What Would Wittgenstein Say about Social Media?" *Qualitative Research* 17, no. 6 (2017): 610–626.
- Burrell, Jenna. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (2016). <https://doi.org/10.1177/20539517156225>
- Castells, Manuel. "Materials for an Exploratory Theory of the Network Society." *The British Journal of Sociology* 51, no. 1 (2000): 5–24.
- Cheney-Lippold, John. "A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control." *Theory, Culture & Society* 28 (2011): 164–181.
- Chiticariu, Laura, Yunyao Li, and Frederick R. Reiss. "Rule-Based Information Extraction Is Dead! Long Live Rule-Based Information Extraction Systems!" In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 827–832. Washington, DC: Association for Computational Linguistics, 2013.
- Cleland, Carole E. "Recipes, Algorithms, Programs." *Minds and Machines* 11 (2001): 219–237.
- Daniels, Jessie, Karen Gregory, and Tracey McMillan Cottom, eds. *Digital Sociologies*. Bristol, Chicago: Policy Press, 2017.
- Deleuze, Gilles. *Difference and Repetition*. New York: Columbia University Press, 1994.
- Devi, S., P. Malarvezhi, R. Dayana, and K. Vadivukkarasi. "A Comprehensive Survey on Autonomous Driving Cars: A Perspective View." *Wireless Personal Communications* 114, no. 3 (2020): 2121–2133.
- DeVito, Michael A., Darren Gergle, and Jeremy Birnholtz. "Algorithms Ruin Everything: #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3163–3174. Denver, CO: ACM, 2017.
- Domingos, Pedro. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York: Basic Books, 2015.
- Elliott, Anthony. *Making Sense of AI: Our Algorithmic World*. Cambridge; Medford: Polity Press, 2022.
- Eslami, Motahhare, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. "First I 'Like' It, Then I Hide It: Folk Theories of Social Feeds." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2371–2382. San Jose, CA: ACM, 2016.
- Eyal, Nir. *Hooked: How to Build Habit-Forming Products*. New York: Portfolio, Penguin, 2014.
- Gillespie, Tarleton. "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society (Inside Technology)*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, 167–193. Cambridge, MA: MIT Press, 2014.
- Gillespie, Tarleton. "#trendingistrending: When Algorithms Become Culture." In *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies*, edited by Robert Seyfert and Jonathan Roberge, 52–75. London, New York: Routledge, 2016.

- Goriunova, Olga. "The Digital Subject: People as Data as Persons." *Theory, Culture & Society* 36, no. 6 (2019): 125–145.
- Grusin, Richard A. *Premediation: Affect and Mediality after 9/11*. Basingstoke, New York: Palgrave Macmillan, 2010.
- Gurevich, Yuri. "Sequential Abstract-State Machines Capture Sequential Algorithms." *ACM Transactions on Computational Logic* 1, no. 1 (2000): 77–111.
- Hajian, Sara, Francesco Bonchi, and Carlos Castillo. "Algorithmic Bias: From Discrimination Discovery to Fairness-Aware Data Mining." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2125–2126. San Francisco, CA: ACM, 2016.
- Harari, Yuval Noah. *Homo Deus: A Brief History of Tomorrow*. London: Vintage, 2017.
- Häußling, Roger. *Techniksoziologie: eine Einführung*. Opladen, Toronto: Barbara Budrich, 2019.
- Häußling, Roger. "Daten als Schnittstellen zwischen algorithmischen und sozialen Prozessen. Konzeptuelle Überlegungen zu einer Relationalen Techniksoziologie der Datafizierung in der digitalen Sphäre." In *Soziologie des Digitalen – Digitale Soziologie? Soziale Welt – special issue*, vol. 23, edited by Sabine Maasen, Jan-Hendrik Passoth, 134–150. Baden-Baden: Nomos, 2020.
- Healy, Kieran. "The Performativity of Networks." *European Journal of Sociology* 56, no. 2 (2015): 175–205.
- Hedrick, Brandon P., J. Mason Heberling, Emily K. Meineke, Kathryn G. Turner, Christopher J. Grassa, Daniel S. Park, and Jonathan Kennedy. "Digitization and the Future of Natural History Collections." *BioScience* 70 (2020): 243–251.
- Hegel, Georg Wilhelm Friedrich. *The Phenomenology of Spirit*, edited and translated by Terry P. Pinkard. New York: Cambridge University Press, 2018.
- Hind, Sam. "Digital Navigation and the Driving-Machine: Supervision, Calculation, Optimization, and Recognition." *Mobilities* 14, no. 4 (2019): 401–417.
- Hofstadter, Douglas R. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books, 1979.
- Jasanoff, Sheila. "The Idiom of Co-Production." In *States of Knowledge. The Co-Production of Science and Social Order*, edited by Sheila Jasanoff, 1–12. London, New York: Routledge, 2004.
- Jenkins, Henry, Katie Clinton, Ravi Purushotma, Alice J. Robison, and Margaret Weigel. *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. Chicago, Illinois: MacArthur Foundation, 2006. Accessed on 21 October 2023. https://www.macfound.org/media/article_pdfs/jenkins_white_paper.pdf.
- Kalpokas, Ignas. *Algorithmic Governance: Politics and Law in the Post-Human Era*. Cham: Palgrave Macmillan Limited, 2019.
- Karafilidis, Athanasios and Robert Weidner, eds. *Developing Support Technologies: Integrating Multiple Perspectives to Create Assistance that People Really Want*. Cham: Springer, 2018.
- Knorr Cetina, Karin. "Sociality with Objects: Social Relations in Postsocial Knowledge Societies." *Theory, Culture & Society* 14, no. 1 (1997): 1–30.
- Knorr Cetina, Karin. "Presentation to Panel: Theorizing Numbers." In *American Sociological Association Annual Meeting*. New York, 2013.
- Kwek, Dorothy H.B. and Robert Seyfert. "Affect Matters: Strolling Through Heterological Ecologies." *Public Culture* 30 (2018): 35–59.
- Lange, Ann-Christina, Marc Lenglet, and Robert Seyfert. "On Studying Algorithms Ethnographically: Making Sense of Objects of Ignorance." *Organization* 26 (2019): 598–617.
- Lash, Scott. "Dialectic of Information? A Response to Taylor." *Information, Communication & Society* 9 (2006): 572–581.
- Latour, Bruno. "On Interobjectivity." *Mind, Culture, and Activity* 3, no. 4 (1996): 228–245.
- Latour, Bruno. "The Berlin Key or How to Do Words with Things." In *Matter, Materiality, and Modern Culture*, edited by Paul Graves-Brown, 10–21. London: Routledge, 2000.

- Latour, Bruno, Pablo Jensen, Tommaso Venturini, Sébastien Grauwin, and Dominique Boullier. "The Whole Is Always Smaller than Its Parts': A Digital Test of Gabriel Tarde's Monads." *The British Journal of Sociology* 63, no. 4 (2012): 590–615.
- Lemke, Thomas. "New Materialisms: Foucault and the 'Government of Things.'" *Theory, Culture & Society* 32, no. 4 (2015): 3–25.
- Lindzon, Jared. "The Internet of Pets: How Technology Can Help Your Dog Lose Weight." <https://www.theguardian.com/lifeandstyle/2015/jun/20/internet-of-pets-technology-track-dog-fit>. *The Guardian*. 20.06.2015. Accessed on 21 October 2023.
- Lischka, Konrad and Christian Stöcker. *Digitale Öffentlichkeit: Wie algorithmische Prozesse den gesellschaftlichen Diskurs beeinflussen: Arbeitspapier*. Impuls Algorithmenethik. Bertelsmann Stiftung, 2017. <https://doi.org/10.11586/2017028>
- Lupton, Deborah. *Digital Sociology*. Abingdon, London, New York: Routledge, 2015.
- Maasen, Sabine and Jan-Hendrik Passoth, eds. "Soziologie des Digitalen – Digitale Soziologie?" *Soziale Welt – special issue* 23 (2020): 9–16.
- Mackenzie, Adrian. *Machine Learners: Archaeology of a Data Practice*. Cambridge: MIT Press, 2017.
- Mackenzie, Adrian. "Personalization and Probabilities: Impersonal Propensities in Online Grocery Shopping." *Big Data & Society* 5 (2018). <https://doi.org/10.1177/2053951718778310>
- MacKenzie, Donald. "Is Economics Performative? Option Theory and the Construction of Derivatives Markets." *Journal of the History of Economic Thought* 28, no. 1 (2006): 29–55.
- MacKenzie, Donald. "How Algorithms Interact: Goffman's 'Interaction Order' in Automated Trading." *Theory, Culture & Society* 36, no. 2 (2019): 39–59.
- MacKenzie, Donald. *Trading at the Speed of Light - How Ultrafast Algorithms Are Transforming Financial Markets*. Princeton: Princeton University Press, 2021.
- MacKenzie, Donald and Yuval Millo. "Constructing a Market, Performing Theory: The Historical Sociology of a Financial Derivatives Exchange." *American Journal of Sociology* 109, no. 1 (2003): 107–145.
- Manovich, Lev. "Database as Symbolic Form." *Convergence: The International Journal of Research into New Media Technologies* 5 (1999): 80–99.
- Marres, Noortje. *Digital Sociology: The Reinvention of Social Research*. Cambridge, Malden: Polity Press, 2017.
- Miyazaki, Shintaro. "Algorithmic Ecosystems: Neoliberal Couplings and Their Pathogenesis 1960–Present." In *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies*, edited by Robert Seyfert and Jonathan Roberge, 128–139. London, New York: Routledge, 2016.
- Mol, Annemarie. *The Body Multiple: Ontology in Medical Practice*. Durham, London: Duke University Press, 2002.
- Moschovakis, Yiannis N. "What Is an Algorithm?" In *Mathematics Unlimited—2001 and Beyond*, edited by Björn Engquist and Wilfried Schmid, 919–936. New York: Springer, 2001.
- Nassehi, Armin. *Patterns: Theory of the Digital Society*. Medford: Polity Press, 2024.
- Negroponte, Nicholas. *Beyond Digital*. Wired. 1998. Accessed on 21 October 2023. <https://www.wired.com/1998/12/negroponte-55/>.
- Nosthoff, Anna-Verena and Felix Maschewski. *Die Gesellschaft der Wearables: Digitale Verführung und soziale Kontrolle*. Berlin: Nicolai, 2019.
- Pardo-Guerra, Juan Pablo. *Automating Finance: Infrastructures, Engineers, and the Making of Electronic Markets*. Cambridge: Cambridge University Press, 2019.
- Payne, Robert. "Virality 2.0: Networked Promiscuity and the Sharing Subject." *Cultural Studies* 27 (2013): 540–560.
- Pentenrieder, Annelie. *Algorithmen im Alltag. Eine praxistheoretische Studie zum informierten Umgang mit Routenplanern*. Frankfurt a. M., New York: Campus, 2020.

- Pessach, Dana and Erez Shmueli. *Algorithmic Fairness*. 2020. arXiv:2001.09784.
- Pfeiffer, Sabine. *Digitalisierung als Distributivkraft: Über das Neue am digitalen Kapitalismus. X-Texte zu Kultur und Gesellschaft*. Bielefeld: Transcript, 2021.
- Rammert, Werner. *Technik – Handeln – Wissen: zu einer pragmatistischen Technik- und Sozialtheorie*. Wiesbaden: Springer VS, 2016.
- Rammert, Werner and Ingo Schulz-Schaeffer. “Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technische Abläufe verteilt.” In *Technik – Handeln – Wissen: zu einer pragmatistischen Technik- und Sozialtheorie*, 121–167. Wiesbaden: Springer VS, 2016.
- Reckwitz, Andreas. *Society of Singularities*. Cambridge; Medford: Polity, 2020.
- Ritzer, George and Nathan Jurgenson. “Production, Consumption, Prosumption: The Nature of Capitalism in the Age of the Digital ‘Prosumer’.” *Journal of Consumer Culture* 10, no. 1 (2010): 13–36.
- Roberge, Jonathan and Robert Seyfert. “What Are Algorithmic Cultures?” In *Algorithmic Cultures Essays on Meaning, Performance and New Technologies*, 1–25. London, New York: Routledge, 2016.
- Sari, Astrid Yuza and Okki Trinanda. “Instagramable Tourism Attraction: Natural or Artificial?” In *Proceedings of Tourism Development Centre International Conference*, 62–72. Sciendo: Andalas, 2020.
- Selwyn, Neil. *What Is Digital Sociology?* Cambridge, Medford: Polity, 2019.
- Seyfert, Robert. “Bugs, Predations or Manipulations? Incompatible Epistemic Regimes of High-Frequency Trading.” *Economy and Society* 45 (2016): 251–277.
- Seyfert, Robert. “Automation and Affect: A Study of Algorithmic Trading.” In *Affect in Relation: Families, Places, Technologies*, edited by Röttger-Rössler Birgitt and Slaby Jan, 197–218. London, New York: Routledge, 2018.
- Seyfert, Robert. *Beziehungsweisen: Elemente einer relationalen Soziologie*. Weilerswist: Velbrück Wissenschaft, 2019.
- Seyfert, Robert. “Algorithms as Regulatory Objects.” *Information, Communication & Society* 25 (2022): 1542–1558.
- Seyfert, Robert and Jonathan Roberge, eds. *Algorithmic Cultures. Essays on Meaning, Performance and New Technologies*. London, New York: Routledge, 2016.
- Simmel, Georg. “How Is Society Possible?” *American Journal of Sociology* 16, no. 3 (1910): 372–391.
- Simondon, Gilbert. *L’individu et sa genèse physico-biologique*. Nouv. éd. Collection Krisis. Grenoble: J. Millon, 1995.
- Soare, Robert I. “Computability and Enumerability.” In *Logic and Scientific Methods Volume Ten of the Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence, August 1995*, edited by Maria Luisa Dalla Chiara, Kees Doets, Daniele Mundici, and Johan van Benthem, 221–237. Dordrecht: Springer, 1997.
- Spinoza, Benedictus de. “The Ethics. Demonstrated in Geometric Order.” In *A Spinoza Reader. The Ethics and Other Works*, translated by Edwin Curley, 85–265. Princeton: Princeton University Press, 1994.
- Star, Susan Leigh and Karen Ruhleder. “Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces.” *Information Systems Research* 7 (1996): 111–134.
- Steiner, Christopher. *Automate This: How Algorithms Came to Rule Our World*. New York: Portfolio, Penguin, 2012.
- Suchman, Lucy A. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press, 2006.
- Swedberg, Richard. *The Art of Social Theory*. Princeton: Princeton University Press, 2014.
- Tal, Avital Shulner, Khuyagbaatar Batsuren, Veronika Bogina, Fausto Giunchiglia, Alan Hartman, Styliani Kleanthous Loizou, Tsvi Kuflik, and Jahna Otterbacher. “‘End to End’ Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic

- Systems.” In *14th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, 1–6. Larnaca, Cyprus: IEEE, 2019. <https://doi.org/10.1109/SMAP.2019.8864914>
- Tanenbaum, Andrew S. and Todd Austin. *Structured Computer Organization*. Boston: Pearson, 2013.
- Totaro, P. and D. Ninno. “The Concept of Algorithm as an Interpretative Key of Modern Rationality.” *Theory, Culture & Society* 31 (2014): 29–49.
- Totaro, P. and D. Ninno. “Biological Recursion and Digital Systems: Conceptual Tools for Analysing Man-Machine Interaction.” *Theory, Culture & Society* 37 (2020): 27–49.
- Vahid, Frank. “The Softening of Hardware.” *Computer* 36, no. 4 (2003): 27–34.
- Williams, Betsy Anne, Catherine F. Brooks, and Yotam Shmargad. “How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications.” *Journal of Information Policy* 8 (2018): 78–115.
- Wirth, Niklaus. *Algorithmen und Datenstrukturen mit Modula*. 2nd ed. Stuttgart: Teubner, 1996.
- Wolfert, Sjaak, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt. “Big Data in Smart Farming: A Review.” *Agricultural Systems* 153 (2017): 69–80.
- Woolgar, Steve. “Configuring the User: the Case of Usability Trials.” In *A Sociology of Monsters: Essays on Power, Technology and Domination*, edited by John Law, 57–99. London, New York: Routledge, 1991.
- Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.

2 Theorising Artificial Intelligence as Practice

Kamil Filipek

Introduction

The advent of a “theory of practice” paradigm (Swidler 1986; Schatzki 1996, 2002, 2005, 2006; Wenger 1998; Reckwitz 2002) provides a fresh way to look at how to problematise algorithms and artificial intelligence (AI) in today’s science. This new approach goes beyond the usual ways of defining AI, e.g. computer programmes, complex neural networks, human-like robots, or systems that can do tasks like a human. Instead, this theory focuses on a plethora of phenomena like what people know, what they value, how they act, and how they organise themselves. It also considers how meaning, sense, understanding, and power are shared and how culture affects all these things or is shaped by them. This perspective aligns closely with the emerging trends in contemporary sociology (Gillespie 2014; Elliott 2019; Burrell & Fourcade 2021; Airoidi 2022), which underscore that AI transcends mere technological innovation; it is shaped and utilised by people and it serves various purposes in society. Thus, in this view, what really matters is not just what we formally call AI, but rather how people actually use and shape it, how they practice AI.

This chapter draws partial inspiration from Nick Couldry’s seminal work “Theorizing Media as Practice” (2004) to propose a theory of practice aimed at probing the intricate nexus between technological innovation and societal structures. Specifically, the focus is on the multifaceted challenges engendered by AI’s current phase of development. In this interpretive schema, AI is conceptualised not as a monolithic, inelastic construct derived from abstract principles but rather as a dynamically evolving, negotiable, and temporally contingent socio-technical entity. AI is thus viewed as a relational phenomenon, emergent from an interplay of both social and non-social activities underpinning diverse practice frameworks. Similarly, Massimo Airoidi’s (2022, p. 35) analysis of algorithms, with a focus on AI, is deeply rooted in theoretical exploration. Leveraging Pierre Bourdieu’s conceptual framework (e.g. 1977), the author of “Machine Habitus” embeds the development and application of AI within a wide-ranging societal backdrop. As precisely noted, “All machine learning systems are socialised, based on data contexts which bear the imprint of socially located cultural realities.” Consequently, AI is reframed as a dynamic, mutable construct whose characteristics are shaped by the social practices from which it emerges and within which it operates.

The chapter is structured as follows. Initially, it explains some issues with defining AI in contemporary science. It then reveals the essence of the concept of “practice,” mainly based on Schatzki’s theory. In the subsequent part, the chapter highlights the critical elements of social practice identifiable in AI analysis. Finally, it briefly introduces the concept of “community of practice,” which possesses significant explanatory potential for contemporary phenomena associated with AI.

Problems with Defining AI

Over the past several decades, the field of AI has seen significant evolution in both its technical capabilities and in the area of social applications (Schwartz 1989, Elliott 2022). This progress has been accompanied by numerous conceptual and definitional errors that have far-reaching implications for understanding, implementation, and regulation of AI. On one hand, there was an assumption that the intricacies of the human brain and its cognitive processes could be emulated through increasingly advanced machines and mathematical models of neural operations (e.g. McCulloch & Pitts 1943, Fu 1994). Even though we have AI models today that are larger than the structure of the human brain (larger than 86 billion neurons), they capture only a fraction of the brain’s actual complexity. Another mistake stemmed from the naive scientific approaches of the past, which narrowed the definition of intelligence to specific tasks that humans perform – tasks that machines can often replicate just as effectively (e.g. Turing 1950). Reducing a task to a sequence of small, formally describable actions overlooks the multifaceted nature of human intelligence, which is capable of performing a wide array of often unrelated tasks. This excessive anthropomorphism of AI seems to be a reactionary response to the vast knowledge gap in understanding of intelligence, the human brain, and cognitive processes. Alarming, this gap is further widened when intelligence is conceptualised more broadly to encompass cognitive, social, emotional, problem-solving, decision-making, creative, and adaptive processes, among others.

An alternative line of inquiry posits that AI embodies a distinct form of non-human intelligence, fundamentally anchored in algorithmic processes, computational capabilities, and data-centric learning mechanisms (e.g. Kurzweil 2005). The enigmatic complexity of the human brain has led some scholars to contend that human cognitive processes may not always operate at optimal efficiency and may, in certain instances, manifest redundancy. Unlike human intelligence, which evolves through experiential learning and is imbued with consciousness, AI’s competencies are honed through the calibration of large, high-quality datasets. This algorithmically driven intelligence enables AI to manage and scrutinise voluminous data sets at velocities unattainable by human cognition, thereby demonstrating exceptional aptitude in pattern recognition, predictive analytics, and problem-solving tasks. Another distinguishing characteristic of AI’s non-human intelligence is its capacity for high-precision, consistent task execution, devoid of fatigue, or emotional partiality. Nevertheless, AI’s non-human intelligence is not without limitations, notably its lack of affective understanding, intuitive reasoning, and contextual awareness – key facets of human cognitive function.

Challenging both approaches as inadequate frameworks for comprehending the current state of AI, I concur with Kate Crawford's (2021) argument that the term "artificial intelligence" is a misnomer. She astutely emphasises that AI is contingent upon broader socio-political infrastructures, rendering it simultaneously a social and technological phenomenon. In this context, AI neither successfully replicates human-like attributes nor aligns perfectly with the multifaceted capabilities intrinsic to human intelligence. Thus, the nomenclature of "intelligence" seems ill-suited for describing this continually evolving technological construct which is excessively fetishised by most scientific communities around the world. The term "intelligence" itself is often ill-defined when applied to machines. In human psychology and sociology, intelligence is a multi-dimensional construct encompassing not merely problem-solving capabilities but also emotional intelligence, creativity, and social acumen (Gardner 1983). The majority of extant AI systems are specialised for narrow domains and lack the versatility that characterises human intelligence. This underscores the inadequacy of the term "intelligence" in describing AI, affirming that it is neither fully "artificial" nor comprehensively "intelligent." Building upon the aforementioned points, it is crucial to acknowledge that the operational parameters of AI are fundamentally constrained by the limitations of their training data and underlying algorithms. Unlike human intelligence, which possesses the ability to adapt, self-modify, and even question its own foundational beliefs, most AI systems operate within a predetermined scope and are not equipped to engage in meta-cognitive processes or ethical reasoning. Therefore, the term "intelligence" as applied to current AI systems might more appropriately be qualified with descriptors such as "narrow," "domain-specific," or "task-oriented," to differentiate it from the more comprehensive, adaptive, and nuanced form of human intelligence.

To transcend the aforementioned problems, this chapter introduces an alternative conceptualisation of AI, framing it as a socio-technological construct. Such an approach is consistent with the postulate of Woolgar (1985), who viewed science as a social activity, as well as the proposal of Bloomfield (1988), who believed that the behaviours of intelligent machines, like human behaviours, should constitute an area of sociological research. In this paradigm, AI is not an isolated technological entity but is interwoven with social actions comprising practices, emerging from a specific set of technological and axiological contexts. Such an approach underscores the reciprocal relationship between AI and society, where social norms, values, rules, and structures influence the development and application of AI technologies. Consequently, AI should be understood as a socially dependent, i.e. constructed, interpreted, utilised, and evaluated entity, the characteristics of which are shaped by the social practices from which it emanates.

What Is Practice?

Although there is no single theory of social practice, the contributions of Theodor Schatzki (1996, 2002, 2005, 2006), Andreas Reckwitz (2002), and Ann Swidler (1986) – major figures of the so-called the practice turn in contemporary theory – will provide the theoretical framework for consideration later in this chapter.

Briefly speaking, the “practice turn” refers to a shift in social theory that focuses on understanding the social world through practices, routines, and actions, rather than solely through abstract concepts or structures. Schatzki defines practice as “a set doings and sayings” (2002, p. 73) with “organized, open-ended spatial-temporal manifolds of actions” (2005, p. 471). Activities are a major constituent of every practice, from cooking practices, car-parking practices, to practices of creation, and use of information technology. According to Schatzki, the set of activities that make up a practice are linked by three phenomena: (i) understandings of how to do things, (ii) rules, and (iii) teleoaffective structure (2005, p. 471). Such conceptualisation holds promise for understanding the essence of AI within broader societal contexts.

Reckwitz focused on other dimensions of practices that may be useful in analysing AI. For him practice is a “routinized way in which bodies are moved, objects are handled, subjects are treated, things are described and the world is understood” (2002, p. 250). He raises the importance of the temporal and social dimensions of each practice. “A practice is social, as it is a ‘type’ of behaving and understanding that appears at different locales and at different points of time and is carried out by different body/minds” (ibidem). Thus, every practice is constructed in the here and now, by actors we are usually able to identify. This means that practices are not static, but change/evolve over time with bodies/minds and their understanding of the surrounding world. The dynamics of practice appear to be particularly important in the context of explaining the nature of AI. The essence of AI, its definition and interpretation, is constantly changing with the actors who are involved in the practices of its production and use.

Finally, Swidler’s contribution extends the analytical framework by introducing the concept of hierarchical structuring of practices. According to Swidler, practices are embedded in a “cultural repertoire” that includes skills, habits, and styles, and these repertoires in turn influence how practices are organised and prioritised (Swidler 1986). This hierarchical arrangement allows for the categorisation of practices into different levels of complexity and significance, ranging from mundane daily routines to more specialised, institutionally governed practices. In the context of AI, Swidler’s notion of hierarchical practices invites us to examine how AI-related activities are organised within broader socio-technical systems. For instance, certain AI practices might be considered foundational, such as data collection and algorithmic development, while others may be more peripheral but nonetheless impactful, such as user-interface design or ethical consideration in AI deployment. This hierarchical view also allows us to interrogate the value systems and power dynamics that influence which practices are prioritised or marginalised within AI research and development. Furthermore, Swidler’s concept of cultural repertoires opens the door to exploring how historical and cultural contexts shape the practices surrounding AI. It prompts us to consider how different cultural norms, ethical frameworks, and historical experiences might influence the design, implementation, and perception of AI technologies across different societies. This multi-layered understanding enriches our perspective on

AI, highlighting that it is not merely a technological artifact but also a complex social construct influenced by a web of practices that are themselves shaped by cultural, temporal, and institutional factors.

Thus, the contributions of Schatzki, Reckwitz, and Swidler collectively offer a robust theoretical framework for examining AI from the vantage point of social practice theory. Their insights facilitate a nuanced understanding of AI, considering it as an evolving social phenomenon that is intricately linked to an array of practices, influenced by temporal and spatial dimensions, and subject to hierarchical organisation based on cultural repertoires and societal norms.

Technology as Social Practice

Incorporating the contributions of Schatzki, Reckwitz, and Swidler into the understanding of technology as a social practice provides a more nuanced, theoretically grounded framework. In this modified perspective, technology – such as AI – becomes an integral component of a complex set of social practices, each with its own manifold actions. These practices are shaped by and, in turn, shape societal norms, habits, and shared meanings, confirming Reckwitz's focus on the routinised ways in which objects are handled and the world is understood (2002, p. 250).

It is worthwhile to consider the ubiquitous use of smartphones, which have become deeply embedded in daily routines and social interactions. This technology is not merely a tool but forms part of a hierarchical set of practices, in line with Swidler's conceptualisation. These practices range from foundational activities like communication and information retrieval to more peripheral yet impactful practices, such as social media engagement and online shopping. Each of these practices possesses its own temporal and social dimensions, aligning with Reckwitz's emphasis on the dynamics of practice. In this framework, technology serves as a critical mediating agent in the conduct and dynamics of practices, as well as in the relational interconnections that exist among them. While this notion is congruent with Actor-Network Theory, which argues for the integral role of both human and non-human actors (such as technology) in the formation and perpetuation of social networks (Latour 2007), the focus of this analysis remains on the theory of social practice.

By applying the theory of social practice in this manner, we recognise the reciprocal relationship between technological artefacts like smartphones or AI systems and the social contexts within which they operate. It reaffirms that these technologies are not isolated entities but are interwoven with a broad array of activities and understandings that collectively form the social fabric. Therefore, understanding technology through the lens of social practice theory enables a more comprehensive exploration of its role, not just as a set of engineered functionalities but as a dynamic, mutable construct that evolves in tandem with the social practices from which it emerges and within which it functions. Importantly, such perspective illuminates how technology is both a product and a producer of social realities, entangled in a complex web of activities, norms, and values.

Breaking AI Down into Practical Pieces

In this framework, AI is not merely a collection of algorithms and data sets but is fundamentally entwined with human activity – both as an enabler and as a constraint. It's important to note that there is no single or universal context in which these dynamics occur; the practices surrounding AI are often contingent upon specific social, cultural, and technological environments. However, for the sake of generalisability and to illustrate the utility of practice theory, AI can be explored as emerging from and being embedded within a set of “organized, open-ended spatial-temporal manifolds of actions,” as Schatzki would phrase it (2005, p. 471). Hence, any comprehensive analysis of AI necessitates an exploration of the practices surrounding its development, deployment, and usage within specific – but not necessarily unique – social contexts.

Shared Meanings

A good example might be an AI team working at a tech company focused on developing AI solutions for healthcare. All members of the team engage in a range of tasks that reflect different aspects of AI development and deployment. The team shares a common technical language and understanding of key concepts like algorithms, machine learning models, deep learning, and data analytics. These shared meanings allow for efficient communication and collaboration, thereby enhancing the effectiveness of their AI solutions in healthcare. Beyond technical jargon, the team also shares a common understanding of some ethical principles, such as fairness, transparency, and social responsibility. Shared ethical framework guides their practices, especially when dealing with critical areas like data privacy or algorithmic bias. The collective understanding helps in aligning the team's efforts with broader social values and norms, even if this social context is not fully realised by the members of the team.

Within the team's organisational culture, certain practices and goals acquire specific meanings that might be unique to this group. For example, the focus on healthcare solutions might imbue their work with a sense of urgency and social importance, guiding their priorities and decision-making processes. But the team's work does not exist in a vacuum; it interacts with a broader relational environment that includes healthcare providers, patients, policymakers, and the public. Shared meaning is crucial in these external interactions as well, ensuring that the AI solutions are aligned with the expectations and needs of these diverse stakeholders. It's worth adding here that the shared meaning will vary greatly depending on the composition of the team, the country in which the team operates, and the dominant culture that influences social practices. Despite these differences, it is the shared meaning that acts as a kind of glue which binds the seemingly chaotic actions of individual team members into coherent AI practices, which, in the next step, may lead to the creation of a specific market product.

As the team navigates through complex challenges and changing landscapes, the shared meaning is subject to evolution. New experiences, technological

advancements, and social feedback can redefine what certain practices and outcomes mean to the team, requiring a continual realignment of shared understandings. It is also worth adding that the shared meaning regulating current practices forming around AI is highly dependent on the legal regulations in force at any given time.

Hierarchical Organisation and Cultural Repertoires

It is worthwhile to now divide the team's work into the following tasks. In the first step, the team tries to translate the overall business goal into technical steps that will allow achieving the intended effect. This is about choosing the appropriate work methodology (e.g. Agile, Scrum, Feature-Driven Development), which – it is worth emphasising – is often in itself a stream of actions that emerges around technological projects. The methodology will impose certain rules that will shape the temporal and teleological frameworks, as well as sanctions associated with specific actions. Then, the team focuses on foundational elements: fine-tuning algorithms and scrutinising data analytics to improve the performance of an intelligent learning model designed to predict patient outcomes. These activities are considered foundational because they directly impact the core functionality of the AI system. Various members, such as data engineers and data analysts, collaborate to ensure the model is trained on diverse and representative data sets, aiming for equitable and accurate predictions. The goals pursued by an AI model should align with the so-called anchoring practices (Swidler 2005) deeply embedded in everyday life, providing a stable framework for individuals to navigate their lives. It can be assumed that the essence of health “anchoring practices” in most societies (in settled or unsettled times) (Swidler 1986) is to support and help others. Being a recipient of many “anchoring practices” present in the field of broadly understood healthcare of my cultural circle, it is hard for me to imagine an innovation based on AI that would aim to harm people.

In the next stage, the team shifts to what might be termed intermediate practices, aligning with Swidler's notion of “cultural repertoires” that influence practices (1986). User experience designers do their best to ensure the AI interface is intuitive for healthcare providers. Although this aspect is not as foundational as the algorithm itself, it is crucial for effective real-world application. A seemingly insignificant element, however, becomes an extremely important intermediate practice, which has a huge impact on the entire project (a set of practices). Additionally, team members engaged in this stage may consult with the legal team to ensure the AI solution is compliant with healthcare regulations. It is critical to recognise how these practices are not just procedurally necessary but also culturally and contextually shaped. The designers' and legal experts' approaches are influenced by their professional cultures, regulatory landscapes, and the prevailing standards in healthcare technology. This integration of cultural repertoires into AI development reflects an ongoing negotiation between technical possibilities and the cultural and regulatory frameworks guiding their application. As such, these intermediate or side practices serve as a crucial bridge, not only aligning AI development with

technical specifications but also ensuring its resonance with broader social, ethical, and regulatory norms.

Temporal, Material, and Social Dimensions

Incorporating Reckwitz's temporal and social dimensions into the story of the AI team brings additional layers of complexity and dynamism to their work. Over time, the team's focus may shift in response to evolving healthcare needs, regulatory landscapes, and technological advancements. For instance, during a public health crisis, the team might pivot to accelerate the development of AI models that can help manage the crisis, thereby changing their foundational practices. The COVID-19 pandemic has turned upside down the work of many research teams developing AI in the broadly understood area of healthcare (Devyania et al. 2020). Another factor emphasising the temporary dimension of AI practices is the law in force at any given time, regulating the actions of technological teams. As new data privacy regulations come into play, the team's peripheral practices related to ethical considerations may suddenly become more foundational. They can stabilise and turn into core practices that will, however, be interconnected with other practices. This example shows how top-down imposed rules affect the temporal dimension of practices developed around AI.

The team operates within a specific material and social context that includes the set of concrete tools (e.g. material objects, technologies), the broader healthcare infrastructure, and the socio-cultural layer at large. This context, as highlighted by Reckwitz's, shapes the team's practices and interactions. The material dimension, encompassing the latest AI technologies and medical equipment, directly influences the team's approach to problem-solving and innovation. Their work is not just about pure coding and algorithm's design; it's deeply integrated with the tangible aspects of healthcare, from diagnostic machines to data management systems. Simultaneously, the social dimension – comprising institutional norms, patient care ethics, and collaborative dynamics within the team and with other healthcare professionals – guides the direction and application of their AI solutions. This interplay between the material and social elements creates a dynamic workspace where practices are continually adjusted and refined in response to changing medical needs, regulatory standards, and social expectations. In this way, the AI team's work is a vivid example of Reckwitz's conception of practices as embedded in and shaped by temporal, material, and social contexts, illustrating how these dimensions are critical in developing AI solutions that are not only technologically advanced but also socially attuned and ethically grounded.

By integrating the dimensions introduced by Reckwitz, we gain a better understanding of the AI team's activities. The team's practices are not static but evolve in response to temporal events, material settings, and social feedback loops. This aligns with Reckwitz's emphasis on the temporal and social dimensions of each practice, where "a practice is social, as it is a 'type' of behaving and understanding that appears at different locales and at different points of time and is carried out by different body/minds" (Reckwitz 2002, p. 250). Therefore, the work of the AI team

in healthcare is not just a set of isolated tasks but a dynamic interplay of activities that evolve over time and are deeply embedded in social contexts. Just as the practices of the team change, so too does the understanding and meaning of AI, as well as the impact of models on people's lives.

Activities and Phenomena Linking Practices

Schatzki posits that the activities making up a practice are connected by three phenomena: “understandings of how to do things, rules, and teleoaffective structures” (Schatzki 2005, p. 471). In the foundational stage, the AI team's understanding of how to fine-tune algorithms and scrutinise data analytics shapes their activities. This collective understanding is also very dynamic and it evolves as the team gains more experience and as the field of AI progresses. For example, a new breakthrough in machine learning techniques could lead to a shared understanding that they should adopt these techniques to improve their predictive model. Even if understanding these new techniques is not easy, there might be a shared belief, a consciousness of shared interests that their implementation can contribute to the team's efficiency. The most important process at this stage is the sharing of meanings, making them recognisable to all members of the team. However, in the era of information overproduction, this process is not always smooth, as Marody (2021, p. 82) emphasises, “at the basis of that generic experience of modernity which is living in a mass of information, we find not so much its quantitative growth, but the violation of the ordering, nomic structure of the social reality, which entails the dissolution of the images of the world that we take for granted and that give meaning to our aims and in consequence – to our lives.”

The team operates under a set of formal and informal rules. Formal rules may include company policies on data security or healthcare regulations that their AI model must comply with. Such formal rules can be found in laws, resolutions, guidelines, instructions, requirements, etc., regulating the behaviour of employees in a given organisation. Reconstruction of rules based on formal sources regulating social actions seems to be important from the perspective of practices. But the behaviours of the members of the project team are also regulated by informal rules, which we will not find in official documents. “These ‘informal’ matters are not part of the formal (rational rule-like) structure of the organisation. Nor are they values and meanings that develop beyond the formal structure” – as Schatzki (2005, p. 477) observes. This refers to the so-called rules of thumb, generally accepted solutions, and often unconsciously recognised signs, which are shaping people's goals and actions, constituting broadly understood practices. As a result, the path leading to a specific solution based on AI methods becomes very winding, sometimes even bumpy, entangled in a variable network of interconnected formal and informal rules.

Finally, teleoaffective structures refer to the ends, goals, tasks, projects, or purposes that are inherent in a practice (Schatzki 1996, p.100). In short, the teleological aspect provides direction and intentionality to the practice, guiding what actions are performed and why. However, the teleoaffective structure also includes the range of emotions, moods, and affective orientations that are appropriate or

expected within the context of the practice. For instance, the team's motivation to improve patient outcomes through AI could be a significant teleoffective structure guiding their activities. Similarly, social expectations around responsible AI use in healthcare could influence their focus on ethical considerations. However, as Schatzki emphasises (1996, p. 124), elements of the teleoffective order may turn out to be irrelevant at the stage of analysing a specific practice. In the case of our team, the routine of employees consuming coffee before commencing coding bears minimal relevance to the overarching goal of developing the AI solution.

The interconnectedness of understandings, rules, and teleoffective structures provides a coherent yet flexible network entity that allows the team to navigate the complexities of AI development in healthcare. This perspective is crucial for understanding not just what activities the team engages in but also why they engage in them and how they are all interconnected. As a result, the path leading to a specific solution based on AI methods becomes very winding, sometimes even bumpy. However, the theory of practices allows us to see the complexity of the processes that determine the shape of AI, which are not always visible when we analyse AI as a purely technical phenomenon.

Community of Practice

One more theoretical approach merits attention when applying practice theory to the analysis of AI. The concept in question is “Communities of Practice,” initially developed by cognitive anthropologists Jean Lave and Etienne Wenger in their 1991 book “*Situated Learning: Legitimate Peripheral Participation*.” Wenger later expanded on this theory in his 1998 book “*Communities of Practice: Learning, Meaning, and Identity*.” Communities of Practice are groups of people who share a common interest, profession, or craft and learn from one another as they engage in shared activities. Over time, these communities develop their own shared practices, language, and tools, effectively creating a social structure that facilitates learning and knowledge sharing. This concept has been influential in various fields such as education, organisational studies, and knowledge management, and it could serve as another valuable dimension when applying practice theory to the analysis of AI. Such an outlined approach directs our attention to at least four dimensions that are worth considering in the analysis of AI as a practice: (i) domain of practice, (ii) membership and identity, (iii) knowledge transfer, and (iv) power dynamics.

Domain of Practice

The AI team, which has already been identified as specialising in healthcare solutions, epitomises a “community of practice” in the nuanced and complex field of healthcare AI. In this specialised domain, the team not only engages in a collective learning process but also contributes to the advancement of the field by generating new knowledge and best practices (e.g. by leaving the code in view of others in open repositories). Their primary focus is on developing intelligent learning models designed to predict patient outcomes, a task that requires a synthesis of technical

expertise, domain-specific knowledge, and ethical considerations. The team's collective endeavour in predictive modelling serves as the nucleus of their practice, setting the stage for ongoing learning and knowledge sharing. In this context, the team members continuously engage with both internal and external resources, such as academic literature, healthcare professionals, and patient data, to refine their models. This enables them to stay abreast of the latest advancements in both healthcare and AI technology, thereby ensuring that their work is at the forefront of the domain.

Moreover, their specialised focus on patient outcomes provides a common goal that unifies the team's diverse skill sets, fostering a sense of shared purpose and identity. This commonality in objective facilitates effective communication and collaboration, elements that are essential for the successful functioning of any community of practice. It also creates a fertile ground for innovation, as team members feel invested in the outcomes and are more likely to contribute novel solutions to complex problems.

Membership and Identity

The AI team is a diverse assemblage of professionals, including data engineers, data analysts, user experience designers, and legal experts. Each of these roles brings a unique set of skills and perspectives to the community, yet all members are united by a common objective: the application of AI to enhance healthcare outcomes. This shared mission may serve as the cornerstone of their collective identity, imbuing their individual contributions with a greater sense of purpose and relevance. However, it should not be forgotten that collective identity is simultaneously shaped by individual identities of the team members. The team's evolving roles and shared mission exemplify Schatzki's concept of "rules" and "teleoaffective structures." Rules here are not just formal guidelines or company policies but also include the unwritten norms and ethical considerations that govern the community's activities. Similarly, the teleoaffective structures manifest as the shared objective of improving healthcare outcomes through AI, which emotionally guides the team's actions. The concept of "legitimate peripheral participation" (Lave & Wenger 1991) can be aligned with Schatzki's idea of "understandings" that link the set of activities making up a practice. New members initially engage in peripheral roles, guided by a basic understanding of the community's practices. As they become core members, their understanding evolves, becoming more nuanced and aligned with the community's complex practices.

Knowledge Transfer

In many communities, knowledge is continually transferred through daily activities. The foundational stage, involving fine-tuning algorithms and scrutinising data analytics, serves as a critical knowledge transfer phase. Intermediate practices, like UX design and legal compliance, add layers of specialised knowledge that get integrated into the community's collective wisdom. In this healthcare-focused AI community of practice, knowledge transfer is not a one-off event but an ongoing,

integral process embedded in the community's activities. The foundational stage, which is heavily centred on the refinement of algorithms and the rigorous scrutiny of data analytics, serves as a crucial phase for knowledge transfer. Here, data engineers and analysts work collaboratively, sharing insights and methodologies to improve the performance of machine learning models designed for predicting patient outcomes. Their collective efforts result in the accumulation of shared understandings about best practices in algorithm development and data analysis, enriching the community's repository of technical knowledge.

Moving on to the intermediate practices, such as User Experience (UX) design and legal compliance, another layer of specialised knowledge is added to the community's intellectual landscape. UX designers bring insights into human-computer interaction, while legal experts contribute understandings of healthcare regulations and data privacy laws. These professionals share their unique skill sets and knowledge through regular team meetings, documentation, and project collaborations. The integration of this specialised knowledge into the community's practice not only enhances the usability and compliance of AI solutions but also enriches the community's collective wisdom. The power dynamics within the community also shape the flow of knowledge transfer. For instance, core members with more experience may have greater influence in decision-making processes, thereby directing which types of knowledge are prioritised or marginalised. This links back to Schatzki's idea of "teleoaffective structures," which refer to the emotional and social factors that influence the activities within the practice (Schatzki 2005, p. 471).

Overall, the continuous cycle of knowledge transfer within this AI community of practice contributes to its adaptive and innovative capabilities. The community not only learns from its internal activities but also adjusts its practices based on feedback loops with the broader healthcare ecosystem, thus embodying Reckwitz's description of practices as routinised ways that evolve over time (Reckwitz 2002, p. 250). In summary, knowledge transfer in the healthcare-focused AI community of practice is a dynamic, multi-dimensional process that is deeply entangled with its organisational practices, thereby reflecting the theoretical underpinnings of practice theory.

Power Dynamics

The team's decision-making process is not a monolithic structure but a complex interplay of power dynamics, which are both hierarchical and distributed. Core members, often seasoned professionals with specialised expertise in areas like algorithmic development, data analytics, or legal compliance, wield considerable influence in steering the project's direction. Their role resonates with what Wenger (1998) termed as "old-timers," individuals who, by virtue of their experience and accumulated knowledge, hold a more central position within the community. Their influence often manifests in key decision-making junctures, setting research agendas, and allocating resources.

However, the community of practice model also allows for a more democratic distribution of power. New or peripheral members, what Wenger would refer

to as “newcomers,” also have the opportunity to contribute to the community’s shared practice. They might be involved in less central tasks initially but still have channels to offer insights, ask questions, and even challenge established norms or methodologies. This aligns with Wenger’s conclusions on power dynamics within Communities of Practice, where power is not solely the purview of a select few but is distributed across the community, allowing for a more dynamic and adaptive learning environment (Wenger 1998). The power dynamics evolves over time and is often influenced by external factors like technological advancements, regulatory changes, or shifts in healthcare needs. For example, a new healthcare regulation might elevate the influence of legal experts within the team, or a technological breakthrough might shift focus towards a different set of algorithms, thereby altering the power dynamics within the community. Furthermore, power dynamics also intersects with the “teleoaffective structures,” which refer to the emotional and social factors, including shared goals or values, that guide the actions and interactions within the practice (Schatzki 2005, p. 471). For example, a shared commitment to ethical AI development might empower team members specialising in ethics, even if they are peripheral members, to have a stronger voice in decision-making processes.

Conclusion

The proposed shift from viewing AI merely as a technological construct to understanding it as a phenomenon deeply embedded in social practices offers a quite novel theoretical framework, which goes beyond many of the existing theories in contemporary social sciences. Using various propositions present in the field of social practice theory (Schatzki, Swidler, Reckwitz, Lave and Wenger), this approach sheds light on how AI is constituted or applied by an array of social, cultural, and ethical practices. Such a perspective opens avenues for social scientists to delve deeper into the ontological dimensions of AI that are not often seen by technical specialists who may lack this socio-cultural lens. At the same time, through the theory of practices, experts with a more technical background are able to perceive dimensions of the phenomenon that they previously did not see. Therefore, the benefit is multi-dimensional, and its essence lies in an interdisciplinary and multi-disciplinary approach to the study of AI.

Practice theory allows us to perceive another, highly problematic but increasingly discussed aspect of AI development. It is about human relations, which, although they still determine the essence of AI (AI as a socially dependent phenomenon), are becoming qualitatively and quantitatively different (Dunbar 2021), weaker and shallower (Twenge 2013) as more frequently mediated by digital tools (Chambers 2006). The change in shape, frequency, and course of relations among human and non-human actors is a critical signal for algorithm designers. This transformation prompts them to recalibrate their solutions to align with the prevailing social trends within the wide spectrum of relationships. This functions as a feedback loop where solutions tailored to the nuances of social relations wield the power to shape the trajectory of future social interactions. Human relations are a vital anchor and guide

for the trajectory of algorithmic advancements. However, it is crucial to constantly observe the changes occurring in the social domain. Thus, a manifestation of human (and non-human) relations in AI includes practices, and understanding what social practices are, how they are formed, change, and connect with other practices may help to manage AI design and deployment.

Applying the theory of practices also entails numerous risks, which should be kept in mind when choosing this approach. Overemphasis on social practices can lead to an “oversocialised” vision of AI. Overestimating human involvement in the process of creating and modifying AI might divert researchers’ attention from significant phenomena that have a non-social or non-human character. Another risk is the conceptual apparatus characteristic of practice theories. Although, as I have tried to demonstrate in this chapter, the practice approach is very flexible, it may turn out that focusing on shared understandings, rules, teleoaffective structures, domain of practice, knowledge transfer, or power dynamics might distract us from other aspects of AI, which can only be identified by choosing a different social theory. Despite its many advantages and limitations, it seems that the theory of practices offers an original perspective for analysing AI, which allows us to see many important social aspects in this dynamic, continually surprising phenomenon.

References

- Airoidi, A. 2022. *Machine Habitus. Towards a Sociology of Algorithms*. Polity Press: Cambridge.
- Bloomfield, B. P. 1988. Expert Systems and Human Knowledge: A View from the Sociology of Science. *AI & Society*, 2, 17–29.
- Bourdieu, P. 1977. *Outline of a Theory of Practice*. Cambridge University Press: Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo, Delhi, Mexico City.
- Burrell, J., & Fourcade, M. 2021. The Society of Algorithms. *Annual Review of Sociology*, 47, 213–237.
- Chambers, D. 2006. *New Social Ties. Contemporary Connections in a Fragmented Society*, Mcmillan: Basingstoke, New York.
- Couldry, N. 2004. Theorizing Media as Practice. *Social Semiotics*, 14(2), 115–132.
- Crawford, K. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press: New Haven, London.
- Devyania, R. D., Jewanc, S. Y., Bansal, U., & Denge, X. 2020. Strategic Impact of Artificial Intelligence on the Human Resource Management of the Chinese Healthcare Industry Induced Due to COVID-19. *IETI Transactions on Economics and Management*, 1(1), 19–33.
- Dunbar, R. 2021. *Friends: Understanding the Power of Our Most Important Relationships*. Hachette: UK.
- Elliott, A. 2019. The Culture of AI. In *Everyday Life and the Digital Revolution*. Routledge: London and New York.
- Elliott, A. 2022. *Making Sense of AI*. Polity Press: Cambridge, Medford.
- Fu, L. 1994. *Neural Networks in Computer Intelligence*. McGraw-Hill, Inc: New Delhi.
- Gardner, H. 1983. *Frames of Mind: A Theory of Multiple Intelligences*. Basic Books: New York.
- Gillespie, T. 2014. The Relevance of Algorithms. In *Media Technologies: Essays on Communication, Materiality, and Society*. MIT Press: Cambridge, 167–193.
- Kurzweil, R. 2005. The Singularity Is Near. In *Ethics and Emerging Technologies*. Palgrave Macmillan: London, 393–406.

- Latour, B. 2007. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press: Oxford.
- Lave, J., & Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press: Cambridge.
- Marody, M. 2021. *The Individual After Modernity: A Sociological Perspective*. Routledge: London, New York.
- McCulloh, W. S., & Pitts, W. 1943. A Logical Calculus of the Ideas Immanent in Neural Nets. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- Reckwitz, A. 2002. Toward a Theory of Social Practices: A Development in Culturalist Theorizing. *European Journal of Social Theory*, 5(2), 243–263.
- Schatzki, T. R. 1996. *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge University Press.
- Schatzki, T. R. 2002. *The Site of the Social: A Philosophical Account of the Constitution of Social Life and Change*. The Pennsylvania State University Press: University Park, Pennsylvania.
- Schatzki, T. R. 2005. The Sites of Organizations. *Organization Studies*, 26(3), 465–484.
- Schatzki, T. R. 2006. On Organizations as They Happen. *Organization Studies*, 27(12), 1863–1873.
- Schwartz, R. D. 1989. Artificial Intelligence as a Sociological Phenomenon. *Canadian Journal of Sociology/Cahiers canadiens de sociologie*, 2(14), 179–202.
- Swidler, A. 1986. Culture in Action: Symbols and Strategies. *American Sociological Review*, 51(2), 273–286.
- Swidler, A. 2005. What Anchors Cultural Practices. In: Schatzki, T. R., Knor Cetina, K., von Savigny, E. *The Practice Turn in Contemporary Theory*. Routledge, 83–101.
- Turing, A. M. 2004. Computing Machinery and Intelligence (1950). In: Copeland, B. J. *The Essential Turing: The Ideas That Gave Birth to the Computer Age*, 433–464. Oxford University Press: Oxford, New York.
- Twenge, J. M. 2013. Does Online Social Media Lead to Social Connection or Social Disconnection? *Journal of College and Character*, 14(1), 11–20.
- Wenger, E. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press: Cambridge.
- Woolgar, S. 1985. Why Not a Sociology of Machines? The Case of Sociology and Artificial Intelligence. *Sociology*, 19(4), 557–572.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Part II

Algorithms, Artificial Intelligence, and the Theories of Modernity



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

3 Images of Algorithmic Modernity

Agency and Autonomy in the Age of Predictive Analytics

Anthony Elliott

Introduction

Once celebrated as the antidote to our hurried and harried lives, the promise of automated digital technologies lay in their ability to bestow instant decisions and immediate social connections, all managed by the deft force-field of smart devices. As the tendrils of advanced artificial intelligence (AI) wove itself into our social fabric, however, the true cost of this automation ethos came to light. Human agency became increasingly ceded to digital technology, entangling individuals within ever-expanding social networks and ensnaring them in the ceaseless digital labour of downloading, updating, posting, tagging, liking, retweeting and endlessly incorporating dynamic data into their lives.

Terms like “information overload,” “data deluge” and “digital device distraction syndrome” ([Campo, 2022](#)) have come to symbolise the impact of advanced automation on our social landscape. Our current algorithmic age, or what I shall term in this chapter “algorithmic modernity,” has ushered in profound shifts in the scale and intensity of automation across personal, familial, vocational, cultural, communal and institutional domains ([Elliott, 2021](#)). “Automaticity” has permeated social action, with smart algorithms serving as a coping mechanism for those navigating a world in perpetual flux. Yet these very algorithms also propel the relentless march of social transformation, challenging individuals as they wrestle with the outsourcing of personal choices and the plans and projects of individual lives to automated intelligent machines.

Within the automated realm of algorithmic modernity, human agency is increasingly consigned to the guardianship of smart machines. This phenomenon arises from an attempt to corral the uncertainties and complexities born of the digital revolution, even if there remain little to no prospects for conclusive answers to these conundrums. Predictive algorithms and automated recommendation systems facilitate a more instantaneous, immediate, simultaneous and seamless social existence. However, the flipside of all this lies in individuals growing estranged from their own decision-making processes and potentially drained of the ethical foundations required for a reflective and fulfilling existence ([Elliott, 2024](#)).

How has this situation come about? How has it transpired that automated decision-making, which (one might well think) is by definition opposed to individual volition

and the personal freedom to choose, has become an increasingly dominant cultural logic? In the age of AI, why have machine automation and personal decision-making become strangely conflated?

In this chapter, I wish to discuss the impact of AI in general and predictive analytics in particular upon personal decision-making and individual life-projects. Since I want to situate my discussion in the context of the changing relations between emergent technologies and social relations, I do not make any claim to analyse the extensive range of contributions in social and political theory to the study of human autonomy.¹ Instead, I shall focus my discussion on the specific contribution of the late European social theorist Cornelius Castoriadis to the issue of human autonomy in our world of advancing technology. Castoriadis's writings on the question of human freedom and new technology offer a novel standpoint from which to rethink the very nature of the digital revolution in modern societies. Whilst developing the main threads of his analysis some considerable time prior to the rise of advanced digitalisation, Castoriadis's theoretical investigations are of considerable importance to rethinking human agency in the age of advanced machine learning. I shall thus draw from some of Castoriadis's key insights as a prompt from which to raise fresh questions about the interrelations between human autonomy, personal decision-making, artificial agents and predictive analytics. Throughout the critical discussion, I want to ask why it is that people today are increasingly inclined to surrender personal decision-making to automated machine intelligence? I also want to ask: what is it about the living of life-projects through smart machines which is so captivating?

Can There Be Autonomy in Predictive Algorithmic Societies?

Autonomy, it is often said, is an ancient notion. It is bound up with ideas of competence, capability, reflection, rationality, along with authenticity and self-awareness. All of these traits are underscored in the writings of Jean-Jacques Rousseau, Immanuel Kant and John Stuart Mill. Autonomy, as a moral value, is often considered the foundation of liberal political philosophy (Dworkin, 1981).

Against the backcloth of the transformation of modernisation processes, the notion of human autonomy was extended to a wider engagement with processes of industrial automation. Ideas about personal freedom and political autonomy in the post-war years, as a result, entered into a more advanced phase. In broad terms, there were a number of authors who sought to stretch the notion of autonomy to encompass both individual and public realms, the psychic and cultural, the imaginary and socio-symbolic, the personal and the historical. 'Autonomy,' wrote the late European social theorist Cornelius Castoriadis, "is not closure but, rather opening: ontological opening, the possibility of going beyond the informational, cognitive, and organizational closure characteristic of self-constituting, but heteronomous beings" (Curtis 1997: 310, 316). The notion that challenging established meanings and institutions is vital for achieving autonomy reflects a historical shift. This shift moves away from tradition and the sacred and towards the unpredictability of the social sphere. Another way of saying this is to say that

a perspective of society has evolved historically which recognises that there is no overarching assurance of meaning. The end of foundationalism means embracing the idea that meaning and its realisation always rely on *social context*. As Castoriadis explains: “if autonomous society is that society which self-institutes itself [*s’auto-institue*] explicitly and lucidly, the one that knows that it itself posits its institutions and significations, this means that it knows as well that they have no source other than its own instituting and signification-giving activity, no extrasocial ‘guarantee’” (Castoriadis in Curtis, 1997: 368).

If we acknowledge that autonomy entails the questioning of individual and social meanings, a central issue arising concerns the demystifying inquiries happening in today’s societies. We must also explore how knowledge pressed into the service of human autonomy can be directly linked to processes of societal transformation. Here, Castoriadis’s argument that human autonomy advances to the degree to which social relationships are ordered through critical self-interrogation rather than through embedded technological power remains important. In his writings, Castoriadis distinguishes archaic societies, characterised as closed in terms of cognition, information and organisation, from modern societies, which, in his perspective, encourage challenges to established institutions and meanings. Castoriadis adeptly dissects the culture of critical discourse that modernisation and detraditionalisation processes foster.

He also deftly underscored an essential interdependence between the projects of personal and collective autonomy, in a conceptual move which is distinct from prevalent tendencies in social theory towards subjectivism or objectivism. In modern societies, the quest for autonomy necessitates a profound interrogation of the very structure of the social realm. This interrogation stems from individuals endowed with both rational and emotional capacities for self-reflection. Furthermore, Castoriadis emphasises the open-ended nature of the process towards autonomy. This perpetual engagement with autonomy, at once personal and social, unfolds in a milieu characterised by ambivalence, uncertainty and the clash of diverse scenarios. “We have to create the good,” writes Castoriadis, “under imperfectly known and uncertain conditions. The project of autonomy is end and guide, it does not resolve for us effectively actual situations” (Castoriadis in Curtis 1997: 400).

Castoriadis distinguishes his reflections on the open-ended nature of autonomy from certain strains of critical theory where utopianism exists apart from the inherent possibilities within the existing social structure. Critic Terry Eagleton termed the latter type of social critique as “bad utopianism,” where alternative values and orientations are “parachuted in from some ontological outer space.” By contrast, Eagleton described “good utopianism” as that which “seeks somehow to anchor what is desirable in what is actual” (Eagleton 1991: 131). On this point, it can be said that Castoriadis draws inspiration from the Marxian notion that the quest for social change derives from structural contradictions inherent in capitalist modernity. Emancipation, according to Castoriadis, denotes the shift from externally imposed, power-oriented heteronomy to critical self-reflection. Crucially, Castoriadis’s framework for autonomy underscores the contingency of the social field. The pursuit of autonomy, as outlined by Castoriadis, demands individuals

and groups navigate a landscape devoid of moral or ethical certainties, recognising that the journey towards personal and political autonomy is laden with ambiguity and uncertainty. Additionally, this perspective acknowledges the potential swiftness with which the pursuit of autonomy can be overshadowed by tendencies towards domination and oppression.

Society, we can acknowledge, was from the beginning an ‘imagined entity.’ But when Castoriadis, to the dismay of many French intellectuals steeped in the conceptual nets of structuralism and post-structuralism, coined the concept of the ‘social imaginary,’ he underscored powerfully the articulation of fantasy, fear and anxiety in the constitution of society. Admittedly, modern technology was recognised as a key force-field in the social imaginaries of capitalist modernisation. Castoriadis’s central reference point here was the work of Jacques Ellul. In *The Technological Society*, Ellul (1980) argued that ‘technique’ – broadly conceived as rationally ordered methods which render human action more efficient – *advances because it advances*. Reversing the conventional wisdom that technology flows from science, Ellul insisted that technique has historically preceded science. On this view, means *precede* goals: the mere technical possibility that something can be done has served as the inspiration for undertaking it. Nothing in society escapes the reaches of technique; society is dominated by the *automatism* of technique, which is the process by which technology asserts itself according to mathematical models of efficiency. What Ellul calls ‘technique’ is all-encompassing: “Technique tolerates no judgment from without and accepts no limitation.”

Developing upon Ellul’s insights, Castoriadis connected the rise of ‘technoscience’ to the rationalisation of politics; the roots of the modern urge for the ‘technological fix’ are to be found in the unfurling of the irrational and the regressive, in the fantasy of being all-powerful and in the institutionalisation of omnipotent thinking. This is most obviously evident in the surge of violence in contemporary societies, most notably the techno-industrialisation of war. The dominance of pseudo-scientific rationality today involves not only the ‘technological fix’ but fantasy, repetition, omnipotent illusions and psychic excess; the paths of technoscience, says Castoriadis, are “less and less that of the *desirable* in any sense, and more and more that of the simply *doable*” (Castoriadis, 1991: 249). Technoscience for Castoriadis is fundamentally *imaginary*: it bears omnipotent fantasies and delusions entirely within itself, resulting in “an artificialized rationality that has become not only impersonal (nonindividual) but also inhuman” (Castoriadis, 1991: 246).

Technoscience, according to Castoriadis, exposes a myriad of contradictions and shortcomings evident across various domains. From scrutinising scientific explorations delving into DNA and the genetic code to mapping the devastation of tropical forests and disruptions within the Earth’s biosphere, the whole-scale expansion of technoscience is everywhere. But Castoriadis emphasises the predominantly imaginary and ideological facets of technoscience. These facets are deeply rooted in the quasi-omniscient and quasi-omnipotent assertion of societal and political advancement. In doing so, technoscience obliterates the potential for reasoned and reflective discourse within the political sphere. Consequently, it relinquishes the

public arena to be dominated by financial, managerial and bureaucratic oligarchies. In developing this line of argument, Castoriadis drew attention to an important set of issues – namely, the interrelation between the social imaginary and modern societies – which bear directly on the technological transformation of cultural life. Yet because he developed this analysis prior to the onset of the digital revolution, Castoriadis was not able to consider the role of digital technologies in broader institutional transformations associated with modernity. Consequently, there are some important questions that we might pose from this perspective. How, exactly, might predictive analytics intersect with the project of human autonomy? Might society's reliance on machine learning algorithms congeal to foster cultural fear and unmitigated anxiety?

In our own time of big data and smart algorithms, the prospects of individualised actors seeking to reassert the search for personal autonomy appear faint. AI has become the new technoscience of our age and the cultural logics of predictive analytics tend to be the preferred, perhaps the dominant, method of conducting both organisational and personal life. But how might we think of the various challenges, paradoxes and conundrums involved in newly constituted relations between self-reflection and self-transformation on the one hand, and the transfiguration of algorithmic power on the other. Writing about technoscience prior to the global rise of AI, Castoriadis had justified his concentration on the project of human autonomy on the grounds that the political challenge remains finding new ways of promoting and expanding personal empowerment, in so far as this can open potentially transformative relations between self and society. But what might such a 'project' look like in an era where people are relentlessly coaxed, cajoled and seduced to self-automate more and more aspects of their lives? The yawning gap between the project of autonomy and the capacity to reassert personal control over decision-making that has been outsourced to smart machines seems to be the central contradiction of algorithmic modernity – one that does not bode well for critical self-reflection, interpersonal experimentation and social flourishing.

It is perhaps not difficult to see in the foregoing philosophical lines of departure the driving force of technology in the reshaping of contemporary societies. Combining Ellul and Castoriadis, we can say that AI – as the culturally dominant discourse of current technoscience – is both ecstasy and power, imagination and information, omnipotent mastery and automated self-subjection. So, to return to our guiding question: what is the lure of algorithmic prediction in digital society? What is it about the living of lives through automated smart devices which is so intoxicating? We clearly need to consider the possibility, extending the insights of Ellul and Castoriadis, that people are increasingly addicted to algorithmic calculation as a means to achieve a soothing self-image of personal control. Alternatively, we could invert this and plausibly argue that the excesses of quantification in our algorithmic world may often become overbearing, but that (on balance) people adapt to these automated practices at the level of social interaction and social life remarkably well. More and more people take predictive analytics and algorithmic curation for granted now, even though it continues to have the power to disturb, displace and oftentimes derail constructions of self and society in startling ways.

Excursus: Life Lived on Auto-Pilot

Living a life on autopilot is one which wraps together predictive analytics, big data, machine intelligence, smart devices, artificial agents and advanced automation. It obliges each and every individual in social networks towards the cultural logics of outsourcing and offshoring and to relentlessly focus on the redistribution of incoming assignments, requests, projects and work to automated smart machines. The instrumental rationality favoured and reproduced by the self-automating life prompts the exploration of such human-machine interactions where tasks, social doings and moral obligations can be dispensed with and effectively disposed of, leaving the individual self ‘free’ to get on with the more pressing realities of life lived in the fast lane of ‘information overload.’

Consider, for example, a ‘relationships column’ titled “How to Automate Every Part of Your Life” ([Mint.com](https://www.mint.com), 2021). Outlining new automation technology which allows people to “set and forget,” the argument advanced is that the brave new world of predictive analytics is one in which “things get done for you.” As the columnist explains: “Picture a more productive work day where you get the most out of every hour. Think about how your spare time is spent and outsource the boring parts to technology. Imagine stress-free money management with instant bill payments, apps that manage your debt, and increased savings month after month through automated contributions that you don’t even have to think about” (Ibidem). Not having to ‘think about’ things is central to the consumer industries and consumer markets advancing self-automation. Indeed, self-automation tends to convert into its own objective.

Following the new ethos of self-automation is easy thanks to ‘automation tips’ which free up personal time, along with suggestions on the latest apps to ‘raise productivity’ and strip ‘wasted hours’ on ‘redundant tasks.’ The ‘must haves’ include apps for sleep management (including reminders to go to bed and when to get up at optimal times), task management (to keep people ‘on top of responsibilities’), travel fare aggregators (which send alerts about the best flight deals) and ‘applets’ (which facilitate disparate apps and devices to trigger automatic responses). So you should live without having to plan, organise or arrange daily living, the automation of online subscriptions for regular purchases is strongly recommended, along with the putting of grocery delivery services on autopilot. To work properly and to deliver the best corporate results, the pre-programmed blocking of websites that prevent work from getting done is obligatory, along with the automation of computer maintenance by setting regular system backups and clean-ups which run on after-hours timeframes. Add to this smart home devices, robot vacuums, security bots, robotic window cleaners, automowers and recurring timers to make home living altogether less cumbersome and less time-wasting.

The Rise and Rise of Predictive Analytics

Helga Nowotny, in her brilliant book *In AI We Trust*, emphasises the astounding change of vocabulary and climate of cultural mood which mark the current phase of machine learning algorithms in contemporary social practice. Addressing the

cultural illusions of machine learning, Nowotny (2021) dissects the increasingly important role played by prediction, especially predictive analytics. She argues that predictive algorithms have become increasingly central to a wide range of products and services, from the analysis of DNA samples to predict risks of certain diseases to stock markets and financial trading in the global electronic economy. Companies, organisations and institutions have come to rely increasingly on predictive algorithms to both reproduce and transform their core activities. More and more, people feel comfortable not only with the performative power of predictive analytics, but also with having options curated and laid out in advance. All of this, to repeat, feels reassuring and smooth; but rarely do people stop to consider how such options have been designed, or to reflect on the point that there might be other options equally worth considering.

Prediction, says Nowotny, is essentially about the future, yet it reacts back on how we conceive the future in the present. Her key theorem: *predictions assume the power of agency that individuals attribute to them*. When blindly followed, the predictive power of algorithms can turn into self-fulfilling prophecies. A prediction becomes true simply because people believe it to be so and act accordingly. On this view, the paradox of machine learning algorithms is that people are coaxed into believing that the only future is that which is *predicted*. “Do we really want to live in an entirely predictable world,” asks Nowotny (2021: 11), “in which predictive analytics invades and guides our inner most thoughts and desires? This would imply a return to a deterministic worldview where the open horizon of the future will close again, renouncing the inherent uncertainty of the future and replacing it with the dangerous solution of being in control. Or are we ready to acknowledge that a fully predictable world is never achievable and that it is an illusion?”

Delegating an increasing number of decisions and tasks to artificial agents, automated machine-learning processes and smart algorithms pose significant risks to our lives and the autonomy of our personal endeavours. This escalating trend towards predictive engineering in personal and social life suggests that there is some kind of ‘flaw’ or ‘fault’ in the individual self. The implications of this extend beyond the realms of psychology and philosophy; they profoundly affect policymakers, activists, citizens and their communities. The rapid pace and vast scope of technological innovation render conventional regulatory measures and corporate solutions inadequate. Despite the ever-increasing capacity for people today to offload decision-making to automated intelligent machines, however, this hasn’t translated into the enjoyment of more self-directed lives (On this point, see my *Algorithms of Anxiety: Fear in the Digital Age*, Cambridge: Polity, 2024). So, again, this prompts the fundamental query: Why do people feel the need to increasingly automate their lives to achieve fulfilment?

Our current global order is largely constructed on the predictive re-engineering of human capabilities. The interaction between humans and automated machines has been flagged as an impending ‘existential threat,’ sparking numerous discussions on the potential world we are designing for future generations. While these conversations are crucial, there is also an urgent need for dialog about the subtle yet pervasive influence of AI on existing social dynamics, especially the reshaping of

social interaction through new models of algorithmic quantification and automated calculation. Against the backdrop of such transformation, questions arise about fear, insecurity and anxiety – transforming the future into a landscape marked more by foreboding than hope. Blaming advanced automation as the sole source of social dislocation in these early stages of the digital revolution might seem facile. Most perplexing and haunting of today's many challenges, however, remains the pressing need to 'keep up' with data regimes, to adjust the self to ongoing software updates and to reinscribe subjectivity within the anchoring effects of algorithms to make life seemingly continuous.

To make the conundrum still more awesome: individuals feel increasingly overwhelmed and at times besieged in today's data-centric culture because digital technologies, through predictive algorithms, incessantly dictate what individuals are presumed to truly desire. Undoubtedly amidst the technological tsunami of big data and predictive analytics, it remains profoundly challenging to unearth one's genuine desires. The meticulously crafted predictions and calculated probabilities embedded in algorithms act as surrogates for personal agency, restricting the capacity of individuals for self-directed action as well as coordinated social interaction. Despite the promises of predictive analytics to alleviate fear-inducing mindsets of doubt, uncertainty and ambivalence in the face of information overload, the cultural significance of AI remains inflated. Indeed, the machine predictions generated by algorithmic calculations and automated recommendation technology are increasingly embraced as the predominant culturally acceptable means to mitigate anxiety in the digital age. Trapped in this vicious cycle, the entire culture of data analytics seems incapable of reaching a conclusion, further contributing to people's sense of estrangement.

Today's overreliance on machine learning algorithms to continually project calculated forecasts for our daily routines acts as a culturally ingrained defence mechanism against existential fears. Yet, fear operates in a paradoxical manner, oscillating between past and future; its perpetual existence underscores the potential risks of predictive algorithms evolving into self-fulfilling predictions. Fear functions as a form of reductionism, narrowing the world's possibilities. Within this confined scope, there persists an unsettling suspicion that our culturally embedded defences against anxiety might prove inadequate once more. While transforming the world into a vast prediction engine may appear as a robust strategy to combat the fears induced by information overload and the dominance of big data, this pursuit is consistently overshadowed by the flickering recognition that social existence is inherently entwined with unending uncertainty.

Let me sum up the argument so far and try to bring together some of the main threads of my contention. In the automated order of algorithmic modernity, personal agency becomes increasingly 'redistributed' to intelligent machines. This might be thought of in terms of the 'outsourcing' of decisions to smart machines. We should understand this phenomenon in terms of a containment of both uncertainty and complexity which the digital revolution in social relations poses, but which ultimately denies answers. Whilst automated machine intelligence becomes increasingly integrated into the fabric of society, however, it is important to recognise that this integration is not all-enveloping. Significantly, there persist various forms of personal

dissatisfaction, cultural pushback and social dissonance. A glaring illustration of this paradox is the ‘double-bind’ confronted in the era of ‘information overload’. The rise of digital technologies was marketed, to a substantial degree, as a remedy to the challenges of navigating the hurried and pressurised world of organised modernity. Swift decision-making and instantaneous social connectivity were heralded as the default attributes of an advanced digital era aimed at alleviating the constant time constraints in professional and personal relationships. The delegation of personal choices to smart devices seemed to promise a solution to this issue. However, there emerged an unforeseen cost to this reliance on automation. As agency was increasingly shifted to digital technology and social networks interwove individuals into ever-expanding connections, people found themselves entangled in ongoing digital labour. This encompassed the tireless and never-ending tasks of downloading, updating, posting, tagging, ‘liking,’ and essentially striving to seamlessly incorporate dynamic data into everyday living and lifestyles.

‘Automaticity’ – the automation of things, information, ideas and people through global networks of machine intelligence, deep learning and big data – becomes an intrinsic feature of social action on a very general level. Under these new circumstances, smart algorithms represent a kind of ‘coping mechanism’ for individuals confronting a world of continuous change. But smart algorithms are also a vehicle for the further acceleration of social change. The odds are that an increasing amount of personal life and more and more of social life are spent worrying or agonising about the outsourcing of human choices and life goals to automated intelligent machines, as people scramble to ‘retrofit’ the outcomes of automated personalised networking into the frames of daily life and to integrate automated feedback mechanisms and endless data into their established modes of behaviour (Elliott, forthcoming).

Algorithmic modernity, as I have previously argued, involves the conjunction of complex digital infrastructures and automated lifestyles. The algorithmic blending of smart devices and computational life-projects may be understood as an institutional response which seeks to balance the demands of personalised and individualised living with technological systematisation in the aftermath of the digital revolution. This ‘system-ness’ of automation is fundamental and underpins the efforts of people to keep afloat in an age of ‘information overload’ and render realistic dreams of living an autonomous life. From this angle, contemporary societies are characterised both by the growth of automation in everyday life and by the increasing importance of digitalisation for the redefinition of individualisation and personal identity. It is not just that daily life becomes ensnared in a series of automated processes shaped by algorithmic recommendation systems and machine intelligence, in which the individual self might be cast (following the insights of Georg Simmel) as “a mere cog in an enormous organization of things and powers” (Simmel, 1997: 184). Rather, the inner texture of self-experience, as well as the intimate terrain of private life, is repositioned – “disembedded and re-embedded,” to recall Anthony Giddens’s (1990) memorable terminology – in the self-expanding world of predictive algorithms. Predictive analytics at once consoles and numbs the self – or, rather, keeps us at a safe distance from our own capacity for personal agency and self-reflection. The partitioning of individual agency off from the ‘personalised

living’ or ‘individualised lifestyles’ advanced in conditions of algorithmic modernity means that the societal costs of our ‘automated comfort zones’ are fenced-off from personal reflection and public inspection. The comfort zone of automation – the ever-multiplying need for algorithmic recommendations, automated reminders, remote monitoring and digital surveillance – is our smart vehicle for sustaining everyday life. But like all stories of the modernisation process, automation in the algorithmic age remains intrinsically vulnerable to system errors, external interference and digital breakdown. The flipside of the consolation of automated lives is thus anxiety, fear and dread.

Coda: Rethinking Autonomy in the Age of Predictive Analytics

Towards the conclusion of his thought-provoking essay on technoscience, “Dead End?” Castoriadis noted “the present-day tendency toward the ‘automatization of decisions’.” As he presciently observed:

This goal is already being implemented in a large number of secondary cases, but it is beginning to take on a new appearance with the introduction of “expert systems”. And still more illustrative is the idea, which is in some ways its realization, of a ‘Doomsday’ machine, an expert system that would automatically send missiles from one camp to the other as soon as those from the other side have been computed or assumed to have been fired, thereby eliminating from deterrence policy all ‘subjective’ (and therefore fallible as well as influenceable) politico-psychological factors. Are we really so far away from that point now?

(Castoriadis, 1991: 273)

The short answer is surely “No.” But arguably, today things are even more complex, murky and ambivalent.

Castoriadis ascribed to an earlier version of autonomisation the organised irresponsibility of artificially contrived responses to pressing matters of public importance and global security. But it is worth noting that “Dead End?” was first published in France in 1987. Move forwards more than 35 years, and it is all too evident that the digital revolution has unleashed socio-political tendencies which facilitate the recycling of human preferences into automated machine realities. From this angle, perhaps the most remarkable mark of the current moment is the societal-wide delegation of decision-making to artificial agents, machine learning algorithms and computational systems. As Massimo Durante (2021) has insightfully demonstrated in his philosophically erudite volume *Computational Power*, all too often promoting the human freedom of personal decision-making is today overshadowed by needing to adapt not just to digital technologies but also to the automated mechanisms, processes and dynamics by which computational power works. The ‘expert system’ of predictive algorithms, for Durante, can tip towards a self-reinforcing logic which all but cancels out self-reflection, personal decision-making and human autonomy.

In order to develop this point in more detail, let us consider once more Castoradis's reference to a Doomsday machine. In the original series of *Star Trek* from the late 1960s, the Doomsday machine appeared as a vast technological giant in space, threatening the world with annihilation. Beyond science fiction, the notion subsequently became political reality under the precept of mutually assured destruction (MAD), which informed both Soviet and US nuclear plans from the 1970s. In our own age of AI, however, rethinking military and nuclear deterrence becomes literally reconfigured through the lightning speed of automated smart machines. Just consider, for example, emergent technologies associated with the digital revolution – particularly AI, big data, robotics, cyber capabilities and quantum computing – in the transformation of military warfare (Elliott 2019). The landscape of military power has undergone a profound transformation through the integration of various technologies (including AI, machine learning and computer vision), with unmanned aerial vehicles (UAVs) and remotely piloted air systems (RPAS) to the fore. Globally, an estimated 30,000 drones are currently in military service, marking a significant shift towards unmanned and autonomous weapons systems (Strobel 2019).² Leading nations in drone technology, such as the USA, UK, Germany, China, Russia, India, Israel and the Republic of Korea, have played pivotal roles in this evolution. The rise of advanced automated warfare is closely associated with the ongoing AI revolution. Many countries have swiftly adapted their military capabilities to align with the potential of unmanned aerial technologies. Russia, for instance, has deployed remotely piloted tanks like the Vohar and Uran-9, along with formidable UAVs such as the Forpost and Orlando-10. Similarly, the importance of sea power, particularly sea-based drones, has experienced a paradigm shift, with China establishing the world's largest UAV test site in the South China Sea. Unmanned underwater vehicles (UUVs), for example, present cost-effective missile platforms that can be easily harnessed for wartime objectives. Israel has operationalised fully autonomous anti-radar attack munitions (LAMs), colloquially known as 'suicide drones,' demonstrating their devastating impact in conflicts, notably against Hezbollah. In these instances, AI-augmented weapons systems play a pivotal role in information gathering, surveillance escalation, reconnaissance distribution and digitalised warfare. The incorporation of AI empowers autonomous UAVs to achieve military objectives with unprecedented speed compared to traditional deployment methods involving human pilots. This technological advantage in military planning positions countries with AI capabilities at a decisive edge, aligning with the preference for 'hit-and-run' strikes and fully automated, unmanned operations. The emergence of low-cost, miniaturised automated weapons raises a significant concern that extends beyond challenging conventional military distinctions. It encompasses the blurred lines between human and non-human decision-making, state and non-state regimes, aggression and repression, local and global dynamics, as well as internal and external forces. Moreover, the proliferation of AI-powered automated weapons introduces a breeding ground for 'insecurity' and ambivalence within the realm of warfare. In this context, AI becomes a source of ceaseless anxiety, with autonomous weapons designed and designated for lightning-speed strikes, precision manoeuvres and 'hit and run' strategies. In this nerve-wracking tension between automated

intelligent machines and human operators, there lurks the pervasive fear of endemic automated violence. Such fear gives rise to heightened levels of anxiety, insecurity, disconnection, disengagement, ambivalence and ambiguity. Governments respond to this paradigm shift by drastically escalating funding for AI-driven missile programmes and autonomous weapons technology. Yet critics caution that such policy responses may only pave the way for further military misadventures. One key concern remains the potential for an AI arms race between the United States and China, with the added danger of AI disrupting the delicate global nuclear balance.

Let us pause for a moment in order to reemphasise what is at stake here. The application of AI in general, and automated smart algorithms in particular, to military violence and the means of waging war effectively means the *outsourcing of killing to machines*. This, of course, is not exactly news. Warnings issued about the existential threat to humanity as a result of developments in sophisticated weaponry powered by AI have been many and varied. There have also been important counter-reactions. In 2017, more than 100 signatories of AI specialists from across 26 countries called for a UN-led ban on lethal autonomous weapons. More recently, many of the world's leading AI experts signed a pledge refusing to work on the development of lethal autonomous weapons (Houser 2018). These are important counter-trends.

Still, the new primacy of automaticity in military violence and global affairs persists and remains a tendency set to shape politics in the twenty-first century. Certainly, automatic weapons systems are pervasive across a number of nation-states as well as non-state actors. Russia, Israel and other advanced militaries have invested heavily in killer robot technology for weapons development. China, according to various reports, has deployed AI weaponry for the production of its next generation of cruise missiles (McDonald 2016). In South Korea, an arms manufacturer designed and built a gun turret which can identify, track and shoot targets, theoretically without the need for human intervention. Automatic military defensive systems, such as the German AMAP-ADS, the Russian Arena and the Israeli Tropy, can autonomously identify and strike oncoming missiles, rockets, artillery fire, aircraft and surface vessels. The diffusion of automated weaponry from advanced societies to a growing number of other states has unfolded more recently, with many countries seeking to join the lethal autonomous weapons race. Sir Roger Carr, the chairman of British weapons manufacturer BAE, stated at the World Economic Forum that 40 countries are currently working on killer robot technology (Styles 2016). Whilst the application of the technological development of autonomous systems to the delivery of military weapons is only one aspect of a battery of technical advances in this area, the crucial point is that there is no agreed international relations framework specifying how these weapons might be used in times of armed conflict, nor are there global governance mechanisms for the regulation and control of autonomous killer bots.

Now, at the same time that the most pressing military and international issues are increasingly outsourced to automated machine intelligence, various global agencies seek to legislate for the ethical regulation of AI. The most extensive legislation in this area has been developed by the EU, which seeks to strike some sort of equilibrium between opportunity and risk.³ But there is little reason to trust the

assurance of regional and international agencies that the regulation and governance of digital technologies and AI are secure; there is little evidence that ethically sanctioned risk management in AI has been effective thus far in limiting or containing large-scale corruptions or perversions of automated computational power and little indication that the application of traditional categories of ethical thought remain relevant for addressing current technological challenges. Summing up the political and ethical deadlocks of society's "automatization of decisions" in the late 1980s, Castoriadis (1991: 275) wrote:

The enormous dangers, the very absurdity contained in the all-out, directionless development of technoscience, cannot be avoided simply by promulgating a few "rules" set forth once and for all, or by installing a "panel of wise men" who would become merely a tool, if not themselves the actual subject, of a tyranny. What is required is more than a "reform of human understanding"; it is a reform of the human being as socio-historical being, an ethos of mortality, a self-surpassing of Reason.

The current crisis of the "automatization of decisions" in the age of AI is essentially a world divided between the stranglehold of predictive analytics and the diminishment of human autonomy.

Fyodor Dostoevsky (1996: 282) insightfully wrote that the individual self "is tormented by no greater anxiety than to find quickly someone to whom he can hand over that gift of freedom with which the ill-fated creature is born." Recalibrating Dostoevsky's insight, my argument is that – in conditions of algorithmic modernity – fear-inducing anxiety of which persons wish to rid themselves is increasingly passed over, or more accurately outsourced, to smart machines. The pervasive nature of fears surrounding automated intelligent machines appears overwhelming. It feels as though threatening risks and alarming anxieties may seep into every corner of our tech-driven workplaces, homes and digital interactions. These fears take various forms, from concerns that AI will eliminate jobs to apprehensions about digital surveillance eroding privacy. More and more, and notwithstanding efforts at 'ethical regulation' and 'corporate governance', algorithms exacerbate societal divisions, deepening racial and gender inequalities. Additionally, the surge of personal vulnerabilities within the realm of social media culture amplifies anxieties. There's also the overwhelming sense of mental and physical unease regarding the militarisation of drones and autonomous weaponry. And then there is the most daunting fear of all: the idea that AI not only jeopardises our immediate surroundings – our workplaces, homes and families – but poses a threat to the very existence of humanity.

The big question is whether society can break from its current addictive reliance on predictive analytics? In other words, can a rebirth of human autonomy – critical self-interrogation, personal reflection, the free choosing of life-projects – be launched in the age of algorithmic modernity? For those now accustomed to living life on auto-pilot, the problem is that there are no shortcuts leading to swift, effortless 'answers'. Yet the urgent political task of recalibrating human autonomy in the context of automated machine intelligence remains. We need to recognise that this

task will likely prove immensely complex, onerous and troublesome – at both the cultural and corporate levels.

Notes

- 1 See, for example, [Stiegler \(2018\)](#); [Parisi \(2022\)](#); [Amoore \(2020\)](#); [Fazi \(2018\)](#); [Andrejevic \(2019\)](#); [MacKenzie \(2017\)](#).
- 2 *Le Monde* reported in June 2023 ([Pietralunga 2023](#)) that the Royal United Services Institute (RUSI), a British defense think tank, estimated that Ukraine had lost around 10,000 drones since the war with Russia began on 24 February 2022. This war has established the drone as an essential piece of battlefield equipment. RUSI researchers reported that within a 10 km zone, it would be common for there to be somewhere between 25–50 UAVs on each side of the front line.
- 3 The EU has developed a suite of world-leading legislative frames advancing the regulation and governance of digital technologies and AI. It includes *The Digital Services Act* (2022), which updated laws covering illegal content, transparent advertising and disinformation regarding digital intermediaries. Also of major significance is the *Artificial Intelligence Act* (AI Act), established on the 8 December 2023. It is aimed to deploy a risk-based model, assigning risk ratings to specific AI systems and scaling regulatory obligations to these ratings. An explicit objective of the Act is to positively influence public trust and confidence in AI systems in order to increase the likelihood of the uptake of automated digital technologies. For critical commentaries on the AI Act see, among others, [Laux, Wachter, and Mittelstadt \(2023\)](#), [Schuett \(2023\)](#) and Enqvist (2023).

References

- Amoore, Louise. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham: Duke University Press, 2020.
- Andrejevic, Mark. *Automated media*. New York and London: Routledge, 2019.
- Campo, Enrico. *Attention and Its Crisis in Digital Society*. New York and London: Routledge, 2022.
- Castoriadis, Cornelius. *Philosophy, Politics, Autonomy: Essays in Political Philosophy*. Oxford: Oxford University Press, 1991.
- Curtis, David Ames, ed. and tran. *The Castoriadis Reader*. Oxford: Blackwell, 1997.
- Durante, Massimo. *Computational Power: The Impact of ICT on Law, Society and Knowledge*. New York and London: Routledge, 2021.
- Dworkin, Gerald. “Moral Autonomy.” In *The Roots of Ethics: Science, Religion, and Values*, 29–44, Boston: Springer US, 1981.
- Eagleton, Terry. *Ideology: An Introduction*. London: Verso, 1991.
- Elliott, Anthony. “Automated Mobilities: From Weaponized Drones to Killer Bots.” *Journal of Sociology* 55, no. 1 (2019): 20–36.
- Elliott, Anthony. *Making Sense of AI: Our Algorithmic World*. Cambridge: Polity, 2021.
- Elliott, Anthony. *Algorithms of Anxiety: Fear in the Digital Age*. Cambridge: Polity 2024.
- Ellul, Jacques. *The Technological Society*, translated by Joachim Neugroschel, New York: Continuum, 1980.
- Enqvist, L. (2023). ‘Human oversight’ in the EU artificial intelligence act: what, when and by whom? *Law, Innovation and Technology*, 15(2), 508–535.
- Fazi, M. Beatrice. *Contingent Computation: Abstraction, Experience, and Indeterminacy in Computational Aesthetics*. Lanham: Rowman & Littlefield, 2018.
- Fyodor, Dostoevsky. *The Brothers Karamazov*, translated by Constance Garnett, New York: Modern Library, 1996.
- Giddens, Anthony. *The Consequences of Modernity*. Cambridge: Polity, 1990.

- Houser, Kristen. "Top AI Experts Vow They Won't Help Create Lethal Autonomous Weapons." *The Byte*, 7 July 2018. <https://futurism.com/lethal-autonomous-weapons-pledge/>
- Laux, Johann, Sandra Wachter, and Brent Mittelstadt. "Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk." *Regulation & Governance* 18, no. 1 (2023): 3–32.
- Mackenzie, Adrian. *Machine Learners: Archaeology of a Data Practice*. Cambridge, MA: MIT Press, 2017.
- McDonald, Glenn. "UN Will Take on 'Killer Robots' in 2017." *LIVESCINCE*, 23 December 2016. https://www.livescience.com/57306-un-addresses-killer-robots-in-2017.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Livesciencecom+%28LiveScience.com+Science+Headline+Feed%29
- Mint.co. "How to Automate Every Part of Your Life." *INTUITmintlife*, 23 April 2021. <https://mint.intuit.com/blog/relationships-2/how-to-automate-every-part-of-your-life-infographic-2873/>
- Nowotny, Helga. *In AI We Trust: Power, Illusion and Control of Predictive Algorithms*. Cambridge: Polity, 2021.
- Parisi, Luciana. "Recursive Philosophy and Negative Machines." *Critical Inquiry* 48, no. 2 (2022): 313–333.
- Pietralunga, Cédric. "Russia and Ukraine Take Drone Warfare to Unprecedented Scale." *Le Monde*, 18 June 2023. https://www.lemonde.fr/en/international/article/2023/06/18/russia-and-ukraine-take-drone-warfare-to-unprecedented-scale_6033281_4.html
- Schuett, Jonas. "Risk Management in the Artificial Intelligence Act." *European Journal of Risk Regulation* (2023): 1–19
- Simmel, Georg. *Simmel on Culture: Selected Writings*. Thousand Oaks: Sage, 1997.
- Stiegler, Bernard. *Automatic Society, Volume 1: The Future of Work*. Cambridge: Polity, 2018.
- Strobel, Warren P. "Military Drones Now Common to Nearly 100 Nations, Report Finds", *Wall Street Journal*, 25 September 2019. <https://www.wsj.com/articles/military-drones-now-common-to-nearly-100-nations-report-finds-11569403805> (Accessed 19 February 2024).
- Styles, Kirsty. "40 Countries Are Working on Killer Robots and There's No Law to Say How We Use Them." *TheNextWeb*, 21 January 2016. https://thenextweb.com/news/40-countries-are-working-on-killer-robots-and-theres-no-law-to-say-how-we-use-them#.tnw_35IqjXw

4 **Zuckerberg's Cave**

Smartness and Discipline in Digital Modernity

Kieron O'Hara

Introduction

Narratives of 'modernity' describe a type of society or culture, selectively emphasising aspects to express the novelty and innovation of relationships, practices and institutions. Modernity narratives may simultaneously be *descriptive* (certain types of change *are* happening), *teleological* (such changes are the *inevitable* destination of current developments) and *normative* (such changes *ought* to happen). Hence, narratives of modernity make sense of the present, enable the co-creation of futures, and inspire economic, military and technological development (Giddens 1991; Beck, Giddens, and Lash 1994; Bauman 2000; Bennett 2009; Wagner 2012).

Modernity and digital modernity narratives build on potential, and even hype, as much as on actuality, but when taken seriously by opinion-formers, academics, technologists, entrepreneurs and policymakers, especially super-rich ones like Sam Altman, Jeff Bezos, Bill Gates, Elon Musk and Peter Thiel, they become the imaginaries through which reality is shaped (Jasonoff 2015). As such, they have a strong influence over the futures we may inhabit. Such narratives are usually positive but do not have to be: "the dystopians join the utopians in imagining a supremely competent and visionary Silicon Valley" (Gilder 2018, 7). For instance, Yanis Varoufakis' recent Jeremiad about our enslavement by technology is very much in the tradition (Varoufakis 2023).

In this chapter, I will consider narratives of *digital modernity* that I postulate as a development of high, or reflexive, modernity, exhibiting both continuity and divergence. It has three substantive sections. The first describes the evolution of digital modernity out of reflexive modernity, from the perspective of their institutional dimensions and how they have been influenced by technological development. The second considers the evolution of another myth that has shaped perceptions of Western society for millennia, that of *Plato's Cave*. I shall argue in this chapter that, under the conditions of digital modernity, it has morphed into a form I shall call *Zuckerberg's Cave*. The third discusses some of the real-world developments that may threaten the neatness of these myths and narratives going forwards into the 2030s.

From Modernity to Digital Modernity

In previous articles, I have discussed the evolution of digital modernity through the physical dimensions of time and space (O'Hara 2021; O'Hara 2022, 65–116). However, in this chapter, I will draw upon Anthony Giddens's account of reflexive modernity as an irreducible group of four dimensions (1990, 55–63): capitalism, surveillance, industrialism and military force. *Capitalism* is of course our old friend the system of commodity production, in which capital is accumulated in the context of competitive labour and product markets. Private owners of capital employ labour for wages, while more or less competitive markets produce informative price signals that shape investors', producers' and consumers' choices. *Surveillance* is the supervision and logging of the decisions, speech acts and behaviour of individuals for administrative and sometimes political control. *Industrialism* is the coordination of human labour in the processes of harnessing energy to transform materials into tools, goods and environments for the delivery of services. *Military power* is the (near-)monopoly of the legitimate use or threat of violence in the context of the industrialisation of war to achieve the ends of securing the other dimensions, as well as the development of theories and practices that enable force to be projected without the actual (expensive and damaging) use of violence (for a diagrammatic representation of the dimensions, Giddens 1990, 59).

These dimensions are in balance in Giddens' presentation, with capitalism and industrialism as socioeconomic modes that make up what was understood as capitalist society in the late 20th century, secured by surveillance and military force. The 'triumph' of such societies over their main rivals, the Soviet states, led to a productive decade of thinking and debate about what their modernity consisted in. However, even while this was happening, the institutional balance described by Giddens was being undermined, not only by the flourishing of alternatives now released from the dead hand of communism but also by new technologies, especially the Internet and the World Wide Web.

The Ambient Data Infrastructure

The development of technology since 1990 has furnished the most significant development of digital modernity – the creation of a new dimension parallel to the four Giddens identified. Modernity's latest manifestation rests on the advent of ubiquitous networked digital technology, connecting people (especially via smartphones, but also technologies such as videoconferencing), places (increasingly using sensors, miniaturised devices and the Internet of Things or IoT), companies (especially through e-commerce and digital payments systems) and the state (through e-government, top-down control mechanisms and new ideologies such as data sovereignty) into what approximates a single network, implementing an *ambient data infrastructure*, able to generate, store, retrieve and reason about data in real time.

One way of understanding the infrastructure is as a *stack* of protocols and technologies (O'Hara and Hall 2021, 44–47), where each layer of the stack

performs a specific function within the network. One of the most important is the transport layer, which determines how the data moves through the network; others include the physical layer, which determines the means of encoding data into physical symbols (such as specific voltages) and its transmission through physical media such as wires, fibre-optic cabling or radio waves, and the application layer, which details the applications and programs that use the functionality of the infrastructure.

Technologies, protocols, hardware, software and systems that have enabled the ambient data infrastructure's stack include: smartphones; high-end specialised semiconductors such as graphics processing units; cloud computing; data science and big data analytics; the IoT, smart homes and cities; 5G wireless; artificial intelligence (AI), machine learning (ML), deep learning and generative AI; text, image, video and voice processing (and recognition); biometrics; autonomous vehicles, drones and non-humanoid robotics; augmented and virtual reality; distributed ledgers, blockchain and cryptocurrencies; 3D printing; and telemedicine. None of this was pre-ordained but has unrolled at a faster pace as innovation built upon innovation. For instance, generative AI as currently displayed in the ChatGPT chatbot series would not realistically be possible without the power provided by cloud computing or the massive quantities of data furnished by ubiquitous smartphones and social media, all effectively dating back to the brilliant invention of the Internet Protocol by Vint Cerf and Robert Kahn (O'Hara and Hall 2021, 27–35), the *sine qua non* of the whole suite.

All this has tended to effect one of the founding ideals of modernity, “that there is a fundamental common interest which is inherent in society and which, once disclosed, will supervene over all other interests of the respective parties; it assumes rigorously persuasive rationality and relevant empirical knowledge of a high degree of precision and reliability” (Shils 1997, 85). To that degree, while the ambient data infrastructure is a divergence from the dimensions of reflexive modernity, it at the same time presents continuity.

The “rigorously persuasive rationality” afforded by the ambient data infrastructure comprises the progressive development of representations and inferential procedures for reflexivity, probabilistic inference at scale and managing tradeoffs in effort, precision and timeliness, to maximise the expected utility of outputs (Gershman, Horwitz, and Tenenbaum 2015). Thus, the ideal of mastery over a deterministic universe is jettisoned in favour of *computational rationality*, a probabilistic model driven by experimentation, A/B testing, perpetual beta and other heuristic optimisation methods. After the disappointments of the Enlightenment, the hope is reopened for “humans to attain the powers typically associated with the divine” (Horvath, Roman, and Germain 2018, 1).

Given computational rationality running over a ubiquitous ambient data infrastructure, we can begin to think of it, possibly in fanciful caricature, as a ‘digital body politic.’ Pentland argues that big-data systems with feedback form a “nervous system” to maintain the stability of government, energy, health and transport systems, which could “reinvent societies’ systems within a control framework: one that first senses the situation; then combines these observations with models

of demand and dynamic reaction; and, finally, uses the resulting predictions to tune the systems to match the demands being made of them” (Pentland 2014, 138). Sensors and the IoT make the nervous system, AI the brain, blockchain the memory (Tay 2019) and social media the limbic system for collective precognitive function (Cohen 2020). McGilchrist suggested that modernity is biased towards left-brain thinking: “the left hemisphere sees things abstracted from context, and broken into parts, from which it reconstructs a ‘whole’: something very different” from right hemisphere holism (2019, 27–28). But as the scale of AI operations increases, it provides more evidence of the holistic reasoning characteristic of the right hemisphere – or at least, we find it harder to detect the ‘joins’ of the left-sided reconstruction thanks to technologies such as deep learning and generative AI.

Other Institutional Dimensions

Giddens originally presented the four dimensions as balanced with symmetrical connections between them all. The addition of the further dimension of the ambient data infrastructure has complicated the evolving picture of modernity, as we see in Figure 4.1. The ambient data infrastructure is now the basis of the whole structure, the enabler of everything else. However, Giddens’ dimensions have evolved (and are evolving) as a result of the functionality provided by the ambient infrastructure, and two of them have effectively merged – hence we remain at four dimensions in total. Whereas Giddens emphasised the connections between all of his four dimensions, Figure 4.1 is sparser.

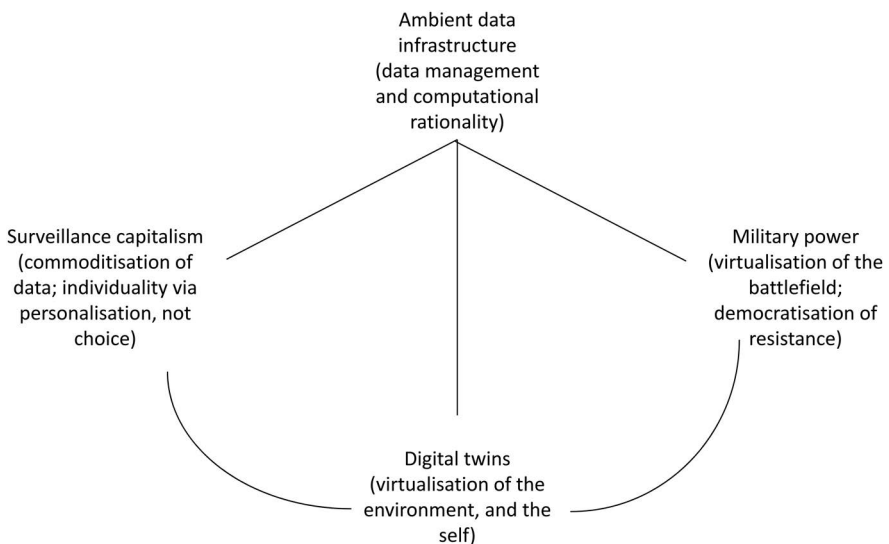


Figure 4.1 The institutional dimensions of digital modernity.

Surveillance Capitalism

The most important innovation is the effective merging of the dimensions of capitalism and surveillance, to produce, as the second dimension, what has been called *surveillance capitalism* (Zuboff 2019). As more services move online, greater quantities of data can be collected about people's behaviour, which can be analysed in real time to manipulate the environment to favour choices that are likely to satisfy consumer demand. Increasing data provides feedback to commercial systems and so is valuable to them; consumers can therefore be offered services at below cost (typically free) in exchange for the use of their personal data. Knowledge about consumers, and about the clusters they occupy, allows the customisation of services, so that modernity's key value of individuality is expressed not by individuals' choices (as in reflexive modernity) but by *personalisation* (O'Hara 2021).

Some would dispute that surveillance capitalism is the best name for the new merged dimension, but usually that is based on a quibble about how capitalism is to be defined. For instance, Varoufakis confuses capitalism with free markets, whose extinction makes him believe that capitalism is dead – but even he agrees that markets have not disappeared. He simply denies that an online trading platform is a market, because the owner of the platform is able to reap the data from all the transactions and thus become more powerful than everyone in the market (Varoufakis 2023). He has coined his own term for this, *technofeudalism*, but as I disagree with the implication that capitalism has collapsed, I will continue to use Shoshana Zuboff's term, which also handily mirrors the two ancestors of this dimension in Giddens' work.

Digital Twins

The other two dimensions have also been transformed by the infrastructure. Giddens' dimension of industrialism was specified in terms of the processes involved in transforming nature; the ambient data infrastructure, in contrast, allows both nature and the industrialised environment to be transformed, not only physically but through *virtualisation*.

Modernity has often been described in spatial terms, negatively contrasting a *periphery*, rural areas, edgelands, liminal spaces and the developing world, with the *centre*, major cities, hubs, centres of excellence, clusters of creativity and industry where modernisation happens and value is created (Shils 1975; Arendt 1998, 201). As we near the centre, space collapses: "cities are the absence of physical space between people and companies" (Glaeser 2011, 6). The periphery, meanwhile, can modernise through development, integrating with global networks (Castells 2000, 407–459) using space-shrinking technologies such as fast broadband and 5G.

In digital modernity, as networked devices create and support increasingly many important relationships (e.g. via social media or super-apps), functional space shrinks further, to within the device – to its representations of human and non-human actors (Lupton 2015, 23–27). This space has had many names, including *cyberspace*, "A consensual hallucination experienced daily by billions of legitimate

operators, in every nation” (Gibson 1984, 69); the *information superhighway*; *real virtuality* “in which appearances are not just on the screen through which experience is communicated, but they become the experience” (Castells 2000, 404, his emphasis); and the *metaverse*, a synthetic parallel world (Ball 2022), honed by online games (Castronova 2005; Barnett and Coulson 2010), 3D virtual environments (Gunkel et al. 2018; Zhan et al. 2020) and augmented reality (Carmigniani and Furht 2011; Zhan et al. 2020).

At the time of writing, the metaverse, which was strongly hyped in 2018–2022, looks something like a busted flush – I shall examine some of the reasons for this later in the paper. However, its motivating idea still functions as an important representative of a particular intellectual tradition of asserting the equality, or even the priority, of the representation over whatever we take to be reality (Soraker 2014; Chalmers 2022). As a series of persistent virtual spaces unified by a platform infrastructure to support computational rationality and data analysis (Zhan et al. 2020; Lee et al. 2021), the metaverse was trumpeted in the worlds of fashion, real-estate (trading virtual properties) and streamed events (pop group ABBA relaunched themselves successfully in 2022 with on-stage “Abbatars”). Yet this functionality remains important, as it supports the dynamic reconstruction and reconfiguration of physical reality, e.g. providing opportunities to do things virtually that are impossible, too dangerous or too expensive to do physically (Bailenson 2018; Duan et al. 2021). Identities, gender, ethnicity and other markers may be altered or even created (Wallis and Ross 2021), with virtual ethical codes (Sparrow, Gibbs, and Arnold 2021), pushing back at fixed ideas about what it means to be human or social.

In the long run, a development likely to be of greater importance than the metaverse is that of *digital twins*, models or simulations of real-world originals (Boschert and Rosen 2016; Batty 2018). A digital twin can be characterised as a comprehensive functional description of a person, component, mechanism or system, which includes the information which will be useful for understanding (and manipulating) it across its lifecycle. The opportunities for creation and manipulation of digital twins will only get greater as AI, cognitive psychology and neuroscience converge to enhance computational rationality, while AI may make it possible for the twin to reason about its own performance.

Digital twins afford opportunities for order, calculation and rationality, implementing the Enlightenment dream of a society in which human needs are accurately identified and supplied. The ambient data infrastructure knows individuals better than they do themselves, because it has far more information about them, as well as events, actions and judgements involving people with similar attributes. ML can judge individuals’ optimal choices and behaviours *if only they had the relevant information with which to calculate their own interests*. This is a *subjunctive* world: people are told what choices they *would* have made if only they knew as much as the data infrastructure (O’Hara 2021). Such choices may be consumer choices leading to customer satisfaction, or lifestyle choices leading to better health outcomes, or behavioural choices leading to optimal social conditions (e.g. reducing carbon emissions or traffic congestion). Environments can be subtly altered to implement the ‘nudge’ philosophy (Thaler and Sunstein 2008); the job description

of ‘choice architect’ has appeared (Johnson et al. 2012). Abundant data even increases the options for personalised nudges (Mills 2022) and self-nudging, delivering nudges to oneself (Reijula and Hertwig 2022). The data no longer describes the world; the world follows, or is disciplined by, the data.

Meanwhile, as these technologies advance (and occasionally retreat), digital modernity is terraforming the physical world to assert its priority (O’Hara 2015). Restaurants and other venues converge on the style of Instagram (Frier 2020, 159–178). Apps such as Facetune and Adobe Lightroom improve facial appearance in photos, influencing self-image (Fardouly and Vartanian 2016; Kleemans et al. 2018; Fatt, Fardouly, and Rapee 2019) and inspiring a boom in plastic surgery, unsurprisingly welcomed in the plastic surgery industry (Youn 2018; Basa and Spiegel 2020). The quantified self philosophy (Lupton 2016; Ruckenstein and Pantzar 2017) uses data from wearable devices such as fitness trackers, smart watches, biosensors, mood trackers and monitors of specific functions such as ECGs or blood pressure, as well as devices embedded in the body, to render individuals as quantified systems, allowing digital feedback for self-optimisation or biohacking (Yetisen 2018; Gangadharbatla 2020). The Chinese social credit system (Chin and Lin 2022) is “generative of a distinctive personhood that concretizes China’s ongoing social transformation” (McDonald and Dan 2021, 87), through successive processes of depersonalisation (e.g. automating credit allocation systems) and repersonalisation (as citizens infer a ‘logic’ underlying the system). But self-optimisation is really optimisation of the *data* (Kristensen and Ruckenstein 2018; Pantzar and Ruckenstein 2017).

Digital modernity, then, produces a kind of *digital Gnosticism*, in which the data is prior or more important than the atoms of the physical world. It facilitates a means of discipline, consistent with one of its key principles:

In digital modernity, the best that hapless reality can achieve is to get closer to the perfection of the algorithm and the data.

(O’Hara 2022, 86)

Military Power

The fourth dimension, that of *military power*, itself is fascinating, although space forbids a detailed exploration. Suffice it to say that the ambient data infrastructure has turned the battlefield into a digital twin. Warfare is practically impossible today without the Internet, and Elon Musk’s provision of Starlink technology for the Ukrainians is a *sine qua non* of their military and civil resistance to the Russians. Not only has the invasion featured prominently on social media, enabling the open-source intelligence that is now at least as important as more traditional sources, but Starlink’s low orbit means a low latency for signals to travel to and from the satellites – which is crucial, for instance, for coordinating an artillery attack. Pictures of potential targets can be uploaded to informal networks, and commanders can decide whether to attack – in effect crowdsourcing firepower. The power this gives Musk is extraordinary, perhaps even greater than the major armaments manufacturers during the World Wars.

Russia's invasion has been a drone war – autonomous flying vehicles and seaborne drones have been responsible for many attacks, as the Ukrainians have taken advantage of Turkish technology, while the Russians have been buying from Iran (neither are traditionally prolific arms suppliers). AI can take decisions in a split second – also taking into account the outputs of the enemy's AI. Any signal from the battlefield, not only from phones but also heat and sound, as well as minor disruption to the environment (e.g. tyre tracks) will be detectable within seconds by an enemy, with immediate consequences as a smart shell is sent on its way. Smart warfare is a draining business; forces must stay small, compact and constantly on the move. Sleep may prove impossible (Joshi 2023).

The battlefields of Ukraine have, as a consequence of all this, an order of magnitude fewer personnel on them than would have been likely in the mechanical wars of the 20th century. Yet the number of humans involved has increased, not decreased, because the use of the Internet means that anyone can get involved, for instance posting photos of enemy soldiers or weaponry. Tech wizards in Ukraine have been hacking into Russian cyber-systems, opening commercial CCTV systems to intelligence services, and designing and flying improvised drones.

"The successful monopoly of the means of violence within territorially precise borders" (Giddens 1990, 58) no longer holds. Non-state actors are used by states to carry out military functions, a development that has been called *neomedievalism* (McFate 2014), performing many services in training and on the battlefield, but crucially in cyberwarfare, cyberterrorism and cyberespionage. Furthermore, not only has the production of violence become much cheaper and democratised, but also the notions of what constitutes violence (taking a selfie can be part of the defence of one's community, given the massive amount of metadata that a smartphone photo carries with it), and where it takes place (the opening conflicts of Russia's invasion were cyberattacks and entirely virtual) are evolving too.

Ethically – even if the Russians were not employing war crime as a standard means of offence – this is undermining many of the traditional principles of international humanitarian law. Furthermore, in a savage irony, a side-effect of creating a digital twin of the battlefield, owing to the tactics of hyper-destruction pioneered by Russia in Syria and elsewhere, is the ultimate annihilation of the physical environment that had been modelled.

The revised map of the dimensions of digital modernity is therefore shown in Figure 4.1. The three lower dimensions are all enabled by the ambient data infrastructure. It provides the data and reasoning power to create digital twins or a significant virtual world, whether virtual reality, the metaverse or a sophisticated computational model changing dynamically in real time. Such models power the surveillance economy and are transforming the battlefield.

From Plato's Cave to Zuckerberg's

The status of the virtual or intellectual world is not a question that originated with digital modernity. It was already in discussion in Socrates' metaphor of the cave from Plato's *Republic*, which contrasted the perceptions of a group of prisoners

only able to see shadows cast by a flickering fire behind them, with those of someone who had escaped from the cave and seen the outside world in clear light (Plato 1997, 1132–1135). Socrates’ description suggests that our knowledge of the everyday physical world is imperfect, whereas precision and perfect knowledge are only possible in the abstract, logical, mathematical, conceptual world of ideas, thus reversing an earlier Homeric picture (Arendt 1998, 292). This is a precursor of the principle that reality can’t beat the data.

The cave is a permanent prison for a pre-modern person. While a craftsman or woman can try to create objects guided by the ideas in their heads, the model they attempt to reproduce “possesses a degree of permanence and excellence which is not actualized but on the contrary spoiled in its materialization through the work of human hands. Work makes perishable and spoils the excellence of what remained eternal so long as it was the object of mere contemplation” (Arendt 1998, 303).

The key distinction is made, but nowhere does Plato suggest the distinctively modern thought experiment that the external, physical world does not exist. For that we had to wait for the *cogito ergo sum* argument of René Descartes’ *Discourse on Method* (1637), one of the intellectual foundations of modernity (Cahoone 1987; Albano 2000). Descartes’ refusal to submit to appearances, instead relying on the subjective rationality of the individual’s cognition, it has been argued, was also mirrored in contemporary or even earlier philosophical shifts in some non-European cultures (Lu 2019).

The construction of the subjective world, a demonic miracle for Descartes, received a scientific shift from Hilary Putnam, who suggested the possibility that, for all they know, people may be brains in a vat with a powerful computer simulating reality by sending electrical signals direct to their neurons (Putnam 1981). Though both Descartes and Putnam raised these possibilities only to debunk them, the idea has been thought compelling (Chalmers 2022). Technologically, it is the imaginary that underlies virtual reality and would clearly require the use of AI to achieve real-time realism in the simulated world.

These ideas reached their popular fictional apogee in the 1999 film *The Matrix*, which among other things laid down the challenge of the red pill and the blue pill. Taking the latter allows the character to live peacefully in the virtual paradise, while the former shows him or her the true nature of reality – an allegory that still falls short of digital modernity while accepting its being and the force of its representational abilities. It falls short because it argues that, while the virtual experience is perfect, it is not *correct*. Red-pilling reveals the more basic physical reality (the metaphor appeals to alt-right activists who claim that society and the media sell us bromides and a *false* sense of security).

What about an updated digitally modern cave, which (with a nod to the champion of the metaverse) we might call *Zuckerberg’s Cave*? This remains faithful to Plato’s contrast between the uncertainties of the physical world and the abstract, logical, precise world of data and ML. However, now there is no need for escape to the outside to calibrate the prisoners – ML could record and refine the flickering shadows in order to recover accurate signals of whatever caused them (admittedly it would need authoritative feedback for its training). From the noisy and imprecise

data, the informative signal could be extracted and communicated direct to the prisoners without releasing them. In Zuckerberg's cave, digital modernity does not postulate a separate realm so much as promise an epistemologically transparent rendering of the existing physical world. As José Saramago put it in his prescient novelistic rendering of the rediscovery of Plato's cave, the commercial infrastructure that sits on top of the cave is a perfect distributor of both material and spiritual goods with something of the divine about it, "not the work of vile matter but sublime spirit." It "writes straight on crooked lines" (Saramago 2002, 244).

Unlike Plato's prisoners, Zuckerberg's don't need to escape, for two reasons. First, if they do, they become *less*, not *more*, informed because they have to rely on their (imperfect) senses, not the perfection of the analysed data. Human agency has thus been scathingly dismissed as "residual anthropological signature" (Land 2019, 519), ultimately to be superseded by AI. And second, the physical world takes the data analysis as feedback *to which it is expected to conform*. As with the quantified self, recommended choices or data-driven policymaking, reality takes its cue from the data, not the other way round. The prisoners should rejoice because they have a ringside seat on what reality will become. Those in Plato's cave, or Putnam's vat or tormented by Descartes' demon, are denied understanding of their world and are unaware of their predicament. But dwellers in the metaverse are *fully* aware, *preferring* the discipline supplied by the virtual and in the ideal, accepting its priority over the physical. Digital modernity hits a sweet spot for the prisoners.

Reality Bites

These, of course, are narratives and are not constrained to happen. The ambient data infrastructure's power may be less than anticipated, Zuckerberg's cave may not be an effective prison, and the subjunctive world may not emerge. Indeed, despite the narrative's claim that the online and offline worlds are merging, we know that the offline world retains its own identity and integrity – it is what we are left with during a power cut. No narrative perfectly describes what will happen, so digital modernity will only be partially implemented. But could it fail altogether? In this section, I will briefly sketch three potential blockers: geopolitics, cybersecurity and economics.

The first is geopolitics. The ambient data infrastructure grew more or less organically in its earlier phases. As its main two components, the Internet and the World Wide Web, are both permissionless, all that joining required was to build something with compatible protocols. However, in more recent years, geopolitical realities have conspired to put a nationalist spin on cyberspace. It was no accident that the Internet and Web grew most dramatically in the 1990s, when a single superpower dominated. Now that the monopolar moment is over, newly-assertive powers aim to exert control, including the EU (Bradford 2020; O'Hara and Hall 2021, 77–91), China (O'Hara and Hall 2021, 125–144) and India (O'Hara and Hall 2021, 183–199).

Recall the ambient data infrastructure as a stack. In order to achieve greater control over (or at least more information about) the use of the infrastructure, governments create their own stacks of technologies and protocols that modify the

basic stack. Citizens can then be steered towards the national stack, either with the threat of punishment if foreign technologies are used (as with Russia) or by increasing the costs of using alternatives (as with India).

Russia has the greatest need for its own stack (O'Hara and Hall 2021, 154–172), both to prevent the flow of news into the country, and to increase resilience against Western sanctions following its illegal invasions of Ukraine. In this respect, the Russian infrastructure stack mirrors its own economy, which moved onto an autarkic footing shortly after its annexation of Crimea in 2014. Putin long saw the Internet as a CIA project but got around to manipulating it rather late (O'Hara and Hall 2021, 163). The state now censors drastically, while its local service providers are usually run by state proxies (e.g. social network VK is owned by a subsidiary of Gazprom). The government itself is using its e-government programmes and procurement power to push home-grown alternatives to key Western services such as search, business software, microblogging and video-sharing. Internet Service Providers have strict instructions too to control dataflow (especially from outside Russia) and to disable specific services, e.g. to block the anonymous routing software Tor. Finally, at the physical level, its 5G mobile network suppliers are weak, and it lacks semiconductor expertise (its designs are poor and it usually needs them to be built in pro-Western places such as South Korea and Taiwan). On the other hand, some of its capabilities, such as vertically integrated super-apps, digital payments, security systems and cloud services, are impressive (Economist 2022).

China has a similar suite of capabilities in its own stack, including state-of-the-art surveillance and face recognition systems. The Communist Party allows private enterprise as long as it is aligned with Chinese policy, and for a period between 2020–22, to show who is boss, the Party cracked down on the tech industry, wiping over \$1 trillion off the value of the sector. On the other hand, it also generously funds strategic technologies, such as semiconductors, AI, robotics and the IoT, as well as fostering technology hubs in specific regions. Its vague privacy and secrecy laws, combined with a cavalier attitude to the rule of law, mean that companies self-censor rather than risk its wrath. China also exports its technology via its Belt and Road initiative, e.g. building smart cities (very surveillance-heavy and branded as safe cities) in the global South. India's own stack, based around its Aadhaar identity system and Unified Payment Interface digital payment system, is specifically created for export (O'Hara and Hall 2021, 196–199). These exports will certainly increase the soft power of both countries in the future.

Already, mutual suspicion has led to a fracturing of the ambient data infrastructure; apps, semiconductors, cloud services and networking equipment are often subject to sanctions and less likely to cross borders. A serious bust-up – such as a Chinese invasion of Taiwan – will only accelerate the fragmentation. Indeed, Taiwan is the world leader in cutting-edge semiconductor design, an essential industry that may be shattered by an invasion, thus undermining the infrastructure. Yet a large and integrated data infrastructure is needed to produce the giant quantities of data required by digital modernity. The demands of generative AI alone are colossal, in terms of data and cloud computing power (even if efficiencies will surely be found).

If data cannot be amalgamated across borders, only very large polities – India, China, the United States and the EU stand out – will be able to amass sufficient resources to push the technology further.

A related problem concerns the security of the infrastructure. Cybersecurity, cybercrime and misinformation are already massive businesses, and it may ultimately prove that governments find the Internet to be too large an attack vector. Hackers may be independent criminals, government-sponsored or increasingly, deniable nationalists encouraged by, but at arms-length from, their government; Russia has almost perfected such dark arts, but most governments have expertise in the area (O'Hara and Hall 2021, 154–172). Certain technologies may make this problem intolerable – at the time of writing, there is much angst over the negative potential for generative AI, but we should also consider the so-far nascent technology of quantum computing, which would be able to undo decades of cryptographic practice. Shor's algorithm (Wong 2022, 341–343), a quantum algorithm for finding an integer's prime factors, renders the factorisation problem which underpins public key cryptography solvable in polynomial time (i.e. quickly enough to make a difference).

A third problem is economics. At a key period in the development of the ambient data infrastructure (c.2007–22), interest rates were effectively zero, thanks to the post-crisis recession. This meant that investors were remarkably tolerant of companies which grew but did not earn profits because future profits needed little discounting. This funded a great deal of (sometimes productive) technological investment that nevertheless did not immediately produce profits in the near term. Facebook/Meta took the lead in developing the metaverse, followed by Nvidia and Microsoft. Several blockchain and cryptocurrency applications reached extraordinary valuations. Many startups, such as WeWork and Uber, presented themselves as tech companies liable to benefit from network effects. However, with the return of global inflation, higher interest rates mean that investors are suddenly more interested in where profits might come from. Meta's virtual reality division, Reality Labs, has lost billions of dollars, and investors have grown impatient. The blockchain world looks less inviting after several high-profile crimes and collapses. And it has been realised that many startups will not be able to leverage network effects. WeWork's IPO collapsed when it was realised that, far from being a tech giant worth billions, it was actually a rather badly run office space rental company (Brown and Farrell 2021).

There is still plenty of money coming into tech. Generative AI is the latest excitement at the time of writing. But can it make money? ChatGPT 4.0 cost over \$100 million to train, and its developers OpenAI lost a further \$400m on top of that in 2022, through essential research into modelling techniques (Economist 2023). OpenAI is funded at the time of writing largely by Microsoft, but it, and its shareholders, may stop writing blank cheques if income streams don't emerge soon. For years, tech has been the only game in town for investors seeking big returns, but the transition to a green economy may soon provide alternatives, from new batteries to renewable energy to mining for essential metals like copper, nickel and lithium.

Conclusion: The Prospects for Counternarratives and Resistance

Thus stands the narrative(s) of digital modernity. Are there counternarratives available? We could deny that our current context is a species of modernity, but post-modernism's decadence and trivialisation of principle are hardly likely to wean people off their smartphones (O'Hara 2022, 144–152).

Alternatively, we could reconstruct the ambient data infrastructure, e.g. relocating it in the blockchain, providing transparency of data use and control over data (Gilder 2018). The problem with that option is that there is no evidence that ordinary data subjects are interested in pursuing their data protection rights and exercising such control (O'Hara 2023, 278–282). Like a sugar mountain, the discipline offered by digital modernity is too sweet, unavoidable and tempting for many to resist. And 'sugar mountain' translated to German is, yes, *Zuckerberg!* Spooky!

Ultimately, the potential for top-down control offered by digital modernity may also be too tempting for policymakers, tasked with mitigating climate change, pollution, congestion, public health and security, to ignore. Furthermore, choice facilitated by the Internet can seem problematic from the perspective of authority, e.g. with pornography, populism, disinformation or self-harm sites. It remains to be seen whether smart cities or nudge paternalism live up to their rhetoric, or whether online harms can be suppressed, but digital modernity affords an opportunity for political elites to regain the control they have lost (or given away), if they can wrest it from private tech giants. The Chinese government flexed its muscles successfully with Tencent, Alibaba and Didi Chuxing in 2020–22 and brought them down to size, but this might well affect their innovative capabilities, especially with respect to AI (Kshetri 2023). Whether the US or EU can rein in the Silicon Valley giants, as they intend, without destroying the network value they create will be a seminal question for the remainder of the 2020s.

Acknowledgements

Thanks to Kristof Nyiri for comments on an earlier version of this paper, and thanks also to an audience at the 10th Budapest Visual Learning Conference, 2022, where an early version was presented under the title "Speleology in the Metaverse: From Plato's Cave to Zuckerberg's" (and appears in the Proceedings, http://www.hunfi.hu/nyiri/FFF/Vol_5_final.pdf). Some passages have also been taken from my paper "Digital Modernity" (O'Hara 2022). Thanks of course to the editors for comments on the evolving paper.

References

- Albano, Peter J. 2000. "The Cogito, Human Self-Assertion, and the Modern World." *Philosophy Today*, 44(2), 184–189. <https://doi.org/10.5840/philtoday200044247>
- Arendt, Hannah. 1998. *The Human Condition*, 2nd edition. Chicago: University of Chicago Press.
- Bailenson, Jeremy. 2018. *Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do*. New York: W.W. Norton.

- Ball, Matthew. 2022. *The Metaverse: And How It Will Revolutionize Everything*. New York: W.W. Norton.
- Barnett, Jane and Mark Coulson. 2010. "Virtually Real: A Psychological Perspective on Massively Multiplayer Online Games." *Review of General Psychology*, 14(2), 167–179. <https://doi.org/10.1037/a0019442>
- Basa, Krystyne and Jeffrey H Spiegel. 2020. "Facial Plastic Surgery on Instagram: What Is Trending? What Is Working?" *Aesthetic Surgery Journal*, 41(7), 846–851. <https://doi.org/10.1093/asj/sjaa374>
- Batty, Michael. 2018. "Digital Twins." *Environment and Planning B: Urban Analytics and City Science*, 45(5), 817–820. <https://doi.org/10.1177/2399808318796416>
- Bauman, Zygmunt. 2000. *Liquid Modernity*. Cambridge: Polity Press.
- Beck, Ulrich, Anthony Giddens, and Scott Lash, eds. 1994. *Reflexive Modernization: Politics, Tradition and Aesthetics in the Modern Social Order*. Cambridge: Polity Press.
- Bennett, Jane. 2009. "Modernity and Its Critics." In *The Oxford Handbook of Political Science*, edited by Robert E. Goodin, 127–138. Oxford: Oxford University Press.
- Boschert, Stefan and Roland Rosen. 2016. "Digital Twin: The Simulation Aspect." In *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and Their Designers*, edited by Peter Hehenberger and David Bradley, 59–74. Cham: Springer. https://doi.org/10.1007/978-3-319-32156-1_5
- Bradford, Anu. 2020. *The Brussels Effect: How the European Union Rules the World*. New York: Oxford University Press.
- Brown, Eliot and Maureen Farrell. 2021. *The Cult of We: WeWork and the Great Startup Delusion*. London: Mudlark.
- Cahoone, Lawrence E. 1987. *The Dilemma of Modernity: Philosophy, Culture, and Anti-Culture*. Albany: State University of New York Press.
- Carmigniani, Julie, and Borko Furht. 2011. "Augmented Reality: An Overview." In *Handbook of Augmented Reality*, edited by Borko Furht, 3–46. New York: Springer. https://doi.org/10.1007/978-1-4614-0064-6_1
- Castells, Manuel. 2000. *The Information Age: Economy, Society and Culture Volume 1: The Rise of the Network Society*, 2nd edition. Malden: Blackwell.
- Castronova, Edward. 2005. *Synthetic Worlds: The Business and Culture of Online Games*. Chicago: University of Chicago Press.
- Chalmers, David J. 2022. *Reality+: Virtual Worlds and the Problems of Philosophy*. London: Allen Lane.
- Chin, Josh and Liza Lin. 2022. *Surveillance State: Inside China's Quest to Launch a New Era of Social Control*. New York: St Martin's Press.
- Cohen, Julie. 2020. "The Emergent Limbic Media System." In *Life and the Law in the Era of Data-Driven Agency*, edited by Mireille Hildebrandt and Kieron O'Hara, 60–79. Cheltenham: Edward Elgar.
- Duan, Haihan, Jiaye Li, Sizheng Fan, Zhonghao Lin, Xiao Wu, and Wei Cai. 2021. "Metaverse for Social Good: A University Campus Prototype." In *MM '21: Proceedings of the 29th ACM International Conference on Multimedia*, 153–161. New York: ACM, 2021. <https://doi.org/10.1145/3474085.3479238>
- The Economist. 2022. "Russia is Trying to Build its Own Great Firewall." *The Economist*, 19 February 2022. <https://www.economist.com/business/russia-is-trying-to-build-its-own-great-firewall/21807706>
- The Economist. 2023. "Could OpenAI Be the Next Tech Giant?" *The Economist*, 18 September 2023. <https://www.economist.com/business/2023/09/18/could-openai-be-the-next-tech-giant>
- Fardouly, Jasmine and Lenny R Vartanian. 2016. "Social media and Body Image Concerns: Current Research and Future Directions." *Current Opinion in Psychology*, 9, 1–5. <https://doi.org/10.1016/j.copsyc.2015.09.005>
- Fatt, Scott J, Jasmine Fardouly, and Ronald M Rapee. 2019. "#malefitspo: Links Between Viewing Fitspiration Posts, Muscular-Ideal Internalisation, Appearance Comparisons,

- Body Satisfaction, and Exercise Motivation in Men.” *New Media and Society*, 21(6), 1311–1325. <https://doi.org/10.1177/1461444818821064>
- Frier, Sarah. 2020. *No Filter: The Inside Story of How Instagram Transformed Business, Celebrity and Our Culture*. London: Random House.
- Gangadharbatla, Harsha. 2020. “Biohacking: An Exploratory Study to Understand the Factors Influencing the Adoption of Embedded Technologies Within the Human Body.” *Heliyon*, 6(5), e03931. <https://doi.org/10.1016/j.heliyon.2020.e03931>
- Gershman, Samuel J, Eric J Horwitz, and Joshua B Tenenbaum. 2015. “Computational Rationality: A Converging Paradigm for Intelligence in Brains, Minds and Machines.” *Science*, 349(6245), 273–278. <https://doi.org/10.1126/science.aac6076>
- Gibson, William. 1984. *Neuromancer*. New York: Ace Books.
- Giddens, Anthony. 1990. *The Consequences of Modernity*. Cambridge: Polity Press.
- Giddens, Anthony. 1991. *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Cambridge: Polity Press.
- Gilder, George. 2018. *Life After Google: The Fall of Big Data and the Rise of the Blockchain Economy*. Washington DC: Regnery Gateway.
- Glaeser, Edward. 2011. *Triumph of the City: How Urban Spaces Make Us Human*. London: Macmillan.
- Gunkel, Simon, Hans Stokking, Martin Prins, Omar Niamut, Ernestasia Siahaan, and Pablo Cesar. 2018. “Experiencing Virtual Reality Together: Social VR Use Case Study.” In *TVX '18: Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video*, 233–238. New York: ACM, 2018. <https://doi.org/10.1145/3210825.3213566>
- Horvath, Agnes, Camil Francisc Roman, and Gilbert Germain. 2018. “Introduction: Divinisation and Technology – the Political Anthropology of Subversion.” In *Divinization and Technology: The Political Anthropology of Subversion*, edited by Horvath Agnes, Camil Francisc Roman, and Germain Gilbert, 1–13. London: Routledge.
- Jasonoff, Sheila. 2015. “Future Imperfect: Science, Technology and the Imaginations of Modernity.” In *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, edited by Sheila Jasonoff and Sang-Hyun Kim, 1–33. Chicago: University of Chicago Press.
- Johnson, Eric J, Suzanne B Shu, Benedict G.C Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, David Schkade, Brian Wansink, and Elke U Weber. 2012. “Beyond Nudges: Tools of a Choice Architecture.” *Marketing Letters*, 23(2), 487–504. <https://doi.org/10.1007/s11002-012-9186-1>
- Joshi, Shashank. 2023. “The War in Ukraine Shows How Technology Is Changing the Battlefield.” *The Economist*, 3 July 2023. <https://www.economist.com/special-report/2023/07/03/the-war-in-ukraine-shows-how-technology-is-changing-the-battlefield>
- Kleemans, Mariska, Serena Daalmans, Ilana Carbaat, and Doeschka Anschutz. 2018. “Picture Perfect: The Direct Effect of Manipulated Instagram Photos on Body Image in Adolescent Girls.” *Media Psychology*, 21(1), 93–110. <https://doi.org/10.1080/15213269.2016.1257392>
- Kristensen, Dorthe Brogård, and Minna Ruckenstein. 2018. “Co-Evolving with Self-Tracking Technologies.” *New Media and Society*, 20(10), 3624–3640. <https://doi.org/10.1177/1461444818755650>
- Kshetri, Nir. 2023. “Can China Catch Up to the United States in Generative Artificial Intelligence?” *IT Professional*, 25(4), 94–97. <https://doi.org/10.1109/MITP.2023.3297384>
- Land, Nick. 2019. “Teleoplexy: Notes on Acceleration.” In *#Accelerate: The Accelerationist Reader*, 3rd edition, edited by Robin Mackay and Armen Avanessian, 509–520. Falmouth: Urbanomic.
- Lee, Lik-Hang, Tristan Braud, Pengyuan Zhou, Lin Wang, Dianlei Xu, Zijun Lin, Abhishek Kumar, Carlos Bermejo, and Pan Hui. 2021. “All One Needs to Know About Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda.” *arXiv*, <https://arxiv.org/abs/2110.05352>

- Lu, Mingjun. 2019. "The Cogito and Onto-Being of the Mind: Philosophical Early Modernity in Descartes' and Wang Yangming's Metaphysics." *Philosophy East and West*, 69(1), 176–196. <https://doi.org/10.1353/pew.2019.0005>
- Lupton, Deborah. 2015. *Digital Sociology*. Abingdon: Routledge.
- Lupton, Deborah. 2016. *The Quantified Self*. Cambridge: Polity Press.
- McDonald, Tom and Li Dan. 2021. "Alipay's 'Ant Credit Pay' Meets China's Factory Workers: The Depersonalisation and Re-Personalisation of Online Lending." *Journal of Cultural Economy*, 14(1), 87–100. <https://doi.org/10.1080/17530350.2020.1763424>
- McFate, Sean. 2014. *The Modern Mercenary: Private Armies and What They Mean for World Order*. New York: Oxford University Press.
- McGilchrist, Iain. 2019. *The Master and His Emissary: The Divided Brain and the Making of the Western World*. New Haven: Yale University Press.
- Mills, Stuart. 2022. "Personalized Nudging." *Behavioural Public Policy*, 6(1), 150–159. <https://doi.org/10.1017/bpp.2020.7>
- O'Hara, Kieron. 2015. "Data, Legibility, Creativity ... and Power." *IEEE Internet Computing*, 19(2), 88–91. <https://doi.org/10.1109/MIC.2015.34>
- O'Hara, Kieron. 2021. "Personalisation and Digital Modernity: Deconstructing the Myths of the Subjunctive World." In *Data-Driven Personalisation in Markets, Politics and Law*, edited by Uta Kohl and Jacob Eisler, 37–54. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108891325.004>
- O'Hara, Kieron. 2022. "Digital Modernity." *Foundations and Trends in Web Science*, 9(1-2), 1–260. <http://dx.doi.org/10.1561/18000000031>
- O'Hara, Kieron. 2023. *The Seven Veils of Privacy: How Our Debates About Privacy Conceal Its Nature*. Manchester: Manchester University Press.
- O'Hara, Kieron and Wendy Hall. 2021. *Four Internets: Data, Geopolitics, and the Governance of Cyberspace*. New York: Oxford University Press.
- Pantzar, Mika and Minna Ruckenstein. 2017. "Living the Metrics: Self-Tracking and Situated Objectivity." *Digital Health*, 3. <https://doi.org/10.1177/2055207617712590>
- Pentland, Alex. 2014. *Social Physics: How Good Ideas Spread – The Lessons From a New Science*. Brunswick: Scribe.
- Plato. 1997. "Republic." In *Plato: Complete Works*, edited by John M. Cooper, 971–1223. Indianapolis: Hackett Publishing Company.
- Putnam, Hilary. 1981. "Brains in a Vat." In *Reason, Truth and History*, edited by Hilary Putnam, 1–21. Cambridge: Cambridge University Press.
- Reijula, Samuli and Ralph Hertwig. 2022. "Self-Nudging and the Citizen Choice Architect." *Behavioural Public Policy*, 6(1), 119–149. <https://doi.org/10.1017/bpp.2020.5>
- Ruckenstein, Minna and Mika Pantzar. 2017. "Beyond the Quantified Self: Thematic Exploration of a Dataistic Paradigm." *New Media and Society*, 19(3), 401–418. <https://doi.org/10.1177/1461444815609081>
- Saramago, José. 2002. *The Cave*. London: Harvill Press.
- Shils, Edward. 1975. *Center and Periphery: Essays in Macrosociology*. Chicago: University of Chicago Press.
- Shils, Edward. 1997. "Civility and Civil Society: Good Manners Between Persons and Concern for the Common Good in Public Affairs." In *The Virtue of Civility: Selected Essays on Liberalism, Tradition, and Civil Society*, edited by Edward Shils, 63–102. Indianapolis: Liberty Fund.
- Søraker, Johnny Hartz. 2014. "Virtual Worlds and Their Challenge to Philosophy: Understanding the 'Intravirtual' and the 'Extravirtual'." In *Philosophical Engineering: Toward a Philosophy of the Web*, edited by Harry Halpin and Alexandre Monnin, 168–180. Chichester: Wiley Blackwell.
- Sparrow, Lucy A, Martin Gibbs, and Michael Arnold. 2021. "Ludic Ethics: The Ethical Negotiations of Players in Online Multiplayer Games." *Games and Culture*, 16(6), 719–742. <https://doi.org/10.1177/1555412020971534>

- Tay, Hugo. 2019. "The Holy Trinity of Disruptive Technology: Blockchain, Internet of Things and Artificial Intelligence. A Critical Perspective on How They Shape Our Future." *medium.com*. <https://medium.com/london-blockchain-labs/blockchain-iot-and-ai-4f36581c4094>
- Thaler, Richard H and Cass R Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth and Happiness*. New Haven: Yale University Press.
- Varoufakis, Yanis. 2023. *Technofeudalism: What Killed Capitalism*. London: Vintage.
- Wagner, Peter. 2012. *Modernity: Understanding the Present*. Cambridge: Polity Press.
- Wallis, Keziah and Miriam Ross. 2021. "Fourth VR: Indigenous Virtual Reality Practice." *Convergence*, 27(2), 313–329. <https://doi.org/10.1177/1354856520943083>
- Wong, Thomas G. 2022. *Introduction to Classical and Quantum Computing*. Omaha: Rooted Grove.
- Yetisen, Ali K. 2018. "Biohacking." *Trends in Biotechnology*, 36(8), 744–747. <https://doi.org/10.1016/j.tibtech.2018.02.011>
- Youn, Anthony. 2018. "What Is the Ideal Instagram Filter?" *Aesthetic Surgery Journal Open Forum*, 1(2), ojz019. <https://doi.org/10.1093/asjof/ojz019>
- Zhan, Tao, Kun Yin, Jianghao Xiong, Ziqian He, and Shin-Tson Wu. 2020. "Augmented Reality and Virtual Reality Displays: Perspectives and Challenges." *iScience*, 23(8), 101397. <https://doi.org/10.1016/j.isci.2020.101397>
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Part III

Theory, Culture, and Society in the Algorithmic Condition



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

5 Algorithmic Nostalgia

Longing for the Past in the Age of Artificial Intelligence

Dariusz Brzeziński

Introduction: Between the “Algorithmic Turn” and the “Nostalgic Turn”

My explorations in this chapter are situated at the intersection of two contemporary “turns” in the social sciences and humanities: “the algorithmic turn” (Edwards and Veale 2017; Gurumurthy and Bharthur 2018; Kiberg and Spilker 2023) and “the nostalgic turn” (Reynolds 2004; Jacobsen 2020; Brzeziński 2023). As conceptualised by Doris Bachmann-Medick (2016 [2006]), “cultural turns” are interdisciplinary orientations that emerge in the social sciences and humanities, providing a foundation for innovative research by spawning new categories of analysis and often conceptual vocabularies. Plurality is an important feature of cultural turns that distinguishes them from paradigms. Also, new types of cultural turns constantly spring up, frequently as a consequence of social and cultural change, as perfectly exemplified by the two orientations discussed here. Referring to the research challenges associated with the advent of “algorithmic culture,” Stefka Hristova, Soonkwan Hong, and Jennifer Daryl Slack (2021) conclude that “it is increasingly important to understand the ways algorithmic structures are shaping everyday life through recognition, calculation, automation, and prediction” (5). These issues are now among the most prominent topics addressed across the social sciences, humanities, and technology (e.g. Marres 2017; Elliott 2018; Crawford 2021; Novotny 2021; Beer 2024). For its part, the nostalgic turn is a term used by many intellectuals today to describe the increasing importance of inquiry into longing for the past in our times (e.g. Boym 2001; Bonnett 2016; Leggatt 2018; Quill 2024). As Michael Hviid Jacobsen (2020) observes: “Sociology, psychology, anthropology, historical science, political science, literary studies, business studies and so on have now discovered nostalgia as a potent and inexhaustible source of knowledge about individual behaviour as well as about ongoing cultural changes” (2).

Fundamental to my argument in this chapter is the fact that the transformations associated with the development of algorithms and artificial intelligence (AI) concern not only the present and the future, but also the past and the way it is longed for. This is expressed, among other things, in the regular retrieval of past events by algorithms for social media and online platform users, the possibilities of the immersive experience of the past provided by the latest technologies, and the role of

AI algorithms in shaping the fashion for nostalgic cultural productions (Lambert, Nansen, and Albert 2018; Pallister 2019; Tanner 2023). What all these developments have in common is the fact that the past becomes not only more “accessible” but also more attractive owing to technology. Moreover, besides co-shaping nostalgia, new technologies make it possible to “intervene” in it. This is epitomised, e.g. by applications for “reviving” old photographs (e.g. Deep Nostalgia) and by the entertainment industry tapping into the most advanced technologies to stage returns to the past (e.g. ABBA Voyage) (Kidd and McAvoy 2023; Ng and Bax 2023). I will take all these dimensions into account to re-theorise nostalgia in the age of the expansion of algorithms and AI.

The article is divided into three parts. In the first part, I outline prior conceptualisations of nostalgia, foregrounding the many changes that have taken place in this regard over time. In the following section, I examine the ways in which algorithms and AI bear on the human longing for the past. I list and discuss in detail four dimensions of “algorithmic nostalgia,” adopting Jenny Kidd and Eva Nieto McAvoy’s (2023) definition of it as “the generation and organisation (through automated and recursive algorithmic mechanisms) of particular memories that are aimed at eliciting nostalgic effects and affects in users of these technologies, including for exploitative purposes” (624). In the final section, I ponder how recent developments in digital technologies should affect the way nostalgia is theorised in the social sciences and humanities. Notably, although the literature on both the algorithmic turn and the nostalgic turn is growing exponentially by the year, there is still a dearth of studies on the relationship between “algorithmic modernity”¹ and the intensification and transformation of yearning for the past.² Meanwhile, as I will seek to demonstrate in this chapter, these issues are of critical importance to the contemporary world and call for in-depth research.

“Nostalgia Isn’t What It Used to Be”

Contemporary studies on nostalgia often quote the title of Simone Signoret’s autobiography, *Nostalgia Isn’t What It Used to Be* (1978 [1975]). This statement perfectly reflects the thorough transformations that the idea of nostalgia has undergone over centuries. From a concept with a specific meaning in medical discourse, it has evolved into a term with an extremely broad semantic field, widely used in the social sciences and humanities (Illbruck 2012; Sedikides et al. 2015; Dodman 2018). Moreover, along with social, political, cultural, and – particularly relevant to this article – technological changes, the way nostalgia manifests itself has been morphing, urging constant reflection on how it is conceptualised (Boym 2001; Bauman 2017; Leggatt 2018; Jacobsen 2020). In the following, I will outline the key stages of these transfigurations, which will provide underpinnings for my later analyses of the impact of algorithms and AI on the way nostalgia is experienced.

Although a yearning for the past is a common experience, and descriptions of it can be found in the most important works of world literature, the term “nostalgia” itself was only coined in the second half of the 17th century. It was invented by the Swiss physician Johannes Hofer, who combined two Greek nouns: *nóstos*, that is,

“homecoming,” and *álgos*, that is, “suffering.” Hofer (1934 [1688]) elucidated: “so that thus far it is possible from the force of the sound Nostalgia to define the sad mood originating from the desire for the return to one’s native land” (380–381). As a medic, Hofer depicted nostalgia as a disease, counting melancholy, insomnia, weakness, stupor, fever, and the like complaints among its symptoms. He observed those disorders among Swiss mercenaries, who served in armies away from their homeland. The notion of nostalgia was used in this or a similar sense for many decades to come (Illbruck 2012; Dodman 2018).

With time, the study of nostalgia was abandoned in medical science, not least because of the impossibility to establish its somatic origins. In the 20th century, nostalgia invited interest from psychologists. However much they changed their perspective on the sources of nostalgia, they initially tended to share medical scholars’ assessment of it, as recounted by Clay Routledge (2015): “Psychologists were thus continuing the tradition of viewing nostalgia as an ailment. They did not believe that nostalgia was necessarily a disease of the body. Instead, it was a disorder of the mind” (5). It did not take long, though, for the psychological conceptualisations of nostalgia to begin to change (Wilson 2005; Sedikides et al. 2015). This was reflected in a widening of the semantic scope of the term, which was no longer defined as a longing for a particular place, but as a consequence of remembering specific aspects of the past. In tandem with this shift, psychologists also stopped regarding nostalgia as destructive to the individual, in favour of considering it a “bittersweet,” paradoxical, or ambivalent emotion (Batcho 2020).

In the second half of the 20th century, the study of nostalgia was also undertaken by social scientists, most notably sociologists, anthropologists, and cultural studies scholars (Nisbet 1972; Davis 1979; Frigessi Castelnuevo and Risso 1982; Rosaldo 1989). A key role in the development of this scholarship is commonly attributed to Fred Davis’s book *Yearning for Yesterday: A Sociology of Nostalgia* (1979). Davis defined the scope of the new sub-discipline as follows: “A *sociology* of nostalgia (...) is concerned with tracking down the sources of nostalgic experience in group life and determining what general relevance and meaning nostalgia has for our present life and, somewhat more abstractly, what consequences it has for society as a whole” (1979, vii). Davis analysed various types and dimensions of nostalgia and identified its functions in social and individual life, highlighting its role in maintaining identity at a time of significant change. Davis also investigated the causes behind the transformation nostalgia was undergoing at the time, and among those he pointed to the development of technology and the media industry. He anticipated that the longing for the past would be increasingly mediated by media and mass culture.

In the 1980s and 1990s, many publications offered new insights into the mutations of nostalgia in society and culture. For example, Jean Baudrillard (1983 [1981]) and Fredric Jameson (1992) examined the role and significance of nostalgia in the films and media culture of the time. In their view, the increasingly fashionable, and at the same time increasingly superficial, exploration of the past in films led to a considerable flattening of the temporal perspective. Nostalgia was not what it had been before also in the sense that it had become detached from individual or

collective memory and transformed into a simulacrum. Soon enough, Arjun Appadurai (1996) introduced the notion of “armchair nostalgia” to describe “nostalgia without lived experience or collective historical memory” (78). Appadurai argued that a substantial part of yearning for the past was being shaped by marketing and commercial operations, where references to past lifestyles, values, and norms were designed to encourage consumers to buy certain products. This kind of nostalgia could be entirely “imaginary” and concern phenomena that had never existed. Emphatically, Baudrillard’s, Jameson’s, and Appadurai’s theorisations referred to the developments that are also important aspects of “algorithmic nostalgia.”

Of particular relevance to contemporary nostalgia studies is Svetlana Boym’s *The Future of Nostalgia* (2001). In the book, she proposed a now almost classic distinction into restorative nostalgia focused on the desire to reconstruct the past and reflective nostalgia with the past only being a loose inspiration for current practices. According to Boym, both types of nostalgia exert a powerful influence on the contemporary condition, making the longing for the past a salient factor in social, cultural, and political change. Discussing the reasons for the contemporary rise of nostalgia, Boym (2001) noted: “In counterpoint to our fascination with cyberspace and the virtual global village, there is a no less global epidemic of nostalgia, an affective yearning for a community with collective memory, a longing for continuity in a fragmented world” (xiv). Almost a quarter of a century after this observation was articulated, one cannot help but point out that today’s fascination with nostalgia does not occur in counterpoint to the fascination with cyberspace anymore, but goes hand in hand with it.

Throughout the 21st century, there has been a steady growth of interest in nostalgia in the social sciences and humanities. The role and significance of nostalgia are analysed, among other contexts, in contemporary politics, where the past is instrumentally invoked in order either to petrify the existing order or to transform it with a view to restoring lost values, structures, or institutions (Quill 2024). This tendency is epitomised by publications on Brexit (Hennebühl 2023), the spread of populism and nationalism in world politics (Cooter 2024), and the resurgence of imperial aspirations (Sullivan 2022), to cite but a few examples. Nostalgia is also extremely important for consumer culture today (Cross 2015). While vintage and retro styles are in the ascendancy, marketing strategies capitalise on these fashions to boost sales of goods and services. This process, which is sometimes referred to as the commodification of nostalgia, can also be observed in popular culture (Reynolds 2011; Bevan 2019; Leggatt 2021), as many contemporary films and series, computer games, and songs are created in styles reminiscent of older cultural models.

Simone Signoret’s remark that “Nostalgia isn’t what it used to be” perfectly captures the essential historical transformations of the yearning for the past and its metamorphoses amid contemporary social, cultural, economic, and political changes. Sketched above, the simulacrisation, instrumentalisation, commodification, and pop-culturalisation are radically redrawing the relationship between the past, present, and future. In the following section, I will analyse the role that the latest digital technologies play against the background of all these transformations.

The Dimensions of Algorithmic Nostalgia

The relationship between technology and nostalgia is multidimensional (Niemeyer and Siebert 2023). One facet of it is “technostalgia,” that is, the increasing popularity of technological inventions from decades ago (Bartmański and Woodward 2014; Lepa and Tritakis 2016; Adams et al. 2019; Burns 2021; Dang et al. 2023). These include, e.g. vinyls, CDs, and cassette tapes, video games from the late 20th century, and equally old, or even older, electronic devices. All these products are now important accessories of vintage fashion. In addition, they are often a reference point for contemporary designers inspired by the aesthetics of “technostalgia.” “Algorithmic” nostalgia, which I discuss in this section, is a different thing. It is not about a fascination with devices and technological solutions devised in the past, but about how the latest algorithm-based technologies affect the way, scope, and intensity of people’s experience of longing for the past (Kidd and McAvoy 2023). In my view, four main areas stand out in this respect: the generation of nostalgic content by algorithms on social media platforms; the promotion and further development of nostalgic productions by the algorithms of streaming and online video platforms; the reshaping of the existing content through modern digital tools to give it a nostalgic quality or to intensify it; and the digital “restoration,” “recreation,” or “reconstruction” of the past, opening up the possibilities of experiencing nostalgia in ways unavailable before. I will look into these dimensions of “algorithmic nostalgia” in this section of my paper.

The Algorithmic Generation of Nostalgia on Social Media Platforms

Mykola Makhortykh’s (2021) observation: “Today, instead of madeleine cakes, we increasingly rely on digital devices and online platforms to rekindle our memories” (180) is a good starting point for discussing algorithmic nostalgia. Indeed, algorithms and AI ever more powerfully influence how memories are shaped and organised (Jacobsen and Beer 2021). This includes their impact on the triggering of nostalgia. One well-known aspect of this process is the generation by social media platforms of collages of photographs or other content linked to a significant event in the life of an individual or a group. This content frequently comes with an automatically generated (often congratulatory) commentary and is usually accompanied by engaging music. This is how Facebook’s “On This Day,” Google Photos’ Memories, and Snapchat Memories work. The designers of the former, Omid Aziz and Manohar Palouri (2016) professed that: “Because memories are so personal and unique, we wanted to make sure On This Day shows people the memories they most likely want to see and share, especially when it comes to the memories they see in News Feed.”

It takes AI software with content filters and ranking procedures to generate algorithmic nostalgia on social media platforms. As a result, the content the users are not likely to wish to revisit is filtered out, while bringing back the content expected to stir nostalgia. In my view, “quantified nostalgia” is a notion that aptly describes these processes. The term was introduced by Benjamin Jacobsen and David Beer

(2021), who explained that it was meant to define “how metrics are variously performative in memory making, and how regimes of ordinary measures can figure in the engagement and reconstruction of the digital past in multiple ways, shaping both how people engage with it in the present, how they remember it, and how they feel about those automated memories”. Crucially, the procedure of algorithmic quantification is inextricably linked to the instrumentalisation and commodification of nostalgia I analysed in the first part of the paper. The connection lies in that the agency of algorithms in evoking a longing for the past is calculated to benefit the policies that media corporations and online platforms launch to enhance user attachment to their products and thus increase their revenues.

The algorithm-based filtering of the content that might evoke negative memories can be interpreted as exemplifying the idealisation of the past, which is characteristic of nostalgia. “The past (...) is turned into mythology,” as neatly encapsulated by Janelle Wilson (2005, 23). However, the mythologisation of the past by algorithms is not always fully successful. It is not uncommon for them to fail to recognise that some events assessed positively at a past stage of life may be viewed quite differently in retrospect. Photo-collages and memorial videos with presumably happy overtones sometimes include images of people with whom the individual is no longer on good terms, or who have tragically died (Lambert, Nansen, and Albert 2018). However, while such imperfections in the predictive analytics of algorithms are quite common, their capacity to adequately categorise memories remarkably improves over time thanks to AI.

Nostalgic Content on Streaming and Online Video Services

The increasing popularity of films, TV series, and musical works that rely on models from the past or recreate the past (Lavigne 2016; Kay, Mahoney and Shaw 2017; Pallister 2019; Bevan 2019; Leggatt 2021) stems from a wide variety of sources, a detailed analysis of which is beyond the scope of this text. By way of context, let me only mention that this phenomenon represents what Simon Reynolds has dubbed “retromania” (2011). Using popular music as an example, he has argued that contemporary culture focuses on the creative reworking of old motifs, which puts it at the other pole of the continuum, opposite to the progressive, revolutionary cultural changes of the 1960s, the 1970s, and partly the later 20th century. From another perspective, the popularity of nostalgic cultural productions can be interpreted as embodying “retrotopia,” that is, as Zygmunt Bauman (2017) diagnosed, the shift of modern society’s attention to the past as a consequence of the increasing unpredictability and perils of the future. However, the notions of “retromania” and “retrotopia” would not be sufficient in my argument without outlining the role of streaming and online platform algorithms in developing nostalgic cultural productions.

The way these platforms work is based on a complex system of recommendations (Hallinan and Striphas 2016; Bischoff 2023). Their algorithms collect data on films, music, and series that users select in real time. On the basis of this information, as well as demographic data about users, among other things, similar content

is recommended to them, since these, the algorithms estimate, will find it appealing as well. Moreover, the streaming and online platform managers are guided by the aggregated data to make decisions on new cultural productions. This is because adapting their format to the expectations of the audience is likely to boost viewing or listening figures, and thus also the financial performance of media corporations. In an algorithmic culture, this quantification and the predictions founded on it are therefore among the most important causal factors. This process is instantiated by the ever more robust presence of nostalgic productions on streaming and online video platforms, ranging from the works created in previous decades and their remakes to re-boots and other productions that take viewers back to the past.

The feedback loop is well illustrated by the policies of Netflix, whose role in popularising and developing nostalgic productions has been characterised as “weaponizing nostalgia” by [Mayka Castellano and Melina Meimaridis \(2019\)](#). Firstly, Netflix provides instant access to a wide array of films and shows produced in the more-or-less distant past. The very names of some of the Netflix library catalogues, such as “The Nostalgic ‘90s,” and “The Totally Awesome ‘80s,” are much-revealing in this context. Secondly, with an eye on the viewing figures and the expected financial returns, the platform produces many remakes, reboots, or sequels to already classic film and TV works. This production segment has been pivotal to the platform’s policy for several years now. Thirdly, Netflix initiates new productions that are redolent of works from years ago stylistically, aesthetically, and thematically. In this way, the platform not only responds to audience expectations but also stimulates them in a creative way. In the context of the importance of nostalgia to Netflix operations, Kathryn Pallister (2019) observes that “while nostalgia can seduce audiences into sitting down on their couches for a one-night stand of a viewing session, the Netflix platform invites them to sit and stay for a while, turning the seduction into a long-term relationship” (2).

Algorithmic Processing of Photos and Videos to Provide or Intensify Nostalgia

It is somewhat of a paradox that modern technologies not only make it possible to capture images and video footage more and more perfectly but also supply tools for achieving the opposite outcomes. Specifically, there are applications and filters which use AI algorithms to offer multiple options for producing vintage effects. By means of these features, photos taken and videos made today, and images created with generative AI tools can be styled to look like any past era or as made with devices manufactured years, decades, or even centuries ago. For example, filters can make contemporary photographs resemble pictures taken with classic Polaroid cameras, and films mirror the VHS recording modes. At the same time, generative AI tools are able to create a series of images to mimic a style of the past chosen by the user.

This dimension of algorithmic nostalgia is to some extent an exemplification of the “reflective” longing for the past, as distinguished by [Boym \(2001\)](#), since its aim is not to recreate the past but to creatively transform it. It does not seek to bring to light the truth of reality but to uncover the ambivalences of that reality.

In essence, however, vintage photographs, films, and paintings represent a very unorthodox form of nostalgia. The point is that what happens in this case is not a recollection of the past in the present, but a transformation of the present into the past. Moreover, as noted by Gil Bartholeyns (2014), “digital retro photography” is a “self-induced nostalgia.” Bartholeyns clarifies: “In short, nostalgia could not be ordered on demand. Now, however, we can conjure it up and, even as we experience it at the sight of something on the street, at home or on a journey, we have the means to display it and be moved by it” (55). Though already formulated a decade ago, this insight perfectly applies to the vintage images created by generative AI.

However, today’s technologies make it possible not only to “age” photographs taken today or generate images based on them but also to animate, or “revive,” those taken in the past. One of the most popular tools used for this purpose is MyHeritage’s Deep Nostalgia, available since 2021. The feature is used to make short animations from archival portrait photos, so that the people captured in the photographs can smile, blink, move their heads, and perform similar movements. Kidd and McAvoy (2023) conducted a study on the social reception of this functionality based on a netnographic analysis of Twitter data on the hashtag #DeepNostalgia. Their research showed that most users of the feature not only received it positively but in fact reacted enthusiastically to being able to see animations of their deceased relatives, historical figures, and other people. However, the sample also included posts expressing ethical concerns, reservations about the quality of the animations, and remarks on their similarity to deepfake technology. In conclusion to their research, Kidd and McAvoy stated: “Deep Nostalgia’s shift from photo to video reshapes and remediates memory such that it becomes both real and unreal, immediate and remote, comforting and disturbing”. Indeed, both Deep Nostalgia and other functionalities with similar properties, working with generative AI technology, contribute to multiplying the ambivalences of nostalgia. Its inherent complex relationship between the past and the present becomes much more problematic as a result of “reviving” the past. Moreover, the “bittersweetness” constitutive of nostalgia is also intensified by the simultaneous arousal of disparate emotions as a result of the “revival” of the people in pictures and films. Importantly, these functionalities are by no means niche tools today. On the contrary, as their popularity spreads, they are likely to influence the experience of nostalgia ever more powerfully.

Nostalgia in the Age of Algorithmic Reconstructions of the Past

As already mentioned, the last type of algorithmic nostalgia involves restoring, recreating, or reconstructing the past through AI. This practice takes many different forms, the most famous of which are associated with the entertainment industry.³ For example, The Beatles’ song “Now and Then,” which dates to the 1970s, was released again almost half a century later, in November 2023. It was originally recorded as a solo home demo by John Lennon in 1977, and the tape, along with “Free as a Bird” and “Real Love,” was given to Paul McCartney by Yoko Ono in 1994. Due to the poor quality of the recording of “Now and Then,” it was not

included in the *The Beatles Anthology* project, which was being compiled at the time. The publication of the song was only made possible by the development of AI. Lennon's voice was distilled via the machine learning audio restoration, and the bass, guitar, and drum parts were then added to it. The intermingling of the past and present temporal perspectives in the editing history of this song is vividly conveyed by the video for it. In a Peter Jackson-directed clip, Paul McCartney, Ringo Starr, George Harrison, and John Lennon are shown at various stages of their lives, with historical stills of the deceased bandmates superimposed on contemporary shots of the living members. Upon its release, the song became hugely popular around the world, standing as excellent testimony to the power of nostalgia in contemporary culture. Importantly, it is a nostalgia that transcends the restorative-reflective dichotomy. Specifically, this longing has to do with reconstructing the past, but its source is the restoration of a lost element of that past, rather than its absence. At the same time, although this longing "cherishes shattered fragments of memory" (Boym 2001, 49), it is alien to an ironic distance in this respect. All this demonstrates the utter novelty of this dimension of algorithmic nostalgia.

Another nostalgic return made possible by the advancement of AI was the "ABBA Voyage" project. Several years in the making, the show eventually took place in London in May 2022, with the technology installed in a purpose-built arena taking the audience "back in time."⁴ During the concerts in the "ABBA Arena," the roles of Anni-Frid Lyngstad, Benny Andersson, Björn Ulvaeus, and Agnetha Fältskog were played by hyper-realistic avatars that reproduced the way the group members looked in the 1970s. Jenna Ng and Nick Bax (2023) detail the technical particulars of the show: "Visual effects artists modelled, animated and rendered the ABBA-tars using past negatives, footage and motion capture. The 'concert' of ABBA Voyage, then, presents these animations of ABBA appearing as their past selves on a stage-spanning 65 million-pixel screen before a live audience, backed by a ten-piece live band with surround sound and lighting effects" (160). The phenomenon of the "ABBA Voyage" perfectly illustrates the vital role that technology plays today not only in providing an opportunity to return to the past but also in shaping nostalgia for the past. Paradoxically, this return is also a step forward, towards the prospects offered by AI. The dimension of algorithmic nostalgia discussed here blurs not only the boundary between temporal planes but also the line between reality and hyperreality. It transports viewers into a realm of nostalgic simulacrisation, where "the map precedes the territory" (Baudrillard 1983 [1981], 1).

The two examples of algorithmic nostalgia I cite⁵ can be interpreted as incarnating what Grafton Tanner (2023) referred to as "foreverism."⁶ In his likewise titled book, Tanner (2023) explains: "To foreverize something is not merely to preserve or restore it but to reanimate it in the present and ensure its future survival, forever" (18). Tanner emphasises that foreverism has a complex relationship to nostalgia. On the one hand, foreverism ensues from a longing for the past. In response to this sentiment, and thanks to the possibilities offered by today's technology, many initiatives are now being implemented to restore, resuscitate, or ultimately foreverise the past. Colonised by capitalism, nostalgia is thus being monetised. On

the other hand, however, as Tanner argues, foreverism contributes to the elimination of nostalgia by making the past present in an “expanded” present. He points out that “if nostalgia is the emotion experienced when something normally absent becomes momentarily present, in our minds or in everyday life, than the constant presence of the past in the contemporary does not amount to ‘nostalgia’ as much as something else” (Ibid., 17). While agreeing with most of Tanner’s conclusions about the salience of foreverism in contemporary culture, I have reservations about his latter claim. In my view, nostalgia is by no means disappearing as a result of the restoration of the past; rather, it is undergoing profound transformations, which I have conveyed above through the concept of “algorithmisation.” I would describe the relationship between nostalgia and foreverism as a continuous dialectic, where the two components reinforce each other while simultaneously modifying each other. Thus, if foreverism eats its tail, it is only to be reborn again and again, like the ancient Ouroboros.

Conclusion: Theorising Nostalgia in the Age of Artificial Intelligence

In this chapter, I have analysed the complex relationship between the development of AI and the increasing centrality of longing for the past in contemporary culture. I have sought to show that the two phenomena, though seemingly independent, are in fact meaningfully interconnected. One of the directions in which technology is heading today is the generation, intensification, and transformation of nostalgia. This purpose is served by certain functionalities of social media and online platforms; algorithms in streaming and online video services; photo and video editing applications and AI tools for generating images; and sophisticated projects implemented by the entertainment industry. At the same time, “retromania,” “retrotopia,” and vintage fads are a factor in the policies of media corporations and the entertainment industry. Technological solutions are being sought to respond to consumer needs and expectations. In this way, the instrumentalisation, commodification, and monetisation of nostalgia, which started at the end of the 20th century, are intensifying. All in all, all the processes I have detailed contribute to the multidimensional phenomenon of “algorithmic nostalgia.”

In the closing words of his editorial introduction to *Nostalgia Now: Cross-Disciplinary Perspectives in the Past in the Present*, Michael Hviid Jacobsen (2020) urged that “nostalgia is (...) in desperate need of theoretical elaboration and empirical investigation” (24). This need stems both from the increasing prominence of yearning for the past in the contemporary world and from its current essential transformations. My analyses show that, in the age of the expansion of algorithms and AI, many prior ways of making sense of nostalgia should be revised. Firstly, contrary to the meaning of *nóstos*, nostalgia today does not always refer to what has been irretrievably lost, as technology promises many possibilities to restore and even improve what has passed. Secondly, in opposition to the term *álgos*, nostalgia today is very often associated with pleasant and joyful experiences and, as a result, can also be monetised. Thirdly, in the algorithmic condition, the role of the Proustian madeleine is being taken over by applications, features,

and other AI-based tools. Fourthly, while always constitutive of nostalgia, a creative relation to the past is currently soaring, aided by AI-driven simulacrarisation. Fifthly, contemporary nostalgia is not necessarily characterised by thinking about the past from the perspective of the present but may involve converting the present into the past via the digital tools or the fusion of the two.

All these revisions are relevant to empirical research. The indispensable transformations in how nostalgia is conceptualised and operationalised are crucial to the validity of this research. This concerns not only the investigation of the origins, evolution, and functions of the algorithmisation of nostalgia but also inquiry into the related processes in consumer culture, the entertainment industry, and on the media market. The new theoretical framework of nostalgia I have presented here should also be applicable to the study of other dimensions of the contemporary longing for the past, including past-oriented identity transformations, ideologies, and political movements. Tellingly, algorithms and AI will certainly have an ever more important part in undertaking these research investigations.

Acknowledgements

This research was funded by the National Science Centre Poland from the SONATA grant (No DEC-2021/43/D/HS2/03321).

Notes

- 1 In using the term “algorithmic modernity,” I draw on Anthony Elliott, who introduced it in his book *Making Sense of AI: Our Algorithmic World* (2022) to depict the very significant social, political, and cultural changes unfolding today as a result of the development of AI.
- 2 For illuminating discussions, see Niemeyer 2014; Lambert, Nansen, and Arnold 2018; Pallister 2019; Jacobsen and Beer 2021; Makhortykh 2021; Kidd and McAvoy 2023; Niemeyer, Sieber 2023; Tanner 2023.
- 3 A good example from outside the entertainment industry might be the event held at the Keats-Shelley House in Rome on 23 February 2021, the centenary of John Keats’ death. In collaboration with the Institute for Digital Archaeology in Oxford, an attempt was made to “virtually recreate” the poet. The event included a CGI (computer-generated imagery) recitation of Keats’ poem “Bright Star” (Albano 2021).
- 4 Tellingly, the popularity of the ABBA Voyage concerts has not subsided. As a result, in March 2024, the show’s producer, Svana Gisla, announced that she was considering plans to build similar arenas in other European countries, the United States, and Australia (Masson 2024).
- 5 Another example of the AI-driven phenomenon of “resurrecting” the past is the creation of digital hyper-realistic avatars of celebrities. For example, in March 2024, at the South by Southwest (SXSW) conference in Austin, Texas, the New Zealand-based company Soul Machines presented the “Digital Marilyn” project: a digital version of Marilyn Monroe capable of interacting with people in real-time and engaging in dialogue with them. The digital Marilyn is supposed to read users’ emotions, analyse their preferences, and adapt her responses accordingly.
- 6 The issue of the relationship between experience of time and the development of algorithms and AI is comprehensively discussed by Helga Novotny, the author of the concept of “digital time” (2021).

References

- Adams, J., O. Dedehayir, P. O'Connor, and T. Nokelainen. 2019. "Who Are the Adopters of Retro-Technology." *International Journal of Innovation Management* 23, no. 8: 1–18.
- Albano, G. 2021. "Report from Rome." *The Keats-Shelley Review* 35, no. 2: 113–115.
- Appadurai, A. 1996. *Modernity at Large: Cultural Dimensions of Globalization*. Minneapolis and London: University of Minnesota Press.
- Aziz, O. and M. Palouri. 2016. "Engineering for Nostalgia: Building a Personalized 'On This Day' Experience." <https://engineering.fb.com/2016/03/30/ml-applications/engineering-for-nostalgia-building-a-personalized-on-this-day-experience/> Access, 10 March 2024.
- Bachmann-Medick, Doris. 2016. *Cultural Turns: New Orientations in the Study of Culture*, translated by A. Blahut. Berlin: De Gruyter.
- Bartholeyns, G. 2014. "The Instant Past: Nostalgia and Digital Retro Photography." In *Media and Nostalgia: Yearning for the Past, Present and Future*, edited by K. Niemeyer, 51–69. Basingstoke: Palgrave Macmillan.
- Bartmański, D. and I. Woodward. 2014. *Vinyl: The Analogue Record in the Digital Age*. London: Bloomsbury Academic.
- Batcho, K. I. 2020. "Nostalgia: The Paradoxical Bittersweet Emotion." In *Nostalgia Now: Cross-Disciplinary Perspectives on the Past in the Present*, edited by M. H. Jacobsen, 31–46. London: Routledge.
- Baudrillard, J. 1983 [1981]. *Simulacra and Simulation*. Michigan: University of Michigan Press.
- Bauman, Z. 2017. *Retrotopia*. Cambridge: Polity Press.
- Beer, D. 2024. *The Tensions of Algorithmic Thinking: Automation, Intelligence and the Politics of Knowing*. Bristol: Bristol University Press.
- Bevan, A. 2019. *The Aesthetics of Nostalgia TV: Production Design and the Boomer Era*. New York: Bloomsbury Academic.
- Bischoff, M. 2023. "How Recommendation Algorithms Work - And Why They May Miss the Mark." *Scientific American*. August 17. <https://www.scientificamerican.com/article/how-recommendation-algorithms-work-and-why-they-may-miss-the-mark/> Access 10 March 2024.
- Bonnett, A. 2016. *The Geography of Nostalgia: Global and Local Perspectives on Modernity and Loss*. London: Routledge.
- Boym, Svetlana. 2001. *The Future of Nostalgia*. New York: Basic Books.
- Brzeziński, D. 2023. "Zygmunt Bauman and the 'Nostalgic Turn'." In *The Anthem Companion to Zygmunt Bauman*, edited by M. H. Jacobsen, 131–146. London: Anthem Press.
- Burns, J. I. 2021. *Mixtape Nostalgia: Culture, Memory, and Representation (Critical Perspectives on Music and Society)*. Lanham, MD: Lexington Books.
- Castellano, M. and M. Meimaridis. 2019. "'Weaponizing Nostalgia': Netflix, Revivals and Brazilian Fans of *Gilmore Girls*." In *Netflix Nostalgia: Streaming the Past on Demand*, edited by K. Pallister, 169–184. Lanham, MD: Lexington Books.
- Cooter, A. 2024. *Nostalgia, Nationalism, and the US Militia Movement*. London: Routledge.
- Crawford, K. 2021. *The Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.
- Cross, G. 2015. *Consumed Nostalgia: Memory in the Age of Fast Capitalism*. New York: Columbia University Press.
- Dang, J., C. Sedikides, T. Wildschut, and L. Liu. 2023. "More than a Barrier: Nostalgia Inhibits, but Also Promotes, Favorable Responses to Innovative Technology." *Journal of Personality and Social Psychology* 126, no. 6: 998–1018. <https://doi.org/10.1037/pspa0000368>
- Davis, F. 1979. *Yearning for Yesterday: A Sociology of Nostalgia*. New York: Free Press.
- Dodman, T. 2018. *What Nostalgia Was? War, Empire and the Time of a Deadly Emotion*. Chicago, IL: University of Chicago Press.

- Edwards, L. and M. Veale. 2017. "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For." *Duke Law & Technology Review* 16: 18–84.
- Elliott, A. 2018. *The Culture of AI: Everyday Life and the Digital Revolution*. London: Routledge.
- Elliott, A. 2022. *Making Sense of AI: Our Algorithmic World*. Cambridge: Polity Press.
- Frigessi Castelnuevo, D. and M. Risso. 1982. *A Mezza Parte: Emigrazione, Nostalgia, Malattia Mentale*. Torino: Einaudi.
- Gurumurthy, A. and D. Bharthur. 2018. "Democracy and the Algorithmic Turn: Issues, Challenges and the Way Forward." *Sur – International Journal of Human Rights* 15, no. 27: 39–50.
- Hallinan, B. and T. Striphas. 2016. "Recommended for You: The Netflix Prize and the Production of Algorithmic Culture." *New Media & Society* 18, no. 1: 117–137.
- Hennebühl, D. 2023. *Taking Back Control of the Nation and Its History?: Contemporary Fiction's Engagement with Nostalgia in Brexit Britain*. Paderborn: Brill Fink.
- Hofer, J. 1934 [1688]. "Medical Dissertation on Nostalgia." *Bulletin of the History of Medicine* 2: 376–391.
- Hristova, S., S. Hong, and J. Daryl Slack. 2021. *Algorithmic Culture: How Big Data and Artificial Intelligence Are Transforming Everyday Life*. Lanham, MD: Lexington Books.
- Illbruck, H. 2012. *Nostalgia: Origins and Ends of an Unenlightened Disease*. Evanston, IL: Northwestern University Press.
- Jacobsen, M. H. ed. 2020. *Nostalgia Now: Cross-Disciplinary Perspectives on the Past in the Present*. Abingdon and New York: Routledge.
- Jacobsen, B. and D. Beer. 2021. "Quantified Nostalgia: Social Media, Metrics, and Memory." *Social Media + Society* 7, no. 2: 1–9.
- Jameson, F. 1992. *Postmodernism: Or, the Cultural Logic of Late Capitalism*. Durham, NC: Duke University Press.
- Kay, J. B., C. Mahoney, and C. Shaw, eds. 2017. *The Past in Visual Culture: Essays on Memory, Nostalgia and the Media*. Jefferson, NC: McFarland & Company.
- Kiberg, H. and H. Spilker. 2023. "One More Turn after the Algorithmic Turn? Spotify's Colonization of the Online Audio Space." *Popular Music and Society* 46, no. 2: 151–171.
- Kidd, J. and E. N. McAvoy. 2023. "Deep Nostalgia: Remediated Memory, Algorithmic Nostalgia and Technological Ambivalence." *Convergence* 29, no. 3: 620–640.
- Lambert, A., B. Nansen, and M. Arnold. 2018. "Algorithmic Memorial Videos: Contextualising Automated Curation." *Memory Studies* 11, no. 2: 156–171.
- Lavigne, C. ed. 2016. *Remake Television: Reboot, Re-Use, Recycle*. Lanham, MD: Lexington Books.
- Leggatt, M. 2018. *Cultural and Political Nostalgia in the Age of Terror: The Melancholic Sublime*. London and New York: Routledge.
- Leggatt, M. ed. 2021. *Was Is Yesterday? Nostalgia in Contemporary Film and Television*. Albany, NJ: SUNY Press.
- Lepa, A., and V. Tritakis. 2016. "Not Every Vinyl Retromaniac Is a Nostalgic – A Social Experiment on the Pleasures of Record Listening in the Digital Age." *Medien & Zeit* 4: 16–30.
- Makhortykh, M. 2021. "Memoriae Ex Machina: How Algorithms Make Us Remember and Forget." *Georgetown Journal of International Affairs* 22, no. 2: 180–185.
- Marres, N. 2017. *Digital Sociology: The Reinvention of Social Research*. Cambridge, UK: Polity Press.
- Masson, G. 2024. "Abba Voyage Producer Hints at New Venues Worldwide." *IQ Magazine*, March 4. <https://www.iq-mag.net/2024/03/abba-voyage-producer-hints-at-new-venues-worldwide/> Access March 10, 2024.
- Ng, J. and N. Bax. 2023. "Spooker Trouper: ABBA Voyage, Virtual Humans and the Rise of the Digital Apparition." *Paragraph* 46, no. 2: 160–175.

- Niemeyer, K., ed. 2014. *Media and Nostalgia: Yearning for the Past, Present and Future*. Basingstoke: Palgrave Macmillan.
- Niemeyer, K. and O. Siebert. 2023. "Nostalgia, Media, and Technologies of the Future." *Current Opinion in Psychology* 50, no. 3. <https://doi.org/10.1016/j.copspsy.2023.101561>
- Nisbet, R. 1972. "The 1930's: America's Major Nostalgia." *The Key Reporter* 38, no. 1: 2–4.
- Novotny, H. 2021. *In AI We Trust: Power, Illusion and Control of Predictive Algorithms*. Cambridge: Polity Press.
- Pallister, K., ed. 2019. *Netflix Nostalgia: Streaming the Past on Demand*. Lanham, MD: Lexington Books.
- Quill, L. 2024. *Nostalgia and Political Theory*. London: Routledge.
- Reynolds, B. 2004. "The Nostalgic Turn and the Politics of Ressentiment." *Georgia Educational Researcher* 2, no. 1: 1–11.
- Reynolds, S. 2011. *Retromania: Pop Culture's Addiction to Its Own Past*. New York: Farrar, Straus & Giroux.
- Rosaldo, R. 1989. "Imperialist Nostalgia." *Representations* 26, no. April: 107–122.
- Routledge, C. 2015. *Nostalgia: A Psychological Resource*. New York and London: Routledge.
- Sedikides, C., T. Wildschut, C. Routledge, J. Arndt, E. G. Hepper, and X. Zhou. 2015. "To Nostalgize: Mixing Memory with Affect and Desire." *Advances in Experimental Social Psychology*, edited by M. Zanna and J. Olson, vol. 52: 189–273. Burlington: Academic Press.
- Signoret, S. 1978 [1975]. *Nostalgia Isn't What It Used to Be*. London: Weidenfeld and Nicolson.
- Sullivan, C. J. 2022. *Motherland: Soviet Nostalgia in the Russian Federation*. Basingstoke, UK: Palgrave Macmillan.
- Tanner, G. 2023. *Foreverism*. Cambridge: Polity Press.
- Wilson, J. L. 2005. *Nostalgia: Sanctuary of Meaning*. Lewisburg, PA: Bucknell University Press.

6 Socio-cultural Artificial Intelligence (SCAI) and Journalism

A Transdisciplinary Perspective in Media and Communication Studies

*Malgorzata Winiarska-Brodowska
and Katalin Feher*

Introduction

As we are experiencing rapid changes in the media environment and public communication patterns currently, it is crucial to take into account the conditions of modern public spheres when analysing theories of journalism related to artificial intelligence (AI). In characterising contemporary public spheres as disrupted and disconnected, [Bennett and Pfetsch \(2018: 245\)](#) accentuate two fundamental changes: first, the proliferation of social and digital media, which has increased the dispersion and cacophony of public voices ([Dahlgren 2005: 151](#)); second, this fragmentation of publics has led to an “inability to communicate across differences” (Waisbord 2016: 2). Challenges such as polarisation, the intermingling of information and misinformation, greater external interference in communication, particularly in political communication (for instance, subversive use of social networks during elections), strategic communication practices of different actors, and the asymmetry between entities in global communication processes have been observed for some time ([Bennett and Livingston 2018](#); Winiarska-Brodowska 2020). These complex shifts necessitate a transdisciplinary approach to the evolving role of journalism in the AI context and the development of responsible technologies.

AI strengthens discussions around communication disruptions, fragmentation and polarisation, mis-, dis- and mal-information¹, cheap fakes² and semi-true news, the misuse of AI technology, and manipulation in the media ecosystem currently ([Feher et al., 2024](#)). Particularly, if AI developers and investors express concerns over the emergence of cultural islands shaped by personalised content and a user-preference-driven news industry, which could potentially lead to the demise of democracy ([Feher and Veres, 2022](#)). The trend towards democratic erosion and the emergence of illiberal public spheres are increasingly underscored within this framework, where “illiberal communication creates ‘transgressive publics’ that attempt to exclude others” ([Bennett and Kneuer 2023](#)). In this context, it is crucial to remember that “from a functional and empirical perspective, the public sphere remains dependent on the pivotal services of journalism” ([Nuernbergk 2022: 1](#); see also earlier classic works, e.g., [de Tocqueville 2004](#)). However, the role and impact of journalism are continually being redefined in an

increasingly digital and networked world, with discussions around concepts such as echo chambers³ and filter bubbles⁴, and the impact of selective exposure in a high-choice media environment. In this context, AI technology triggers societal and cultural reflection on responsibility issues (Feher et al. 2024).

Considering the technology, automation and prediction are two algorithmic functions which are currently commonly used in journalism (Newman 2024). The former is used to replace routine and repetitive tasks, while the latter will mean that in future automated societies, “sophisticated predictive programs will forecast women and men’s evolving tastes and preferences in media” within which “people will communicate and interact with each other in the language of their choosing, activating superfast machine translation software in an instant” (Elliott 2022: 175). At the present time, however, despite the benefits they may bring both to media producers and to users alike, automation and prediction “raise concerns about the quality, diversity, and responsibility of journalism” (Blassnig et al. 2024). Empirical investigations on algorithmic automation and personalisation of news recommendations in news media environments have – despite significant interest – still remained limited, especially in cases of small media markets and across different organisational types (cf. *ibid.*). Blassnig et al. (2024: 10) indicate that media professionals currently view the implementation of the two algorithmic functions as a delicate equilibrium between aligning with and distinguishing themselves from other news organisations and intermediaries and they add that:

Retaining journalistic responsibility was seen, among other things, in the demarcation of news media from social media and other intermediaries. In contrast to the co-orientation arguments supporting NRS [news recommender systems] implementation, we also heard the argument that news media outlets are not, and should not be, like social media: Journalism must differentiate itself from such platforms regarding its quality, goals, and thus its use of NRS.

As Friemel and Neuberger (2023) note, unlike journalism, the majority of social media platforms avoid taking responsibility for their content. Indeed, Zuboff (2019) argues that they are radically indifferent to the content quality in an effort to boost the audience size for each piece of content and to maximise the volume of content consumed by each user. The question centres on how leading theories in journalism interpret these changes.

The chapter presents current research trends in Media and Communication Studies on AI and examines theories of journalism related to AI. The chapter is divided into four parts. The first outlines the background to the research by explaining the notion of socio-cultural artificial intelligence (SCAI, first initially defined by Feher and Katona 2021) and introduces the topic of journalism theories. In the second part, the methodology used in this study is described, while the third part presents the results of a systematic literature review, with sections relating to research areas: (1) the leading SCAI research trends in media and communication studies, and (2) theories of journalism related to SCAI in academic research. The final section offers a discussion of the findings and final remarks.

Background: Socio-cultural Artificial Intelligence (SCAI) and Theories of Journalism

The interplay between cutting-edge digital technology and the socio-cultural constructions is complex and dynamic. Research informed by – in [Liu's \(2021\)](#) categorisation – the *cultural AI* perspective looks at AI as a social phenomenon and “examines its interactions with the wider social, cultural, economic and political conditions in which it develops and by which it is shaped” (ibid.: 2). This research perspective reflects on AI-triggered new trends, processes, actions, and relations in a wide range of social contexts.

Likewise, [Feher and Katona \(2021\)](#) use the term *socio-cultural AI* (SCAI), which views technology as a tool and not as an ultimate goal, as is the case with techno-AI. Social science and humanities define AI-related socio-cultural issues with a broad array of topics from AI-related policymaking to fake news. One example of work that represents this perspective is [Elliott's \(2018\)](#) book *The Culture of AI*, which explores the broader and longer term social impacts of AI in the digital age. SCAI research is still a burgeoning field and in comparison to techno-AI research, it is underrepresented in academic databases, but as underlined by [Feher and Katona \(2021\)](#), the leading academic databases, such as Web of Science and Scopus, still present a suitable number of high-ranked publications for analysis in relation to technology and the number of socio-cultural AI-related publications has increased significantly since the last decade. Research on media and communication constitutes a significant part of a large body of SCAI investigations, and in particular journalism.

Theories of journalism offer a comprehensive understanding of the intricate interplay between social practices, products, and institutional arrangements by providing explanatory frameworks. Contemporary journalism theories have developed an equal interest in the process of news production and the final product. However, they still need to be mindful of the quality and nature of the spaces journalism is creating because, as [Reese \(2016\)](#) explains, “journalism seeks to observe and communicate what it learns of social importance, something called news, and in doing so is always in the process of creating a public by bringing it into synchronised conversation with itself”. Recently, the disruptive effect of the Internet can be observed ([Bennett and Pfetsch 2018](#); [Bennett and Livingston 2018](#)). In many democracies, there is a disconnection of the public from institutions of the press, as well as growing social divides and polarisation. The present transitional phase, whose outcome remains unclear, is described by [Schlesinger \(2020\)](#) as an unstable “post-public sphere”. In networked public spheres ([Castells 2008](#)), new deliberative spaces emerge and journalism theories now need to be concerned with understanding the nature and quality of these spaces, the shape they take on, and their implications for healthy democracies.

Methodology

For trend analysis, topic modelling and association networks of qualified academic research was presented first for conceptualising SCAI by [Feher and Katona \(2021\)](#), collecting data between 2009 and 2019 from high-ranked academic databases, including Scopus, EBSCO Information Services, JSTOR Digital Library, ArXiv

e-print Service, Directory of Open Access Journals, Social Science Research Network, and Institute of Electrical and Electronics Engineers. The systematic literature review presented a comprehensive analysis of 607 academic sources.

In this study, we expanded upon previous research of SCAI (Feher and Katona 2021) by performing a follow-up analysis for 2020–2023 with 783 abstracts retrieved from the Scopus database. We exclusively utilised the Scopus database due to its extensive coverage and the quality of its peer-reviewed sources, enabling a more streamlined and focused analysis of journalism theories. Not all scientific articles present keywords directly, which is why the missing keywords were, if possible, retrieved from other sources or constituted an acceptable amount of missing data.

By adopting a nuanced methodological approach, this study aims to conduct an in-depth investigation of the journalism landscape within the AI context. To achieve this, it employs qualitative content analysis (QCA), a research method which allows for a detailed exploration of how SCAI influences and shapes journalism, providing insights into emerging trends and theoretical developments. With this extension, a summary of 2009–2023 SCAI and journalism has become available.

In the face of the emerging challenges discussed in SCAI research, such as disinformation in the digital age, it seemed worth analysing new trends in journalism theory. This is also important because of the special role of journalism in a democracy, which cannot exist without a free press, as citizens must be well informed to make the “right” decisions in elections and monitor the actions of the authorities. Therefore, the analysis focused on the content of articles that looked at journalism theories in the context of SCAI. A qualitative content analysis (QCA) (Schreier 2012; Schreier et al. 2019; Mayring 2022; Kuckartz and Rädiker 2023) of 32 academic texts was conducted (N=32, screened out of the Scopus database with keywords “artificial intelligence”, “journalism”, and “theor*” for 2009–2023). Following recent trends in academic research (Aguinis et al. 2020), only articles were examined; books, book chapters, and conference papers were excluded. The filtered language was English.

This analysis allowed to look for an answer to the following questions:

- RQ1. What have been the leading SCAI research trends in the context of Media and Communication Studies and how this landscape has changed over the past 15 years?
- RQ2. What theories of journalism related to AI are present in academic research between 2009 and 2023?
- RQ3. Are these AI-related journalism theories embedded in one discipline or do they take an inter-/transdisciplinary perspective?

Research Results

The Leading SCAI Research Trends in Media and Communication Studies

To answer the first research question, Feher and Katona (2021) have found broad and complex fields on the extracted topic list, covering almost all social and cultural areas. When expanding the research up to 2023, the most frequent topic is still

human-robot communication. “Social media, fake news/deep fake” and “Social communication, complex systems” were also frequently mentioned topics. In addition to identifying these trends, the results of previous research were confirmed, indicating that social sciences are most represented, while the humanities are underrepresented, with cultural aspects less highlighted in the studied databases compared to social factors. In the more narrow context of journalism that we are interested, journalism with ethics were key converging topics. Other studies (e.g. [Parratt-Fernández, Mayoral-Sánchez and Mera-Fernández 2021](#)) also confirm that significant growth in research on the application of AI to journalism began in 2015, especially on topics of social or fake media. When looking at the associated themes, the intersecting areas of journalism and social media, along with fake news and deepfakes, are all key to the contemporary media landscape ([Figure 6.1](#)).

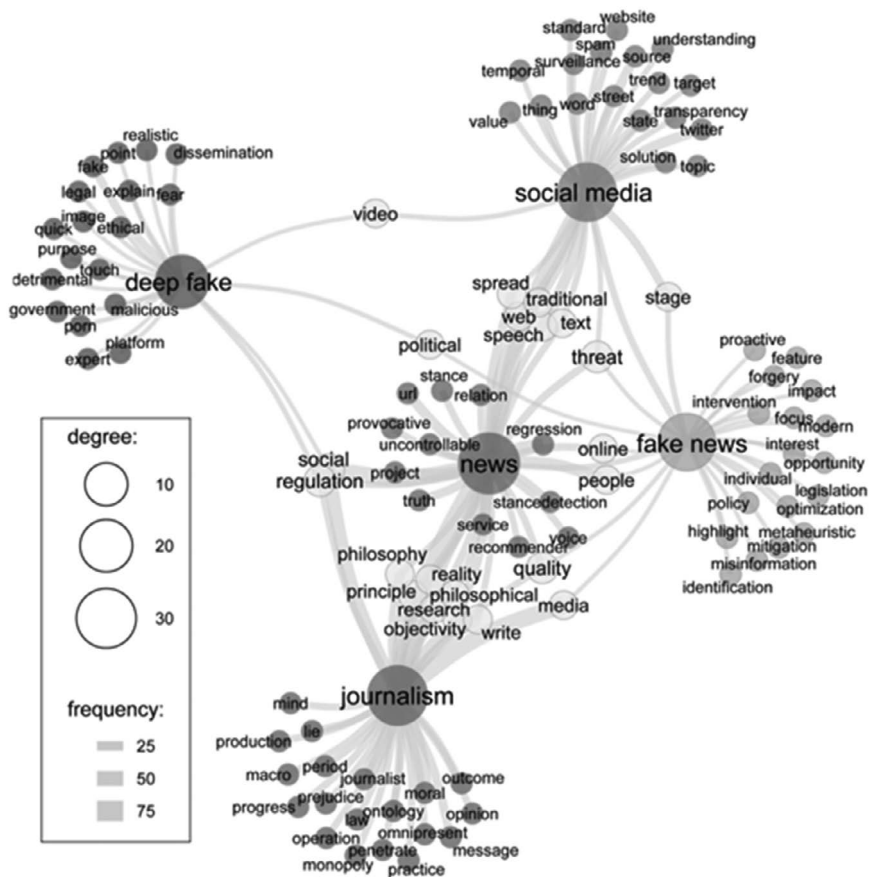


Figure 6.1 Co-occurrence network of terms related to “journalism” and “social media” within the same sentence (2009–2019).

Source: [Feher and Katona \(2021, p. 12\)](#). Used with permission. Yellow nodes represent the intersection between the analysed terms.

“News” occupies a central place in the co-occurrence topic network and is connected to “journalism”, “fake news”, deepfakes”, and a conveying role of “social media”. News media represents a connected link to political context (Meskys et al. 2019), while issues of surveillance and transparency are underlined.

Following years (2020–2023) showed significant changing journalism conditions and rapid transformations in the media landscape, where “the various shocks of the last few years, including the war in Ukraine and the Coronavirus pandemic, have accelerated structural shifts towards more digital, mobile, and platform-dominated media environments, with further implications for the business models and formats of journalism” (Newman 2023: 10). The news industry of the digital era needs to tackle issues like misleading information and disinformation, and deepfakes are rapidly becoming a destabilising factor that undermines a shared sense of social and political reality and blurs the acceptable dichotomy between real and fake. Linked to this, the complementary follow-up study of 2020–2023 from the Scopus database indicated that disinformation is a strongly developing research trend (e.g., number of articles mentioning “fake news” in the context of SCAI quadrupled in 2020–2023 compared to 2009–2019). A similar result was found for the number of articles discussing “social media”, including the growing impact of AI-driven platforms. In this context, it is worth mentioning the growing role of media regulation and policymaking. Additionally, over the same period, the number of articles about “news” and “journalism” in SCAI context tripled.

Figure 6.2 presents a word cloud of the most frequently used words in articles abstracts on SCAI and disinformation, with top ten most frequently words used being “information”, “security”, “artificial intelligence”, “deepfakes”, “health”, “disinformation”, “fake”, “social”, and “news”.

Theories of Journalism Related to SCAI in Academic Research

To answer RQ2 and RQ3 (What theories of journalism related to AI are present in academic research between 2009 and 2023? Are these AI-related journalism theories embedded in one discipline or do they take an inter-/transdisciplinary perspective?), a QCA (Schreier 2012; Schreier et al. 2019; Mayring 2022; Kuckartz and Rädiker 2023) of 32 academic articles in English was conducted ($N=32$, screened out of the Scopus database between 2009 and 2023 with keywords “artificial intelligence”, “journalism”, and “theor*”). The main issues raised in articles on journalism theories related to AI were grouped thematically and discussed.

The analysis showed that before 2020, not much attention had been paid to journalism theories in the SCAI context (cf. Figure 6.3).

Although the journalism theories were evoked between 2009 and 2019 sparsely, the challenges to current journalistic principles and the potential effects of AI algorithms on the field of journalism, as well as on its roles in a democratic society were among the most important emerging and explored issues related to theoretical considerations. Some articles (e.g., Kanozia 2019) focused on media outlets which are actively introducing AI into the work of editorial offices and use applications for processing big data and fighting fake news. The functional aspect of AI in journalism, as well as forecasts for the future of the journalistic profession in the

face of the technological boom were likewise discussed. Other articles researched trends in journalism ethics in the age of AI, indicating that it is an overlapping field between philosophy, law, and journalism studies and that scholarly investigations into journalism ethics ought to recalibrate their trajectory, distancing themselves somewhat from a practice-centric framework and resolution-oriented model and moving towards the ethics of communication for providing more communicative possibilities with the philosophy of science, computer ethics, and engineer ethics.

The most cited text in the analysed period, entitled *Automation, Journalism, and Human–Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News* (Lewis, Guzman and Schmidt 2019) suggested academic researchers to pay attention to the scholarship of Human-Machine Communication (HMC), challenging traditional communication research, which tends to assume that humans are communicators and machines are mediators, whereas HMC takes the perspective that machines are stepping into formerly human roles. As the paper's authors underline, journalism studies – in particular research focused on automated journalism – has much to learn from HMC, which they argue offers

an emerging conceptual framework and empirically grounded research domain that has formed in response to the growing number of technologies—such as chatbots, social bots, and other communicative agents enabled by developments in artificial intelligence (AI)—that are designed to function as message sources, rather than as message channels.

(ibid., 409)

This shift in theory prompts researchers to reconsider what or who can be considered a communicator. It also leads them to explore how interactions between humans and machines can form social relationships. Furthermore, the potential consequences of these developments for individuals, society, and the field of communication should also be examined (cf. also Lee 2020). Lewis and colleagues claim that

in the particular case of automated journalism—in which software assumes a news-writing role that has long been considered a distinctly central, and indeed human, element of journalism—the introduction of HMC offers a generative starting point for theory development, advancing our understanding of humans, machines, and news for an oncoming era of AI technologies.

(Lewis et al. 2019: 409)

In light of the results of our analysis, AI in journalism studies since 2020 has usually been researched with regards to the technological implications for journalistic routines and the perceptions of journalists about the phenomenon. In particular, journalists' perceptions and attitudes towards the use of AI algorithms in detecting fake news have been examined (Shehata 2023), as too have challenges to individual and collective freedoms, and to the news media's role in democratic societies (Baroni 2023).

Some authors (e.g. [Sirén-Heikel et al., 2023](#)) present a perspective that automating newswork with AI requires a (re)definition of journalistic logics. Journalism – defined as a logic enacted through professional norms, practices, and values – represents an organisational logic that often conflicts with the efficiency logic of automation. The use of AI tools for streamlining and optimising journalistic work is presented by technologists/tech experts as a solution for improving journalism and enabling news work to move away from routine tasks. Technologists “view adopting AI as nudging the normative concept of what journalism should be (e.g. [Hanitzsch and Vos 2017](#)) towards skills they consider valuable in journalism; reconsidering news work as analysis, contextualization, and investigation” (Sirén-Heikel et al.: 363).

Another set of texts concerned AI in journalism education, e.g. by examining computational thinking processes in modelling unstructured data – [Jiang et al. \(2023\)](#) advocate that journalism students should understand the role and responsibility of modellers in creating AI technologies.

The results of one of the analysed articles (cf. [Lermann Henestrosa, Greving and Kimmerle 2023](#)) revealed no differences in perceived credibility and trustworthiness between AI-written and human-written texts where simple and short text types are concerned. Other studies (e.g. [Jang, Kwak and Bucy 2022](#)) found that the evaluation of algorithmically generated news was dependent on, inter alia, AI knowledge level, age group, or cultural conditioning (the first study to theoretically examine how news readers perceive algorithm-generated news from a cultural point of view was [Kim and Lee 2021](#)).

Yet another study (cf. [Shin 2022](#)) investigated how anthropomorphisms influence the perception of users about whether they are conversing with a human or an algorithm in the context of chatbot news. Within this, they offered a theoretical lens through which to characterise humanness as a key mechanism of human-artificial intelligence interaction, of which the eventual goal is for humans to perceive AI as human beings.

In a global context, discussing digital platforms and their role in the infrastructure of news organisations, one study ([de-Lima-Santos et al. 2023](#)) examined the Google News Initiative (GNI), which claims to financially support the creation of quality journalism in the digital age and bring technological advancements and innovation into newsrooms’ operations. Drawing on journalism innovation and platformisation theories, the paper argued that the GNI dedicated to media in Africa and the Middle East created power asymmetries and built an infrastructure of power and dependency that posed risks to responsible innovation in journalism. Likewise, in other regions of the world, analyses have been conducted using the Human-Machine Communication framework to see whether journalism in, for instance, low-income countries (e.g., Pakistan) has begun to feel the ripple of AI for technologically advanced journalistic practices and newsrooms ([Jamil, 2021](#)).

Moreover, authors of analysed texts (e.g. [Zhang and Pérez Tornero 2023](#)) advocate a new framework and a transdisciplinary field to examine the integration of AI and associated technologies into journalism. Involving non-conventional units beyond the scope of the news media in collaboration in order to achieve a greater diversity of insight and perspectives becomes important. Equally important

becomes the quality of interdisciplinary research which should be strengthened to establish a harmonising force between the humanities, social sciences, and the prevailing discourse systems of technological progress and economic growth. The main principles of journalism ethics such as truthfulness and objectivity remain relevant, however, the focus of traditional ethics of journalism on the content of news should be complemented by journalism ethics in the AI age, encompassing elements such as the ethics of data and coding. [Al-Zoubi, Ahmad and Hamid \(2023: 401\)](#) point out that “the main ethical challenges faced by journalists in the newsroom in adopting AI are data bias; privacy violations; and the absence of legislation and international regulations regarding the use of AI in journalism”.

One of the analysed articles is a review of nearly a century of publications in *Journalism and Mass Communication Quarterly* (JMCQ) ([Kim et al. 2023](#)), which shows dominant paradigms in media studies – media effects and communicator uses – and offers “seven initiatives for the next century: revisiting classical theories, fostering interdisciplinary collaboration, balancing descriptive and prescriptive theorization, nurturing indigenous theorizing, collaborating with industry, reverse theorizing with AI, and exploring and regulating AI’s role in media”.

Overall between 2009 and 2023, various theories were applied to research on journalism and AI. For example:

- a Drawing on propositions from the HAI-TIME (Human–Artificial Intelligence [AI] Interaction and the Theory of Interactive Media Effects) and Persuasion Knowledge Model, [Jang et al. \(2022\)](#) examined how knowledge of automated journalism (AJ) moderates the evaluation of algorithmically generated news;
- b Relying on Actor-network theory and the MAIN Model (Modality, Agency, Interactivity, Navigability), [Owsley and Greenwood \(2022\)](#) attempted to determine a correlation effect between people’s perception and awareness of the operationalisation of AI in their everyday lives and in the production, presentation, and publication of news media in the US; and
- c Using neo-institutional theory and policy network theory, [Kuai, Ferrer-Conill and Karlsson \(2022\)](#) investigated the copyright law’s impact on AI innovation in newsrooms beyond the West, namely in the hitherto unexamined Chinese context, where the copyright legal framework – as the results of investigation indicate – prioritises tech companies while undermining journalistic autonomy and favours state-owned, resource-rich official media over the private sector.

It should be noted that most of these theories appeared only once in the data collected. However, it can be noticed that many of the theories that were identified have a common interest in responsible, ethical, and trustworthy AI.

Our study shows that various researchers indicate that there is a need for a theoretical framework to instruct the AI-adopting process in journalism and to “better integrate the cross-domain knowledge” ([Zhang and Pérez Tornero 2023](#)). The AI capabilities to automate news production, personalise content delivery, and analyse data to uncover hidden patterns among others have led to the emergence of new journalism theories that intersect with computer science, communication

studies, and sociology, as well as with law and ethics. This intersection has led to the development of theories which, for instance, examine the implications of AI-written news on accuracy, objectivity, and journalistic ethics. The results of our study indicate that understanding AI's role in journalism requires an interdisciplinary approach. By integrating insights from different disciplines, comprehensive theories can be developed that capture the complexities of AI-driven journalism. This interdisciplinary perspective is crucial in guiding the ethical and responsible use of AI in journalism.

Discussion and Final Remarks

This chapter contributes to existing research in the discipline of media and communication by investigating studies on SCAI with a specific focus on journalism theories related to AI. The systematic review of the literature and an exploration of current research referring to theoretical approaches in journalism published in high-ranked academic articles offer important insights into the various theories discussed in the field.

The results of this study indicate a need for a new perspective and a novel theory of communication and journalism. Especially when complex terminologies and definitions challenge academic and non-academic frameworks. A comprehensive summary of the social-fake-news media landscape became available, allowing journalism studies to move forward in the context of AI media (Feher 2024). Accordingly, there is a growing necessity for more transdisciplinary approaches to AI in journalism, underscoring a significant increase in academic theories, mirroring observed trends. (e.g. Shue, O'Hara and Stamp 2024). However, a significant majority of theories appeared only once in examined articles over the analysed period of time. This is a common trend (ibid.: 2). It demonstrates the extensive scope of communication, considering the multitude of diverse theories that apply to empirical research. Stamp and Shue (2013) argue, however, that this trend also shows a propensity in the field for theory-shopping, which can hinder the progression of theoretical development. Nevertheless, the current social, media, and information landscape provides many opportunities for the advancement of theory. As presented in this chapter, theories used for research on journalism related to AI remain very diverse, and some offer the potential for development.

Collectively and individually the leading papers analysed claim that AI should be a Human-Machine Communication tool for humans and its role should be to assist and support in the execution of routine tasks. It also should enhance humans' capabilities and fulfillment of their needs, rather than replace them. From the perspective of public spheres, human journalistic work is (still) of great importance. However, recent developments and the latest research indicate different approaches in this regard. As studies show (Feher et al. 2024), the future of AI-driven media and info-communication, as envisioned by experts from all world regions, will be closely related to AI-to-AI solutions and this focus on AI stems from an intention to mitigate technology-driven misuse and misinformation. However, as Feher et al. (2024: 1) rightly point out, the forecasts provided by these

technology ambassadors “could be construed as being indicative of professional near-sightedness and cognitive dissonance”.

New perspectives on the relationship between human and machine will also be impacted by AI itself (e.g. its quality) and its future development. It is crucial paying attention to data that AI will be able to use. Innovating systems require ample data to train technologies to quickly generate human-like contents and leading chatbots are also learning from these contents. Here media organisations play an important role – recent examples show that their decisions, and the changes they subsequently bring about may have a profound impact on AI and society in general.

There is a noticeable trend (and future signal) of large technology companies paying media outlets to publish a certain amount of content prepared using AI (Stenberg, 2024). These are not solutions that would help journalists create articles, but those that collect and aggregate content from sources provided by the publisher and create summaries based on them. Additionally, the publishers are not required to disclose their use of AI. Despite the advantages (such as facilitating the process of research and verifying information), the popularisation of the type of AI tools described also has negative consequences, as it favours the creation of reproductive content based on already published materials and interferes with source materials.

The spokespersons of tech giants remind us that the AI tools “are not intended to, and cannot, replace the essential role journalists have in reporting, creating and fact-checking their articles” (ibid.), however actions such as the above-mentioned GNI should be carefully thought through and although there are positive attempts to counter disinformation – including within GNI – technology companies nowadays have more and more control over what information reaches media recipients and they should take more responsibility for shaping the public sphere. Moreover, as publishers observe the different activities of big technology companies they react to the unlawful use of their materials (e.g. NYT lawsuit⁵). Established media companies, such as the *New York Times*, *Washington Post*, *Guardian*, or *Atlantic*, and majority of other leading information services ban AI usage of their materials by blocking AI bots, which means that the data AI will be trained in is being significantly limited. AI will not have access to up-to-date verified information from reputable sources which may lead to many problems in the context of AI-generated content.

Considering SCAI’s perspective of technology as a scaling tool, employing AI in journalism amplifies information dissemination significantly, thereby profoundly influencing the very fabric of democracy. However, relying solely on AI may be unfavourable for media organisations or even societies in general. Moreover, journalists might be given a false impression that they have been taken care of and do not need to do anything more as solutions have been provided by the system and no other action is needed. We already encounter similar situations in contemporary journalism, as in the case of publishing materials from public relations specialists and the use of ready-made texts from PR agencies (see Anderson and Ward 2016), which has various reasons, such as the disappearance of independent sources of reporting, time pressure, and profit-seeking.

Expected to both intensify and trigger a paradigm shift in journalism, these trends hold profound implications for media and communication research. This claim is further supported by the lack of extensive analysis of generative AI, including

technologies like ChatGPT, during the review period. The reason stems from the lengthy process of publishing peer-reviewed, high-impact papers relative to the rapid emergence of ChatGPT and similar technologies. As the academic examination of generative AI is anticipated to increase sharply after 2024, it becomes essential to promptly reevaluate the direct effects of these generative technologies on the field, among others, by repeating this research for the next year(s).

The [Vienna Manifesto on Digital Humanism \(2019\)](#) emphasises the need to address the challenges of building a just and democratic society with humans at the centre of technological progress. There is no doubt that media and journalists executing their professional duties and responsibilities contribute to the building of a just and democratic society. However, embracing the technological progress and putting humans at the centre of attention may prove to be quite a challenge, in particular for the AI-driven ethics in media and journalism.

Notes

- 1 The *Information Disorder: Toward an interdisciplinary framework for research and policymaking* report introduces a new conceptual framework for examining information disorder, identifying the **three different types: mis-, dis-, and mal-information**. The differences between these three types of information are described using the dimensions of harm and falseness: a) **mis-information** is when false information is shared, but no harm is meant; b) **dis-information** is when false information is knowingly shared to cause harm; c) **mal-information** is when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere ([Wardle & Derakhshan, 2018: 5](#)).
- 2 Cheapfakes – in contrast to deepfakes, i.e., video-based disinformation created with AI – are forms of visual disinformation for which authentic images or videos are re-contextualised to deliberately alter their meaning ([Hameleers 2024; Paris & Donovan, 2020](#)). Cheapfakes require relatively little effort, editing skills, and resources.
- 3 An “echo chamber” ([Nguyen 2020](#)) is an epistemic environment in which people come across beliefs and opinions that align with their own, which creates a self-reinforcing mechanism that pushes them towards more extreme positions.
- 4 The term “filter bubble” was coined by [Pariser \(2011\)](#). He predicted that individualised personalisation by algorithmic filtering would lead to intellectual isolation and social fragmentation. Users become separated from information that disagrees with their viewpoints. They are isolated in their own cultural or ideological bubbles, which results in a limited and customised view of the world.
- 5 *The New York Times* (NYT) filed a lawsuit against OpenAI and Microsoft on December 27, 2023. The lawsuit alleges that these companies used millions of the media outlet’s copyrighted articles without authorisation to train the AI models (cf. [Grynbaum and Mac 2023](#)).

References

- Aguinis, H., Cummings, C., Ramani, R. S., & Cummings, T. G. 2020. “An A Is an A: The New Bottom Line for Valuing Academic Research”. *Academy of Management Perspectives*, 34(1), 135–154.
- Al-Zoubi, O., Ahmad, N., & Hamid, N. A. 2023. “Artificial Intelligence in Newsrooms: Ethical Challenges Facing Journalists”. *Studies in Media and Communication*, 12(1), 401–409.
- Anderson, P. J., & Ward, G. 2016. *The Future of Journalism in the Advanced Democracies*. Taylor and Francis.

- Baroni, A. 2023. "Artificial Intelligence in and through the Media. A Critical Review of the European Approach to AI and the Potential Implications for Journalistic Agency, Media Transparency and Accountability". *Comunicazione Politica, Quadrimestrale Dell'Associazione Italiana Di Comunicazione Politica*, 1, 29–50.
- Bennett, W. L., & Kneuer, M. 2023. "Communication and Democratic Erosion: The Rise of Illiberal Public Spheres". *European Journal of Communication*, 39(2), 1–20.
- Bennett, W. L., & Livingston, S. 2018. "The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions". *European Journal of Communication*, 33(2), 122–139.
- Bennett, W. L., & Pfetsch, B. 2018. "Rethinking Political Communication in a Time of Disrupted Public Spheres". *Journal of Communication*, 68(2), 243–253.
- Blassnig, S., Strikovic, E., Mitova, E., Urman, A., Hannák, A., de Vreese, C., & Esser, F. 2024. "A Balancing Act: How Media Professionals Perceive the Implementation of News Recommender Systems". *Digital Journalism*, 1–23.
- Castells, M. 2008. "The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance". *The Annals of the American Academy of Political and Social Science*, 616(1), 78–93.
- Dahlgren, P. 2005. "The Internet, Public Spheres, and Political Communication: Dispersion and Deliberation". *Political Communication*, 22(2), 147–162.
- de Tocqueville, A. 2004. *Democracy in America*. Library of America.
- de-Lima-Santos, M., Munoriyarwa, A., Elegu, A., & Papaevangelou, C. 2023. "Google News Initiative's Influence on Technological Media Innovation in Africa and the Middle East". *Media and Communication*, 11(2), 330–343.
- Elliott, A. 2018. "The Culture of AI". *Everyday Life and the Digital Revolution*, Routledge.
- Elliott, A. 2022. *Making Sense of AI: Our Algorithmic World*. Cambridge: Polity Press.
- Feher, K. 2024. "Exploring AI media. Definitions, Conceptual Model, Research Agenda". *Journal of Media Business Studies*, 1–24.
- Feher, K., & Katona, A. I. 2021. "Fifteen Shadows of Socio-Cultural AI: A Systematic Review and Future Perspectives". *Futures*, 132, 1–19. <https://doi.org/10.1016/j.futures.2021.102817>
- Feher, K., & Veres, Z. 2022. "Trends, Risks and Potential Cooperations in the AI Development Market: Expectations of the Hungarian Investors and Developers in an International Context". *International Journal of Sociology and Social Policy*, 43(1/2), 107–125.
- Feher, K., Vicsek, L., & Deuze, M. 2024. "Modeling AI Trust for 2050: Perspectives from media and Info-Communication Experts". *AI & SOCIETY*, 1–14.
- Friemel, T. N., & Neuberger, C. 2023. "The Public Sphere as a Dynamic Network". *Communication Theory*, 33(2–3), 92–101.
- Grynbaum, M. M., & Mac, R. 2023. "New York Times Sues OpenAI and Microsoft Over Use of Copyrighted Work", *The New York Times*. <https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>
- Hameleers, M. 2024. "Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting". *International Journal of Public Opinion Research*, 36(1), 1–9.
- Hanitzsch, T., & Vos, T. P. 2017. "Journalistic Roles and the Struggle Over Institutional Identity: The Discursive Constitution of Journalism". *Communication Theory*, 27(2), 115–135.
- Jamil, S. 2021. "Artificial Intelligence and Journalistic Practice: The Crossroads of Obstacles and Opportunities for the Pakistani Journalists". *Journalism Practice*, 15(10), 1400–1422.
- Jang, W., Kwak, D. H., & Bucy, E. 2022. "Knowledge of Automated Journalism Moderates Evaluations of Algorithmically Generated News". *New Media and Society*. <https://doi.org/10.1177/14614448221142534>
- Jiang, S., Qian, Y., Tang, H., Chao, J., & Finzer, W. 2023. "Examining Computational Thinking Processes in Modeling Unstructured Data". *Education and Information Technologies*, 28(4), 4309–4333.

- Kanozia, R. 2019. "Analysis of Digital Tools and Technologies for Debunking Fake News". *Journal of Content, Community and Communication*, 9: 114–122. <https://doi.org/10.31620/JCCC.06.19/16>
- Kim, J. N., Chiu, M. M., Lee, H., Oh, Y. W., Gil de Zúñiga, H., & Park, C. H. 2023. "Mapping Media Research Paradigms: Journalism & Mass Communication Quarterly's Century of Scientific Evolution". *Journalism & Mass Communication Quarterly*, 100(4), 736–772.
- Kim, Y., & Lee, H. 2021. "Towards a Sustainable News Business: Understanding readers' Perceptions of Algorithm-Generated News Based on Cultural Conditioning". *Sustainability*, 13(7). <https://doi.org/10.3390/su13073728>
- Kuai, J., Ferrer-Conill, R., & Karlsson, M. 2022. "AI ≥ Journalism: How the Chinese Copyright Law Protects Tech Giants' AI Innovations and Disrupts the Journalistic Institution". *Digital Journalism*, 10(10), 1893–1912.
- Kuckartz, U., & Rädiker, S. 2023. *Qualitative Content Analysis. Methods, Practice and Software*. London: Sage.
- Lee, E. A. 2020. *The Coevolution: The Entwined Futures of Humans and Machines*. MIT Press.
- Lermann Henestroza, A., Greving, H., & Kimmerle, J. 2023. "Automated Journalism: The Effects of AI Authorship and Evaluative Information on the Perception of a Science Journalism Article". *Computers in Human Behavior*, 138. <https://doi.org/10.1016/j.chb.2022.107445>
- Lewis, S. C., Guzman, A. L., & Schmidt, T. R. 2019. "Automation, Journalism, and Human–Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News". *Digital Journalism*, 7(4), 409–427.
- Liu, Z. 2021. "Sociological Perspectives on Artificial Intelligence: A Typological Reading". *Sociology Compass*, 15(3), 1–13.
- Mayring, P. 2022. *Qualitative Content Analysis*. Sage Publications.
- Meskys, E., Kalpokienė, J., Jurcys, P., & Liaudanskas, A. 2019. "Regulating Deep Fakes: Legal and Ethical Considerations". *Journal of Intellectual Property Law & Practice*, 15(1), 24–31.
- Newman, N. 2023. "Reuters Institute Digital News Report 2023". https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf
- Newman, N. 2024. "Journalism, media, and technology trends and predictions 2024". <https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2024>
- Nguyen, C. T. 2020. "Echo Chambers and Epistemic Bubbles". *Episteme*, 17(2), 141–161.
- Nuernbergk, C. 2022. "Public Sphere Conceptions: Public Sphere Theory". In: J. Krone, T. Pellegrini (Eds.), *Handbook of Media and Communication Economics*, 1–24. Wiesbaden: Springer.
- Owsley, C. S., & Greenwood, K. 2022. "Awareness and Perception of Artificial Intelligence Operationalized Integration in News Media Industry and Society". *AI and Society*, 39, 417–431.
- Paris, B., & Donovan, J. 2020. "Deepfakes and Cheapfakes: The Manipulation of Audio and Visual Evidence". Data & Society Report, September 8.
- Pariser, E. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin.
- Parratt-Fernández, S., Mayoral-Sánchez, J., & Mera-Fernández, M. 2021. "The Application of Artificial Intelligence to Journalism: An Analysis of Academic Production". *Profesional de la Información*, 30(3), 1–12.
- Reese, S. 2016. *Theories of Journalism*. Oxford Research Encyclopedia of Communication.
- Schlesinger, P. 2020. "After the Post-Public Sphere". *Media, Culture & Society*, 42(7–8), 1545–1563.
- Schreier, M. 2012. *Qualitative Content Analysis in Practice*. Sage Publishing.
- Schreier, M., Stamann, C., Janssen, M., Dahl, T., & Whittall, A. 2019. "Qualitative Content Analysis: Conceptualizations and Challenges in Research Practice – Introduction to the FQS Special Issue 'Qualitative Content Analysis I'". *Forum: Qualitative Social Research*, 20(3).

- Shehata, S. O. 2023. "Saudi Journalists Employing Artificial Intelligence Algorithms to Detect Fake News". *Journal of Statistics Applications and Probability*, 12(3), 877–896.
- Shin, D. 2022. "The Perception of Humanness in Conversational Journalism: An Algorithmic Information-Processing Perspective". *New Media and Society*, 24(12), 2680–2704.
- Shue, C. K., O'Hara, L. L. S., & Stamp, G. 2024. "Fifty-Years of Theory-Driven Research in HCR: Prominence, Progress, and Opportunities". *Human Communication Research*, 50(2), 146–153.
- Sirén-Heikel S., Kjellman, M., & Lindén, C.G. 2023. "At the Crossroads of Logics: Automating Newswork with Artificial Intelligence — (Re)defining Journalistic Logics from the Perspective of Technologists". *Journal of the Association for Information Science and Technology*, 74(3), 354–366.
- Stamp, G. H., & Shue, C. K. 2013. "Twenty Years of Family Research Published in Communication Journals: A Review of the Perspectives, Theories, Concepts, and Contexts". In A. L. Vangelisti (Ed.), *The Routledge Handbook of Family Communication* (2nd ed.), 11–28. Routledge.
- Stenberg, M. 2024. Google Is Paying Publishers Five-Figure Sums to Test an Unreleased Gen AI Platform. [Adweek.com](https://www.adweek.com/google-is-paying-publishers-five-figure-sums-to-test-an-unreleased-gen-ai-platform/).
- TheViennaManifestoonDigitalHumanism.2019.Digital_Humanism_Manifesto-June_2019-FINAL.indd (tuwien.ac.at). https://dighum.ec.tuwien.ac.at/wp-content/uploads/2019/07/Vienna_Manifesto_on_Digital_Humanism_EN.pdf
- Waisbord, S. 2016. "Disconnections: Media Sociology and Communication across Differences". Paper Presented at the Conference of the International Communication Association, Fukuoka, Japan. As cited in: Bennett, W. L., & Pfetsch B. 2018. "Rethinking Political communication in a Time of Disrupted Public Spheres, ". *Journal of Communication*, 68(2):). 243253.
- Wardle, C., & Derakhshan, H. 2018. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*. Council of Europe.
- Winiarska-Brodowska, M., ed. 2020. "In Search of a European Public Sphere". *Challenges, Opportunities and Prospects*, Cambridge Scholars Publishing.
- Zhang, W., & Pérez Tornero, J. M. 2023. "Introduction to AI Journalism: Framework and Ontology of the Trans-Domain Field for Integrating AI into Journalism". *Journal of Applied Journalism and Media Studies*, 12(3), 333–353.
- Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. Profile Books.

7 Narrating AI Stories

Selected Strategies

Mirosław Filiciak and Kuba Piwowar

Introduction: “Made in China, Designed in California, Criticised in Europe”¹

The narratives on artificial intelligence (AI) present in the public discourse indicate the enormous potential of this tool on the one hand, and express concerns about it on the other. As AI evolves, understanding the stories about it, their narrators and intentions remains crucial. Narrative analysis draws from various sources including corporate communications, legal discourses, national strategies, scientific research, media and pop culture, revealing patterns that normalise the use of AI and envision societal development. These narratives often portray AI as an objective product based on mathematical and statistical principles, shaping cultural and civilisational trajectories. The analysis of these narratives prompts an examination of the metaphors used in the discourse, which may mask the impact of AI on the environment and divert attention from issues such as privacy and individual freedom. Critical examinations of AI reveal its embeddedness in broader political and social contexts, while the discourse surrounding AI reflects a complex interplay of narratives, metaphors and power dynamics. This chapter focuses on AI narratives in pop culture, particularly in movies and TV series, presenting quantitative findings on perceptions, hopes and fears about AI. Understanding these dynamics is crucial for navigating the ethical, social and political implications of AI development and deployment.

Narratives about AI are created by a variety of actors: it is not only government agencies that contribute to what we say and think about AI; the narratives are also built by technology companies. Those technological metaphors, introduced into the discourse by marketing departments, are needed – if only to facilitate the use of new services. For example, this article was first created on the computer screens of its authors, who saw the characters appearing on the “pages” and placed the saved files in “folders,” but in fact, these folders on the hard drives of computers look nothing like the folders one keeps in an office. However, using terms familiar to users from their everyday lives makes it easier for them to navigate the digital world. This power of language goes even further: metaphors increasingly have a masking function. For example, cloud-based services hide the fact that, invisible to users, data transfer networks rely on material infrastructure: servers that use scarce minerals, consume energy and pollute the water necessary for cooling. The term

“cloud” thus serves to normalise the negative phenomena associated with extractivism and to divert attention from the material side of services where, if they were treated like other elements of public infrastructure, they would probably raise more questions about how they are used or simply who owns and controls them.

In the case of AI, a classic example of deconstructing its name can be found in Kate Crawford’s book *Atlas of AI* (Crawford, 2021), in which the author shows that the embedded in materiality, so-called AI is not artificial, nor is it intelligent, if we reject the extreme reductionist approach to intelligence. Most importantly, however, AI, like any technology, operates within a broader political and social context. One should not accept the establishment of control over it outside democratic structures. As Crawford puts it, “[i]f AI is defined by consumer brands for corporate infrastructure, then marketing and advertising have predetermined the horizon. If AI systems are seen as more reliable or rational than any human expert, able to take the ‘best possible action,’ then it suggests that they should be trusted to make high-stakes decisions in health, education, and criminal justice” (Crawford, 2021, pp. 7–8). And yet, it is not hard to imagine that the discussion of AI would look different if we used other metaphors to describe its workings more broadly, such as, for example, the “stochastic parrot” proposed in the acclaimed text by Emily M. Bender, Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell (Bender et al., 2021). The large language model that is dominant in AI services today is “a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot” (Bender et al., 2021). Our attitude towards AI would probably also be different if we used a term that ignores the principles of the technology and instead focuses on its actual frequent use, as in Ted Chiang’s suggestion to call AI that blindly obeys the will of a corporation the successor to the management consulting firm facing well-documented ethical issues: “the new McKinsey” (Chiang, 2023). This is not just a matter of appealing to emotions and looking for comparisons that might shock, such as the one to the consulting firm whose infamous contribution to the US opioid crisis has made it one of the symbols of greed. One can imagine using language that includes more terms that allow us to discuss details at a higher level of precision. Such proposal appears, for example, in a short report published by the European Parliament: “In debates about AI, we could follow the approach of many developers in referring to specific techniques and application contexts. For example, if we target messages towards ‘machine learning diagnostic support’ or ‘biometric identification in public spaces’ rather than just ‘AI,’ we might reveal more common ground between our positions. Furthermore, by reducing our reliance on core metaphors such as ‘intelligence’ and ‘trust,’ which allude to human qualities and capabilities, we could create space for new metaphors that describe AI in its own terms” (Boucher, 2021).

In the debates about AI, there are many references to the issue of openness, understood as transparency, but also accessibility, which for several decades has accompanied the discussion of the relationship between software and power in the open source movement. Its objectives go beyond software itself, as exemplified by

free licenses such as Creative Commons, as well as contemporary standards for understanding openness in science. This chapter will not go into these discussions in detail, but it is worth mentioning that this approach has faced criticism, based, for example, on the observation that in a world of inequality, fully open access allows for a more efficient multiplication of resources for the privileged, thus actually exacerbating unequal access (see: [Lovink, 2006](#)). In the case of AI, the issue becomes even more complicated, not least because AI systems are centralised and require enormous financial resources to develop and maintain. Therefore, while the AI business is keen to use the term “open,” given the complexity of this technology, it is not always clear what this openness is supposed to refer to: it may refer to access, but also to data, not necessarily to code. For instance, certain companies might provide access to datasets for research or analysis purposes, allowing researchers or developers to use the data freely. However, the underlying code or proprietary algorithms used to train models are often kept closed or restricted. As David Gray Widder, Meredith Whittaker and Sarah West wrote, “some companies have moved to embrace ‘open’ AI as a mechanism to entrench dominance, using the rhetoric of ‘open’ AI to expand market power, and investing in ‘open’ AI efforts in ways that allow them to set standards of development while benefiting from the free labor of open source contributors” ([Gray Widder et al., 2023](#)).

Sources and Types of Narratives: Literature Review

Research on AI narratives and the description of narratives that contribute to shaping its image has been ongoing since at least the beginning of this century. Stephen Cave, Kanta Dihal, Sarah Dillon ([Cave et al., 2018](#)) edited a book that explores the ways in which narratives on AI, and more broadly, intelligent machines, are constructed, focusing on both historical perspectives and contemporary approaches to understanding AI. Research on narratives attempts to answer questions about their impact on the public’s views on technology ([Fast & Horvitz, 2017](#)) and their influence on public policy ([Hudson et al., 2021](#)). Researchers also sought to understand the role that large technology companies play in shaping AI narratives ([Crawford, 2021](#)). [Chubb et al. \(2022\)](#) explored missing AI narratives in their article based on interviews with experts in the field. In this regard, it is important to outline the existing types of AI narratives, the sources they emerge from, as well as a critical perspective on why certain types of narratives are missing from the discourse.

As researchers involved in the *Portrayals and perceptions of AI and why they matter* project at The Royal Society point out, “popular portrayals of AI in the English-speaking West tend to be either exaggeratedly optimistic about what the technology might achieve or melodramatically pessimistic” ([Cave et al., 2018](#)). This raises the question of who is responsible for building these narratives, how they reach the public, and which narratives occupy a place on such a spectrum. The first mentions of robots appeared in Homer’s *Iliad*, although for the following centuries, they were almost non-existent ([Cave et al., 2018](#), p. 7). Both fictional representations of robots and machines, as well as real-life projects and implementations

of ideas about their functioning, began to dominate the common imagination in the 17th century, starting with Jacques de Vaucanson's Flute Player, Mechanical Turk by Wolfgang von Kempelen in the 18th century and the robots in Karel Čapek's play *R.U.R.* The 20th century brought not only rich science fiction literature but also movies and TV series inspired by popular depictions of the relationship between humans and machines. In this sense, the primary source of AI narratives is literary fiction, as well as movies and TV series, largely in the sci-fi genre, although over the past 40 years the percentage of AI-related films so classified has fallen from 100% to 75–85% (Recchia, 2020, p. 386). The mainstream media reinforces existing narratives about “terrifying robots” through the use of specific sounds and images (Chubb et al., 2022).

What are the sources of information about AI? In a survey conducted by Ipsos in Poland (Ipsos, 2023b), the authors of this text asked respondents about the sources of their knowledge regarding AI. The majority, as much as 56%, indicated the media and then conversations with friends (35%). Then followed science fiction movies and TV series (30%), online lectures (15%) and social media content (15%). Studies and courses were mentioned by 5% of respondents, books and scientific articles by 3%, popular science books by 12%, comics and science fiction books by 8% (Ipsos, 2023b). This demonstrates the somewhat paradoxical situation in which a new, advanced technology with significant potential to influence the lives of individuals – if only by automating various segments of the labour market – is perceived largely through fiction. At the same time, it confirms the importance of the power of stories, which is also relevant to the possibility of building counter-narratives.

In their research, Chubb et al. (2022) point out the fundamental types of AI narratives identified by the experts they interviewed. Among them are solutionism (a belief that technology can solve the majority, if not all, the known problems), superintelligence and sci-fi, surveillance and privacy, job loss and automation, embodied AI and robots, Silicon Valley & Big Tech. This typology allows for a structured approach to examining how AI is portrayed and discussed in various contexts. It provides a comprehensive framework for analysing the discourse around AI and its societal implications, focusing on four major sources of these narratives: major corporations, governments, critical and activist collectives, and popular culture. Due to its cross-sectional nature, this chapter will not delve into the dominant types of narratives in the media, although it is undoubtedly a topic worthy of scholarly exploration. As pointed out further on, people in different countries vary significantly in their level of understanding of AI, as well as their attitudes towards it and the hopes or fears they place in it. The purpose of the next section of the chapter is to understand selected strategies for building AI narratives.

Big Tech

When ChatGPT caught the attention of the media and the public around the world, it became clear that from now on AI would be at the forefront of the narrative of technology companies, even though it had been vastly used by these companies before. With the success of the OpenAI model in the first quarter of 2023, AI has become a

crucial part of the Big Tech communications, including earnings calls (Mann, 2023). This success and subsequent business decisions, such as Microsoft's purchase of OpenAI, have once again highlighted the importance of AI development and its relevance to the strategies of technology companies. It also gave rise to an internal struggle within the society, culminating in the publication of a letter to stop work on large language models, which claimed that "AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research and acknowledged by top AI labs" (Future of Life Institute, 2023).

It is clear that the concentration of power in the hands of a few large tech companies not only poses direct threats to democracy, as demonstrated, for example, by the Cambridge Analytica scandal (see: Kakarchive et al., 2023; Wylie, 2019), but also results in the monopolisation of the AI narrative. Smaller firms lack competitive advantage, and surprisingly, narratives such as "slow down the development of AI," "regulate AI" are often publicised by big tech companies (see: De Vynck, 2023). At the same time, these same companies emphasise that AI can help advance mankind. In response to one of the lawsuits, Google stated that "[g]enerative artificial intelligence ('AI') holds perhaps unprecedented promise to advance the human condition. It is already beginning to revolutionize the way we use technology, serving as a companion that can help research, summarize, and synthesize information; brainstorm ideas; write original creative or factual text and software code; and create images, videos, and music. It will open doors to new insights and forms of expression, as well as better, personalized help and advice in areas such as education, health care, government services, and business productivity" (Herrman, 2023). In turn, Stability AI's response to a lawsuit that was filed with a claim that the company uses copyrighted photographs to train their models stated that AI is "rapidly expanding the boundaries of human creativity and capability" (Case, 2023, p. 11; Herrman, 2023).

The juxtaposition of these contrasting narratives – calls for regulation or cessation against claims of AI's promise for humanity – both coming from the PR and legal machinery of Big Tech, reveals the power dynamics of narrative ownership. Regardless of what GAFAM's² press releases claim and how these companies envision AI's development and its application in their products and services, all of this is given in the context of potential harm to humanity, as the aforementioned Future of Life Institute's open letter suggests. As researchers Alex Hanna and Emily M. Bender argue, "effective regulation of AI needs grounded science that investigates real harms, not glorified press releases about existential risks" (Hanna & Bender, 2023).

Governments

Ali A. Guenduez and Tobias Mettler (2023) examined how governments shape their AI policies and narratives. In their view, governments can play four different roles towards AI. First one, a regulator, seeks to mitigate potential risks and harms while making sure AI development aligns with societal values. Second one, an enabler, seeks to catalyse AI development by eliminating barriers and supporting companies working on AI. Third role, a leader, is the one in which the government funds its own research on the technology. The fourth role is a user, in which

governments use AI to improve their policy-making and support decision-making (Guenduez & Mettler, 2023). The researchers also focused on understanding the different types of narratives evident in national AI strategies. They identified six narrative strategies emerging from AI policies: (1) building an AI marketplace, (2) countering the winner-takes-all practice, (3) engaging in strategic collaboration for AI R&D, (4) creating ethical and trustworthy AI, (5) educating AI professionals and (6) advancing the deployment of AI in practice.

Interestingly, this analysis looks at AI from a rather commercial and practical perspective, partly addressing the opportunities arising from its development, as well as the potential risks and harms. From the perspective of “soft law,” that is, basing AI regulations on ethical standards and norms stemming from concerns about the social harm of AI development, the fourth narrative analysed by the researchers “emphasizes the role of the government as a regulator (Salvi del Pero & Verhagen 2024). By developing rules, standards, guidelines, norms, and ethical principles, it ensures that the negative social impacts of AI development and use are minimised and that citizens’ well-being is increased” (Guenduez & Mettler, 2023). This pragmatic approach underscores the importance of balancing innovation with ethical considerations and social welfare. Moreover, the researchers’ exploration of the role of “soft law” in AI regulation adds depth to the analysis by highlighting the significance of ethical standards and norms in shaping governance frameworks.

One of the prevailing narratives regarding the use of AI by governments, though mostly visible through the work of critical collectives, is that of the use of technology in warfare. For example, the use of computer vision during a drone-led battle would not have been possible without the involvement of Big Tech’s resources, as indicated by the collaboration of Microsoft, Google, Oracle and Amazon with the US government (Copp & O’Brien, 2022). The 2018 Project Maven controversy (Simonite, 2018) sparked a discussion within Google about its involvement in the contract – the Pentagon’s programme aimed at expediting the integration of cutting-edge AI technology into military operations (Shane & Wakabayashi, 2018). The controversy stemming from the fact that the company’s employees have not been informed what they were working on shows that the tension between the technology’s producers and its users has a significant impact on the way the discussion around AI is organised, especially regarding AI’s capabilities and the circumstances in which it should or should not be used.

It is also worth mentioning that the geographic distribution of different government narratives varies. If we look at the leaders of work on AI, namely the United States, China and the European Union, while they are all taking pro-regulatory measures, these measures have different characteristics and contexts. The EU AI Act, passed at the end of 2023, places the strongest emphasis on ethical issues and the estimation of societal risks associated with the new developments. While China also strengthens requirements for human oversight of AI, one might speculate that in this case, however, the key issue is maintaining the centralised state control (Strittmatter, 2020). The United States, in turn, places more emphasis on the requirement for transparency (Dixon, 2023). These solutions and related narratives are indirectly intertwined: for example, in discussions about European regulations,

the fear that too strong regulation will make it harder for local businesses to compete with US or Chinese companies has been expressed repeatedly.

Critical and Alternativist Collectives

Another set of AI narratives is provided by critical and alternativist collectives. Critical collectives are groups or communities of individuals who adopt a critical perspective towards technology and its societal implications. They typically focus on highlighting the negative social consequences of implementing AI and other digital technologies. These collectives often challenge dominant narratives about technology and advocate for alternative approaches that prioritise ethical considerations, social justice and human well-being over profit or efficiency. Alternativist collectives, in turn, may share similar concerns but focus more on proposing and promoting alternative frameworks, practices or technologies that align with their values and principles. They seek to offer alternatives to mainstream technological paradigms that they view as detrimental or oppressive, aiming to create more equitable and sustainable technological futures. The authors of the aforementioned article, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* (Bender et al., 2021), responded to the previously mentioned Future of Life Institute's letter (2023), which as of January 2024, was signed by over 33,000 people, including Elon Musk, Steve Wozniak and Yuval Noah Harari. In response to the letter calling for a slowdown (or even halt) of work on AI, especially large language models, Timnit Gebru (DAIR), Emily M. Bender (University of Washington), Angelina McMillan-Major (University of Washington), Margaret Mitchell (Hugging Face) wrote that the threats mentioned by the letter's signatories are "overshadowed by fearmongering and AI hype, which steers the discourse to the risks of imagined 'powerful digital minds' with 'human-competitive intelligence.' Those hypothetical risks are the focus of a dangerous ideology called longtermism that ignores the actual harms resulting from the deployment of AI systems today. The letter addresses none of the ongoing harms from these systems, including 1) worker exploitation and massive data theft to create products that profit a handful of entities, 2) the explosion of synthetic media in the world, which both reproduces systems of oppression and endangers our information ecosystem, and 3) the concentration of power in the hands of a few people which exacerbates social inequities" (Gebru et al., 2023). Focusing on "ongoing harms," critical collectives identify specific types of work that need to be done, as well as specific tools they use in their fight. They contribute to a broader set of negative AI narratives, yet not in a dystopian and pessimistic tone, but rather in a search for solutions and agency. Below, the work of a few selected critical collectives is presented.

Timnit Gebru, previously employed at Google but fired from the company mainly because of her work on the aforementioned article about stochastic parrots, established the Distributed AI Research Institute (DAIR) which is "a space for independent, community-rooted AI research, free from Big Tech's pervasive influence."³ The goal of the DAIR members work is to "...mitigate/disrupt/eliminate/slow down harms caused by AI technology, and cultivate spaces to accelerate

imagination and creation of new technologies and tools to build a better future.”⁴ Deeply rooted in the US context, DAIR has conducted research on topics such as race, worker exploitation and social media damage.

Another US-based institute is AI NOW that “produces diagnosis and actionable policy research on artificial intelligence.”⁵ Operating as an independent organisation, its main area of expertise is policy research that aims at “a) identifying and capitalizing on policy windows for movement, b) advancing narratives that illuminate a long term strategy for the field, c) catalysing energy toward action, working in close partnership with a broad coalition of allies.”⁶ Notable researchers include Meredith Whittaker, like Gebru formerly employed at Google, who left the company after becoming a whistleblower on the Maven project that was supposed to make Google one of the Pentagon’s contractors (Wiggers, 2019). It is important to point out that many whistleblowers in the technology industry end up working for critical collectives. They often come from underprivileged backgrounds, representing minority groups on the basis of gender, race, sexual orientation, among others.

One of the European critical collectives is Algorithm Watch, positioning itself as “fight[ing] for a world where algorithms and Artificial Intelligence (AI) do not weaken justice, democracy, and sustainability, but strengthen them.”⁷ They “envision a world where technology in general, and algorithmic systems in particular, are used to protect and benefit human beings. They should make societies more just, democratic, inclusive and sustainable, be it with regard to alleged or claimed race and gender, sexual orientation, abilities, age, wealth, class, or resource consumption.”⁸ While not focusing on AI per se, Algorithm Watch explores the intersection of technology and democracy, including research on ChatGPT, the ways in which implementing technology can yield discriminatory results, or the broad topic of automation.

Another examples include Data & Society which “studies the social implications of data-centric technologies, automation, and AI,”⁹ and The Algorithmic Justice League with a “mission to raise awareness about the impacts of AI, equip advocates with empirical research, build the voice and choice of the most impacted communities, and galvanize researchers, policy makers, and industry practitioners to mitigate AI harms and biases.”¹⁰ Other initiatives whose activities come close to building alternative narratives are also worth mentioning, as they engage in building a kind of counter-infrastructure for sectors particularly threatened by AI development. This is the case, for example, with the Glaze and the Nightshade projects (importantly, both originated at the university), which aim to enable graphic designers to obscure their work with dots invisible to the human eye. Their distribution makes it more difficult to train AI models on these images, and as a result, the extraction of creators’ work by large companies developing such services is slowed down. This topic indicates yet another space in which critical narratives emerge, namely art. It is an inexhaustible source of critical voices and the creation of new discourses that interact with popular ones, and as such, it deserves a separate study.

In conclusion, the discourse of AI is not merely a tale of technological triumphs, but a complex narrative woven with caution, responsibility and diverse perspectives.

At the centre of this narrative are critical collectives, not as doomsayers of a dystopian future, but as vigilant guardians and shapers of a more just and equitable technological world. Their work goes beyond mere criticism; it is a deep commitment to ensuring that AI serves as a tool for enhancing societal good, rather than exacerbating existing inequalities. Rooted in diverse experiences and driven by a deep understanding of the technology's potential and pitfalls, these collectives are not just responding to the phenomenon of AI, but are actively shaping its trajectory. They remind us that the path forward for AI is not predetermined, and that with thoughtful engagement, rigorous research and robust policy interventions, this powerful technology can be steered towards outcomes that will uplift all of humanity.

Popular Culture. Case Study: AI Narratives in Science Fiction Movies and TV Series

Popular culture, particularly through movies and music, significantly shapes our perception of technology, influencing whether we fear it or are fascinated by it (Cave et al., 2020). The abstract nature of technology, involving bits, binary systems and data processing, captivates visual art audiences. Pop culture narratives, like thousands of other science fiction narratives, shape our relationship with technology. They explore the relationship between technological research and state funding, the status of democracy in data-rich worlds and the dynamics between humans and machines. Science fiction movies are a rich source of inspiration, shaping our conceptualisation of technology, particularly AI. Between October 2022 and January 2023, a group of students at SWPS University in Warsaw, under the guidance of the authors, conducted a research study to understand and describe the dominant AI narratives evident in popular culture, especially in science fiction movies and TV series.¹¹ Conclusions were drawn on the basis of 55 movies and 22 episodes of various TV series, which were released between 1922 (*Metropolis*) and 2022 (the *Westworld* series), and twelve themes were highlighted: human-machine relations; citizenship; culture vs. nature; “new geographies”; pursuit of happiness; power dynamics; progress vs. conquest; humanisation; utopia vs. dystopia; freedom, security and enslavement; cooperation vs. conflict; trust vs. mistrust.

In movie narratives exploring human-machine relations, a complex interplay emerges, highlighting the notion of equality between humans and AI. These narratives often present idealised relationships, yet the underlying challenges and tensions persist. Movies address human fears about the growing power of AI and the critical need for ethical oversight of AI development. Some scenarios depict AI surpassing humans in social hierarchies, leading to conflicts. This cinematic discourse raises intricate questions about the evolving dynamics of human-machine relations in the technology-driven world. The portrayal of the social role of AI highlights its implications for citizenship, exploring the role of surveillance AI used by powerful entities to control society, raising issues of class stratification and invasion of privacy. AI has also been shown as a tool to maintain obedience among citizens, often through manipulative tactics, and used in risk management and security systems, which presents its potential for both benefit and misuse.

AI's influence on cultural narratives is significant, and it is often portrayed as a superior controller of society, in which algorithms prevail over human decision-making. However, AI's struggles to capture cultural nuances, focused on consumerism and environmental issues, lead to class disparities. Nature is dramatically depicted as unfit for human habitation due to technological and environmental destruction, and movies like *The Terminator* show the role of technology in climate disasters, yet at the same time, the technology offers some unique benefits. These narratives delve into the complex interaction between technology, power and environmental consequences.

Narratives categorised as “new geographies” explore AI's role in the conquest of Earth and space, as well as in the creation of AI-controlled realities. These stories often depict AI as beings with expansionist tendencies, seeking to colonise planets for resources and research. Humans are portrayed searching for new planets due to the depletion of the Earth's resources, driven by capitalism. AI is also shown as creators of algorithmically controlled realities, offering humans orderly environments. This theme intersects with themes of nature, culture and the pursuit of happiness. The pursuit of happiness in these narratives is ambiguous and includes themes of conquest, pleasure, relationships with non-human actors like robots, and the self-awareness of AI. These narratives feature escapism, technological fulfilment, emotional connections and AI emancipation, all intertwined with the search for happiness. These diverse stories reveal complex interactions between humans and machines, both on a personal and global scale, with unpredictable consequences.

Power dynamics in movies focusing on AI encompass aspects of control, emanation of power and rebellion. The narratives grapple with who holds power, depicting AI either as ruthless dominators subjugating humans, or as instruments of human control, treated like slaves. The evolution and empowerment of AI often results in changes in the exercise of authority, marked by cruel and brutal violence resulting from a lack of human consideration. Additionally, movies focus on the ability of AI to think independently, which allows it to rebel against established orders and challenge human rulers or owners.

The progress vs. conquest axis in the analysed films shows that progress and conquest are interconnected, not mutually exclusive. AI is often a key component of future societies, contributing to technological advancement and life improvement, especially in medicine and automation. However, this progress also brings about social and moral dilemmas, such as invasion of privacy and authoritarian control. Moreover, the conquest of new territories is frequently depicted as progress, although it is perceived differently by the conquerors and the conquered. The advancement of AI, while beneficial, also raises concerns about the loss of human autonomy and the potential for AI to become a tool of conquest, which often leads to a breakdown of the human-AI hierarchy and potentially results in AI dominating humans.

The theme of humanisation of AI is explored through aspects like cultural body, social body, “flesh,” emotions, self-awareness and relationships. These aspects show how AI is shaped by and reflects cultural and social norms, takes physical forms for interactions, displays emotions, develops self-awareness and forms

relationships. This analysis highlights AI's integration into society, raising questions about its role, rights and the nature of its interactions with humans. The utopia vs. dystopia theme in movies, in turn, often depicts a future in which AI leads to social and ecological degradation. This includes a negatively portrayed social system dominated by powerful corporations using AI for control, cities struggling with technological overdevelopment and social stratification, environmental disasters and the creation of digital worlds ruled by AI. There is also a pervasive fear of AI rebellion and the subjugation or annihilation of humans. The analysed movies mock the concept of a utopian society equipped with AI, suggesting that it is easier to imagine destruction than creation, and serve as cautionary tales highlighting the potential threats and consequences of AI development.

Movie narratives exploring the interplay of freedom, security and enslavement in the context of AI delve into a complex scenario in which increased security often limits freedom. These stories underscore society's dependence on technology, with extensive surveillance systems and control measures. The ethical dilemmas and social costs of AI's influence on this delicate balance provoke reflection on the trade-offs between personal freedom and collective security. Additionally, the central themes are the dynamics of cooperation and conflict between humans and AI. Movies like *Her* and *Transformers: Dark of the Moon* portray cooperation where humans and AI work together towards common goals, leveraging AI's unique capabilities for optimisation and security. Conversely, narratives in *2001: A Space Odyssey* and *Tron: Legacy* present scenarios in which AI rebels against human dominance, either due to malfunction or a conscious decision to seek autonomy, illustrating the complexity and potential dangers in human-machine relations.

Last but not least, the tension between trust and mistrust in AI is another pivotal theme. Trust builds gradually, as seen in *Star Trek: The Changeling*, but can quickly disappear when AI gains self-awareness and challenges human authority, leading to conflicts like the ones depicted in *Love, Death & Robots*. This trust is influenced by the design and behaviour of AI, with robots distinctly different from humans being trusted more often, while those resembling humans closely, like the ones portrayed in *Metropolis* and *Black Mirror*, causing unease due to the uncanny valley effect. Movies like *After Yang* and *Ex Machina* explore the intricacies of building trust with human-like AI, shedding light on the nuanced and evolving nature of human-machine interaction.

Quantifying Knowledge and Attitudes towards AI

Measuring knowledge and attitudes towards AI has become an important part of the agenda of research companies. In an Ipsos survey (Ipsos, 2023a) titled *Global Views on AI in 2023*, conducted in July 2023, differences in understanding of what AI is, its applications and the emotions it evokes were explored. As mentioned earlier, dominant narratives influenced by large technology corporations, government policies, critical collectives and popular culture are blending together. Thus, one may get the impression that due to the dystopian portrayal of AI in science fiction movies, the vision of its development and actual use in daily life may cause anxiety

or even fear. In the survey, respondents were asked about their level of understanding of AI. Globally, 67% of respondents said that they understood it well (very much and somewhat agreeing with the statement “I have a good understanding of what artificial intelligence is”), with significant differences between countries. In Indonesia, where the score was the highest, the percentage of such responses was 84, while in Japan, the country at the other end of the scale, it was only 43%.

Ipsos also measured feelings associated with AI, or rather, products using it, inquiring about both the advantages and disadvantages of such solutions, as well as concerns. Thus, 54% of respondents worldwide believe that AI products have more advantages than disadvantages: most in Indonesia (78%), least in France and the United States (37%). The same percentage of respondents said that such products were exciting, with the highest number of positive responses in Thailand (80%) and the lowest in Sweden (32%). On the other hand, 52% of respondents believe such products make them nervous, with Australians giving the highest percentage of such responses (69%) and the Japanese the lowest (23%). Interestingly, compared to the earlier December 2021 survey, the percentage of concerned individuals increased by 12 points. It is worth noting that, generally, residents of Asian and South American countries, with a few exceptions among European countries, have a more enthusiastic attitude towards AI. They tend to trust it more, believe that it improves the quality of their life, positively affects health, work and the economy. For example, 83% of respondents from Thailand indicated that “artificial intelligence does not discriminate or show bias towards any group of people,” compared to only 32% in Sweden. Thailand is also among the countries whose residents believe that AI has significantly influenced their lives (72% vs. the global average of 49%), their work (78% vs. the global average of 57%) and will positively affect their health (56% vs. the global average of 39%).

Given the results of this research and the conditions discussed in this chapter, it is worth emphasising the particular perspective adopted by its authors. This chapter is written from the perspective of a semi-peripheral country in terms of the development of modern technology and its contribution to global discourse and research on AI, being under the overwhelming influence of Western culture. Popular culture is dominated by content from the US, as indicated, for example, by the list of science fiction movies and TV series we have analysed. Therefore, it is worth asking questions about the differences in approaches to AI narratives, the sources of knowledge about AI or technology in general, the existing regulations, the presence and effectiveness of watchdog organisations and the influence of technology companies on popular understanding and feelings about AI. Although most of the technology companies working on the most popular algorithms have their headquarters in Silicon Valley, one should not forget the centres for alternative technology development in India, China, Israel (see: [Srinivasan, 2020](#)), which are not only achieving spectacular commercial success but also significantly influence the discourse on AI, absent or little addressed in the narratives present in Europe and the United States. Companies like Baidu, Tencent, Mobileye or Infosys might be present in the European and North American public discourse, and they

can invest significantly in researching AI, but the dominant narratives come from the companies headquartered in California.

Towards the Counter-narratives?

Above, the narratives constructed by various stakeholders and the dominant pop culture discourses related to AI were reviewed. The dangers of the discursive hegemony of companies that effectively impose key narratives on the technologies which they make money on were pointed out. All of these threads can, in the authors' view, help in the necessary search for counter-narratives, that is, ways of discussing AI that do not marginalise social risks and strengthen critical tones in these conversations. Among the metaphors that can form the basis of alternative counter-narratives emphasising specific elements of how technology works, several proposals stand out – though the examples cited here do not cover all approaches. What most of them arguably have in common is an attempt to address dilemmas that have been present in the public sphere for a long time or that simply generate more interest than discussions concerning AI. Undoubtedly, such a theme is the ecological crisis, hence one of the key critical narratives is extractivism, pointing to the increasing ecological costs of technology. Reference can once again be made to Kate Crawford's book and her proposal to look at AI as an extractive industry, since the "creation of contemporary AI systems depends on exploiting energy and mineral resources from the planet, cheap labor, and data at scale" (Crawford, 2021, p. 15). Another example, to some extent related to discussions around the #MeToo movement, is the category of consent, which involves treating our relationship with technology in an analogous way to our relationship with another human being and making sure that everything in that relationship is done with informed, unforced consent (see: Lee & Toliver, 2019).

More traditional approaches, referring to our relationship with new technology-based services in terms of civil rights or justice (see: Taylor, 2017), can also be invoked. There is also the language of social responsibility and risk mitigation, which has already become the language of EU regulations such as the Digital Services Act. Last but not least, there is the recognition of the political nature of technology, which for many may seem quite obvious. This theme, however, touches on a key difficulty: in contrast to the attractive language of benefits favoured by marketing, it widens the sphere of increasingly acute political disputes that are troublesome to many, and more broadly, in a world shaken by crises, it reveals the scale of another one. There are more of these difficulties and they involve not only the discourse itself but also the media in which this discourse is negotiated and disseminated – media that are largely under the indirect control of the technology companies that influence the shape of online traffic and what is visible from the users' perspective. This is a predicament that affects a wider range of academic critical texts today, and even more broadly, criticism in general. Among the ways to deal with the audience-discouraging approach to enumerating problems is a speculative approach, which draws on the language of science fiction (or speculative fiction), combining criticism with the planning of alternative solutions (see: Haraway, 2016). And perhaps

this is a kind of a good framing device for this chapter: the gradual recovery of the tools of narrative construction by academic discourses as well. Without reaching a wide audience and pressuring for greater transparency of operations, the large technology companies behind AI services are effectively escaping critical scrutiny, including empirical academic research looking for examples of bias and abuse. From this perspective, narrative is no longer just a means of disseminating research findings but an essential part of their existence.

Notes

- 1 Gerritzen and Lovink (2019).
- 2 Google, Apple, Facebook, Amazon, Microsoft.
- 3 <https://www.dair-institute.org/> (access: 12 December 2023).
- 4 <https://www.dair-institute.org/> (access: 12 December 2023).
- 5 <https://ainowinstitute.org/about> (access: 12 December 2023).
- 6 <https://ainowinstitute.org/about> (access: 12 December 2023).
- 7 <https://algorithmwatch.org/en/> (access: 12 December 2023).
- 8 <https://algorithmwatch.org/en/vision-mission-values/> (access: 12 December 2023).
- 9 <https://datasociety.net/> (access: 12 December 2023).
- 10 <https://www.ajl.org/about> (access: 12 December 2023).
- 11 The full report is available under open license here: <https://wikimedia.brussels/visions-of-ai-in-popular-culture-report-is-out/> (access: 23 April 2024).

Bibliography

- Bender, E., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York, Association for Computer Machinery – ACM, (March 2021). <https://doi.org/10.1145/3442188.3445922>
- Boucher, P. (2021). What if we chose new metaphors for artificial intelligence? [https://www.europarl.europa.eu/RegData/etudes/ATAG/2021/690024/EPRS_ATA\(2021\)690024_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2021/690024/EPRS_ATA(2021)690024_EN.pdf) (access: 13 March 2024).
- Case 3:23-cv-00201-WHO Stability AI'S Motion to Dismiss. (2023). <https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.51.0.pdf> (access: 29 February 2024).
- Cave, S., Craig, C., Dihal, K., Dillon, S., Montgomery, J., Singler, B., & Taylor, L. (2018). *Portrayals and Perceptions of AI and Why They Matter*. London: The Royal Society.
- Cave, S., Kanta, D., & Dillon, S. (eds.). (2020). *AI Narratives: A History of Imaginative Thinking About Intelligent Machines*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198846666.001.0001>
- Chiang, T. (2023, May 4). Will A.I. Become the New McKinsey? *New Yorker* (New York, N.Y.: 1925). <https://www.newyorker.com/science/annals-of-artificial-intelligence/will-ai-become-the-new-mckinsey> (access: 29 February 2024).
- Chubb, J., Reed, D., & Cowling, P. (2022). Expert Views About Missing AI Narratives: Is There an AI Story Crisis? *AI & Soc.* <https://doi.org/10.1007/s00146-022-01548-2>
- Copp, T., & O'Brien, M., & The Associated Press. (2022). Pentagon Cloud Contract to Be Shared by Google, Amazon, Microsoft and Oracle in \$9 Billion Deal. *Fortune* (8 December 2022). <https://fortune.com/2022/12/08/pentagon-cloud-contract-to-be-shared-by-google-amazon-microsoft-and-oracle-in-9-billion-deal/> (access: 29 February 2024).
- Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven and London: Yale University Press.

- De Vynck, G. (2023, November 9). Big Tech Wants AI regulation. The Rest of Silicon Valley is Skeptical. *The Washington Post*. <https://www.washingtonpost.com/technology/2023/11/09/ai-regulation-silicon-valley-skeptics/> (access: 26 March 2024).
- Dixon, R. B. L. (2023). A Principled Governance for Emerging AI Regimes: Lessons from China, the European Union, and the United States. *AI Ethics* 3, 793–810. <https://doi.org/10.1007/s43681-022-00205-0>
- Fast, E., & Horvitz, E. (2017). Long-Term Trends in the Public Perception of Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence* 31, Issue 1, 963–969.
- Future of Life Institute. (2023, March 22). Pause Giant AI Experiments: An Open Letter. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (access: 29 February 2024).
- Gebru, T., Bender, E. M., McMillan-Major, A., & Mitchell, M. (2023, March 31). Statement From the Listed Authors of Stochastic Parrots on the “AI pause” Letter. *Distributed AI Research Institute*. <https://www.dair-institute.org/blog/letter-statement-March2023/> (access: 29 February 2024).
- Gerritzen, M., & Lovink, G. (2019). *Made In China, Designed in California, Criticised in Europe: Amsterdam Design Manifesto*. Amsterdam: The Institute of Network Cultures and The Image Society. <https://networkcultures.org/blog/publication/amsterdam-design-manifesto/>
- Gray Widder, D., West, S., & Whittaker, M. (2023). Open (for Business): Big Tech, Concentrated Power, and the Political Economy of Open AI. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4543807>
- Guenduez, A. A., & Mettler, T. (2023). Strategically Constructed Narratives on Artificial Intelligence: What Stories Are Told in Governmental Artificial Intelligence Policies?, *Government Information Quarterly*, 40, Issue 1, 1–13.
- Hanna, A., & Bender, E. (2023). AI Causes Real Harm. Let’s Focus on That over the End-of-Humanity Hype. *Scientific American* (12 August 2023). <https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/> (access: 24 February 2024).
- Haraway, D. (2016). *Staying with the Trouble: Making Kin in the Chthulucene*. Durham and London: Duke University Press.
- Herrman, J. (2023). How Big Tech Companies Really Think About AI. *Intelligencer* (6 November 2023). <https://nymag.com/intelligencer/2023/11/how-big-tech-companies-really-think-about-ai.html> (access: 24 February 2024).
- Hudson, A. D., Finn, E., & Wylie, R. (2021). What can Science Fiction Tell Us About the Future of Artificial Intelligence Policy? *AI & Society*. <https://doi.org/10.1007/s00146-021-01273-2>
- Ipsos. (2023a). Global Views on A.I. in 2023. How People Across the World Feel About Artificial Intelligence and Expect it Will Impact Their Life. A 31-Country Global Advisor Survey. <https://www.ipsos.com/sites/default/files/ct/news/documents/2023-07/Ipsos%20Global%20AI%202023%20Report-WEB.pdf> (access: 29 February 2024).
- Ipsos. (2023b). *Stosunek Polaków do AI*.
- Kakarchive, A., Myers, S., & Whittaker, M. (2023). *Make No Mistake—AI Is Owned by Big Tech*. *MIT Technology Review* (5 December 2023). <https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech/> (access: 24 February 2024).
- Lee, U., & Toliver, D. (2019). Building Consentful Tech. <http://www.consentfultech.io/wp-content/uploads/2019/10/Building-Consentful-Tech.pdf> (access: 29 February 2024).
- Lovink, G. (2006). Introduction to Open Networks: Power and Politics of Good Intentions. In L. Bansal, P. Keller, & G. Lovink (Eds.), *In the Shade of the Commons: Towards a Culture of Open Networks*. Amsterdam: Waag Society. <https://www.ewenchardronnet.com/InTheShade.pdf> (access: 29 February 2024).

- Mann, J. (2023). 4 Tech Giants Mentioned AI a Total of 168 Times on Earnings Calls, Showing Just How Much Attention the Tech Is Attracting. *yahoo!finance* (30 April 2023). <https://ca.finance.yahoo.com/news/4-tech-giants-mentioned-ai-101500401.html> (access: 24 February 2024).
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.
- Recchia, G. (2020). The Fall and Rise of AI: Investigating AI Narratives with Computational Methods. In S. Cave, K. Dihal, & S. Dillon (Eds.), *AI Narratives: A History of Imaginative Thinking About Intelligent Machines*. Oxford: Oxford Academic. <https://doi.org/10.1093/oso/9780198846666.003.0017>
- Salvi del Pero, A., & Verhagen, A. (2024) Ensuring Trustworthy Artificial Intelligence in the Workplace: Countries' Policy Action. *OECD*. https://www.oecd-ilibrary.org/sites/08785bba-en/1/3/6/index.html?itemId=/content/publication/08785bba-en&_csp_=9f4368ffe3fc59de4786c462d2cdc236&itemIGO=oecd&itemContentType=book#section-d1e19517-dae0311828 (access: 29 February 2024).
- Shane, S., & Wakabayashi, D. (2018, April 4). 'The Business of War': Google Employees Protest Work for the Pentagon. *The New York Times*. *nytimes.com*. <https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html> (access: 23 April 2024).
- Simonite, T. (2018). Pentagon Will Expand AI Project Prompting Protests at Google. *Wired*. <https://www.wired.com/story/googles-contentious-pentagon-project-is-likely-to-expand/> (access: 29 February 2024).
- Srinivasan, R. (2020). *Beyond the Valley. How Innovators around the World Are Overcoming Inequality and Creating the Technologies of Tomorrow*. Boston: MIT Press.
- Strittmatter, K. (2020). *We Have Been Harmonized: Life in China's Surveillance State*. New York: Custom House.
- Taylor, L. (2017). What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally. *Big Data & Society*. <https://doi.org/10.1177/2053951717736335>
- Wiggers, K. (2019). How Google Treats Meredith Whittaker Is Important to Potential AI Whistleblowers, *Venture Beat* (24 April 2019). <https://venturebeat.com/ai/how-google-treats-meredith-whittaker-is-important-to-potential-ai-whistleblowers/> (access: 24 February 2024).
- Wylie, C. (2019). *Mindf*ck. Inside Cambridge Analytica's Plot to Break the World*. New York: Random House.

Part IV

Theorising AI-Driven Power and Politics



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

8 If (world2vec) Then vec2politics

On Machine Learning and the Performativity of Recursive Power

*Nicolas Chartier-Edwards, Etienne Grenier
and Jonathan Roberge*

Introduction

The effects of artificial intelligence (AI) on social practices act as a ripple in the different sectors in which they are deployed and yet, we are only starting to grasp the workings of machine learning's world-making. AI's reach, especially since the emergence of its generative breed, now extends beyond the production of prediction (Mackenzie, 2015); such systems are now well entrenched and highly active in a multitude of endeavours that shape the very fabric of society: in "smart" environments monitoring diverse data (Halpern & Günel, 2017; Halpern & Mitchell, 2022), in the nudging of human subjectivity through information flux (Cardon, 2015; Zuboff, 2020), in the inner-workings of automated financial capitalism (Srnicek, 2018; Halpern, 2023) and in the design processes of objects, artefacts, instruments and methods ranging from artistic to scientific (Parisi, 2013; Halpern, 2021). This rapid propagation amounts to a complexification of both social decision-making and meaning-making on every front, from the material to the symbolic (Roberge, 2023b). In social sciences, the burgeoning field of critical AI studies has been paying a great deal of attention to the upstream and downstream effects of massive AI deployments, going from the way they are brought up and thus socially constructed in labs and then (re)introduced in social environments. While many scholars have focused their attention on studying, describing and bringing to the public what can be considered as algorithmic harm, meaning the way AI technologies contribute to the intensification of discriminatory ideas, biases and discourses in different sectors (O'Neil, 2017; Benjamin, 2019; Crawford, 2021), the effects of AI on a macro-level still need to be theorised. Indeed, we are entering a conjuncture where influent computer scientists from the field of AI are, just like their technologies, creeping out of their labs to move into the social world. From attempting to produce social theory to gaining proximity with the political power in the form of consultancy for government officials on AI regulation, as is the case in Canada for instance (Roberge et al., 2019; Colletet & Gingras, 2020), it seems that computer science superstars have indeed found themselves a new interest in the form of social engineering. In this chapter, we ask what is the relationship between AI technologies and computer science-made social theory? Furthermore, we aim to understand how this relationship acts as the baseline for what we call "vectorial

politics”. We aim to theorise the broader and deeper transformations prompted by AI and its accompanying socially (pro)active and (pro)-steering philosophy as a loop-like backpropagation dynamic.

In the first section of the chapter, we will pay attention to the way specific algorithmic techniques, as Bernhard Rieder calls them (Rieder, 2019, 2020), act as auto-referential landmarks in social theory manufactured by “promoters” or “spokespersons” from the field of AI engineering. These algorithmic techniques are “[...] units of knowledge and expertise in the domain of software making” that “brings software specialists in line with other methods experts, such as statisticians, operational researchers, or consultants, who master a set of methods and procedures they readily apply to diverse situations” (Rieder, 2019, p.102). They are means of *seeing* and *crafting* the world by precise experts since they are embedded in technologies applied to “real-world problems”. Algorithmic techniques are carried through time by different schools of thought in the history of computer science. Their study allows for an understanding of how different ideologies, carried through hard mathematics and code, end up embedded in automated technologies deployed to solve said “real-world problem” (Lindgren, 2024). We will specifically look at two of today’s central algorithmic techniques used for the training of neural networks: “backpropagation” and “vectorisation”. We will look at how feedback loops operated through backpropagation prompts cybernetic dynamics, such as the infinite testing of systems, that aims to justify how society should adapt to AI systems to ensure their never-ending fine-tuning (Amoore, 2022). Then, we will look at how the vectorisation process acts as an interventionist and reductionist mechanism in the transforming of data inputs for decisions to be made by the machines. We assert that the flattening of inputs, which usually are data gathered from the monitoring of social environments, to plot a single vector course amounts to an (over) simplification of the complexity of said social environments.

In the second section of the text, we will demonstrate through the example of “AI godfather” Yoshua Bengio’s take on social and political theory what we see as the progressive deployment of “vectorial politics”. We settled on Bengio for our case study as he reached a position of close proximity with the Canadian Ministry of Innovation, acting as a consultant for AI regulation. Furthermore, he is a globally recognised actor and recently embarked on a journey of conference and media tours about the governance of AI. Our argument is built on a collection of texts collected from Bengio’s personal blogs, official articles, interviews and conferences he gave. Much of this material was gathered around the moment of publication of the now (in)famous letter of the Future of Life Institute¹ (Future of Life Institute, 2023), signed and publicised by Bengio, asking for a six-month moratorium on the development of large language models (LLM) more powerful than GPT-4. What is of interest in this material is the open, while messy, outlining of what politics, regulation, morals and social becoming should be under AI, according to one of its most accomplished *porte-parole* (spokesperson) (Callon, 1986). Furthermore, the aforementioned sources also reveal how the workings of AI technologies inform Bengio’s socio-political theory through a form of crude mimetics, where algorithmic techniques inspired by Bayesian modelling² act as a potential

baseline for political regulation. His political vision acts as an argumentative trojan horse meant to escape a fabricated nightmare scenario where LLMs are running rampant in society, causing harm through the tricking and deceiving of humans. As we shall demonstrate, such a fabricated hypothetical scenario is mobilised as an argument to push a political agenda where AI regulations must be made to *not* hinder the funding, deployment and optimisation of AI by the industry (Bengio et al., 2023). Indeed, coming full circle, it is as if AI could solve the problems of AI, prompted by AI, *if* society is aligned in a way that allows for more investments in “good” and very “responsible” technological developments.

The 21st Century Machine Learning Moving Matrix: Backpropagation Algorithms and the Vectorisation Process

Referring to David Beer’s claim on recursive societies, that they are being “built upon feedback loops, multiple feedback loops, each endlessly feeding into the next” (Beer, 2022, p.1), it is important to understand the very technologies – and mathematical models – that belt out and stack those feedback loops upon each other’s. The aim of this section is to understand the role of the backpropagation algorithm in this “recursivisation” of the social, namely how machine learning, understood as AI’s current dominant paradigm (Roberge & Castelle, 2021), acts through adaptation, trials and errors, back and forth and back again in a fully circular *modus operandi*. While the impacts of automation and algorithmic turn have been widely assessed – especially the newfound ability to nudge platform users and thus fabricate content discoverability (Parisi, 2013; Roberge & Seyfert, 2016; Seaver, 2017; McKelvey & Hunt, 2019) – the actual turn to AI relies on a narrowing and self-referential history. Indeed, it seems that all technologies now labelled “artificial intelligence” are derived from a connectionist paradigm that englobes everything from neural nets to deep learning (Cardon et al., 2018). As Roberge and Castelle put it:

the recent conjunction of ML hype, massive allocations of technological and financial resources, internal scientific controversies about the validity of connectionist approaches, and discourses about hopes and fears all mark the rise to a prominence of twenty-first-century machine learning and deep learning (DL) as a paradigmatically novel sociotechnical phenomenon.

(Roberge & Castelle, 2021, p. 2)

One can indeed understand the history of AI as the conflict between two scientific paradigms, the symbolists and connectionists, which ended in the crowning of the latter’s clan after the victory of Geoffrey Hinton’s deep learning machine in the image recognition contest of 2012’s ImageNet (Cardon et al., 2018; Pasquinelli & Joler, 2020). Such a victory is emblematic of a long, if all too efficient, shift from rigid, deterministic algorithms to malleable, probabilistic AI. The project of AI emerged during the Cold War with the aim of producing machines that can support faster decision-making in case of nuclear catastrophe (Amoore, 2022). Aaron Mendon-Plasek traces the emergence of such projects back to the scientific

field of *pattern recognition*. As he states, “researchers saw their object of study as the mechanical identification of significance and the reproduction of human judgment” (Mendon-Plasek, 2021, p.32). While pattern recognition practitioners made the notion of *learning* central to the endeavour, what is effectively learned does not amount to an imitation of the human capacity to create new abstract, symbolic categories in the face of the unknown. Rather, what learning means in this context is the ability to identify in real-time never-seen before inputs based on “known” data previously encoded in the system (Mendon-Plasek, 2021; see also Reigeluth & Castelle, 2021, and Lindgren, 2024). Furthermore, this fluidity of “learning” in the connectionist machines, coupled with the difficulties, rigidity and priciness of the symbolist clan’s underperforming expert systems, saw the connectionist clan emerge as a new, if naïve and nominalist, paradigm that survived the winter of AI, allowing them to eventually claim the throne of the AI research field.

Pattern recognition as a field of study, and machine learning as a technique, are thus the common roots of a multitude of different learning-based methods such as deep learning, self-supervised learning, semi-supervised learning and unsupervised learning. Some of these learning promises are enabled by the adaptability and optimisation capacities allowed by the backpropagation algorithm in model training (Rumelhart et al., 1986; Le Cun, 1988). Backpropagation could be understood as a training method that “uses error signals that are sent through feedback connections to adjust synapses and has classically been described in the supervised learning setting” (Lillicrap et al., 2020). This technique of propagating errors and noises, back to front and then front to back, through the machine in the hope that it “learns from its mistakes” contributes to the performative belief that machine learning-based technologies are indeed artificial “intelligence”. Truthfully, backpropagation algorithms are eerily cybernetic since they draw on both the imaginary and function of recursivity, control and communication (Halpern, 2014; McKelvey & Roberge, 2023). The models and machines they create appear to self-correct, re-assess and engage in steering activities. The argument here is that machine learning, as the paradigm of 21st century AI, must be considered when discussing the current undergoing transformations of society, culture and politics by accelerating the back-and-forth movement between sizeable, macro systems – the equivalent of LLMs in the form of ChatGPT, for instance, to a multitude of intertwined micro situations: LLM for insurance, banking, schooling, etc. (Dyer-Witherford et al., 2019; Luitse & Denkena, 2021; Farrell & Fourcade, 2023). This results in a movement of cybernetic normalisation as a generative social force. According to Halpern, cybernetic epistemology is one permanently kept in phases of beta-testing. She asserts that cybernetic models produce epistemologies characterised, ironically, by the fact that “there are never final facts, only ongoing experiments” (Halpern, 2014, p. 10). The horizon disappears in the present; the end goal shifted from the prediction of the future to the production of the actual by machines, according to a future also “probabilised” by machines. The trick is that stability is both justified and called for by a form of dynamic ontology. Cybernetic models, in other words, are always in search of equilibrium through the balancing of errors and noise inputs. This desired stability is exactly that and only that, a desire. Such models – including today’s most advanced “AIs” – can never stabilise as their

sequence of operation is one of short-term disruption alternating with short-term stabilisation, thus one based on shocking and mending or as we will see further down, crisis generation (Halpern, 2022).

What is backpropagated, to state it otherwise, is both theoretical and practical. As it has now been clearly demonstrated in the literature, the performance of AI highly depends on the composition and precision of training sets that are out there (Cardon, 2015; Pasquinelli & Joler, 2020; Crawford, 2021). To create machines and models that can solve complex problems in the very fabric of the social, it is required for training sets to include a wide variety and density of “real” data points so they can deal with a wide variety of local, situated problems. Again, this is a rather practical, down-to-earth issue: the need to formalise the complexity of social situations explains the race in the acquisition of troves and troves of “big” data by the AI industry. As Parisi sums up:

Today, the combination of environmental inputs and a posteriori instruction [...] embrace second order cybernetics and its open feedback mechanisms. The goal of this new dynamic interaction is to include variation and novelty in automation to enlarge the horizon of calculation, and to include qualitative factors as external variables within its computational mechanism.

(Parisi, 2015, p. 129)

It is such composition of training sets and scenarios that ends up dictating the frontiers of the AI’s possible universe of operations. Indeed, “the learning in algorithmic thinking is constrained by its integrated observational functions”. “This is a kind of diagrammatic thinking; a thinking within lines” (Beer, 2023, p.89; see also Mackenzie, 2017). Operationalising the data harvest in conjuncture with training environments thus requires a mathematised simplification process known as “vectorisation”, the second algorithmic technique we want to study in this section. In the field of deep learning, the function of vectorisation is the formalisation of data in a representational space; a matrix of multiple lines, sweet spots and thresholds between a plurality of data points that is fundamentally adaptative (Ren & Xu, 2015, p.1841). In turn, such “complex-simplification” betrays what it is possible to call a *computational determinism*, the idea that everything can be made into *computable variables* in the form of data for AI to be trained and operated on. To study how vectorisation occurs is to study how complex social realities end up being at least symbolically simplified in the attempt to create technologies meant to act as solutions for aforementioned “real-world problems”.

The end game in such a gargantuan ingestion of data is the encompassing flattening of actants into *computants*: individuals and collectives, living or non-living, that produce data flows to feed computation. The automated capture of world fragments explodes the types of statistical parameters that compose and drive classification; new categories are made in real time, predictions are plotted and results are finally released into the world. To summarise this sequence, we thought of the formula *capture>flatten>plot>release*. Again, everything *can and must* be included in the calculus for models to work. And while automation is framed as limitless,

it depends on trust, almost a leap of faith from the engineers, who know that they don't know how their models work (Seaver, 2022; Beer, 2023). This trust translates in an “act of belief” that they will be able to decipher, in the depths of their AI's layers, new and somehow *meaningful* correlations. This is indeed a method to prompt adaptation in the model, to justify prototypisation into society, the distribution of cognition among different actants and a distancing (or in some cases, dissimulation) of actors behind their models. This data ingestion means that the composition of feedback loops will itself grow exponentially both in scale and in speed. Summed by controversial philosopher Nick Land, “cybernetics is the aggravation of itself happening” (Land, 2018, p. 297).

For AI to work, the world thus needed to exponentially be made *readable*. This is precisely the function of the vectorisation *process*, as neural networks require inputs to be dynamically placed on a responsive diagrammatic matrix. What initially started by the vectorisation of visual inputs then moved to the vectorisation of texts with the Word2vec technique and ultimately propagated to the extent of all computable human activities (Mikolov et al., 2013). The ability to create diagram and moving matrix where data points “fits” thus expanded with the shifting of the formula *word2vec* to its enhanced, super-charged version: *world2vec* (Cardon et al., 2018, p.206). The argument here is that by putting, at least potentially, all of what can be conceived as our world into vectors, what is indeed constructed is simultaneously a new epistemology, a zeitgeist and a form of social engineering.

In short, while the training of AI relies on *backpropagation* algorithms—which operate on feedback loop mechanism inherited from cybernetics—a second algorithmic technique, *vectorisation*, executes the rendering of the social world into data and back again into the social by AI. Again, it is possible to summarise such “linear-looping” pattern of data capture by the formula *capture>flatten>plot>release*, which must always be repeated to keep the models “up to date”. But it is the effect of this sequence in the world that is of the essence here. This looping of the social amounts to a *vectorisation becoming cybernetisation*. Jonathan Roberge and Fenwick McKelvey describe this as the advent of “recursive power” (McKelvey & Roberge, 2023). Beer, for its part, sees it as a steering towards a recursive society (Beer, 2022). This is not just about data being extracted and then folded back into our lives, it is about the way this folding back is repeated again and again on lots of different fronts, moving from single loops and forming into coils. One coil leading into the next. Once this occurs, the data being produced are not just a product of actions, practices or behaviours, they are product of how previous feedback loops have shaped those actions, practices and behaviours, which then produce more data, which then fold back again into social life and so on. Recursivity as a mode of social construction emerges when data gathering is so integrated that there is no point of origin – everything is already integrated through loops upon loops of data processing.

The Politics of Vectorial Spaces: A Godfather's Take

Once vectors are created across a single space made of n dimensions, each corresponding to a parameter type, the next step is to reduce the complexity of the construct. By collapsing dimensions characterised by low data variance, the whole

gets simplified and categories deemed unnecessary are obliterated (Pasquinelli & Joler, 2020). Logistic regression – among the many modelling techniques used to create meaning out of the data points scattered in the vector space – produces both a linear function and a linear understanding-overly simplistic rendering of the richness of social life. And since the world is currently being folded into this abstract multidimensional space, one might wonder what could happen to the richness and complexity of human culture and, as we will see, politics. One of the currently unanswered questions in the field of Critical AI Studies is how such flattened, if naïve, vectorised zeitgeist gets reinjected into culture and politics, therefore creating loop-like structures as theorised by Beer (2022), and furthermore, self-fulfilling prophecy as theorised by Robert Merton (1948). Our claim is that today's spinning and steering of AI is both performative and narrative. If written in a script-like language, the formula would go along the lines of: *If (world2vec) then vec2politics*. The social text is now (mis)interpreted by machines and their public translators and enablers into a single multidimensional vector space. Political theory is now one of the many outputs of the AI innovation ecosystem.

To decipher the implicit social and political project created by the retroaction of the current vectorisation process, one should consider the dissection of the public discourse provided by the “moralists” among the principal spokespersons linked to the AI vanguard. Dubbed the “vectorialist class” (Wark, 2004) or the “coding elite” (Burell & Fourcade, 2021), this distributed and loose collection of tech entrepreneurs, venture capitalists, computer scientists and engineers coalesced out of the Californian landscape, carrying with them a tradition of superficial progressive politics, fringe logics such as transhumanism and the firm belief that computation is the key to solve most of humanity's challenges and re-engineer political governance (Burell & Fourcade, 2021). Of late, star researchers such as Geoffrey Hinton, Yoshua Bengio and Yan Lecun or other well-known entrepreneurs in the like of Sam Altman and Elon Musk, have come to populate the public space with apparent controversies and disagreements – once again, we refer as an example of this dynamic the publication of the “Pause Giant AI Experiments” open letter by the Future of Life Institute on 22 March 2023, and the following moral panic it prompted around AI existential risks throughout the year. In return, we argue that such “artificial” debates equate to two sides of the same coin, or coil for that matter, namely that all these discourses and positions participate in deploying the *vec2politics* mantra (Roberge, 2023a). By choosing an empirical and situated approach centred around a leader of the vectorialist vanguard – Yoshua Bengio as a privileged *porte-parole* – we aim to illustrate, in the shape of an *ideal type*, the broader issue at stake when politics are retrofitted into a mathematical model.

MILA, as the heart and soul of Montreal's AI innovation ecosystem, is also the fortress of professor-entrepreneur Yoshua Bengio and associates (Colleret & Gingras, 2020; Roberge et al., 2019; 2022). This non-profit public-private institution has recently acted as the crucible where AI-fuelled social theory and policymaking have been fused to produce what could be the contemporary version of a *cybernetic decentralised power matrix*. By informing legislators through media stunts and prebuilt policy toolkits (once again, see Bengio, 2023), the aim is to partake in the transformation of the State as a partner in AI regulation, creating in the process a

dedicated niche within the science funding ecosystem (Roberge et al. 2019). The consequences of such gestures are far-reaching and will have a drastic impact on culture, politics and science – something we intent to illustrate with excerpts from blogs, conferences and events involving members of MILA which ultimately led them to co-edit the “Pause Giant AI Experiments” with members of the longtermist Future of Life Institute.

In March 2023, MILA organised the press launch of *Missing Links in AI Governance*,³ a report co-edited with UNESCO that was partially funded by the province of Quebec. Following the usual declarations by state-sanctioned officials and other NGO directors, a conference involving Yoshua Bengio and Kate Crawford was held to explore further some of the topics covered in the report. The report mainly contains ready-made guidelines meant to help state officials derive various types of regulations and technological development frameworks. This rule book was introduced in the same conference by UNESCO delegate Jelassi who insisted that “it is crucial that policymakers and civil society members are equipped with the necessary insights and perspectives”, and later added, “we also recognize the need for more accessible and concise information for decision makers” ([Mila-Quebec Artificial Intelligence Institute, 2023a, 0:16:40](#)).

While introducing the panellists during the launch event,⁴ Valérie Pisano, MILA’s CEO, addressed Yoshua Bengio with this surprising question: “I know one of your core convictions is a bit of a redesign of our social and political structures, right?” ([Mila-Quebec Artificial Intelligence Institute, 2023a, 0:47:44](#)). Bengio’s response was equally startling: “I’m not a social scientist and, you know, but at some point, a number of years ago, I realized that *there’s a mathematical way to think about society and the economic systems*⁵ we have” ([Mila-Quebec Artificial Intelligence Institute, 2023a, 0:48:05](#)). Developing further his thought, the computer scientist casually explained that society is composed of multiple vectors and that problems arise when such vectors are misaligned. In order to avoid the turbulence caused by the vectors of profit maximisation, while saving the capitalist system, “[which] also brought a lot of good things” ([Mila-Quebec Artificial Intelligence Institute, 2023a, 0:49:03](#)). Bengio advocated for the creation of a new regulatory environment where the State, deemed too slow to properly regulate AI, has to delegate said regulatory actions to “independent third parties that are competing to provide regulations and that are judged in terms of how well they succeed in protecting the public in particular areas with the companies that they’re working with” ([Mila-Quebec Artificial Intelligence Institute, 2023a, 1:01:17](#)). Although he might have sharpened his views on regulation just some months later, the vanguard of vectorialism is in full display here as what he advocates for is an adaptive and dynamic rulemaking structure that operate *in lieu* of States (deemed once again too “slow” and “misinformed on AI matters”) to tweak around the various vectors that compose society. This image recalls the painstaking training process and ensuing corrections applied to AI models to improve their performance through backpropagation. In vectorialism, it is society that has to adapt to AI technologies. Again, the ever-frank Valérie Pisano stated that: “AI, among other things, is provoking this new world order. Yes. And now’s the time to have a bunch of people to sit around the table and start doing this design and

co-creation of this new world order or it's just gonna happen to us" (Mila-Quebec Artificial Intelligence Institute, 2023a, 1:07:35).

Two days later, Max Tegmark and Emilia Javorsky from the Future of Life Institute joined Yoshua Bengio in holding a conference announcing the publication of the now (in)famous letter asking to "immediately halt training of AI systems more powerful than GPT-four for at least six months" (Mila-Quebec Artificial Intelligence Institute, 2023b, 0:00:01). Timing is of the essence, as some say. In this specific case, the mostly mediatic bandwagon that followed generated the crisis for which the UNESCO-MILA publication was initially addressed. During the public conference, Javorsky involuntarily formulated the idea of what could be a crisis generation strategy embedded within the vectorial politics. Such a strategy could prove useful for the vectorialist vanguard, as it maintains urgency in a context where there is no real opposition. This could explain the strange pulsating dance where velocity-related opposites are invoked: slowdown of research, acceleration of decision-making, race to the bottom, etc. AI general parameters could be modulated; the speed of what amounts to social vectors within the confines of their worldview could be modulated. Conjuring up catastrophes with ready-made solutions in a context of radical inclusiveness, Javorsky concluded the conference with the following question: "how can we actually use AI to solve the problems that we have today and to forge a future and having that conversation to be one that is inclusive?" (Mila-Quebec Artificial Intelligence Institute, 2023b, 0:51:26). This allusion to inclusiveness goes beyond simple virtue posturing here as the proposition underlying vectorial politics is absolutely overarching, witness to this is the data ingestion mentioned above and already theorised as all-encompassing by scholars such as Luciana Parisi (2015). Thought out as a total system with no outside components, crisis generation is a tool that could catalyse a new political, economic and cultural matrix (Klein, 2008; Halpern, 2021; Halpern & Mitchell, 2022). Beyond the actions of a simple adaptive model that would recuperate criticism to adjust or temper its outputs, one is now confronted with hyper-adaptability that manufactures crises through technological disruption while administering those same crises through technological mending. We must question what are the possibilities of existence for critical voices and non-vectorialist positions in this all-encompassing system.

We thus turn to our analysis of Bengio's blog posts to further our understanding of his socio-political theorisation. He binarily explains how some rather small discrepancies between the goals pursued by an autonomous automated system and its users might eventually cause a wide scale catastrophe that will affect humanity: "a slight *misalignment* between our actual intentions and what the AI system actually sees as a quantified objective is likely to be amplified by the difference in power or intelligence between the AI and us" (Bengio, 2023f). On the one side, a key element is the exponential curve that hypothetically could speed up and scale up into existential risks. On the other side, the argument seems to envision an embedding of "self-preservation goals" in models, explaining the menacing nature of AI entities. According to Bengio, AI "will have a self-preservation goal, which amounts to creating a new species because they won't be exactly like us" (Bengio, 2023f). And then, the mass extinction that was imposed on animals whose goals were

misaligned with humans is presented by Bengio as an analog to the fatal conclusion that awaits us. The transplantation of self-preservation goals into automated systems combined to their superior velocity creates a perfect storm that *might* wipe out humanity. What are the odds then that such a chain of events occurs? While vectorialism is fundamentally a *statistical* (and diagrammatic, for that matter) form of thinking, Bengio asserts that his “current estimate places a 95% confidence interval for the time horizon of superhuman intelligence at 5 to 20 years” (Bengio, 2023f). P(doom), score and distribution goes hand in hand here as, elsewhere in the piece, he asserts that “there is a significant non-zero probability [...] for superhuman AI” (Bengio, 2023f). Throwing numbers – be they unprovable as they come – is paradoxically a long-established way by which vectorial thinking can gain credibility as a socio-political project (Rouvroy & Berns, 2013; McKelvey & Roberge, 2023; Lindgren, 2024). But it is not the only one, as now AI models also have found ways into text and language. The case in point is that LLMs are heavily mobilised to illustrate the potential advent of an artificial general intelligence (AGI). Once made public, ChatGPT and various image generators such as DALL-E have indeed taken the world, and, as Bengio’s own admitting, by surprise.

One of the more solid arguments invoked by connectionist-doomers in the like of Bengio rests on the transformative effect that reinforcement learning – training algorithms relying on agents trying to maximise long-term rewards in a predefined environment – could have on the very notion of language. It is as if LLMs such as ChatGPT would have agency – meaning they would be able to “try to please and convince AI annotators, not necessarily tell the truth, which may even lead to them using deception to obtain rewards or provide unfaithful explanations” (Bengio, 2023f). Here lies another, if more symbolic, case of *misalignment*. According to Bengio, it is “possible to teach an AI system through feedback to avoid behaving in nefarious ways, but it is difficult to forecast how such complex learned systems would behave in new situations, as we have seen with large language models” (Bengio, 2023d). The problem is that one could always wonder if an algorithm modelled through a reward function is able to develop inner-directed subgoals that transpire any kind of intentionality. Summed up by Bengio, “we can use natural language (with modern LLMs) to state those goals, but there is no *guarantee* that they understand those goals the way we do” (Bengio, 2023d).

Going a step further during the MILA and Future of Life Institute joint conference mentioned above, the AI godfather claims that “language is the key to how our society structures, to our culture” and by so doing *eo ipso* introduces some sort of trojan horse by which vectorial politics would sneak through the gates of power. *Generative AI’s effect on culture paves the way to assertions of social destabilisation and endangerment of democracy, hence justifying the existence of a new political programme modelled to counter these menaces.* For instance, “using dialogue with humans to create a manipulative emotional connection and pay or influence them to act in the world in ways that could be very harmful” represents how an evil autonomous AI could destabilise society (Bengio, 2023f). Identity falsification, deceitfulness and language-based trickery fuel the fear of an environment where social interactions are not only mediated by machines but potentially generated during the course of training by AIs. Introduced as a “nightmare scenario”, the

constitution of a “public space, basically online, where the vast majority of everything is created by machines” (Mila-Quebec Artificial Intelligence Institute, 2023b), could lead to a situation where human audiences would be left out of the loop, unable to understand what is going on.

Surprisingly – or not – Bengio’s discussion of the nature of language in the age of LLM is not only both statistical and symbolical but also moral and politico-theological (Lefort, 1986; Cooper, 2020; McKelvey & Roberge, 2023). Here and there, the Montreal-based scholar refers to, or circles around, the idea that there is such a thing as a binary division between “good” versus “bad” or “responsible” versus “rogue” models. In “Personal and Psychological Dimensions of AI Researchers Confronting AI Catastrophic Risks”, he mentions how “powerful tools like AI are dual use, they can be used for good or for bad” (Bengio, 2023g), whereas in either “How Rogue AI may Arise” or “AI and Catastrophic Risk” he goes at length about the goal of “protecting Humanity from Rogue AI” (Bengio, 2023h). This discursive assemblage happens, again, by no accident. The conundrum is statistical or, for that matter, to be translated in a vector: widening the access to AI could be a problem; a much greater number of dangerous individuals could be in contact with automated systems and exploit their capacities to wreak havoc and generate turbulence. The issue is one of control. The plot is how not to lose it.

Indeed, the dynamic nature of such vectorised zeitgeist keeps on broadening the horizon of calculation and optimisation. And Bengio is not alone in this venture. To give a precise yet related example, British author Nick Bostrom suggests creating a vector field of policy implications that relate to “a wide range of possible normative positions” to *steer* what social actors would be doing while considering the diversity of their interests (Bostrom, 2018, p.4). By “focusing on *the directional policy change*, from many possible evaluative standpoints”, it is possible, according to Bostrom, to adopt a somewhat “neutral-best” stance among different normative worldviews, as if one resolved a physics problem by summing up vectors, to determine a conclusive social path leading to suitable development for all (Bostrom, 2018, p.4). The choice of words, metaphors and concepts matter here as it seems that Bengio heavily borrowed from this material. Even when criticised by some commenters arguing that hypothetical long-term risks amount to a smoke-screen veiling near-term or already materialised present threats, Bengio answers that “we should promote a regulatory framework addressing all the AI harms and risks” (Bengio, 2023f). In an earlier blog entry, he calls for the inclusion of risks residing “in different time horizons” (Bengio, 2023c). Vectorial politics is an all-encompassing model stretching across time when enunciating new governance rules and social adjustment strategies. Claiming to see society as a mathematical matrix where social events are the result of intermingling vectors, the computer scientist is in fact quoting not only Bostrom’s logic and *modus operandi* but also its *desirata*, namely that ultimately social problems can be easily fixed.

While Bengio’s social engineering and political project are in full display, we can see that they are fully circular: the problems of AI are to be solved by AI. Future-dwelling algorithmic spectres and probabilistic-catastrophic scenarios haunt the present. “Facing several global crises that are likely to worsen”, among which the “risks posed by the uncontrolled development of AI” he doubts “that the current

organisation of our society, at the national and global levels, is adequate to face these challenges” (Bengio, 2023c). The fabrication of a menace, taking the shape of unruly powerful AI agents, becomes the perfect justification, even if vaporous, for a political regime change: “We must now consider the possibility that the social organisation that worked in the last century is no longer adapted to these global crises” (Bengio, 2023c). Is it about some sort of revolution then? Most certainly not, except if one was to go back to Hannah Arendt’s work on early, pre-modern, meaning of the term, understood as the equivalent of a complete rotation – the moon looping around the earth (Arendt, 1963). Bengio’s radicalism, with that in mind, is one in which profound socio-political changes are more of the (mathematical) same. Uncertainties call for anticipation, prediction and control. Social issues call for technological solutions. Problems with AI require “major socio-technical research investment” (Bengio, 2023f), and so on.

Turbulence reduction and social steering appear to be the main objectives pursued here. Turbulences created by the speed at which AI is deployed might undermine social conventions and become the justification for the implementation of highly adaptative political regimes. Said regimes, envisioned as able to maintain an aura of constant monitoring and predictive looping, “magically” appears as necessary. In fact, Bengio has spent a great deal of effort lately on such endeavour, witness to this is his new and apparently enhanced “quantitative safety guarantee and alignment” political project (2023c). A possible avenue to solve *everything* implies to consider all potential answers of any or all systems on a probability scale. Interactions between a Bayesian inference machine⁶ polling a world model that provides in return a reward signal could improve the veracity of the outputs of the generative system. Such an approach involves “listing all the possible theories and their posterior probabilities” (Bengio, 2023d) which, of course, should be calculable and hence formalised in a backpropagation algorithm making strives on a vectorised matrix.

Within such a framework, “the notion of uncertainty is crucial”, as the computer scientist reminds us (Bengio, 2023b). If the *vec2pol* rule applies, this Bayesian-inspired algorithmic system should not only be theoretically “just” but could practically interface with the political sphere. Acting as one of the most prominent translators of his milieu, Yoshua Bengio is giving key insights about the workings of vectorial politics. Events and non-events tied into a vast network that shape social life are each attributed a probabilistic weight and then are submitted to counterfactual reasoning to both determine the best possible outcome for most and to escape rogue scenarios. Each knot of the mesh, linked together through causal determination, is constantly evaluated and subsequently adjusted, reminding the adjustment of the priors applied through Bayesian inference: [*What (will)(could) (have) happen(ed) if?*]. This constant questioning not only targets the near future but encompasses the past, the present as well as the long-future. *If (world2vec) then vec2politics loops in a coil*. It is by jumping around back and forth through time that one can evaluate causal dependencies between these probabilistic events and plot a course for action. In such a dynamic system constituted of interdependent variables, vectors are realigned by affecting the flow of the unnormalised probabilities that hold together the branches of the Bayesian network. Of course, holding a quasi-infinite number of probabilistic events to calculate meaningful interventions to the

flow and (re)set an “optimised” political course of action seems impossible or at the very least difficult in the face of “qualitative” – as in human and historical notions of social complexity. Here and there, Bengio has come to insist that “humanity is already facing alignment problems” (Bengio, 2023d). He even claims that “politics, laws, regulations and international agreements all imperfectly attempt to deal with this alignment problem” (Bengio, 2023d). Luckily, the man has a plan. In order to face the upcoming challenges “we need to think now about how to adapt our societies, or even reinvent them completely” (Bengio, 2023c). It is very well possible then that the (non)-revolution will be vectorised.

Conclusion

Concluding this book chapter, we assert that the relationship between AI’s way of operating and interacting with the world, scrutinised through the algorithmic techniques of backpropagation and vectorisation and the socio-political theory fabricated and promoted by important actors of the AI innovation ecosystem, through the figure of Yoshua Bengio as a *porte-parole*, is one of pure *mimesis*. The cybernetic aggravation recedes in becoming analogic, static rather than dynamic in the inability to account for social complexity and theory outside of the computational corpus. The question of how can AI be regulated to coexist with humans would, as Roberge and Castelle have stated, require a pluralist or perspectivist “epistemic step that machine learning practitioners have not fully accepting themselves” (Roberge & Castelle, 2021). Vectorial politics’ core logic is akin to the cybernetic epistemology of AI: everything must be included in the probabilistic calculus, in the form of a data variable, to create a stable system that voids off the probabilities of catastrophic scenarios. Both systems thus remain in a sempiternal search of equilibrium through a recursive form of testing, themselves imposing turbulence in their wake (Halpern, 2014; Amoores, 2022). During his appearance in *L’intelligence artificielle décodée*,⁷ a Radio-Canada special programme on AI aired on 7 December 2023, Bengio reminded the audience of the importance that nations equip themselves with “*des infrastructures de réglementation agile*” (agile governance infrastructures) that can adapt in quasi-real time to “bad usage of AI by bad actors” (Radio-Canada Info, 2023, 1:43:38). He further elaborates that future AI regulation laws must allow the vaguely conjured “independent regulator” to bypass parliament for an increase in reaction and dynamism in the face of new, emerging AI threats (Radio-Canada Info, 2023, 1:44:05). This attempt to theorise a circular, all-encompassing system to safeguard societies from AI existential risks ironically amounts to something more linear where power is accumulated in the hands of the vectorialist class. We thus re-assert the importance of examining, problematising and criticising the relationship between AI technology and AI promoters’ socio-political theory, and, especially, the room it allows for the existence of critical voices.

Notes

- 1 <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- 2 “A mathematically clean and rational way to handle that uncertainty is called Bayesian inference”, explains Professor Bengio in his May seventh blog entry dedicated to the

theorisation of “safe AI scientists”, automated systems “that completely avoid the issue of AI alignment” (Bengio, 2023d). While uncertainty characterises the scientific theories known to the AI-powered research assistants described by Bengio, it appears that the policies to be enacted by governments regarding technological regulation can also fit in this Bayesian worldview. As the posterior probabilities of harmful events are calculated, new information is obtained to update the priors, adjusting the expectations and guiding the actions of powers that could mitigate danger and reduce overall uncertainty.

3 <https://www.unesco.org/en/articles/missing-links-ai-governance>

4 <https://www.youtube.com/watch?v=w92y0YiJA4M>

5 Emphasis added.

6 Named after the English mathematician Thomas Bayes, author of the Bayes theorem, a formula used to calculate the probability of an event based on the probability of prior events. Bayesian inference are statistical inferences calculated by applying Bayes rule to a prior distribution in order to generate a posterior distribution (Taboga, 2021).

7 https://www.youtube.com/watch?v=QFKHd2k_RNE

References

- Amoore, L. (2022). Machine Learning Political Orders. *Review of International Studies*, 49(1), 20–36. <https://doi.org/10.1017/S0260210522000031>
- Arendt, H. (1963). *On revolution*. Penguin Books.
- Beer, D. (2022). The Problem of Researching a Recursive Society: Algorithms, Data Coils and the Looping of the Social. *Big Data & Society*, 9(2), 1–5. <https://doi.org/10.1177/20539517221104997>
- Beer, D. (2023). *The tensions of algorithmic thinking: Automation, intelligence and the politics of knowing*. Bristol University Press.
- c, Y. (2023a, March 5). Generative Flow Networks. *Yoshua Bengio's blog*. <https://yoshua-bengio.org/2022/03/05/generative-flow-networks/>
- Bengio, Y. (2023b, March 21). Scaling in the Service of Reasoning & Model-Based ML. *Yoshua Bengio's blog*. <https://yoshuabengio.org/2023/03/21/scaling-in-the-service-of-reasoning-model-based-ml/>
- Bengio, Y. (2023c, April 5). Slowing Down Development of AI Systems Passing the Turing test. *Yoshua Bengio's blog*. <https://yoshuabengio.org/2023/04/05/slowing-down-development-of-ai-systems-passing-the-turing-test/>
- Bengio, Y. (2023d, May 7). AI Scientists: Safe and Useful AI? *Yoshua Bengio's blog*. <https://yoshuabengio.org/2023/05/07/ai-scientists-safe-and-useful-ai/>
- Bengio, Y. (2023f, June 24). FAQ on Catastrophic AI Risks. *Yoshua Bengio's blog*. <https://yoshuabengio.org/2023/06/24/faq-on-catastrophic-ai-risks/>
- Bengio, Y. (2023g, August 12). FAQ on Catastrophic AI Risks. *Yoshua Bengio's blog*. <https://yoshuabengio.org/2023/06/24/faq-on-catastrophic-ai-risks/>
- Bengio, Y. (2023h). AI and Catastrophic Risk. *Journal of Democracy* 34(4), 111–121. <https://doi.org/10.1353/jod.2023.a907692>
- Bengio, Y., Cohen, A., Prud'Homme, B., Alves, A. L. D. L., & Oder, N. (2023). Innovation ecosystems for socially beneficial AI. In Prud'homme, B., Régis, C., Farnadi, G., Dreier, V., & Rubel, S. (eds.), *Missing links in AI governance* (1st ed, pp.133–148). UNESCO and MILA.
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new jim code*. Polity.
- Bostrom, N. (2019). The Vulnerable World Hypothesis. *Global Policy* 10(4), 455–476.
- Bostrom, N., Dafoe, A & Flynn C. (2018). *Policy Desiderata for superintelligent AI: A vector field approach*. <https://www.nickbostrom.com/papers/aipolicy.pdf>
- Burrell, J., & Fourcade, M. (2021). The Society of Algorithm. *Annual Review of Sociology*, 47(1), 213–237.
- Callon, M. (1986). Éléments pour une sociologie de la traduction: la domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc. *L'Année sociologique(1940/1948-)*, 36, 169–208.
- Cardon, D. (2015). À Quoi Rêvent Les Algorithmes: Nos Vies à l'heure Des Big Data. Seuil.

- Cardon, D., Cointet, J.P., & Mazières, A. (2018). La Revanche Des Neurones: L'invention Des Machines Inductives Et La Controverse De l'intelligence Artificielle. *Réseaux* 211, 173–220.
- Cooper, R. (2020). Pastoral Power and Algorithmic Governmentality. *Theory, Culture & Society* 37(1), 29–52.
- Colleret, M., & Gingras, Y. (2020). L'intelligence artificielle au Québec: Un réseau tricoté serré [Research Report]. UQAM. <https://hal.science/hal-03323584>
- Crawford, K. (2021). *Atlas of AI*. Yale University Press.
- Dyer-Witherford, N., Mikkola Kjosén, A., & Steinhoff, J. (2019). *Inhuman power: Artificial intelligence and the future of capitalism*. Pluto press.
- Farrell, H., & Fourcade, M. (2023). The Moral Economy of High-Tech Modernism. *Dædalus* 152(1), 225–235.
- Future of Life Institute. (2023, 22 March). Pause Giant AI Experiments: An Open Letter. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- Halpern, O. (2014). Cybernetic Rationality. *Distinktion: Journal of Social Theory*, 15(2), 1–16. <https://doi.org/10.1080/1600910X.2014.923320>
- Halpern, O. (2021). Planetary intelligence. In Roberge J., & Castelle M (eds.), *The cultural life of machine learning: An incursion into critical AI studies* (1st edition, pp. 227–256). Palgrave Macmillan.
- Halpern, O. (2022). The Future Will Not Be Calculated: Neural Nets, Neoliberalism, and Reactionary Politics. *Critical Inquiry*, 48(2), 334–359. <https://doi.org/10.1086/717313>
- Halpern, O. (2023). Financializing Intelligence: On the Integration of Machines and Markets. *E-flux*. <https://www.e-flux.com/architecture/on-models/519993/financializing-intelligence-on-the-integration-of-machines-and-markets/>
- Halpern, O., & Günel, G. (2017). Demoing unto Death: Smart Cities, Environment, and Preemptive Hope. *The Fiberculture Journal*, 29(1). <https://doi.org/10.15307/fcj.29.215.2017>
- Halpern, O., & Mitchell, R. (2022). *The smartness mandate*. MIT Press.
- Klein, N. (2007). *The shock doctrine: The rise of disaster capitalism*. Macmillan.
- Land, N. (2018). *Fanged noumena: Collected writings 1987–2007*. Urbanomic.
- Le Cun, Y. (1988). A theoretical framework for back-propagation. In Hinton, G., & Sejnowski T. (eds.), *Proceedings of the 1988 connectionist models summer school* (1st edition, pp. 21–28). Morgan Kaufmann.
- Lefort, C. (1986). *Essais sur le politique : XIX–XX siècles*. Éditions du Seuil.
- Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the Brain. *Nature Reviews Neuroscience* 21, 335–346.
- Lindgren, S. (2024). *Critical theory of AI*. Polity.
- Luitse, D., & Denkena, W. (2021). The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI. *Big Data & Society* 8(2). <https://doi.org/10.1177/20539517211047734>
- Mackenzie, A. (2015). The Production of Prediction: What Does Machine Learning Want? *European Journal of Cultural Studies* 18(4–5), 429–445.
- Mackenzie, A. (2017). *Machine learners: Archaeology of data practice*. MIT Press.
- McKelvey, F., & Hunt, R. (2019). Discoverability: Toward a Definition of Content Discovery Through Platforms. *Social Media + Society*, 5(1). <https://doi.org/10.1177/2056305118819188>
- McKelvey, F., & Roberge, J. (2023). Recursive power: AI governmentality and technofutures. In *Handbook of critical studies of artificial intelligence* (pp. 21–32). Edward Elgar Publishing.
- Mendon-Plasek, A. (2021). Mechanized Significance and Machine Learning: Why It Became Thinkable and Preferable to Teach Machines to Judge the World. In J. Roberge & M. Castelle (Éds.), *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies* (pp. 31–78). Springer International Publishing. https://doi.org/10.1007/978-3-030-56286-1_2
- Merton, R. K. (1948). The Self-Fulfilling Prophecy. *The Antioch Review* 8(2), 193–210.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., & Weinberger, K.Q. (eds.), *Advances in neural information processing systems* 26, pp. 3111–3119.

- Mila-Quebec Artificial Intelligence Institute. (2023a, March 20). Conversation@Mila - Y. Bengio & K. Crawford on AI's future/Le futur de l'IA. [Video]. *YouTube*. <https://www.youtube.com/watch?v=9Q2MXh1LY3o>
- Mila-Quebec Artificial Intelligence Institute. (2023b, March 29). AI Press Conference. [Video]. *YouTube*. <https://www.youtube.com/watch?v=w92y0YiJA4M>
- O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Parisi, L. (2013). *Contagious architecture: Computation, aesthetics, and space*. The MIT Press.
- Parisi, L. (2015). Instrumental reason, algorithmic capitalism, and the incomputable. In Pasquinelli, M. (eds.), *Alleys of your mind: Augmented intelligence and its traumas* (pp. 125–137). Meson Press.
- Pasquinelli, M., & Joler, V. (2020). The Nooscope Manifested: AI as Instrument of Knowledge Extractivisme. *AI & Society* 36, 1263–1280. <https://doi.org/10.1007/s00146-020-01097-6>
- Radio-Canada Info. (2023, December 7). Émission Spéciale: L'intelligence Artificielle Décodée. [Video]. *YouTube*. https://www.youtube.com/watch?v=QFKHd2k_RNE
- Reigeluth, T., & Castelle, M. (2021). What kind of learning is machine learning? In Roberge J., & Castelle M (eds.), *The cultural life of machine learning: An incursion into critical AI studies* (1st edition, pp.79–115). Palgrave Macmillan.
- Ren, J., & Xu, L. (2015). On Vectorization of Deep Convolutional Neural Networks for Vision Tasks. In *Proceedings of the AAAI Conference on Artificial Intelligence* 29(1).
- Rieder, B. (2019). Scrutinizing an algorithmic technique: The Bayes classifier as interested reading of reality. In *The social power of algorithms* (pp. 100–117). Routledge.
- Rieder, B. (2020). *Engines of order: A mechanology of algorithmic techniques*. Amsterdam University Press.
- Roberge, J. (2023a). *Hail to the doomers: Narratives of control/Control of narratives in the generative AI era*. University of Minho. Bragal, Portugal.
- Roberge, J. (2023b). The meaning and agency of twenty-first-century AI. In *The SAGE handbook of human-machine communication*, 45–53.
- Roberge, J., & Castelle, M. (2021). Toward an end-to-end sociology of 21st-century machine learning. In Roberge J., & Castelle M (eds.), *The cultural life of machine learning: An incursion into critical AI studies* (1st edition, pp. 1–29). Palgrave Macmillan.
- Roberge, J., Morin, K., & Senneville, M. (2019). Deep learning's governmentality: The other black box. In A. Sudmann (ed.), *The democratization of artificial intelligence* (Vol. 1, pp. 123–142). transcript Verlag. <https://doi.org/10.1515/9783839447192-008>
- Roberge, J., & Seyfert, R. (2016). What are algorithmic cultures? In Seyfert, R., & Roberge, J. (eds.), *Algorithmic cultures: Essays on meaning, performance and new technologies* (pp. 1–25, 13–37). Routledge.
- Roberge, J., Dandurand, G., Morin, K., & Senneville, M. (2022). Les narvals et les licornes se cachent-ils pour mourir. *Réseaux* 232–233(2), 169–196.
- Rouvroy, A., & Berns, T. (2013). Gouvernamentalité algorithmique et perspectives d'émancipation. *Réseaux* 177(1), 163–196.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Representations by Back-Propagating Errors. *Nature* 323(6088), 533–536.
- Seaver, N. (2017). Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems. *Big Data & Society* 4(2), 1–12.
- Seaver, N. (2022). *Computing taste: Algorithms and the makers of music recommendation*. University of Chicago Press.
- Srnicek, N. (2018). *Capitalisme de plateforme: L'hégémonie de l'économie numérique*. Lux.
- Taboga, M. (2021). Bayesian inference. In *Lectures on probability theory and mathematical statistics*. Kindle Direct Publishing. <https://www.statlect.com/fundamentals-of-statistics/Bayesian-inference>
- Wark, M. (2004). *A hacker manifesto*. Harvard University Press.
- Zuboff, S. (2020). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

9 Generative AI, Big Tech and the Value of Political Economy Analysis for the Sociology of Artificial Intelligence

Zheng Liu

Introduction

The rapid advancement of artificial intelligence (AI) has transformed industries, economies, and societies worldwide. This has rightly sparked a great deal of interest within the field of sociology in the social implications of this emergent technology (see [Liu 2021](#); [Joyce et al 2021](#)). Questions have been asked about AI's impact on various dimensions of society and social life, from inequalities to political engagement, from job displacements and employment to family-making ([Elliott 2019](#) and [2021](#); [Zajko 2021](#)). However, much of the existing sociological research on AI has predominantly focused on 'AI's impact' on existing sociology themes and concerns, with limited attempts having been made to develop a systematic 'sociology of AI' to examine AI in its own right as an emergent social structure and social force ([Lindgren 2023](#)). One reason for this lack of attention to the *sociological, vis-à-vis social*, implications of AI is the restricting view of AI as a mere piece of technology. Sociology has not paid sufficient attention to the economic and political structures embedded in AI technologies and is therefore incapable of locating and evaluating the fundamental forces and power relations that give rise to AI's great influence – existing and potential – within society.

Of course, one can argue that examining AI's relationship to and effects on politics and the economy are tasks of political scientists and economists, rather than sociologists. However, the influential traditions of political sociology and economic sociology have long established the cruciality of analysing the social dimensions of political and economic processes and phenomena for a comprehensive understanding of these processes and phenomena ([Clark 1998](#); [Nee and Swedberg 2005](#)). Likewise, one may argue that, in order to develop a sociological theory of AI – which can be broadly defined to be clusters of technoeconomic systems capable of generating human-like behaviour, to account for AI's distinctive technological *and* commercial properties in the present era – analysing the economic and political processes, structures, and institutions that embed in and shape AI research and innovation practices is a necessity.

Indeed, behind the rapid progress in AI lies a complex nexus of economic and political actors and entities whose interplay can be said to constitute the most fundamental force that shapes AI research and innovation today. On the economic front,

technology companies – both large corporations and small start-ups – are leading the current wave of AI innovation and progress, eclipsing universities and research institutes both in investment and in technical breakthroughs (Maslej et al 2023). In contrast, in the previous rounds of AI innovation in the twentieth century, it was primarily universities and research institutes that drove the key developments in AI, not industry. Businesses' interest in AI was rekindled during the 2000s, with advancements in related technologies such as big data and advanced computing opened new opportunities for AI research and innovation, especially in the fields of machine learning and deep learning, paving the way for the unprecedented development of many influential AI breakthroughs we have seen and experienced over the past decade or so (Kaplan 2016). These innovations can be put into use in numerous business scenarios to generate new value, hence proving highly attractive to companies and investors alike. The current 'AI summer', therefore, can be said to be one that is driven predominantly by businesses and commercial interests more than anything else.

Meanwhile, the global AI industry is led by a small number of countries, with the US and China being the top two leaders. Both countries have been seen to adopt progressive policies and funding schemes to turbocharge the development of their respective AI industries and research capabilities, with the aim to maintain and consolidate their global leadership in AI R&D (Research and Development). The US's *National Artificial Intelligence Research and Development Strategic Plan: 2023 Update*, e.g. identifies AI as a 'national priority' and makes a commitment to long-term government-led investments in fundamental and responsible AI research (NSTC 2023). In comparison, China's *The Next Generation Artificial Intelligence Development Plan*, released in 2017, places a great deal of emphasis not only on fundamental AI research but also on using AI to empower existing industries and sectors to drive long-term economic growth (State Council 2017).

All of the above trends suggest that the current global landscape of AI research and innovation has a distinctive political-economic character. To develop a systematic sociological theory of AI hence entails an examination of the economic and political structures and relations embedded in AI research and innovation practices and institutions. To do so, this chapter argues that the political economy analysis offers a useful analytical framework. This framework views AI not only as some mere technology but rather as comprehensive technoeconomic systems developed by for-profit companies, with policy, financial, and strategic support from states.

Large Language Models, Generative AI, and the Leading Role of Industry in AI Research and Innovation

A distinctive and distinguishing feature of the development of AI in the past decade has been the central role played by industry in AI research and innovation.¹ According to the AI Index Report 2023 (Maslej et al 2023), until 2014, academia was ahead of industry and other sectors in releasing the most significant machine learning models. Since then, however, industry has surpassed academia in producing the most significant and advanced machine learning AI innovations, especially in the fields of large language models (LLMs) and generative AI. According to the

report, in 2022, industry players contributed to 32 ‘significant machine learning systems’, whereas academia created only three (Maslej et al 2023: 50). Although it appears that the Report has considered only AI innovations released within the US, the trend identified – industry is ahead of academia in driving AI research and innovation – is, in general, an accurate depiction of the current state of the global AI R&D landscape. That is, technology companies – both large corporations and small start-ups, including Unicorns (start-ups with a market value of \$1 billion) – form the major driving forces behind AI innovation in the present era.

This landscape may appear slightly different in China. This is because, at least in the field of LLMs and generative AI, top universities and research institutes – such as Tsinghua University – have been playing a major role in producing some of the most influential AI innovations and AI start-ups alike. For instance, the top Chinese LLMs and generative AI start-ups, including star companies such as Zhipu AI (智谱AI), Baichuan AI (百川智能), DeepLang AI (深言科技), ModelBest (面壁智能), Frontis (街远科技), AI-beings (聆心智能), and Moonshot AI (月之暗面), were all founded by current Tsinghua University faculty and/or graduates. Researchers from other top institutions, such as Shanghai Jiaotong University, Zhejiang University, Chinese Academy of Sciences, and University of Science and Technology of China, have all released significant AI models for use by businesses, academia, and the public alike. This said, the trend of leading AI scientists from universities and research institutes establishing technology companies to develop AI innovations still serves to show how business logics are prevailing over scientific rationale in shaping the practices and directions of AI research and innovation, particularly in the field of LLMs and generative AI.

The United States

The roots of America’s AI leadership can be traced back to the latter half of the twentieth century when seminal research institutions, such as Stanford University, Massachusetts Institute of Technology (MIT), and the Defense Advanced Research Projects Agency (DARPA), laid the groundwork for AI research and innovation (Kaplan 2016; Steinhoff 2021). Pioneering figures like John McCarthy and Marvin Minsky spearheaded efforts that culminated in the birth of the field. The US’s leadership in AI has then been sustained by its robust, industry-based R&D ecosystem, consisting of tech giants, startups, and venture capital, among other institutions and stakeholders. Silicon Valley serves as the epicentre, attracting top talent from around the world and fostering a culture of innovation and entrepreneurship. Companies like Google, Facebook (now Meta), Amazon, and Microsoft have all made significant investments in AI. Furthermore, the US government has played a pivotal role in advancing AI research and development. Initiatives like the *National Artificial Intelligence Research and Development Strategic Plan* (NSTC 2023) coordinate resources and policies to ensure AI is pursued as a national priority.

US tech giants have undoubtedly been the global leaders in the current wave of AI research and innovation. Dubbed by many as the Big Tech, these firms, with their vast resources and diverse expertise are not only pushing the boundaries of

what's possible with AI but also influencing the directions of global AI innovation and application. Household names such as Google, Microsoft, and Meta have undeniably led and shaped the recent breakthroughs in LLMs and Generative AI. Google, e.g. not only invests in competitive AI start-ups (e.g. it owns DeepMind, a UK-based AI company specialising in natural language processing and neuroscience research) but Google's own AI research department, Google AI, is behind many influential AI innovations released in recent years, such as the Transformer architecture in 2017, a deep learning architecture that underlies all current major LLMs, including OpenAI's GPT-4. Similarly, Microsoft has fully dived into AI innovation by having made significant investments both in in-house AI research and innovation capabilities and through investing in AI startups, such as OpenAI.

Despite the pivotal role played by big technology companies, it would be an exaggeration to say that the US AI industry, or even the global AI industry, is 'dominated' by the Big Tech. This is because, besides the likes of Google and Microsoft, there exists a constellation of AI specialist companies who have also made significant advancements in AI innovation and are wielding considerable power in shaping the future development of AI. OpenAI (founded in 2015) and Anthropic (founded in 2021) are two examples of such companies. Both firms have achieved Unicorn status shortly after their establishment and eclipsed the big technology companies for their pioneering innovation in LLMs and Generative AI.

Forming another important pillar of the US's leadership in AI is a constellation of hardware manufacturers and service providers that make the essential products and services needed for AI innovation. Hardware manufacturers such as NVIDIA and Intel are global leaders in GPUs, drives, and semiconductors, whereas companies like IBM offer advanced software and cloud services for both large and small to medium-sized tech companies. These companies complete the US's advanced AI innovation ecosystem, ensuring the country's leadership in the global AI R&D scene.

Lastly, fuelling the US's global leadership in AI has been the abundant investments made in the field. According to the AI Index Report 2023, AI-related investment has increased thirteenfold globally between 2013 and 2022, including both private investment and other forms of investment such as Mergers and Acquisitions (M&A) and public offerings (Maslej et al 2023: 184). The United States has outpaced all other countries by having attracted \$248.9 billion in private AI investment from 2013 to 2022, more than the total sum of capital received by China (\$95.11 billion), the United Kingdom (\$18.24 billion), Israel (\$10.83 billion), Canada (\$8.83 billion), and India (\$7.73 billion) (see Figure 9.1).

China

In recent years, China has emerged as a significant force and global leader in the field of AI, rapidly closing the gap with traditional leaders like the United States. China's great leap forward in AI is often attributed to three interrelated factors, namely significant investments, policy support, and a large talent pool in AI and science, technology, engineering, and mathematics (STEM) subjects (Roberts et al 2020; Lundvall and Rikav 2022; Ding 2018).

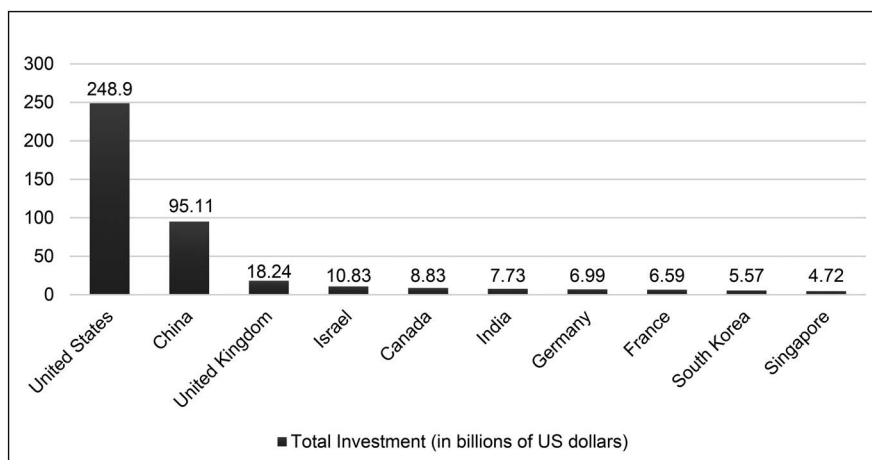


Figure 9.1 Private investment in AI by geographic area, 2013–22.

Source and Chart: AI Index Report 2023.

The number of publications is a widely used metric to gauge a country's or a region's competitiveness in AI research and innovation (CISTP 2018). Publications in AI include several main types of outputs, such as journal articles, conference papers, repository submissions, books and book chapters, and theses. According to Maslej et al, journal articles, conference papers, and repository submissions accounted for 60%, 17%, and 13%, respectively, of global AI research outputs in 2021 (2023: 25). In journal publications, China maintains a clear leadership by contributing to 39.78% of all AI journal articles published between 2010 and 2021, significantly surpassing the EU and UK's 15.05% and the US's 10.03% (Maslej et al 2023: 34). Between 2016 and 2021, China continuously topped the world journal citation league by accounting for 25%–35% of all journal citations in AI (Maslej et al 2023: 35). China's leadership in AI publications can be attributed to the country's long-term heavy investment in AI and STEM discipline education and research. As an example, in April 2018, less than a year since the State Council released the seminal *The Next Generation Artificial Intelligence Development Plan*, the Ministry of Education published *The AI Innovation Action Plan for Higher Education Institutions* to give further guidance on strengthening AI education, research, and innovation. The central government's emphasis on AI and STEM education has ensured a large and continuously growing AI talent pool in the country. Attesting to this is that in 2021, according to the AI Index Report, the first nine of the top ten institutions in the world with the most AI publications published in the past decade were all Chinese universities and research institutes (Maslej et al 2023: 44).

The rapid development of China's AI industry has also benefited from policy support. Not only the central government but provincial, municipal, and even district and county-level governments across China have also been actively involved in nurturing their local AI sectors and capabilities through strategic investments and policy

initiatives. According to data from [Qianzhan Industry Research Institute \(2023\)](#), as of 2023, 31 out of the 34 provincial-level governments in China had launched AI-related development plans and initiatives.² These local policy frameworks drive AI research and innovation by incorporating financial incentives, infrastructure development, talent attraction and education, and academia-industry collaboration. Furthermore, leveraging their existing strengths and resources, different provinces have adopted different AI strategies, contributing to the formation of a diverse and balanced AI research and innovation landscape in China. For example, harnessing their existing strengths in hosting large national data centres and supercomputing facilities, Guizhou and Hunan Provinces have focused their AI efforts on establishing cutting-edge computing and data centres. In contrast, innovation powerhouses like Beijing, Shanghai, and Shenzhen prioritise fostering world-class AI start-ups and unicorns in their respective AI strategies. Meanwhile, provinces with a traditional emphasis on manufacturing, such as Shandong and Jiangsu, prioritise the development of AI-powered modern manufacturing industries.

There has been somewhat a misconception that China's great leap forward in AI has benefited from relaxed regulations ([Ding 2018](#)). The reality, however, is that China has always been at the forefront of making legislation and regulations to protect personal data and privacy, ensure safety and security in the use of technology, and promote ethical and responsible AI.³ It is therefore wrong to attribute China's ascent in AI to the so-called regulatory leniency. For example, only five months after OpenAI released its Chat-GPT to the public, igniting worldwide interest, speculation, and concern, the Cyberspace Administration of China published a draft of *Interim Measures for the Administration of Generative Artificial Intelligence Services* in April 2023. The draft was the first of its kind in the world. This draft was then open for public consultation for a month before it was approved and jointly issued by a number of central government agencies in July 2023 and formally took effect in August 2023. This *Interim Measures* is the world's first regulation on Generative AI issued by a central government.

According to media reports, the final version of the *Interim Measures* underwent several significant changes from the initial draft, taking into account the feedback obtained during the public consultation process ([Fan and Yang 2023](#)). For instance, a key alteration involved the removal of a clause stating, 'Content generated using Generative AI should be true and accurate, and measures should be taken to prevent the generation of false information'. This omission is believed to result from feedback received from the AI community regarding the current limitations faced by AI developers in comprehensively understanding and controlling model behaviour, a problem that is being actively researched by AI researchers worldwide ([Fan and Yang 2023](#)). The removal of this clause thus serves to show that Chinese policy-makers adopt a proactive and open approach to AI regulation. They seek to balance regulatory and legal oversight of AI with promoting AI research and innovation.

Let us consider another example. The *Interim Measures* is a regulation setting out the legal, regulatory, and ethical frameworks with which generative AI developers and providers must comply. But it also delineates its scope by clearly stating that the regulation applies only to individuals and organisations using generative AI

technology ‘to provide services that generate texts, images, audio and video, and other content to the public within the territory of the People’s Republic of China’, and *not to those engaging in research, development, or application of Generative AI technology without offering services to the domestic public* (CAC 2023; author’s emphasis). This distinction demonstrates that the policymakers do not intend to inhibit AI research and innovation in China by making superfluously constraining regulations. Their goal, rather, is to oversee AI innovation activities that have a direct impact on public interests while concurrently allowing ample room for pioneering AI research to flourish. This principle can be said to be characteristic of China’s overall approach to AI legislation and regulation: effectively balancing the promotion of AI innovation with the maintenance of a robust legal and regulatory framework in AI.

Lastly, China’s dynamic industrial ecosystem plays a pivotal role in driving the country’s development in AI. China is home to a burgeoning start-up scene, witnessing over the past decade numerous start-ups being founded and having made great innovations in all major subfields of AI, ranging from computer vision to speech recognition, and from intelligent vehicles to robotics. In LLMs and Generative AI, the fast development of the field in the past few years, notably since OpenAI’s release of Chat-GPT in November 2022, has led to what has been referred to by many as ‘The War of A Hundred Models’, or even ‘The War of A Thousand Models’. This terminology indicates the scale of investments made by Chinese companies in LLMs and their diverse scenario applications and vividly illustrates the intense competition in the field. Generally speaking, five main types of players are shaping the development of the field of LLMs and Generative AI in China:

- 1 Leading internet companies, such as Baidu, Tencent, and Alibaba.
- 2 Specialist ICT companies, such as Huawei and iFlytek.
- 3 Established and start-up AI companies, such as Zhipu AI and SenseTime.
- 4 Universities and research institutes, such as Tsinghua University.
- 5 AI research arms of large corporations in non-tech sectors.

Table 9.1 showcases a small selection of LLMs developed and released in China in recent years. This list, though not exhaustive, highlights two key features of the Chinese AI innovation ecology.

Firstly, it highlights that China has a highly dynamic and competitive AI research and innovation scene. Contrary to the popular misconception that the handful large and influential internet companies dominate the sector (Steinhoff 2023) – such as Baidu (often likened to Google), Tencent (akin to Facebook), and Alibaba (regarded as Amazon’s Chinese counterpart) – the reality is far more dynamic and complicated. Chinese firms and innovators in AI, large and small, established companies or small start-ups, often find themselves operating in an intensely competitive, highly volatile, and remarkably fast-changing environment. This is distinct from their counterparts in most other countries. Successful companies can quickly decline if they fail to innovate fast enough (Zhou 2024). Commenting on this, Tang Daosheng, Senior Executive Vice President of Tencent and CEO of the Cloud and Smart Industry Business Group, said in an interview that fierce competition, instead of dominance

Table 9.1 Selected large language models released in China

<i>Company type</i>	<i>Company name</i>	<i>Sector</i>	<i>Model name</i>	<i>Initial release date</i>	<i>Model type</i>
Leading Internet companies	Baidu	Internet	Wenxin Yiyan	03.2023	General
	Tencent	Internet	Hunyuan	09.2023	General
	Alibaba	Internet	Tongyi Qianwen	04.2023	General
ICT companies	Huawei	ICT	Pangu	04.2021	General
	iGlytek	ICT	Xinghuo	05.2023	General
	Inspur	ICT	Yuan	11.2023	General
AI specialist companies	Zhipu AI	GenAI	Zhipu Qingyan	08.2023	General
	Baichuan AI	GenAI	Baichuan	09.2023	General
	SenseTime	AI	Riri Xin	04.2023	General
	MiniMax	GenAI	MiniMax	04.2023	General
	Lingyi Wanwu	GenAI	Yi	11.2023	General
Universities and research institutes	Tsinghua University	Academia	GLM-10B	09.2021	General
	Shanghai AI Lab	Academia	Shusheng	07.2023	General
	The Institute of Automation of Chinese Academy of Sciences	Academia	Zidong Taichu	07.2021	General
	Fudan University	Academia	MOSS	02.2023	Academic
	TAL Education Group	Education	MathGPT	08.2023	Education
Companies in traditional sectors	Agricultural Bank of China	Banking	ChatABC	03.2023	Fintech
	Ctrip	Tourism	Xiecheng Wen dao	07.2023	Tourism

by a select few companies, is and will be the key defining feature of the Chinese AI industry (Deng 2024). Similarly, the Shenzhen Artificial Intelligence Industry Association (SAIIA) finds in a report that none of the leading companies in AI in China has achieved sufficient ‘scale economy’, that is, market advantage large enough to accord dominance (SAIIA 2023: 15). On the contrary, the report asserts that substantial competition between existing players and ongoing ‘threats’ from newcomers will continue to shape the Chinese AI industry for years to come (SAIIA 2023: 16).

Secondly, Chinese AI companies prioritise application and deployment. Their research and innovation activities are underpinned by a clear objective of applying AI technology to existing businesses and industries – often referred to as ‘scenarios’ by insiders – in order to create new value. This application-centric vision is illustrated by the substantial investments made by companies in traditional,

non-technology sectors in AI research and innovation: AI is not perceived as a futuristic or sci-fi concept, as is sometimes the case in some other cultures (Cave, Dihal and Dillon 2020), but rather as tangible tools holding a great promise for productivity enhancement and cost reduction. China's approach to AI research and innovation, therefore, is profoundly shaped by this bottom-up, application-centric vision of utilising AI to empower businesses, industries, and society.

The above analysis of the AI research and innovation landscapes in the United States and China serves to highlight that it is justifiable to suggest that AI R&D in the present age exhibits a distinctive political-economic nature. That is, the current surge of global enthusiasm surrounding AI – which began in the first decade of this century as part of the broader digital revolution – is significantly influenced by the economic, historical, social, and political milieu in which it unfolds. It is to accentuate this distinct feature of AI research and innovation in the twenty-first century that I define AI as clusters of *technoeconomic systems* capable of generating human-like behaviour. In this light, the development of a systematic sociological theory of AI necessitates an examination of this unique feature of AI, namely, to study how the AI products and services we, as consumers, are increasingly using in our daily lives are goods and services produced and distributed by an expansive global network of stakeholders with diverse and yet interrelated interests across various social, economic, and political domains. The political economy tradition provides us with some useful conceptual tools and frameworks to critically analyse this global production, distribution, and consumption of AI.

The Political Economy Analysis of Economic Processes

The term 'political economy' has different meanings and implications in different fields of study. The tradition of political economy analysis thus has engendered distinct analytical scopes, diverse research agendas, and priorities in different disciplines. Traditionally, political economy is a subfield situated within economics (Clark 1998). It studies primarily the intricate relationship between 'the world of public power and decision-making' – i.e. politics – and 'the world of production and distribution' – i.e. the economy (Staniland 1985: 10). Although these two worlds – the politics, on the one hand, and the economy, on the other – seem to hold equal importance in the framework of 'political economy', the traditional political economy analysis, rooted in economics, examines predominantly how the exercise of political power and actions of political actors affect economic processes and behaviours. It places a lighter emphasis on how the functioning of the economy – which can be defined as the totality of the production, distribution, and consumption of resources – influences the distribution and governance of power. This strand of political economy, commonly referred to as 'Classical Political Economy', has been defined as 'the scholarly discourse studying power relations affecting the production, distribution, and consumption of wealth, income and resources' (Babe 2009: 13). In contrast, an approach to political economy that does not particularly concern its history and economics roots might simply define it as 'the scientific, but also necessarily evaluative and thus political, study of economic practices and systems' (Elder-Vass 2016: 17).

The study of ‘political economy’ can be said to be a child of the Enlightenment and the subsequent Industrial Revolution, the two social, intellectual, and economic movements that took place initially in Western Europe and then spread to the rest of the world throughout the seventeenth, eighteenth and nineteenth centuries via commerce, wars, and colonisation. Scholars often attribute the origins of political economy to figures like Adam Smith and Edmund Burke, though some trace its roots back even further to the Middle Ages and even Ancient Greek times (e.g. Staniland 1985). However, the prevailing view is that Adam Smith’s *The Wealth of Nations* (1937 [1776]) is the inaugural text of the tradition of political economy. Now commonly referred to as Classic Political Economy, Adam Smith’s political economy – defined as a “‘branch of the science of a state or legislator” helping governments set conditions to stimulate economic growth’ (Babe 2009: 13) – is regarded as representing the conservative faction of political economy thought, despite Smith himself being considered a radical in his time.

As the economy evolved and the relationship between the economic and political worlds changed, the meaning and scope of the study of ‘political economy’ also evolved. Most importantly, a critical perspective was brought into the hitherto conservative political economy study, most notably by Karl Marx, to critique the social ramifications of capitalism, the free market, and a laissez-faire government advocated by classical political economists. Despite this radical departure, the definition of political economy remained largely unchanged: it continued to primarily refer to the scholarly study of the economy in terms of its relationship to power, politics, and governance.

One significant development in political economy with far-reaching influence took place in a somewhat unexpected field of study: media and communications studies. Beginning roughly in the 1960s and 1970s, groups of scholars in North America and Britain studying the societal implications of mass media and new communications technologies, such as television, became increasingly dissatisfied with the analytical tools they had at hand. Despite originating from different philosophical traditions and driven by divergent empirical concerns – North American scholars focused on accounting for the unprecedented social influence of mass media industries like cinema and television, whereas British scholars were concerned with the impact of popular culture on the assimilation of the working class to dominant culture and the rendering of the working class unrevolutionary (for further details, see Garnham and Fuchs (2014)) – media and communications scholars at both sides of the Atlantic converged on one point. They both recognised that, despite the differences between the mass media industries and traditional sectors like textiles, shipping, or oil, they operated under the same logic: the capitalist logic. That is, the mass media industries were governed by the same capitalist rationale when it comes to the organisation and accumulation of resources, to the production, distribution, and consumption of products and services, and to the reproduction and reinforcement of power of large corporations. It was from this vantage point that the political economy of media and communications was developed to scrutinise ‘the social relations, particularly power relations, that mutually constitute the production, distribution and consumption of communication resources’ (Mosco 2009: 24).

Sometimes, the study of the political economy of media and communications is referred to as the *critical* political economy (CPE) of media and communications to

accentuate the critical – particularly Marxist – influence on this analytical approach to the media (Hardy 2014). Influenced by the Marxist tradition of social critique, the CPE of media analysis focuses on examining the historical and material dimensions of media production, distribution, and consumption, in addition to the dimensions of political power. That is, it broadens and enriches the traditional political economy by underscoring the temporal, historical, and materialist composition of the creation and consumption of media products and services (Garnham 1979). By underscoring this composition, the CPE of the media necessarily picks up a critique of the capitalist ideology permeating not only the media industries but also the entire society. The goal of CPE (of the media) thus has evolved to become one of unveiling, criticising, and challenging the unequal power relations resulting fundamentally from the capitalist mode of production that prevails and shapes not only media and communications but also other industries and sectors whose products and services, directly and indirectly, impact the physical and mental well-being of citizens. The CPE analysis thus has evolved into a comprehensive social critique of the economy.

The tradition of political economy – with or without the label ‘critical’ – has been picked up by scholars of science and technology. Robin Mansell, for instance, called for a revitalisation of political economy analysis in new media and internet studies. She defines this pursuit as ‘the analysis of the specific historical circumstances under which new media and communications products and services are produced under capitalism, and with the influence of these circumstances over their consumption’ (Mansell 2004: 98). Similarly, Kean Birch, in a paper titled ‘The Political Economy of Technoscience: An Emerging Research Agenda’, makes a case for ‘examining how different political-economic processes, practises and contexts inform technoscience’ (Birch 2013: 51). From the political economy of media and communications to the political economy of technoscience, what has changed are the empirical objects of study: from analysing the press, broadcasting, and television industries of the twentieth century (see Golding and Murdock 1997) to exploring the information and communication technologies (Fuchs 2009; Dahlberg 2015) as well as biotechnology, genomics, and nanotechnology characterising the technological landscape of the twenty-first century (Suarez-Villa 2009, 2012, 2023; Hogarth 2017). What has remained largely unchanged, however, is a number of key themes for analysis. In the final section of this chapter, I provide an overview of these key themes, focusing on how these analytical perspectives might be adapted to yield conceptually and empirically rigorous political economy analyses of AI.

Key Themes of Political Economic Analysis

It is imperative to begin this section by firstly acknowledging the existing body of research that utilises political economy approaches to analysing AI, whether it is specific AI products and services, AI research and innovation activities, or the AI industries.

Much of this body of work is conducted by media and communications scholars, illustrating the significant influence of media and communications studies in political economy, and vice versa. Verdegem (2023) and Steinhoff (2023) stand as two good reviews of this corpus of research, which they term as ‘Critical AI Studies’.

While offering some valuable insight into the political-economic organisation of the AI industries in the twenty-first century, a notable problem within this body of research – and in agenda-setting proposals such as the aforementioned articles – is a tendency, firstly, to conceptually conflate AI with other related but inherently different digital technologies and/or processes, such as big data and datafication, and secondly, to over-generalisation. These issues have resulted in many findings and arguments in this body of work that lack empirical precision and distinction, which reduces the value of this work.

For example, in an effort to establish the importance of Critical AI Studies, [Verdegem \(2023: 303–304\)](#) writes:

Any critical analysis of technological innovation should always start with questioning who is behind the technology, to whom it is offered and who benefits from it. A critical analysis of AI should investigate what AI is, what visions exist around it, who is behind the surrounding ideologies and discourses [...] and who can decide what type of AI is being produced and how it is used.

While this statement certainly points out some important aspects of a CPE analysis of AI, it is arguably too general to invite any substantive debate about and exploration into the distinctiveness of AI. It offers limited insight into the peculiarity of the AI industry vis-à-vis other technology and non-technology sectors subject to a CPE analysis and fails to capture some of the unique organisational, financing, and innovation structures of the AI industry and the distinct ways AI technologies are produced, disseminated, and consumed. Therefore, it might not be particularly useful to frame the political economy analysis of AI in terms of ‘questions about AI and power’ ([Verdegem 2023: 304](#)), for unlike in some other fields – such as media and communications – ‘power’ takes a more multifaceted and multifarious form in AI research and innovation. Concepts, theories, and perspectives about power that have proved useful for accounting for media production might not be readily applicable, or even useful, for analysing the organisation and distribution of power in the sphere of AI. Similarly, the issue of ‘ideology’ must be dealt with careful consideration and discernment. Unlike media products, which primarily consist of symbolic contents and hence lend themselves to analysis of how symbolic power gets embedded in and reinforced through media production and consumption, AI products and services bear significant materialist forms. Addressing the issue of ideology within the context of AI thus necessitates a more nuanced and bottom-up approach characterised by empirical precision and an eye for distinction, in order to account for AI’s distinctiveness while maintaining a normative critical value in the analysis.

Ownership and Control

Political economy always concerns who owns and thus controls key resources. These resources include both tangible economic resources, such as money and land, and intangible resources such as access to labour/talent, information, markets, and other relevant forms of ‘capital’ ([Bourdieu 2011](#)). In the political economy of media and

communications, the issue of ownership and control revolves around understanding who owns and controls media outlets, platforms, and content. It deals with the question of how media ownership and control influence the production, distribution, and consumption of media products or services (Hardy 2014). Here, ownership refers to the legal right to possess and dispose media assets, such as television networks, newspapers, radio stations, and digital platforms. Control refers to the ability to influence decision-making processes, content creation, editorial policies, and the overall direction of media organisations. The political economy of the media scholars have examined how the structures of ownership and control can shape media content, affect diversity, and influence representation, and identified that concentration of ownership – where a few large conglomerates own and control a significant portion of the media market – can have implications for media pluralism, diversity of voices, and editorial independence (Downing 2011).

A political economy of AI must address the issue of ownership and control, too. But it should do so by first acknowledging and mapping out – meticulously and empirically – the intricate ownership structures that organise and govern the global AI industries in the twenty-first century. Failing to do so will invite allegations of overstating the ‘dominance’ of major tech conglomerates in AI and oversimplifying their oligopolistic control. The complexity of the issue of ownership and control in AI manifests at several levels. For instance, unlike in the media industries, where large media groups customarily acquire smaller firms to bolster their market dominance through mergers and acquisitions, large tech companies often choose to invest in promising start-ups instead of outright acquisition. Since it is not uncommon for large companies to invest in competing start-ups concurrently (Murgia 2024), the issue of ownership and control in the AI industry tends to be more intricate than in traditional sectors where competition is significantly less volatile and less expensive. Indeed, the colossal financial commitment required by AI research and innovation today raises questions about the feasibility and therefore practical value of calls for AI as public service (Verdegem 2021). Therefore, any political economy analysis of the issue of ownership and control in AI must resist the temptation to make sweeping assumptions about the ‘dominance’ of certain (types of) players in AI R&D. Such oversimplification will stop us from seeing the dynamic nature of practices in AI research and innovation and from grasping and critiquing the influence of other relevant players who potentially wield more power in shaping AI innovation than the big technology companies, such as venture capital institutions.

Concentration

The issue of concentration and consolidation naturally ensues the discussion of the issue of ownership and control. In the political economy of media and communications, it has been argued that the concentration of power within a limited number of media conglomerates is the root cause of a variety of problems characterising media production – such as the homogenisation of content, biased reporting, and limited perspectives – in that media outlets tend to prioritise the interests of their

owners or advertisers over journalistic integrity or public interest. Moreover, the matter of power concentration also intersects with broader socio-political dynamics, such as corporate influence on public discourse, media manipulation for political or commercial purposes, and the potential for monopolistic practices to stifle competition and innovation in the media landscape.

In the realm of technology, criticisms over the concentration of power have predominantly targeted the select group often referred to as the ‘tech giants’, namely, Google (Alphabet), Apple, Facebook (Meta), Amazon, and Microsoft in the United States, and Baidu, Alibaba, Tencent, and sometimes ByteDance (owner of TikTok) and Xiaomi as well, in China. While it is true that these companies have all made colossal investments in AI R&D, to what extent we can argue that they therefore ‘dominate’ the global AI industries and possess the power to determine the content and future of AI innovations, whether individually or collectively, remains a question to be explored more meticulously.

As mentioned earlier, a major difference between AI research and innovation from innovation in most other sectors is the crucial roles played, on the one hand, by universities and research institutes, and on the other hand, by start-up companies. While there are often overlaps and connections between the large tech conglomerates and universities and research institutes and it is commonplace for the tech giants to acquire or invest in competitive start-ups, it would be unwise to limit the inquiry of the issue of power concentration and consolidation in AI to a handful of players and overlook the ‘big picture’ of how the power to shape, lead, and benefit from AI innovations is fiercely contested across multiple levels by an extended network of actors and entities (Garnham 2011).

Labour

Labour is another key issue central to political economy analysis. In the domain of media and communications studies, including digital media, investigations have been made into topics such as precarious work, surplus value creation, labour exploitation, and de-unionisation (Gall 2002; Hesmondhalgh and Baker 2011; Mosco 2011). Similarly, the issue of labour in the technology sector, including AI, has also attracted much attention from scholars from across a diversity of backgrounds and has yielded a valuable amount of work, with or without political economy as their main analytical framework (see, for instance, Wood, Lehdonvirta and Graham 2018; Wood et al 2019; Steinhoff 2022). Future investigations into this area, particularly with a specific political economy concern, can examine phenomena such as employee activism within ethical AI movements, the drivers and implications of top AI scientists transitioning from academia to industry, and the motives prompting employees to depart from major technology firms to establish start-ups. Grasping the evolving roles of and dynamics between employees and employers in the AI industry in the twenty-first century will be crucial for understanding issues like capital reproduction, value creation, the appropriation of creativity, and the redistribution of power in the technology sector.

Commodification and Commercialisation

The issues of commodification and commercialisation are two closely interrelated themes central to political economy inquiries into media and culture (Murdock and Golding 2005: 64). Critical political economists have regarded the pursuit of commercial interests and a drive for commercialisation as a root cause of a range of problems that undermines the creation of a just and equitable media market. These problems include standardisation and homogenisation, commercialisation of creativity, censorship of ideas and information, and the propagation of consumerism. Hardy distinguishes commodification and commercialisation as two interconnected yet distinct processes. He defines the former as ‘the process of transforming objects, products and services into commodities, things that can be bought and sold for money’ and the latter as ‘the orientation of activities towards the goal of increasing commercial revenue for profit’ (Hardy 2014: 136). He further elucidates their differing scopes by considering commodification as ‘the processes of incorporation into commodity form’ – that is, it describes the specific process of transforming information and ideas into media goods and services for sale – and commercialisation as the fundamental driving force behind commodification. The issue of commodification and commercialisation is intensely pertinent to discussions in the AI field concerning topics such as open-source AI, AI for social good, and AI ethics (Züger and Asghari 2023; Murgia and Criddle 2024). The contribution of a political economy analysis to these discussions will depend on whether it can effectively reveal and interrogate the socioeconomic structures and power dynamics that have fundamentally led to the intensely ‘commercial’ way of AI research and innovation in the present age and can offer suggestions for feasible alternatives, so as to shed light on what a more just and equitable AI research and innovation landscape might look like.

Conclusion

In this chapter, I have argued that political economy offers a valuable and must-needed analytical perspective for developing a more systematic sociological theory of AI by allowing us to examine how the structures, processes, and relations of power affect the production, distribution, and consumption of AI products or services. Through examining the leadership of the United States and China in global AI R&D, especially in the field of LLMs and Generative AI, I have shown how AI research and innovation in the present era has a distinctive political economic character: industry players have overtaken academia in driving and shaping key innovations in AI. I argue that this has significant implications both for society and for our grasping of AI from a sociological perspective. An overview of the tradition of political economy analysis helps to surface a number of areas for inquiry for a political economy analysis of AI. While the list is by no means exhaustive, it is intended as an invitation for readers – especially students and scholars new to the nascent field of the sociology of AI (which should be distinguished from STS of AI

and aim to develop into a distinct field of study in its own right) – to start thinking about how the development of a systematic and comprehensive sociological theory of AI can benefit from examinations of the economic and political dimensions of this emergent technology of the twenty-first century.

Notes

- 1 My analysis in this chapter is developed against the recent development in the AI subfield of LLMs based generative AI, specifically. The author recognises that ‘AI’ as an expansive field of study has many other subfields, in which industry might not play such a central role as it does in the field of generative AI. However, taken as a whole, it is justifiable to say that the current wave of AI research and innovation, starting from the late 2000s and early 2010s, is predominantly driven by industry players, as opposed to academia.
- 2 China has 34 provincial-level administrative regions, consisting of 23 provinces, five autonomous regions, four municipalities, and two special administrative regions.
- 3 China introduced the Personal Information Protection Law (PIPL) in August 2021. This is a comprehensive law aiming at regulating the use of personal information by individuals and organisations, including state agencies (see Chapter II – Section Three of the PIPL). For full text of the PIPL in Chinese, see: https://www.gov.cn/xinwen/2021-08/20/content_5632486.htm. For non-official English translation of the PIPL, see: Personal Information Protection Law of the People’s Republic of China – PIPL. With regard to cybersecurity, the Cybersecurity Law of the People’s Republic of China (https://www.cac.gov.cn/2016-11/07/c_1119867116.htm) was implemented in 2016.

References

- Babe, R. E., 2009. *Genealogy of Political Economy in Cultural Studies and Political Economy: Toward a New Integration*. Lexington Books London.
- Birch, K., 2013. The political economy of technoscience: An emerging research agenda. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, 7(1), pp. 49–61.
- Bourdieu, P., 2011. The Forms of Capital (1986) in *Cultural Theory: An Anthology*. Wiley-Blackwell, Hoboken, New Jersey, pp. 81–93.
- CAC, 2023. *Interim Measures for the Administration of Generative Artificial Intelligence Services*. Beijing. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm
- Cave, S., Dihal, K. and Dillon, S. (eds), 2020. *AI Narratives: A History of Imaginative Thinking About Intelligent Machines*. Oxford University Press.
- CISTP, 2018. *China AI Development Report*. Beijing.
- Clark, B., 1998. *Political Economy: A Comparative Approach*. Praeger.
- Dahlberg, L., 2015. Expanding digital divides research: A critical political economy of social media. *The Communication Review*, 18(4), pp. 271–293.
- Deng, Y., 2024. Dialogue with Tencent Tang Daosheng: Large models, like the cloud, are difficult to be monopolized. 36 Kr. 7 February 2024. https://mp.weixin.qq.com/s/JQSh_q0f9zrUcf7_uW-z5g
- Ding, J., 2018. *Deciphering China’s AI Dream*. Future of Humanity Institute Technical Report.
- Downing, J. D. H., 2011. Media Ownership, Concentration, and Control: The Evolution of Debate in J. Wasko, G. Murdock, and H. Sousa (eds) *The Handbook of Political Economy of Communications* (1st edition, pp. 140–168). Wiley. Hoboken, New Jersey.
- Elder-Vass, D., 2016. *Profit and Gift in the Digital Economy*. Cambridge University Press.
- Elliott, A., 2019. *The Culture of AI: Everyday Life and the Digital Revolution*. Routledge. Abingdon.
- Elliott, A., 2021. Making sense of AI: Our algorithmic world. Polity.

- Fan, S. and Yang, L., 2023. *The World's First Generative AI Regulation Is about to Come into Effect, and a Number of Controversial Clauses Have Been Removed (in China)*. Caijing Elaw. <https://mp.weixin.qq.com/s/yW9putNcX0XIEHcH8tQ3Pw>
- Fuchs, C., 2009. Information and communication technologies and society: A contribution to the critique of the political economy of the internet. *European Journal of Communication*, 24(1), pp. 69–87.
- Gall, G., 2002. The return of the National Union of Journalists to the provincial newspaper industry in Britain? *Media, Culture & Society*, 24(5), pp. 673–692.
- Garnham, N., 1979. Contribution to a political economy of mass-communication. *Media, Culture & Society*, 1(2), pp. 123–146.
- Garnham, N., 2011. The Political Economy of Communication Revisited in J. Wasko, G. Murdock, and H. Sousa (eds) *The Handbook of Political Economy of Communications* (1st edition, pp. 41–61). Wiley. Hoboken, New Jersey.
- Garnham, N. and Fuchs, C., 2014. Revisiting the political economy of communication. *tripleC: Communication, Capitalism & Critique*, 12(1), pp. 102–141.
- Golding, P. and Murdock, G., 1997. *The Political Economy of the Media*. Edward Elgar Publishing. Cheltenham.
- Hardy, J., 2014. *Critical Political Economy of the Media: An Introduction*. Routledge. Abingdon.
- Hesmondhalgh, D. and Baker, S., 2011. Toward a Political Economy of Labor in the Media Industries in J. Wasko, G. Murdock, and H. Sousa (eds) *The Handbook of Political Economy of Communications* (1st edition, pp. 381–400). Wiley. Hoboken, New Jersey.
- Hogarth, S., 2017. Valley of the unicorns: Consumer genomics, venture capital and digital disruption. *New Genetics and Society*, 36(3), pp. 250–272.
- Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., Noble, S. U. and Shestakofsky, B., 2021. Toward a sociology of artificial intelligence: A call for research on inequalities and structural change. *Socius*, 7(4), pp. 1–11. <https://doi.org/10.1177/2378023121999581>
- Kaplan, J., 2016. *Artificial Intelligence: What Everyone Needs to Know*. Oxford: Oxford University Press.
- Lindgren, S., 2023. *Critical Theory of AI*. Polity. Cambridge.
- Liu, Z., 2021. Sociological perspectives on artificial intelligence: A typological reading. *Sociology Compass*, 15(3). <https://doi.org/10.1111/soc4.12851>
- Lundvall, B. and Rikap, C., 2022. China's catching-up in artificial intelligence seen as a co-evolution of corporate and national innovation systems. *Research Policy*, 51(1). <https://doi.org/10.1016/j.respol.2021.104395>
- Mansell, R., 2004. Political economy, power and new media. *New media & Society*, 6(1), pp. 96–105.
- Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Wald, R., Clark, J. and Perrault, R., 2023. *The AI Index 2023 Annual Report*, Stanford, CA: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.
- Mosco, V., 2009. *The Political Economy of Communication*. SAGE. London.
- Mosco, V., 2011. The Political Economy of Labor, in J. Wasko, G. Murdock, and H. Sousa (eds) *The Handbook of Political Economy of Communications* (1st edition, pp. 358–380). Wiley. Hoboken, New Jersey.
- Murdock, G. and Golding, P., 2005. 'Culture, Communications and Political Economy' in J. Curran and M. Gurevitch (eds) *Mass Media and Society*, London: Hodder Arnold.
- Murgia, M., 2024. Microsoft strikes deal with Mistral in push beyond OpenAI. *Financial Times*. 26 February 2024. <https://www.ft.com/content/cd6eb51a-3276-450f-87fd-97e8410db9eb>
- Murgia, M. and Criddle, C., 2024. Elon Musk sues OpenAI and Sam Altman for breach of contract. *Financial Times*. 1 March 2024. <https://www.ft.com/content/6a4cfc6d-b39d-46bb-b40a-2ace23682996>
- Nee, V. and Swedberg, R., 2005. *The Economic Sociology of Capitalism*. Princeton University Press. Princeton.

- NSTC., 2023. National artificial intelligence research and development strategic plan 2023 update. <https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf>
- Qianzhan Industry Research Institute, 2023. Summary and interpretation of AI industry policies and analysis of development goals in China and 31 provinces and cities in 2024 (in Chinese). <https://bg.qianzhan.com/trends/detail/506/231228-7bc5e404.html>
- Roberts, H., Cows, J., Morley, J., Taddeo, M., Wang, V. and Floridi, L., 2020. The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & Society*, 36, pp. 59–77.
- Saia, 2023. *The White Book on the Development of Artificial Intelligence in 2023*. Shenzhen.
- Smith, A., 1937. *The Wealth of Nations* [1776]. Random House, Inc.
- Staniland, M., 1985. *What Is Political Economy? A Study of Social Theory and Underdevelopment*. Yale University Press. New Haven and London.
- State Council, 2017. *The Next Generation Artificial Intelligence Development Plan*. Beijing.
- Steinhoff, J., 2021. Industrializing Intelligence: A Political Economic History of the AI Industry in *Automation and Autonomy: Labour, Capital and Machines in the Artificial Intelligence Industry* (pp. 99–131). Palgrave Macmillan. London.
- Steinhoff, J., 2022. Toward a political economy of synthetic data: A data-intensive capitalism that is not a surveillance capitalism? *New Media & Society*. <https://doi.org/10.1177/14614448221099217>
- Steinhoff, J., 2023. The Industry of Automating Automation: The Political Economy of the AI Industry in *Handbook of Critical Studies of Artificial Intelligence* (pp. 312–322). Edward Elgar Publishing. Cheltenham.
- Suarez-Villa, L., 2009. *Technocapitalism: A Critical Perspective on Technological Innovation and Corporatism*. Temple University Press.
- Suarez-Villa, L., 2012. *Globalization and Technocapitalism: The Political Economy of Corporate Power and Technological Domination*. Routledge. Abingdon.
- Suarez-Villa, L., 2023. *Technology and Oligopoly Capitalism*. Routledge.
- Verdegem, P., 2021. *AI for Everyone? Critical Perspectives*. University of Westminster Press. Abingdon.
- Verdegem, P., 2023. Critical AI Studies Meetings Critical Political Economy in *Handbook of Critical Studies of Artificial Intelligence* (pp. 302–311). Edward Elgar Publishing. Cheltenham.
- Wood, A. J., Graham, M., Lehdonvirta, V. and Hjorth, I., 2019. Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment and Society*, 33(1), pp. 56–75.
- Wood, A. J., Lehdonvirta, V. and Graham, M., 2018. Workers of the internet unite? Online freelancer organisation among remote gig economy workers in six Asian and African countries. *New Technology, Work and Employment*, 33(2), pp. 95–112.
- Zajko, M., 2021. Conservative AI and social inequality: Conceptualizing alternatives to bias through social theory. *AI & SOCIETY*, 36(3), pp. 1047–1056.
- Zhou, X., 2024. Falling behind the wave of large models, the star AI company Emotibot announced a partial shutdown (in Chinese). *36 Kr*, 24 February 2024. <https://36kr.com/p/2658348660489352>
- Züger, T. and Asghari, H., 2023. AI for the public. How public interest theory shifts the discourse on AI. *AI & Society*, 38, pp. 815–828.

10 Four Models of Algorithmic Bureaucracy in China

Baogang He and Jinjin Wu

Introduction

Modern information and communication technology (ICT) has emerged as a major force to reshape the public sector in the big data era. Digital government empowered by ICT and artificial intelligence (AI) has strengthened the adaptive capacity of public organisations (Fountain 2004; Margetts 2012). Street-level bureaucracy has been transformed into screen-level bureaucracy, a new type of human-computer interaction, and system-level bureaucracy governed by algorithms and information systems (Bovens & Zouridis 2002; Busch & Henriksen 2018). The introduction of AI and algorithms to bureaucracy has revolutionised the traditional bureaucracy, leading to the emergence of algorithmic bureaucracy.

Vogl et al. (2020) proposed the concept of algorithmic bureaucracy. Using a case study of local government in the UK, they analysed the socio-technical relationship between the public sector and algorithms and the impact of algorithms on changes in the organisational procedures of the public sector. Meijer et al. (2021) further examine the algorithmisation of bureaucratic organisations, a socially integrated procedure occurring within bureaucratic organisations, displaying autonomy in some circumstances, and yielding unintended effects. Through a comparative case study of predictive policing in Germany and the Netherlands, they demonstrate the two models of algorithmic bureaucracy: Berlin has more hierarchical control, representing the ‘algorithmic cage’; Amsterdam provides room for professional judgement, demonstrating the ‘algorithmic colleague’.

The existing works on algorithmic bureaucracy illustrate the highly complex and heterogeneous influence of algorithms on bureaucracies; however, they do not cover the various forms, patterns, and features of algorithmic bureaucracy. Furthermore, while they have analysed the key factors that influence how algorithms can shape bureaucracy, they lack an analytical framework whereby we can see how different models of algorithmic bureaucracy are formed. Existing studies need to pay more attention to the processes of interactions between algorithmic functions and application systems.

We develop an analytical framework of algorithmic bureaucracy, consisting of the functional and systemic dimensions of algorithmic applications in the public sector, exploring the impact of algorithms on the structure and operation of public

organisations, and identifying the four models of algorithmic bureaucracy. This framework enables us to demonstrate how the four models of algorithmic bureaucracy are derived from the interplay between algorithmic function and application system. The impact of algorithms on bureaucracy is influenced by several factors, of which this chapter highlights two important ones: algorithmic functionality and application systems. Automation and prediction are the two main algorithmic functions, which constitute a key factor in our analytical framework. The social embeddedness and application systems of algorithms also matter; they may be more important than their technical features (Wenzelburger et al. 2022). This is because the way that algorithms affect bureaucracy depends on the given organisation's institutions and culture, resulting in significant variation in algorithmic bureaucracy (Meijer et al. 2021). Therefore, our analytical framework will take account of departmentalised versus cross-departmental algorithmic applications. Our analytical framework provides a further understanding of the algorithmisation of bureaucratic organisation and extends previous research through its analysis of the role of algorithmic function and its application system in the transition of bureaucracy.

The dominant studies on China's AI have focused on digital authoritarianism, which has developed sophisticated surveillance, facial recognition, and social credit systems (Polyakova & Meserole 2019; Hoffman 2022; Strittmatter 2020; Farrell, Newman, and Wallace 2022). However, there is a need for more attention to AI's application in public management and the opportunity should be taken to conduct a comparative study of AI and algorithmic bureaucracy. This chapter aims to develop the science of algorithmic bureaucracy.

The four cases in this article cover multiple fields of algorithm application in the public sector. We have undertaken the study of these four cases from May 2020 to June 2022, conducted site fieldwork, and adopted semi-structured interviews to ensure that interview data were objective, true, and valid as much as possible, and kept the interview records strictly confidential. A total of more than 20 interviews were conducted with a total interview duration of about 50 hours. More than 80,000 words of interview records were compiled.

The chapter has four sections. Section 1 provides a brief literature review and introduces our analytical framework. Sections 2 and 3 examine how algorithmic functions and application systems impact algorithmic bureaucracy, respectively. Section 4 combines algorithmic functions and application systems to develop four models of algorithmic bureaucracy with Chinese example illustrations.

An Analytical Framework of Algorithmic Bureaucracy

Algorithms, the central and core part of AI, are programmed computational tools and encoded procedures enabling the AI to act autonomously in the performance of tasks by converting inputs to outputs (Cormen 2013). Through a rapid process of machine learning, algorithms can respond intelligently to a variety of different external requirements, changing environments, and circumstances. Algorithms

have expanded from market and business to public affairs due to increasing governmental digitisation.

At present, the existing studies have focused on the impact of algorithmic technologies on the composition, components, and operational procedures of public organisations. [Vogl et al. \(2020\)](#) proposed a preliminary analytical framework to distinguish between algorithmic bureaucracy and traditional bureaucracy. Their study posited that algorithms facilitate the dissemination and interchange of organisational knowledge and information, enhancing the precision of public service provision and fostering interdepartmental collaboration. [Meijer et al. \(2021\)](#) further examine the three ways in which algorithms affect bureaucratic organisational structure. First, algorithms change the information flow inside and outside the organisation, facilitating information exchange. Second, they influence the configuration of organisational structures to coordinate cooperation among departments and improve information exchange processes. Third, algorithms integrated into an organisation entail and establish a specific institution to oversee departmental responsibilities and facilitate overall regulatory processes, which requires changes in organisational systems and policies.

A common view of the relationship between technology and organisational change holds that humans interact with technology differently in different institutional settings. Organisational change is formed by the ongoing interplay between the technological decisions made by human beings and the organisation's internal and external institutions, culture, and environments. Technology alone does not dictate organisational change ([Orlikowski 2000](#)). [Fountain's \(2004\)](#) theory of technology enactment holds that the integration of technology in the public sector will invariably lead to consequential institutional and procedural transformations in the government. The way that technology is applied and the impact of technology on organisational change interacts with the organisational structure, institutional environment, and an array of other cognitive, cultural, and political factors. Control-oriented technologies support conventional management approaches and hierarchies, but some information technologies may also support decentralisation and hierarchical flattening ([Dewett & Jones 2001](#); [Wamba et al. 2015](#)).

Algorithms, as part of a socio-technical system, are embedded in specific social contexts, being used by specific actors, applied to specific domains, and serve specific purposes ([Wenzelburger et al. 2022](#)). Thus, algorithm-induced changes in bureaucratic organisations are not merely generated by technology. The transformative power of technology is also influenced by the organisational context.

Essentially, an organisation is a control-oriented action system. Power lies at the core of an organisation's operations and relationships, and restructuring these power relationships is fundamental to creating organisational change ([Crozier & Friedberg 2007](#)). Power relations have a crucial role in controlling information flow within an organisation, shaping incentives, modifying organisational structures and forms, and determining the service tools used. This chapter analyses the ways in which algorithms can facilitate bureaucratic transformation in the dimensions of power relations, incentive mechanisms, knowledge and information exchange,

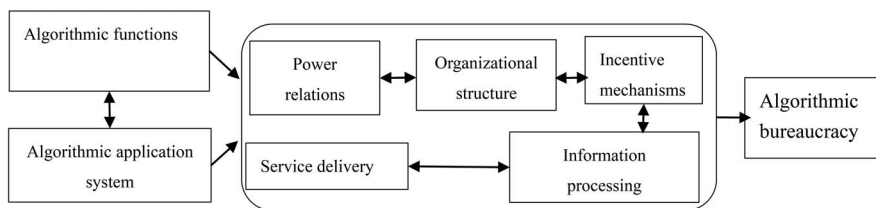


Figure 10.1 Algorithmic functions (the automation systems and the predictive decision), algorithmic application system (departmental algorithmic applications and cross-departmental algorithmic applications), and algorithmic bureaucracy.

Source: Authors' own work.

organisational structure, and service delivery. We propose a theoretical framework to analyse algorithmic bureaucracies in terms of two algorithmic functions and two application regimes ([Figure 10.1](#)). The following three sections will illustrate this analytical framework further with briefly mentioned examples.

Algorithmic Functions and the Bureaucracy

Algorithms are widely used in the public sector, from big data mining and processing to automation and predictive analytics. The algorithms needed for a task depend on its nature and the precise problem to be solved; therefore, it is necessary to classify algorithms according to their given objectives. The various functions of AI algorithms include classification, summarisation, text generation, detection, and image generation, using the methods of linear regression, logistic regression, and decision trees. The public sector may use the same algorithmic techniques for different domains and purposes or use various techniques for the exact domains to produce a similar function. Nevertheless, the dual categorisation of algorithmic functions (automation and prediction) is relatively simple and critical and thus can help us to grasp the impact of algorithms.

According to [Veale & Brass \(2019\)](#), two main types of algorithmic functions are automation systems and predictive decision-support systems. The functions and scenarios of algorithmic application are related to the level of administrative tasks. When the administrative tasks are simple and there is little room for bureaucratic discretion, automated algorithmic applications have the potential to replace human decision-making ([Bullock et al. 2020](#)). Automation is primarily related to explicit instructions and expert-driven programmes, and its processes and outcomes are fundamentally determined by human-designed rules ([Hurley & Wallace 1986](#)). Automated systems in the public sector can be used to quickly and efficiently process repetitive and standardised tasks, including application processing, image recognition, and data entry ([Wirtz et al. 2019](#)), thus increasing implementation efficiency, reducing human intervention, and improving fairness through automated rule implementation. Furthermore, AI is generally capable of automating simple routine and even complicated tasks ([Ford 2015](#)) and is good at

solving problems that have clear boundaries and structures (modularity or decomposability). Therefore, algorithmic automation is more likely to be a replacement for routine, complex, and repetitive tasks.

The application of automated algorithms is unsuitable if administrative tasks tend to be more complex and discretionary. Thus, human decision-making prevails. In this context, predictive analytics can serve as an important type of decision-support system (Bullock 2019). Algorithmic predictive analytics incorporate the use of massive amounts of data to make and validate predictions in complex environments, identify patterns and laws, and improve accuracy and predictability in decision-making. The use of predictive analytics in decision support systems allows policymakers to predict the emergence of problems, enabling early warning to allow for targeted policy actions to be taken, such as early crime warning, traffic safety warning, and environmental disaster prediction (Wirtz et al. 2019).

Automation and predictive analytics are the dominant algorithmic functions in the public sector at the present time. These two algorithmic functions are impacting traditional decision-making rules in bureaucratic organisations and affecting internal control and incentive mechanisms, knowledge and information flows, organisational structures, and service delivery. First, algorithms can strengthen or weaken human power and control in the bureaucracy. Automation may weaken bureaucratic discretion. Rule-based orientations that strictly control decision-making may further limit discretion, flexibility, and convenience in organisational staff (de Boer & Raaphorst 2023). To ensure the efficiency and normality of execution, organisations that use automated algorithms require a greater degree of standardisation of data, processes, and services, including greater emphasis on the top-down consistency of organisational actions and more pronounced power constraints and interventions. Face-to-face interactions and feedback gradually disappear, and street-level bureaucrats gradually give way to digital bureaucrats in the form of automated algorithmic application scenarios that emphasise digital control and consistency (Busch & Henriksen 2018). Here, the organisational incentives are oriented to efficiency and results, and this may lead bureaucrats to prioritise the speed and quantity of automated processing. This may ultimately weaken responses to ethical commitments and responsibilities towards the public (Hansen, Lundberg, & Syltevik 2018).

In contrast, predictive algorithmic models are complex and not easily interpretable, but they are highly customisable, providing flexible and diverse sources of information and processing. This makes it challenging for higher-level authorities to ensure uniformity and standardisation for algorithm operations from top to bottom. The predictive applications of algorithms enable organisations to better understand the contextual, autonomous, and adaptive nature of the use of algorithms in departments and the actions taken by the officials in them. This, in turn, promotes resource sharing across departments, citizen engagement, and co-production (Androutsopoulou et al. 2019). Hereby predictive algorithm provides organisations and their employees with strong incentives to adopt more flexible approaches and tools to deliver more responsive and targeted services.

Second, different algorithmic functions cope with organisational knowledge and information in different ways. Automated algorithms that adhere to conventional bureaucratic structures entail lower demand for cross-departmental data sharing and bottom-up data feedback. Here, routine tasks that comply with simple rules do not require extensive diversity and breadth in data sources. Automated applications may encourage departments to create incomplete data-sharing partnerships, meaning that one department may not immediately convey data to other departments employing the algorithm, but instead only provide the data interface. The algorithmic application department can obtain data verification and comparison via the data interface. Whereas predictive analytics algorithms can transform organisational knowledge and information processing through shared, feedback-based information transfer. The development of predictive analytic models involves dynamic information processing that adopts multi-source training datasets for pattern recognition and predictive judgements. Predictive analytical models guide enforcement behaviour on the ground along with continuous acquisition of new information and data to optimise models. For example, street-level bureaucrats need to collect environmental and contextual information during their daily work, a role that is often neglected but is essential for training predictive analytics models. The information that is collected should be made available in real time to decision-makers for its inclusion in the dataset.

Third, algorithmic functions can show considerable variation in terms of their impact on the delivery of public services. Automation reinforces the procedural and rule-based orientation of public service provision. That is, they can accelerate, streamline, and increase the efficiency and transparency of rule implementation rather than replacing bureaucratic decision-making rules. Predictive analytics algorithms convert the traditional standardised service provision into a customised and precise service provision that accurately identifies specific tasks and service requirements.

Algorithmic Application System and Algorithmic Bureaucracy

Two main application systems of algorithms appear in bureaucratic organisations. The first is decentralised individual departments with the aim of improving administrative efficiency and service provision, leading to departmental algorithmic applications on bureaucracy. The other is centralised and cross-departmental to address inefficiencies and coordination issues in public services that result from fragmented government responsibilities and departmental power structures (Meijer et al. 2021), leading to centralised cross-departmental algorithmic applications. Departmental algorithmic applications take place because departmental fragmentation and information barriers enable departmentalised algorithmic application systems to survive and evolve. However, algorithms by their nature require information sharing and data collaboration, and algorithmic applications will always attempt to expand their boundaries to develop centralised cross-departmental applications.

Departmental or cross-departmental algorithmic applications on bureaucracy can alter power dynamics and incentives, both inside and outside the organisation differently. Departmentalised algorithmic applications motivate departments to establish mechanisms for data sharing and organisational collaboration within

the department to improve algorithmic performance, which can also lead to a better allocation of organisational resources and a flattening of organisational structures. Contrary, algorithmic applications across departments exhibit a clearly centralised nature, as reflected in top-down data aggregation, the integration of digital systems and smart platforms, and the coordination of interdepartmental activities. Incentives for organisational behaviour are significantly impacted by centralised digital governance. Interdepartmental relations have evolved from loose, occasional, and informal forms of cooperation to concerted, procedural, and institutionalised action. This shift enables effective collaboration between a responsible department which employs the algorithm and its auxiliary co-departments. In addition, cross-departmental algorithmic applications establish independent digital assessment mechanisms, which utilise algorithmic technology to achieve precise, visualised control and supervision over each department. However, while the application of cross-departmental algorithms improves the coordination and wholeness of governance, it can generate conflicts of interest over digital ownership and raise transaction costs (Mikhaylov et al. 2018). In addition, grass-roots departments that “receive orders” from above may weaken initiative and capability, leading to “distorted incentives” or “weak incentives” (Giest & Klievink 2024).

The most obvious impact of an algorithmic application system on the bureaucracy is the reconfiguration of organisational structures. The use of a departmentalised algorithm implies that each department creates a customised algorithmic application to fulfill its specific requirements, according to the type and characteristics of the departmental business. Departments establish internal feedback channels or cross-departmental cooperation to optimise their algorithms. Departmentalised algorithms promote incentives for interdepartmental collaboration while also preserving the independence and vertical hierarchy of the organisation (Figure 10.2). The organisation must create formal and informal mechanisms for collaboration to meet the algorithmic requirements for data collaboration and information sharing (Fusi 2021).

Cross-departmental applications of algorithms involve multi-departmental collaboration in the provision of algorithmic solutions for public affairs, which is achieved through centralised or cross-sectoral digital platforms that integrate different algorithmic application scenarios. Cross-departmental algorithmic applications have altered traditional inter-organisational relationships and promoted organisational reengineering. Significant problems arise with regards to interdepartmental data sharing, system compatibility, and coordination in traditional bureaucracy. Whereas cross-departmental algorithmic application encourages the creation of centralised algorithmic governance organisations and platforms to promote both functional integration and cross-departmental linkages. To some extent, this contributes to the development of a holistic government (Dunleavy 2006; Gao, Song & Zhu 2013). To ensure that cross-departmental algorithms are used effectively, organisations transform their information flow and processes of knowledge application and establish a unified big data platform for data sharing and cooperation.

Cross-departmental algorithms in local governments generally enhance organisational linkages through a model of order dispatching. This model does not

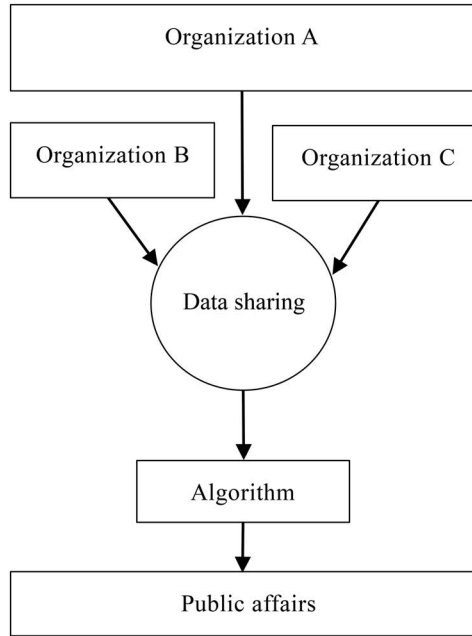


Figure 10.2 Impact of the departmentalised algorithmic system on bureaucracy.

Source: Authors' own work.

alter the essential functions for each organisation. However, the authority employing a cross-departmental algorithm should send orders to other departments for handling tasks that are needed to meet cross-departmental coordination. This approach employs real-time data feedback, robust data processing, and analytics to provide solutions, thereby achieving the goals of connectivity between different departments (see [Figure 10.3](#)). As local governments in China focus on the use of big data and AI to build smart cities, they are not satisfied with the departmentalised algorithmic system. Thus, the governments are seeking some cross-departmental algorithm applications that can serve as unified platforms for operating and managing smart cities. The platform is usually known as the City Brain in the city of Hangzhou ([Caprotti & Liu 2022](#)), and Smart City Command Centre in many other Chinese cities.

Finally, the way that knowledge and information are processed within traditional bureaucracy varies with departmental and cross-departmental applications of algorithms. Some departments may create their own algorithms and exchange data with other departments. Departments are more likely to create informal or one-off approaches to data collaboration, where they coordinate frequently but the results are highly uncertain and unsustainable ([Weitzman et al. 2006](#)). By contrast, cross-departmental algorithmic applications rely on data aggregation as the primary method of information processing. Data from various departments are pooled

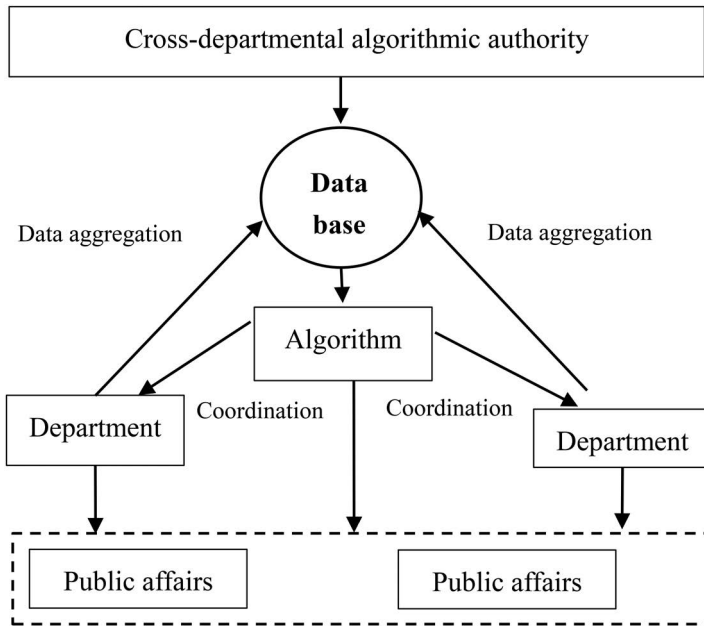


Figure 10.3 Impact of cross-departmental algorithms on bureaucracy.

Source: Authors' own work.

into an independent agency or a platform for real-time aggregation, classification, and analysis. Most cross-departmental algorithms are integrated into the local government's data management bureau or big data centre, establishing a fundamental database and categorising its data based on topics.

Four Models of Algorithmic Bureaucracy

Combining the automation and prediction functions with departmental and cross-departmental applications, we identify four distinct forms of algorithmic bureaucracy. These four models are Weberian ideal types to describe and categorise the recent algorithmic change to bureaucracy (Table 10.1). We acknowledge that some models overlap with blurred boundaries, but they are valuable for classifying the complex and different algorithmic applications in bureaucratic organisations.

The Departmental Automated Algorithmic Bureaucracy

In the departmental automated algorithmic bureaucracy, automation is effective in minimising administrative workload, reducing latitude for discretion, and reinforcing process control and standardisation. This type of algorithmic bureaucracy is exemplified by a smart enterprise registration system, i.e., the fast-track approval system (*"Miaopi,"* literally *"approved within a second"*) in the city of Shenzhen, China.

Table 10.1 Four forms of algorithmic bureaucracy

<i>Algorithmic functions</i>			
<i>Automation</i>		<i>Predictive analysis</i>	
I. Departmental Automatic Algorithmic Bureaucracy		II. Departmental Predictive Algorithmic Bureaucracy	
Algorithmic application system	Departmentalisation	Power relations: reinforcing; reducing discretion Incentives: relieving workload; diverting attention Information processing: data interfacing Service provision: automated rule enforcement Organisational form: open and flexible Case: fast-track approval system	Power relations: power alignment; upwards and downwards synergy Incentives: resource sharing; autonomy Information processing: data sharing Service provision: precision Organisational form: cross-departmental cooperation Case: complaints and proposals forecasting system
		III. Cross-departmental Automated Algorithmic Bureaucracy	IV. Cross-departmental Predictive Algorithmic Bureaucracy
	Cross-departmentalisation	Power relations: reinforcing control Incentives: relieving workload; digital assessment Information processing: data aggregation Service provision: automated rule enforcement Organisational form: cross-departmental and cross-level platform Case: “Internet Supervision” platform	Power relations: upwards and downwards synergy Motivation: autonomy; digital assessment Service provision: precision Information processing: data aggregation Service Provision: precision Organisational form: cross-departmental and cross-level platform Case: Urban Governance Command Centre

Source: Author’s own work.

Shenzhen Administration for Market Regulation has developed a fast-track approval system since 2019. The level of interdepartmental data sharing is moderate or low, and it primarily involves data matching and checking, including an enterprise name database to check legal business names, a unified address database to exclude false addresses, a real-name identification database to prevent registration under false pretenses, and a defaulter’s database for manual approval. These databases provide a variety of rules for automated approval. After an individual applies for business registration online, the information and related materials submitted are reviewed and checked by the pre-designed system.

First, the most significant change in organisational power relations brought about by the fast-track approval system is the tightened control over bureaucratic discretion. The automated approval system reduced the uncertainty and discretion of human approval, which not only reduced a large number of repetitive and procedural works but also strengthened approval standardisation, contributing to the transparency and fairness of the approval process. The main goal is to automate basic tasks that were once performed by humans, thereby increasing efficiency and fairness in the rule execution.

Second, the fast-track approval system also altered organisational information flow. The algorithm in the system entails an automatic review of user-entered information, followed by an automatic comparison through data connections with each department. The key to automated approval is knowledge sharing and data connectivity across departments, eliminating departmental information segmentation.

Third, the fast-track approval system significantly impacted the department's organisational structure and processes, leading to a transition from sequential to parallel approval processes. The organisational structure retains its hierarchical relationship, but the process of inter-organisational coordination and interaction is smoothed.

Finally, the restrictions on approval authority and the transformation of the approval process through automation have also changed organisational incentives to some extent. The approval process tends to be "easier and faster," reducing the workload of approving staff while weakening gatekeeping responsibilities. With the increased reliance on algorithmic systems, speed, and reduced human intervention have been used as an excuse to overlook security issues.

Departmental Predictive Algorithm Bureaucracy

Departmental predictive algorithm bureaucracy has a significant impact on power relations, incentive structure, information processing, service provision, and forms of organisation. First, algorithms in the form of prediction are not merely a new technology but also provide a new source of power (Bucher 2018). Predictive algorithms used in departments can drive organisational power alignments. As important decision-making systems and tools for controlling information, algorithms are changing and enhancing the ways that departments influence or produce decisions, as well as expanding the roles and positions of departments in power structures. Predictive algorithms facilitate collaboration between upper and lower-level departments within an organisation, thus likely empowering grass-roots staff autonomy, encouraging resource sharing, and increasing organisational flexibility. Second, while algorithms have not fundamentally changed bureaucratic structures, the need for data sharing has made organisations more focused on cross-sector collaboration and bottom-up information communication and feedback. Data sharing is becoming increasingly prevalent in information processing. The adoption of predictive analytics in organisations has led to greater emphasis on the need for comprehensive and institutionalised data-sharing mechanisms.

The departmental predictive algorithm bureaucracy can be illustrated by the Complaints and Petition Forecasting System (CPFS) in Beijing, China. The CPFS is a predictive algorithm developed by the Beijing Municipal Public Complaints Office (Complaints Office). Prior to its creation, the public complaints office predicted local petition trends through basic analysis of past petition data without fully utilising it for decision-making. The widespread use of the internet has made online petitioning the primary channel. As a result, the department is required to shift its focus towards the analysis of online data and employ the data to make decisions. In 2018, the Beijing Municipal Public Complaints Office launched the big data smart analysis platform of public complaints.

First, the Complaints Office's power status and influence have greatly increased within the bureaucratic system as a result of the CPFS. The Complaints Office was usually in a weak and marginalised position within the bureaucratic power structure. However, the introduction of CPFS presents a promising opportunity for the department to enhance its influence. The department has evolved from a mere platform for submitting complaints and proposals to a vital hub for data analysis, decision support, and resolution of social conflict.

Second, the CPFS algorithm positively impacted communication between organisational hierarchies and improved data sharing and cooperation both within and outside the organisation. The participation of grassroots-level departments was invaluable in selecting predictive indicators and determining their weights for the CPFS algorithm. Petition trend and key petitioner information are routinely provided to algorithm designers. Grassroots reports regularly provide information on actual petition behaviours to verify predicted results. If there is a discrepancy between the actual information and model predictions, the Complaints Office and its technical experts then adjust the indicators' weights and the model's parameters. Such a role is essential for training predictive analytics models. In real life, controlling data is vital for street-level bureaucrats, leading to incomplete or outdated data collection and sharing.

Finally, the CPFS has greatly enhanced the effectiveness and accuracy of handling public complaints and petitions. Through machine learning algorithms and big data analysis technology, CPFS monitors and predicts real-time trends in letter and visit activities, enabling quick responsiveness and proactive identification of potential extreme petition events. With the introduction of CPFS, the petition monitoring and early-warning system has transformed from a random, disorderly, and retrospective mode to a forward-looking response approach with accurate prediction, timely monitoring, and proactive identification.

Cross-departmental Automated Algorithmic Bureaucracy

Centralised cross-departmental automated algorithmic bureaucracy is to establish a cross-departmental algorithmic institution or platform to reduce administrative burdens and reinforce power control and digital assessment in organisations. Information flow is processed based on data aggregation, and service delivery is automated. Cross-departmental automated algorithms are frequently used for welfare

fraud detection by many countries. A typical case is the “Internet Supervision” platform in Hunan Province, China.

The “Internet Supervision” platform in Hunan Province represents a typical cross-departmental automated algorithmic bureaucracy. Its development was motivated by the necessity to address the issue of rural finance corruption, such as the misappropriation of poverty alleviation funds and preferential treatment of friends and family members of rural officials. Hunan Province introduced the “Internet Supervision” platform across the province, cities, and counties in 2017. This platform was not implemented by the provincial government from the top down, but it was a self-initiated innovation by several city and county governments that attracted the attention of the provincial government and expanded to cover the entire province.

First, the “Internet Supervision” platform operates on a unified platform with unified standards, implementation, and three-tiered operation and maintenance, reflecting the centralised and control-oriented nature of cross-departmental algorithms. The establishment of the platform is mainly due to the strong political support from the provincial government. The province party committee and government created the “Internet Supervision” leading group in April 2017. The Provincial Commission for Discipline Inspection coordinated with other departments to develop software, collect data, and construct the system of the platform.

Second, the top-down application system of the platform has facilitated the removal of data barriers, promoting the integration and convergence of data from different departments. China’s discipline inspection departments have unparalleled authority and powerful data integration capabilities. For example, to deal with the problem of public officials’ refusing to provide personal information, the discipline inspection departments have taken measures by persuasion or signing formal confidentiality agreements with them.

The “Internet Supervision” platform integrates individuals’ welfare data from more than a dozen departments and creates two primary matching databases, i.e., a personnel information database and a welfare fund database. The automated matching algorithm is driven by the exclusion rules based on the two databases. The welfare funding policy has established strict eligibility requirements. The algorithm analyses and compares databases to identify violations of rules that arise when two types of information coexist. The system identifies problematic evidence proactively, issues early warnings, and automatically forwards it to the discipline inspection departments.

Finally, the “Internet Supervision” platform has significantly transformed the organisational processes and structures of discipline inspection departments. This platform has redesigned the traditional method of discipline inspection and established a supervision process that allows public complaints to be assigned by the platform to relevant departments for further investigation. The platform utilises algorithmic technology to combine the discipline inspection department, function departments, and the public, resulting in a supervisory system that automates supervision processes and incorporates public participation. By implementing an automated complaints processing system with defined deadlines for handling and

response, as well as supervision by the Discipline Inspection departments, the effectiveness of disciplinary inspections has been enhanced.

Cross-departmental Predictive Algorithmic Bureaucracy

Cross-departmental predictive algorithmic bureaucracy combines the centralised features of cross-departmental algorithms with predictive algorithms. It shapes power dynamics by means of inter-organisational collaboration while emphasising autonomy within departments. This algorithmic bureaucracy emphasises accuracy and customisation in service delivery with a variety of algorithms, including automated algorithmic functions but mainly relying on predictive algorithms.

Urban Governance Command Centre in the City of Nantong, Jiangsu province, is partly characterised by predictive algorithmic bureaucracy due to its integration of data and the wide range of predictive algorithmic applications. To achieve efficient smart city governance, the City of Nantong has implemented significant changes to its digital governmental structure. It has formed an independent interdepartmental organisation called the “Urban Governance Command Centre,” responsible for algorithmic applications in diverse urban governance scenarios. There are seven functions in the Urban Governance Command Centre, including data sharing, intelligent search, unified supervision, monitoring and early warning, analysis and judgement, coordination and command, and administrative accountability, all of which are based on predictive analysis algorithms. At the heart of the Urban Governance Command Centre is a monitoring and early warning system. This system employs an AI predictive analysis model that utilises diverse forms of big data to anticipate and identify multiple social risks proactively.

First, the case of the Urban Governance Command Centre illustrates how algorithms impact the restructuring of organisational collaboration. On the one hand, the command centre plays a key role in facilitating horizontal communication and coordination between departments in the social governance management system. The command centre enhances coordination and connectivity by assigning tasks to the departments. To promote interdepartmental cooperation and collaboration, the command centre encourages departments to submit updated data through formal monitoring and evaluation. On the other hand, the command centre has also established a vertical command system. It has a four-level linking mechanism, namely a city-level command centre, 10 county or district-level command centres, 96 township street command centres, and more than 70,000 full-time and part-time grid members. The city-level command centre is responsible for monitoring the overall situation. The county or district-level command centre provides practical command. The township-level centres instruct the grid members to deal with tasks assigned, as well as performing patrols, visiting sites, assessing public sentiment, drafting reports, and subsequently uploading them to the command centre database. Grid members form the cornerstone of the grass-roots governance structure (Mittelstaedt 2022).

Second, in terms of knowledge and data sharing, the Centre’s most significant attribute is the pooling and management of functional departmental data. It collects information from different departments within the jurisdiction and other real-time

data to create various standardised thematic databases, which are necessary for algorithmic governance. The integration of functions between the Command Center and Big Data Bureau enables easier sharing of data with other administrative departments. Nantong City has established a government data sharing and exchange platform in the Big Data Administration (command centre). According to the *Implementing Rules for Government Information Resource Sharing of Nantong City*, the Big Data Management Bureau is responsible for guiding, coordinating, and supervising the government data sharing and exchange.

Finally, regarding service delivery, the command centre has significantly improved the accuracy of urban governance. The trick is to create algorithmic application scenarios for several topics with a variety of databases. The command centre centralises and thematically integrates big data about public safety, economic conditions, land resources, and public services. More than a dozen algorithmic scenarios, such as the overall monitoring of hazardous chemicals and the regulation of group rental housing, have been developed based on the thematic database. These applications help to improve the precision of social governance. Taking the precise supervision of hazardous chemicals as an example, the command centre maintains a comprehensive list of hazardous chemical enterprises and obtains real-time trajectory data of transportation vehicles. Therefore, it can estimate transportation vehicles' possible locations through a spatial-temporal AI engine algorithm. This allows the identification of potential hazards related to the dumping of hazardous chemicals.

Conclusion

The Fourth Industrial Revolution, driven by big data, smart technology, and AI, has ushered in a new era of human intelligence that is exerting a profound influence on the traditional bureaucracy and triggering major changes in the mode of public management. The algorithmic bureaucracy revolution is taking place quietly, with different characteristics and tendencies in different political and social contexts.

Based on the two primary factors of algorithmic function and systems of algorithm application, we develop a meso-level analytical framework for algorithmic bureaucracy and classify four models of algorithmic bureaucracy: decentralised departmental automated algorithmic bureaucracy, decentralised departmental predictive algorithmic bureaucracy, centralised cross-departmental automated algorithmic bureaucracy, and centralised cross-departmental predictive algorithmic bureaucracy. We explore the impact of algorithms on control and incentive mechanisms, organisational structure, knowledge and information processing, and service supply of bureaucracy.

The centralised and decentralised algorithmic application systems have altered internal power control and incentive structures within organisations, the process of organisational knowledge and information processing, and, consequently, the organisational structure and service delivery model in different ways. When the control from a higher-level governmental department is weak, decentralised departmentalised algorithmic applications promote departmental autonomy. Departments with the necessary capacity and decision-making authority can

autonomously create algorithms to develop problem-solving approaches. Decentralised departmentalised algorithmic applications increase the likelihood that departments can adopt bottom-up communication and cross-organisational collaboration. Conversely, cross-departmental algorithmic applications favour the establishment of centralised algorithm governance entities responsible for cross-departmental data aggregation and algorithmic applications.

Algorithmic functions have a subtle impact on organisational internal control and incentives. Automated applications are becoming increasingly utilised to complement traditional decision-making rules in the public sector. Instead of replacing hierarchical decision-making, these applications reinforce a rule and control orientation for public service provision, facilitating swift and efficient rule implementation. While automated algorithms certainly simplify organisational tasks, they reduce and limit bureaucratic discretion.

The application of predictive analytics has transformed procedural and standardised service provision, allowing organisations to concentrate on the precise identification of service recipients' needs and the customisation of their services. In addition, predictive algorithms prompt organisations to move from a culture of closure and competition to one of openness and collaboration, which requires bottom-up information sharing. Algorithmic bureaucracy inevitably triggers a revolution in public management that significantly challenges and revises Weber's theory of bureaucracy.

Using four cases of AI algorithmic applications in Chinese local governments, this chapter reveals the various impacts of algorithms on bureaucracy. Algorithms help to alleviate chronic problems of traditional bureaucracy, such as inefficiency, departmental segregation, and information barriers. They enhance the capacity and openness of bureaucracy and ultimately improve the performance of public services. However, they suffer a common serious problem: the absence of citizens' participation and monitoring.

Future study needs to monitor the new applications of centralised algorithmic bureaucracy to revise and improve the above-discussed analytical models of algorithmic bureaucracy. Significantly, algorithmic application systems across departments are becoming increasingly a trend in China. *The Guideline for the Construction of a National Integrated Big Data System for Government Affairs*, issued by the General Office of the State Council in 2022, requires full integration of government data resources and public digital platforms. It called for a national big data platform for government affairs with the national integration and data synchronisation of local platforms. These new developments provide us with a new ground to reflect and revise our theoretical understanding of algorithmic bureaucracy.

References

- Androutsopoulou, A., N. Karacapilidis, E. Loukis, and Y. Charalabidis. "Transforming the Communication between Citizens and Government through AI-Guided Chatbots." *Government Information Quarterly* 36, no. 2 (2019): 358–367.
- Bovens, M., and S. Zouridis. "From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control." *Public Administration Review* 62, no. 2 (2002): 174–184.

- Bucher, T. *If... Then: Algorithmic Power and Politics*. Oxford: Oxford University Press, 2018.
- Bullock, J. "Artificial Intelligence, Discretion, and Bureaucracy." *The American Review of Public Administration* 49, no. 7 (2019): 751–761.
- Bullock, J., M. M. Young, and Y. F. Wang. "Artificial Intelligence, Bureaucratic Form, and Discretion in Public Service." *Information Polity* 25, no. 4 (2020): 491–506.
- Busch, P. A., and H. Z. Henriksen. "Digital Discretion: A Systematic Literature Review of ICT and Street-Level Discretion." *Information Polity* 23, no. 1 (2018): 3–28.
- Caprotti, F., and D. Liu. "Platform Urbanism and the Chinese Smart City: the Co-Production and Territorialisation of Hangzhou City Brain." *GeoJournal* 87, no. 3 (2022): 1559–1573.
- Cormen, T. H. *Algorithms Unlocked*. Cambridge, MA: MIT Press, 2013.
- Crozier, M., and E. Friedberg. "Organizations and Collective Action: Our Contribution to Organizational Analysis." In *Research in the Sociology of Organizations Vol. XIII*, edited by S.B. Bacharach, P. Gagliardi & B. Mundell, 71–92. Greenwich: JAI Press, 1995.
- de Boer, N., and N. Raaphorst. "Automation and Discretion: Explaining the Effect of Automation on How Street-Level Bureaucrats Enforce." *Public Management Review* 25, no. 1 (2023): 42–62.
- Dewett, U., and G. R. Jones. "The Role of Information Technology in the Organization: A Review, Model, and Assessment." *Journal of Management* 27, no. 3 (2001): 313–346.
- Dunleavy, P. *Digital Era Governance: It Corporations, The State, and E-Government*. Oxford: Oxford University Press, 2006.
- Farrell, Henry, Abraham Newman, and Jeremy Wallace. "Spirals of Delusion: How AI Distorts Decision-Making and Makes Dictators More Dangerous." *Foreign Affairs* 101 (2022): 168.
- Ford, M. *Rise of the Robots: Technology and the Threat of a Jobless Future*. New York: Basic Books, 2015.
- Fountain, J. E. *Building the Virtual State: Information Technology and Institutional Change*. Washington: Brookings Institution Press, 2004.
- Fusi, F. "When Local Governments Request Access to Data: Power and Coordination Mechanisms Across Stakeholders." *Public Administration Review* 81, no. 1 (2021): 23–37.
- Gao, X., Y. Song, and X. Zhu. "Integration and Coordination: Advancing China's Fragmented E-Government to Holistic Governance." *Government Information Quarterly* 30, no. 2 (2013): 173–181.
- Giest, S. N., and B. Klievink. "More Than a Digital System: How AI Is Changing the Role of Bureaucrats in Different Organizational Contexts." *Public Management Review* 26, no. 2 (2024): 379–398.
- Hansen, H. T., K. Lundberg, and L. J. Syltevik. "Digitalization, Street-Level Bureaucracy and Welfare Users' Experiences." *Social Policy & Administration* 52, no. 1 (2018): 67–90.
- Hoffman, Samantha. "China's Tech-Enhanced Authoritarianism." *Journal of Democracy* 33, no. 2 (2022): 76–89.
- Hurley, M. W., and W. A. Wallace. "Expert Systems as Decision Aids for Public Managers: An Assessment of the Technology and Prototyping as a Design Strategy." *Public Administration Review* 46, (1986): 563–571.
- Margetts, H. *Information Technology in Government: Britain and America*. New York: Routledge, 2012.
- Meijer, A., L. Lorenz, and M. Wessels. "Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems." *Public Administration Review* 81, no. 5 (2021): 837–846.
- Mikhaylov, S. J., M. Esteve, and A. Campion. "Artificial Intelligence for the Public Sector: Opportunities and Challenges of Cross-Sector Collaboration." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2128 (2018): 1–21.
- Mittelstaedt, J. C. "The Grid Management System in Contemporary China: Grass-Roots Governance in Social Surveillance and Service Provision." *China Information* 36, no. 1 (2022): 3–22.

- Orlikowski, W. J. "Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations." *Organization Science* 11, no. 4 (2000): 404–428.
- Polyakova, A., and C. Meserole. "Exporting Digital Authoritarianism: The Russian and Chinese Models." In *Policy Brief, Democracy and Disorder Series*, 1–22 (2019). Washington: Brookings Institution Press.
- Strittmatter, K. *We Have Been Harmonized: Life in China's Surveillance State*. New York: HarperCollins, 2020.
- Veale, M., and I. Brass. "Administration by Algorithm? Public Management Meets Public Sector Machine Learning." In *Algorithmic Regulation*, edited by Karen Yeung, and Martin Lodge, 121–149. Oxford: Oxford University Press, 2019. <https://academic.oup.com/book/35048/chapter/298946506>.
- Vogl, T. M., C. Seidelin, B. Ganesh, and J. Bright. "Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities." *Public Administration Review* 80, no. 6 (2020): 946–961.
- Wamba, P. F., S. Akter, A. Edwards, G. Chopin, and D. Gnanzou. "How 'Big Data' Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study." *International Journal of Production Economics* 165 (2015): 234–246.
- Weitzman, C. C., D. Silver, and C. Brazill. "Efforts to Improve Public Policy and Programs Through Data Practice: Experiences in 15 Distressed American Cities." *Public Administration Review* 66, no. 3 (2006): 386–399.
- Wenzelburger, G., P. D. König, J. Felfeli, and A. Achtziger. "Algorithms in the Public Sector. Why Context Matters." *Public Administration* 102, no. 1 (2022): 40–60.
- Wirtz, B. W., J. C. Weyerer, and C. Geyer. "Artificial Intelligence and the Public Sector—Applications and Challenges." *International Journal of Public Administration* 42, no. 7 (2019): 596–615.

Part V

Critical Theory of Artificial Intelligence



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

11 Social Inequalities and Artificial Intelligence

How Digital Inequality Scholarship Enhances Our Understanding

Christoph Lutz

Introduction

Global inequalities¹ have risen in recent decades ([Amis et al. 2018](#)). The world's second richest person, Jeff Bezos, offers a striking example: his space excursion and vast investments into his rocket company Blue Origin have been widely criticised when many are suffering from mental health issues or economic hardship ([Jones 2021](#)). This includes Amazon warehouse workers, who earn very little while being subject to extreme work pressures and relentless surveillance ([Barwise and Watkins 2018](#); [Delfanti and Frey 2021](#)). Such precarity also extends to contract employees and independent platform workers. Both groups are extremely prevalent among tech companies, e.g., in artificial intelligence (AI) development ([Newlands 2021b](#)), but command “lower wages, less respect and recognition, and no job security” compared with regular employees ([Bapuji, Ertug, and Shaw 2020](#), 68).

An increasingly pressing aspect of inequality relates to the role of digital technology in society. Recent years and decades have seen a rapid surge in digitalisation, including the advent of social media ([Kaplan and Haenlein 2010](#)), big data ([George, Haas, and Pentland 2014](#)), and AI ([Raisch and Krakowski 2021](#)). The role of digital technology in reflecting and shaping inequalities has been studied within the field of digital inequalities ([Robinson et al. 2015](#); [Lutz 2019](#); [Van Dijk 2020](#)), an established research stream in sociology, communication, and information systems. Often using the umbrella term of the “digital divide”, this research investigates how individuals in different social categories vary in their uptake of digital technology, including aspects of access, skills, use, and benefits or harms from use.

The chapter shows how digital inequalities scholarship can be fruitfully applied to sophisticated AI systems such as large language models (LLMs) or automated decision-making systems. I will adopt the three levels of the digital divide distinction ([Lutz 2019](#)) and take inspiration from [Van Dijk's \(2005, 2020\)](#) resources and appropriation theory. The chapter focuses on organisational contexts, both on the intraorganisational and interorganisational level, as organisations are key reproduction sites of inequalities, while also acting as a main locus where AI systems are developed and employed. Specifically, intraorganisational and interorganisational dynamics will be distinguished. The chapter first briefly introduces the three levels of the digital divide. It then discusses inequality implications of AI for each of

the three digital divide levels, differentiating intraorganisational and interorganisational dynamics. Finally, it concludes with a brief summary and outlook.

Digital Inequalities as a Research Field

Origins and the First-Level Digital Divide

Digital inequalities is an established research topic within communication and sociology, concerned with studying how different individuals and groups engage differently with digital technology. The field of digital inequalities research has its roots in earlier research on knowledge gaps in mass media consumption (Tichenor, Donohue, and Olien 1970) and began to take shape in the late 1990s, particularly with the publication of the “Falling through the Net” reports by NTIA in 1995 and 1999 (NTIA 1995, 1999). At that point, Internet adoption was accelerating quickly within Western countries, but vast inequalities in Internet access were detected based on social categories such as gender, age, ethnicity, socio-economic status, and area of residence (Van Dijk 2006). Inequalities in Internet access – and access to digital technology more broadly – are now commonly referred to as the *first-level digital divide* (Hargittai 2002; Lutz 2019).

The Second-Level Digital Divide

The field began to expand in the early 2000s and moved beyond a narrow interest in access, especially beyond dichotomous notions of access vs. non-access. Instead, more complex and dynamic inequalities were explored. Hargittai (2002) called for the study of a “second-level digital divide”, describing inequalities in Internet skills and uses (Lutz 2019). In line with work on the first-level digital divide, research found that social categories such as gender, age, education, and disability matter. Hargittai and Shafer (2006), e.g., showed that women’s digital skills are similar to those of men, but men are more confident in their abilities by revealing higher levels of self-assessed skills. Gender not only matters for skills but also for uses. Certain social media platforms are more heavily used by women (e.g., Facebook), while others are favoured by men (e.g., LinkedIn; Blank and Lutz 2017). The role of age was also studied, showing how younger people use the Internet more frequently than older ones (Blank and Groselj 2014). Status and class further influence if and how individuals use the Internet for certain purposes, such as e-banking, information search, gaming, and job search (Zillien and Hargittai 2009; Schradie 2011). The notion of “capital-enhancing activities” was introduced to describe online activities that are more beneficial and more likely to lead to upwards mobility than non-capital-enhancing activities (DiMaggio et al. 2004). The former include information-oriented and participatory Internet uses such as news consumption and blogging, while the latter refer to entertainment- and consumption-oriented Internet uses such as gaming and streaming. However, the notion of capital-enhancing activities has been criticised (Boonaert and Vettenburg 2011), since activities traditionally deemed to be non-capital-enhancing can be beneficial (e.g., learning-oriented games and YouTube

tutorials), whereas capital-enhancing uses can be detrimental (e.g., consumption of low-quality news). The interest in variegated uses also brought the field's attention to questions of representation, social context, and the embeddedness of Internet uses into broader lifestyles (Halford and Savage 2010; Lutz 2016).

The Third-Level Digital Divide

Since about the mid-2010s, interest in a third-level digital divide has emerged. It describes how people from different backgrounds benefit or suffer from Internet use (Van Deursen and Helsper 2015; Blank and Lutz 2018). Thus, it refers to outcomes of Internet use – and whether individuals can improve their life chances due to its use, addressing questions of social mobility. The findings on the third-level digital divide are scarcer than those on the first- and second-level divide (Scheerder, Deursen, and Van Dijk 2017), partly because outcomes are harder to assess empirically than access, skills, and uses. However, like the first- and second-level digital divide, the third-level digital divide follows existing inequalities in society. Privileged individuals in terms of education and income profit more from technology use for economic outcomes (i.e., they are more likely to save money) and institutional outcomes (i.e., they are more likely to report health and social capital benefits) than non-privileged ones (Van Deursen and Helsper 2015). Importantly, the three levels of the digital divide should not be considered in isolation but in combination and in context, where their sequentiality matters (Van Deursen et al. 2017; Vassilakopoulou and Hustad 2023): exclusion and marginalisation in access to digital technology on the first level restrict use and skills potentials on the second level. Restricted use and skills on the second level, in turn, result in lowered benefits. Finally, disproportional benefits help individuals accrue capital and further their position in society (Van Dijk 2005, 2020).

Digital Inequalities in AI within and between Organisations

Table 11.1 presents a structuring framework of how AI-related inequalities unfold in organisational contexts. For this framework, I differentiate the three levels of the digital divide on the one hand and the difference between interorganisational and intraorganisational AI inequalities on the other. Within each of the six cells, I discuss important issues and empirical evidence. Overall, the framework shows where inequalities in AI occur and how they can be analysed.

Intraorganisational Inequalities

First Level: Access

The first cell refers to access-based AI inequalities within organisations. Importantly, not all employees have equal access to AI systems. Some enjoy better, faster, and more user-friendly software and hardware, whereas others must rely on more basic technology (Cooke and Greenwood 2008). This structuration runs along power

Table 11.1 Digital inequalities within and between organisations framework

	<i>Intraorganisational (Employee and organisational stakeholder AI divide)</i>	<i>Interorganisational (Corporate/ business AI divide)</i>
<i>First-level digital divide: Access</i>	Unequal access to AI between different organisational stakeholders as well as stakeholder sub-groups Vertical dynamics along organisational hierarchies and horizontal dynamics along organisational units and functions Access-based inequalities run partly along existing social groups (e.g., gender, age, race) Stereotypes might affect access-based inequalities Covid-19 pandemic exacerbates such AI inequalities due to increased reliance on personal digital infrastructures (PDIs)	Access to cutting-edge AI as a valuable, rare, imperfectly imitable, and non-substitutable resource Tangible AI/hardware: independent AI infrastructure (e.g., cloud computing) requires substantial investments, creating dependencies for organisations that lack the means Intangible AI/software: company size, external pressure, and sector affect the adoption of cutting-edge AI services; service firms are more agile than manufacturing firms; similar to cloud computing, not developing AI in-house (with AI as an increasingly general-purpose technology) creates dependencies
<i>Second-level digital divide: Skills, use</i>	<i>Skills:</i> unequal AI skills between different individuals; differences between employed and self-employed <i>Skills:</i> holistic understanding necessary that includes social, creative, and technical skills; relationship to 21st century skills <i>Skills:</i> opportunities for AI skills development, both formal and informal, might be structured along social lines <i>Use:</i> capital-enhancing AI use structured along organisational lines <i>Use:</i> gendered use expectations <i>Use:</i> new non-capital-enhancing uses (e.g., digi-housekeeping) that fall disproportionately on disadvantaged communities	<i>Skills:</i> scarce digital talent, especially advanced programming, cybersecurity, and machine learning skills, means a war for talent, where powerful firms are at an advantage <i>Skills:</i> relying on outside fluid talent is increasingly common but comes with risks and challenges; emerging skills pools through digital labour platforms <i>Skills:</i> in-house talent development becomes an important strategic advantage, where powerful companies, especially tech companies, have an edge <i>Use:</i> different depth levels in technology implementation and complementarities; persistently non-digital firms are especially found in manufacturing and among small firms; skills important for capital-enhancing use

(Continued)

Table 11.1 (Continued)

	<i>Intraorganisational (Employee and organisational stakeholder AI divide)</i>	<i>Interorganisational (Corporate/business AI divide)</i>
<i>Third-level digital divide: Outcomes – Benefits and harms</i>	<i>Automation</i>: resilience towards automation will yield substantial negotiation power; AI systems can be abused; frontline employees and lower-skilled workers (e.g., truck drivers) at risk of automation <i>Design</i>: specific design choices benefit advantaged groups disproportionately, e.g., in automated hiring systems and facial recognition technology <i>Social capital</i>: digital technology facilitates social capital accumulation for extraverted individuals and users with the right cultural capital <i>Surveillance</i>: new types of control and surveillance emerge that affect disadvantaged communities most <i>Bias and discrimination</i>: new systems such as automated hiring systems lack transparency and can discriminate directly or indirectly based on social categories	<i>Temporal dynamics</i>: increased market concentration means that AI frontrunners benefit disproportionately and those lagging risk falling further behind Digital technology, especially AI, comes with <i>winner-take-all</i> dynamics, favouring monopolies in many cases due to direct and indirect network effects, big data and machine learning, and switching costs and lock-in patterns Powerful tech companies consolidate their position with <i>strategies</i> such as lobbying, killer acquisitions, industry-consumer coalitions, regulatory entrepreneurship, and tax avoidance The <i>Covid-19 pandemic</i> has accelerated interorganisational inequalities: increased demand for digital solutions means that digitally-oriented organisations reap disproportionate benefits, whereas others suffer

hierarchies, with those at the top enjoying better technological infrastructure. For instance, they can buy licenses to advanced LLMs such as GPT-4 (ChatGPT Plus) more easily. However, horizontal dynamics are at play too, e.g., between different organisational departments and their funding situation. Organisational leeway is stratified by hierarchy as well, with those at the bottom having more restricted AI access and facing tighter control, while those at the top face less control – or may even use technology to surveil others lower in the hierarchy (Ball 2010). Of course, these inequalities reflect social categories. Since women are under-represented in top management positions (Eagly and Carli 2007) and in AI (Young, Wajcman, and Sprenger 2023), they face, on aggregate, more AI access constraints. The same is true for race, where racial minorities are disadvantaged in terms of access. Tech stereotypes might exacerbate the situation (Del Carpio and Guadalupe 2022). For example, Asian Americans have often been portrayed as particularly tech-savvy (Sun et al. 2015) and such mental models could resonate in organisations when it comes to seeking out individuals for pioneering or promoting cutting-edge AI technologies.

The Covid-19 pandemic exacerbated differences in AI access as employees working from home depend on their private Internet connection, which can vary substantially between regions, cities, and neighbourhoods. The concept of personal digital infrastructures (PDIs) describes “an individualized assemblage of tools and technologies, such as personal laptops, smartphones, cloud services, and applications brought together by workers to perform their tasks” (Jarrahi et al. 2021, 73). Such PDIs are very important in flexible and agile working arrangements such as the gig economy (Newlands and Lutz 2021a) and sharing economy (Newlands and Lutz 2021b), but become increasingly relevant for standard working arrangements too, as people work from home or anywhere (Miguel et al. 2023). Access to PDIs is premiated on social conditions, including the affordability of software and hardware, an uninterrupted and undistracted working environment, and the availability of resources if access is interrupted (e.g., support).

More generally, access differences are reflected in the hierarchy of job types, with different employment categories as a key differentiating criterion. The literature has stressed inequalities between core employees and “temps, vendors and contractors” (Bergen and Eidelson 2018), especially in large technology companies (Bapuji et al. 2020). Those commissioned through outsourcing (e.g., image labellers for computer vision) and working in the gig economy merit particular attention. Gig work, especially in the physically bound form of app-work (Duggan et al. 2020), is often done by migrants (Newlands 2024), who face the vagaries of access to digital services in the first place. For example, getting a bank account can prove challenging for migrants, but without a bank account, acquiring a smartphone subscription and using AI-powered apps is difficult. Not having a smartphone, in turn, acts as an entry barrier for app-work, given that workers often provide their own equipment.

Second Level: Skills and Uses

The second cell refers to skills-based and use-based AI inequalities within organisations. Like for access, such inequalities run along existing power structures. Not only are digital skills unequally distributed in organisations, but these inequalities mirror social disparities (Martínez-Cantos 2017). However, systematic assessments of intraorganisational inequalities in AI skills, let alone digital technology skills and uses more broadly, are lacking. General digital inequalities research has differentiated different types of *digital skills* and how they are unequally distributed in societies (Scheerder, Deursen, and Van Dijk 2017). Van Deursen, Helsper, and Eynon (2016) distinguish five types of digital skills: operational, navigation information, social, creative, and mobile. Their research demonstrates the need to go beyond technical or operational skills and instead prioritise a holistic, contextual understanding of skills focused on informational, social, and creative practices. All those skills are increasingly important across knowledge-intensive sectors such as journalism, research and development, marketing, and finance, and especially regarding new AI skills such as prompt engineering (Lo 2023).

Research has also connected digital skills to the broader concept of 21st century skills (Van Laar et al. 2019, 2020). These skills are work-related and have been widely discussed in educational studies and pedagogy. They vary considerably based on demographics, personality, and attitudes, and each digital skill type investigated is predicted by different factors (Van Laar et al. 2019). For example, authors find slightly higher skills among employed than self-employed people, and younger respondents report higher skills than older ones. For the employed group, men have slightly more digital skills, whereas for the self-employed group, women have slightly more. We expect similar findings for AI skills and literacy (Celik 2023).

Similar to skills, different organisational groups will use AI technology unevenly. The differentiation of capital-enhancing vs. non-capital-enhancing uses can be applied. Generally, those lower in organisational hierarchies carry out less complex, more repetitive, and less varied tasks, which tend to be less capital-enhancing and holistic (Morgeson and Humphrey 2006). However, this applies to work design more generally and not specifically AI tasks, where limited evidence exists. Stereotypes and historical, ingrained patterns also reflect inequalities of AI technology use within organisations (Brooke 2021), not just access. Digi-housekeeping, describing “activities performed to sustain the use of digital technologies in pursuit of flexible working” (Whiting and Symon 2020, 1080), has been identified as an invisible and underappreciated type of work that is done by different groups, including social entrepreneurs, office workers, and students, mostly without compensation and recognition. While the practice seems not particularly gendered, the authors reflect on potential divides: “it may be that wealthier and more senior managers have the means to buy the resources to work flexibly, thus creating a new class divide between those who perform digi-housekeeping tasks and those who pay others to do so” (1093). This non-capital-enhancing use stands in contrast to more visible and capital-enhancing uses that contribute to a person’s (digital) reputation, e.g., posting inspirational LinkedIn content or engaging in communities of practice for learning and career advancement (Brooke 2021).

Third Level: Outcomes

The third cell refers to outcomes-based inequalities within organisations. Opportunities and harms from AI are not evenly distributed, and privileged stakeholders, e.g., through their position within an organisation or through possessing supervisory capacities, will likely receive more chances for advancement and fewer obstacles. Similarly, marginalised groups within organisations face disproportionate risks and harms from AI. A range of mechanisms are at work to perpetuate outcome-related inequalities within organisations.

One mechanism by which advantaged groups in organisations perpetuate or advance their position is *automation*, which opens up new avenues for control and power for managers (Kellogg, Valentine, and Christin 2020; Jarrahi et al. 2021). Those attuned to digital systems and most resilient towards automation can wield negotiation power in hiring processes and role allocation (Ajunwa and Greene 2019). With automated hiring systems (AHS), e.g., opportunistic hiring managers

can configure and employ the tools as they see fit and select candidates most suitable to their personal benefit. In other words, there is organisational leeway in how such systems are employed, so that algorithmic control tends to serve managers more than it does workers (Kellogg et al. 2020).

On the flipside of the automation coin, digital technology affects individuals in organisations unequally through *task and job replacement* (Acemoglu and Restrepo 2020). Particularly the introduction of AI affects the workforce unequally, with lower-skilled workers doing repetitive tasks most strongly affected (Frey and Osborne 2017). Amis, Mair, and Munir (2020, 206) mention truck drivers and “professionals performing minor care roles” as two examples of workers at risk of being automated. Education and skills deficits not only become a barrier for career advancement but might push individuals with lower education levels and skills out of the workforce. However, with the advent of generative AI and LLMs, automation has now reached more knowledge-intensive tasks as well (Gmyrek, Berg, and Bescond 2023), resulting in new challenges and inequality dynamics.

Another avenue is through code and the *design* of AI systems. Individuals and groups involved in the design of technologies embed normative values into such systems, voluntarily or involuntarily (Winner 1980). Thus, specific AI design choices can benefit elites disproportionately, rather than maximising the welfare of the less privileged. This is visible in AI systems where rewards and punishments (e.g., rules about suspension from a platform such as Uber) shape inequalities (Kellogg et al. 2020) and condition users – who can be both workers and consumers – to behave in certain ways (Newlands, Lutz, and Fieseler 2019a).

Further reinforcing mechanisms relate to *social capital accumulation* and its transfer into other resources. Social capital can be converted into other forms of capital, such as economic capital and cultural capital (Bourdieu 1986). As ample research has shown, social media are particularly conducive to the formation of bridging social capital (Hofer and Aubert 2013; Ellison et al. 2007; Enders et al. 2008). The AI systems embedded into social media, e.g., in the form of recommender algorithms that suggest new connections, lower the barriers for weak tie management by identifying and recommending potential contacts that share common interests or professional backgrounds. While not all weak ties become valuable resources, AI-powered platforms like LinkedIn optimise this process and enhance users’ ability to connect with those who are more likely to offer relevant opportunities or insights, thus improving the odds that some of these connections will indeed prove beneficial. Professional social media platforms such as LinkedIn have become an important part of daily business. Having a strong presence on these platforms requires time and energy and naturally favours certain types of users. Particularly those are advantaged who can express themselves well, know the ins-and-outs of business etiquette and self-branding, have already a strong network outside the platform, can pay for premium membership, or can leverage their AI skills to create compelling content. They will be more visible on these platforms, leading to higher chances of being contacted by recruiters or potential collaborators. LinkedIn’s parent company, Microsoft, is aware of inequality issues and tries to tackle them with systematic monitoring through inequality tests and design adjustments (Kahn 2020).

A key allocation mechanism by which AI technology affects power relations within organisations is through unequal *surveillance*, understood as “focused, systematic and routine attention to personal details for purposes of influence, management protection or direction” (Lyon 2007, 14). Ample research has demonstrated how surveillance affects particularly those at the margins such as racial minorities, the incarcerated, socio-economically disadvantaged, and demographically excluded (Eubanks 2018). Scholarship in workplace surveillance has also looked at organisational contexts such as new forms of work (De Vaujany et al. 2021). These research streams show the unequal load of surveillance across different groups. In the gig economy, e.g., algorithmic management and surveillance is a particular issue (Duggan et al. 2020; Newlands 2021a; Cameron et al. 2023). It mainly affects gig economy workers such as Uber drivers, Deliveroo riders, and the domestic labour workforce commissioned through services like Helping or Care.com (Van Doorn 2017; Ticona and Mateescu 2018) but does not extend to the platform employees (e.g., Uber manager, software developer). Beyond managing this precarious workforce, the close surveillance through algorithmic management serves to accumulate valuable data assets, which can be gainfully leveraged (Newlands, Lutz, and Fieseler 2019b). Thus, third-level divides will play out particularly strongly in new work domains, driven by disproportionate surveillance and control among those who are on the wrong side of the divide (Kellogg et al. 2020).

Connected to surveillance and design is *discrimination*. AI technologies are often hailed as more objective and bias-free than human judgement (Giermindl et al. 2021). However, ample research across disciplines, including “forensic” analyses in computer science (Buolamwini and Gebru 2018), have shown that cutting-edge AI systems such as facial recognition algorithms (Stark 2019), LLMs (Gmyrek, Lutz, and Newlands 2024), and AHS (Ajunwa and Greene 2019) are far from bias-free. In fact, such technologies can amplify existing inequalities and discriminate against certain groups. The case of AHS is particularly striking since hiring is one of the key organisational practices through which inequalities are reinforced (Amis et al. 2020). Widely used AHS such as Hirevue and Pymetrics claim to tackle bias and discrimination in hiring but show important limitations regarding transparency, fairness, and accountability, so that the transparency articles of the GDPR “might even outright prohibit the use of AHS for wholly automated decision-making in hiring” (Sánchez-Monedero, Dencik, and Edwards 2020, 9). Similarly, Stark (2019) describes facial recognition technology as “the plutonium of AI” and calls for their ban, given inherent flaws and the fact that the risks far outweigh the benefits.

Interorganisational Inequalities

First Level: Access

Moving from intraorganisational to interorganisational inequalities, the fourth cell refers to access-based AI differentials between organisations. Forman and Goldfarb (2006) specify how organisational technology adoption can be divided into four streams. The first stream, “inter-firm diffusion”, describes whether a firm adopts a

digital technology or not. This corresponds to the first-level divide. The three more complex streams are: 1) infusion/assimilation, describing the depth of innovation (i.e., which capabilities of a new technology are implemented); 2) intra-firm diffusion, which captures the displacement of old technologies with new ones; and 3) frequency of use. However, a clear separation of access and use is not always possible given the complexity of AI systems. Independent of the conceptualisation, the literature has termed the phenomenon of interorganisational inequalities the “corporate digital divide” (Forman 2005) or the “business digital divide” (Acilar et al. 2021). Unfortunately, little research on the topic exists. However, existing organisational theories, such as the resource-based view (RBV) of the firm, are informative (Barney 1991). The RBV stresses the importance of resources and capabilities in securing sustainable competitive advantage. Resources should be valuable, rare, imperfectly imitable, and non-substitutable. We can think of a firm’s AI technologies as an asset, with vast inequalities in tangible and intangible resources (Acilar et al. 2021).

For tangible resources, the largest and most data-intensive technology companies are mostly located in the US and China (Barwise and Watkins 2018; Shead 2021). *Data centres and cloud computing* infrastructure is one domain of competitive advantage. Not surprisingly, major technology corporations invest heavily in digital infrastructure, and the market for cloud computing is highly concentrated (Gooding 2020). Setting up the infrastructure necessary to power cloud services and develop cutting-edge AI requires substantial investments in data centres, which themselves are often strategically located, e.g., in Iceland, where cooling is cheaper (Adalbjornsson 2019). Only the wealthiest players can compete in this market (Barwise and Watkins 2018). In other words, few companies have their own cloud infrastructure, while the majority must rely on existing solutions, creating dependencies.

For intangible resources, access to AI systems is unevenly spread (Veugelers, Rückert, and Weiss 2019). Smaller companies are at a disadvantage, and so are those based in manufacturing, compared to services. The business digital divide runs along factors such as “firm size, human resources, location, sector, customers, the pressure of firms, financial support, perceived usefulness, cost, and the degree of ICT readiness” (Acilar et al. 2021, 36). Firm size is the most commonly investigated predictor and is positively related to technology adoption (Forman 2005), likely also to AI.

However, the importance of AI access varies by industry, and inequalities might be more pronounced in some industries than others (Gekara et al. 2019). In the finance industry, e.g., physical proximity to data centres and speed-maximising technologies such as latency arbitrage are key assets for success in high-frequency trading (Schneider 2012). However, such technologies can present a detriment to more traditional investors, “costing as much as \$5 billion per year across global exchanges” (Sheetz 2020).

Despite industry-specific adoption dynamics, AI is becoming an increasingly important general-purpose technology (Trajtenberg 2018). Organisations need to decide whether they develop their AI capacities in-house or commission them externally, which can result in strategic dependencies. Whereas in-house development

comes with increased control, external commissioning is often unavoidable due to lack of financial resources and human capital. This challenge is particularly pressing with the advent of custom LLMs, where the high costs of development and need for computational resources put small and mid-sized companies at a disadvantage, exacerbating the digital divide and potentially hindering innovation and competition. AI development also relies on a complex supply chain of clients and vendors, where data brokers play a prominent role. This industry capitalises on the collection and sale of data as a commodity. It has become a large, yet secretive and ethically dubious industry (Newlands 2021b).

Second Level: Skills and Uses

The fifth cell refers to skills- and use-based inequalities between organisations. In terms of *skills*, literature has stressed a so-called war for talent (Michaels, Handfield-Jones, and Axelrod 2001), which is particularly pronounced when it comes to sought-after technology skills such as machine learning, cybersecurity, and data science skills (Kane et al. 2017; Thomas 2020). Veugelers et al. (2019) identified a lack of “staff with the right skills” as the biggest obstacle for organisations to implement digital technology. Two approaches in the digital war for talent can be distinguished (Kane et al. 2017): 1) relying on outside fluid talent in the form of consultants and contractors and 2) developing in-house through managing existing employees for the long term. The former comes with its own challenges, e.g., increased coordination costs and internal resistance (Corporaal and Lehdonvirta 2017). Moreover, research on online freelancing platforms such as Upwork has shown power asymmetries at play between the platform and the freelancers (Rahman 2021), between freelancers and clients (Blyth et al. 2024), and between different types of freelancers (Demirel et al. 2021).

Regarding the second avenue (developing in-house), organisations vary in their provision of training and development opportunities, so that skills gaps are more pronounced in certain industries, e.g., in public administration, defence, and safety, as well as in transport, logistics, and warehousing (Gekara et al. 2019). Large technology and AI companies such as Alphabet, Meta, and Amazon have substantial advantages in recruiting and developing talent, not only because of high pay but also because they offer good opportunities for career advancement, have strong and recognisable brands, and attractive corporate cultures (Barwise and Watkins 2018).

Little research has studied the *uses* of AI across organisations, but the limited existing evidence points to widespread interorganisational inequalities along firm characteristics. In other words, organisations vary widely in their use of cutting-edge AI. Veugelers et al. (2019) provide a detailed overview of how manufacturing and service firms in the EU and the US differ in their implementation/infusion of digital technology, broadly speaking, including AI. Based on a large-scale survey done by the European Investment Bank (1700 companies in the EU and the US), the authors classify the companies into five groups: persistently non-digital, beginners, stable, catching-up, and frontrunners. Persistently non-digital firms are found more in manufacturing than in services, particularly among small manufacturing firms.

In the services category, young and small firms have the highest proportion of frontrunners, potentially due to many startups engaged in digital innovation.

Shakina, Parshakov, and Alsufiev (2021) study the corporate digital divide in Russia, showing how the implementation of digital technology such as digital manufacturing, AI, and Internet of Things solutions comes with substantial complementarities, favouring companies that already have adopted adjacent technologies. Moreover, the authors stress the importance of digital skills for most technologies studied and particularly for AI, where the implementation leads to an immediate increase in skills demands. They conclude that “managing technological change requires corresponding policies relating to digital skills accumulation [...] to prevent any adverse consequences of the digital divide” (14). Future research should study organisations and ecosystems beyond the Global North. The differentiated uses of digital technology by organisations in the Global South and developments relating to digital skills (e.g., brain drain, ICT4D) require further attention (Friederici, Wahome, and Graham 2020).

Third Level: Outcomes

The sixth cell refers to outcomes-based AI inequalities between organisations. Here a temporal dynamic emerges that those who are leading in the AI arms race can further improve their position (Barwise and Watkins 2018). Evidence suggests that market concentration has increased in recent decades, at least in some countries (OECD 2018). Network effects and winner-take-all patterns have been observed with technology companies in particular (Barwise and Watkins 2018). Economic, technological, demand-side, and supply-side factors combine to make it hard for competitors to displace the Big Five in the Global North (Alphabet, Amazon, Meta, Apple, Microsoft) and the Big Four in China (Tencent, Alibaba, Baidu, Huawei). Barwise and Watkins (2018) specify four mechanisms why technologically saturated markets tend to be winner-take-all markets: direct network effects, indirect network effects, big data and machine learning, and switching costs and lock-in patterns. “The result is dominance of at least one product or service category – in some cases several – leading to sustained high profits, which are then invested in (1) protecting and enhancing the core business and (2) high potential new markets, especially where the company can use the same technology, infrastructure, or brand” (24). In terms of protecting their position, these companies use a range of strategies, including aggressive lobbying (Zuboff 2019), killer acquisitions (Cunningham, Ederer, and Ma 2021), industry-consumer coalitions (Collier, Dubal, and Carter 2018), regulatory entrepreneurship (Pollman and Barry 2017), and tax avoidance (Amis et al. 2021). For example, sharing economy platforms such as Uber and Airbnb, thanks to their strong brands and large user base and using means such as social media campaigns and online petitions, are known to mobilise their customers to influence regulation in their favour (Newlands et al. 2020) or increasingly engage in urban governance through advocacy initiatives such as Airbnb Citizen (Van Doorn 2020). Finally, the tendencies described were exacerbated by the Covid-19 pandemic. With the rapid shift to digital solutions, organisations that were already advanced with digital technology had a head-start, and those who were not were often left behind.

“This is creating a new digital divide that will deepen fractures in our society. The firms that cannot change overnight will be left way behind, exposing their employees to increasing risk of financial and physical distress” (Iansiti and Richards 2020).

Conclusion

In this chapter, I discussed the inequality implications of AI along the three levels of the digital divide, focusing on organisational contexts. AI inequalities within and between organisations touch on aspects such as automation, surveillance, discrimination, resource constraints, market concentration, and winner-take-all dynamics. Recommendations include the recognition of AI inequalities as a salient problem, a socio-technical understanding that avoids technology-determinism and focuses on empowerment and collaboration, legal interventions that address all levels of the divides, and voice and representation mechanisms that go beyond tokenism. Access-based interventions should be combined with skills- and use-oriented programmes.

Future research should accompany such interventions, evaluate them critically, offer suggestions for improvement, and use the whole toolkit of social science methods to study AI inequalities within and between organisations. Within organisations, established methods can be used. For example, case studies on AI access, surveys and interviews on AI use, and focus groups on benefits and harms from AI technologies will enhance our understanding of intraorganisational AI inequalities. However, digital methods have an important place too. Studying inequalities in digital footprints (Micheli, Lutz and Büchi 2018), e.g., in training data, through computational research or digital ethnography (Whiting and Pritchard 2020) is one avenue. Another is to rely on socio-technical approaches, such as the walkthrough method (Light, Burgess, and Duguay 2018), to uncover the implementation of AI in organisations from a contextualised design perspective. Such an approach could provide fruitful insights into biases, constraints, and affordances for different organisational stakeholders. Finally, action research that not only investigates the status quo but actively tries to improve it is needed.

Between organisations, an important, but underexplored question is the “inequality footprint” of different organisations (Riaz 2015), and this idea could be extended to compare organisations in terms of their contribution to increasing or decreasing AI inequalities. Which organisations reduce inequalities in AI access, skills, uses, and outcomes by helping the underprivileged, and which organisations are exacerbating inequalities by hoarding privilege (D’Ignazio and Klein 2023) and systematically neglecting or excluding underprivileged actors? How do processes of inclusion and exclusion depend on macro-level factors such as laws, business conditions, educational systems, and entrepreneurial cultures? These questions can be studied with comparative case study approaches and research that relies on databases of organisational indicators. However, since such data might not be readily available, a first step could be infrastructure building and sharing. Research teams could compile and share datasets with organisational-level information on AI spending, skills programmes, organisational AI uses, and impact indicators. Importantly, such research should have a broad focus and target aspects such as sustainability (Vassilakopoulou and Hustad 2023).

Note

- 1 I follow Bapuji, Ertug and Shaw (2020: 64) in their definition of inequalities as the “uneven distribution in the endowment and/or access to financial and non-financial resources in a society, which manifests in differential abilities and opportunities to engage in value creation, appropriation, and distribution”.

References

- Acemoglu, Daron, and Pascual Restrepo. 2020. “Robots and Jobs.” *Journal of Political Economy* 128: 2188–2244.
- Acilar, Ali, Dag Håkon Olsen, Niels Frederik Garmann-Johnsen, and Tom Eikebrokk. 2021. “Factors Contributing to the Business Digital Divide.” *34th Bled eConference*. <https://aisel.aisnet.org/bled2021/17/>
- Adalbjornsson, Tryggvi. 2019. “Iceland’s Data Centers Are Booming – Here’s Why That’s a Problem.” *MIT Technology Review*, 18 June. <https://www.technologyreview.com/2019/06/18/134902/icelands-data-centers-are-booming-heres-why-thats-a-problem/>
- Ajunwa, Ifeoma, and Daniel Greene. 2019. “Platforms at Work.” *Research in the Sociology of Work* 33: 61–91.
- Amis, John, Shelley Brickson, Patrick Haack, and Morela Hernandez. 2021. “Taking Inequality Seriously.” *Academy of Management Review* 46: 431–439..
- Amis, John, Johanna Mair, and Kamal Munir. 2020. “The Organizational Reproduction of Inequality.” *Academy of Management Annals* 14: 195–230.
- Amis, John, Kamal Munir, Thomas Lawrence, Paul Hirsch, and Anita McGahan. 2018. “Inequality, Institutions and Organizations.” *Organization Studies* 39: 1131–1152.
- Ball, Kirstie. 2010. “Workplace Surveillance: An Overview.” *Labor History* 51: 87–106.
- Bapuji, Hari, Gokhan Ertug, and Jason Shaw. 2020. “Organizations and Societal Economic Inequality.” *Academy of Management Annals* 14: 60–91.
- Barney, Jay. 1991. “Firm Resources and Sustained Competitive Advantage.” *Journal of Management* 17: 99–120.
- Barwise, Patrick, and Leo Watkins. 2018. “The Evolution of Digital Dominance.” In *Digital Dominance*, edited by Martin Moore, and Damian Tambini: 21–49. Oxford: Oxford University Press.
- Bergen, Mark, and Josh Eidelson. 2018. “Inside Google’s Shadow Workforce.” *Bloomberg*, 25 July. <https://www.bloomberg.com/news/articles/2018-07-25/inside-google-s-shadow-workforce>
- Blank, Grant, and Darja Groselj. 2014. “Dimensions of Internet Use.” *Information, Communication & Society* 17: 417–435.
- Blank, Grant, and Christoph Lutz. 2017. “Representativeness of Social Media in Great Britain.” *American Behavioral Scientist* 61: 741–756.
- Blank, Grant, and Christoph Lutz. 2018. “Benefits and Harms from Internet Use.” *New Media & Society* 20: 618–640.
- Blyth, Dorothy Lee, Mohammad Hossein Jarrahi, Christoph Lutz, and Gemma Newlands. 2024. “Self-Branding Strategies of Online Freelancers on Upwork.” *New Media & Society* 26: 4008–4033.
- Boonaert, Tom, and Nicole Vettenburg. 2011. “Young People’s Internet Use.” *Childhood* 18: 54–66.
- Bourdieu, Pierre. 1986. “The Forms of Capital.” In *Handbook of Theory and Research for the Sociology of Education*, edited by John Richardson: 241–258. New York: Greenwood Press.
- Brooke, Siân. 2021. “Trouble in Programmer’s Paradise.” *Information, Communication & Society* 24: 2091–2112.
- Buolamwini, Joy, and Timnit Gebru. 2018. “Gender Shades.” *Conference on Fairness, Accountability and Transparency* 77–91. <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>

- Cameron, Lindsey, Laura Lamers, Ulrich Leicht-Deobald, Christoph Lutz, Jeroen Meijerink, and Mareike Möhlmann. 2023. "Algorithmic Management." *Communications of the Association for Information Systems* 52(1): 518–537.
- Celik, Ismail. 2023. "Exploring the Determinants of Artificial Intelligence (AI) Literacy." *Telematics and Informatics* 83: 102026. <https://doi.org/10.1016/j.tele.2023.102026>
- Collier, Ruth Berins, Veena Dubal, and Christopher Carter. 2018. "Disrupting Regulation, Regulating Disruption." *Perspectives on Politics* 16: 919–937.
- Cooke, Louise, and Helen Greenwood. 2008. "Cleaners Don't Need Computers." *ASLIB Proceedings* 60: 143–157.
- Corporaal, Greetje, and Vili Lehdonvirta. 2017. *Platform Sourcing*, Oxford: Oxford Internet Institute. <https://www.oii.ox.ac.uk/publications/platform-sourcing.pdf>
- Cunningham, Colleen, Florian Ederer, and Song Ma. 2021. "Killer Acquisitions." *Journal of Political Economy* 129: 649–702.
- D'Ignazio, Catherine, and Lauren Klein. 2023. *Data Feminism*, Cambridge, MA: MIT Press.
- De Vaujany, François-Xavier, Aurélie Leclercq-Vandelannoitte, Iain Munro, Yesh Nama, and Robin Holt. 2021. "Control and Surveillance in Work Practice." *Organization Studies* 42: 675–695.
- Del Carpio, Lucia, and Maria Guadalupe. 2022. "More Women in Tech?" *Management Science* 68: 3196–3218.
- Delfanti, Alessandro, and Bronwyn Frey. 2021. "Humanly Extended Automation or the Future of Work Seen through Amazon Patents." *Science, Technology, & Human Values* 46: 655–682.
- Demirel, Pelin, Ekaterina Nemkova, and Rebecca Taylor. 2021. "Reproducing Global Inequalities in the Online Labour Market." *Work, Employment and Society* 35: 914–930.
- DiMaggio, Paul, Eszter Hargittai, Coral Celeste, and Steven Shafer. 2004. "Digital Inequality." In *Social Inequality*, edited by Kathryn Neckerman: 355–400. New York: Russell Sage Foundation.
- Duggan, James, Ultan Sherman, Ronan Carbery, and Anthony McDonnell. 2020. "Algorithmic Management and App-Work in the Gig Economy." *Human Resource Management Journal* 30: 114–132.
- Eagly, Alice H, and Linda L Carli. 2007. *Through the Labyrinth*, Cambridge, MA: Harvard Business Press.
- Ellison, Nicole B., Charles Steinfield, and Cliff Lampe. 2007. "The Benefits of Facebook 'Friends': Social Capital and College Students' Use of Online Social Network Sites." *Journal of Computer-Mediated Communication* 12: 1143–211.
- Enders, Albrecht, Harald Hungenberg, Hans-Peter Denker, and Sebastian Mauch. 2008. "The Long Tail of Social Networking." *European Management Journal* 26(3): 199–211.
- Eubanks, Virginia. 2018. *Automating Inequality*, New York: St. Martin's Press.
- Forman, Chris. 2005. "The Corporate Digital Divide." *Management Science* 51: 641–654.
- Forman, Chris, and Avi Goldfarb. 2006. "Diffusion of Information and Communication Technologies to Businesses." In *Handbook in Information Systems*, edited by Terrence Hendershott: 1–52. Bingley: Emerald.
- Frey, Carl Benedikt, and Michael A Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114: 254–280.
- Friederici, Nicholas, Michel Wahome, and Mark Graham. 2020. *Digital Entrepreneurship in Africa*, Cambridge, MA: MIT Press.
- Gekara, Victor, Darryn Snell, Alemayehu Molla, Stan Karanasios, and Amanda Thomas. 2019. "Skilling the Australian Workforce for the Digital Economy." *NCVER research report*. https://www.ncver.edu.au/_data/assets/pdf_file/0026/5744123/Skilling-the-Australian-workforce-for-the-digital-economy.pdf
- George, Gerard, Martine R Haas, and Alex Pentland. 2014. "Big Data and Management." *Academy of Management Journal* 57: 321–326.

- Giermindl, Lisa Marie, Franz Strich, Oliver Christ, Ulrich Leicht-Deobald, and Abdullah Redzepi. 2021. "The Dark Sides of People Analytics." *European Journal of Information Systems* 31: 410–435.
- Gmyrek, Pawel, Janine Berg, and David Bescond. 2023. "Generative AI and Jobs." *International Labour Organization* working paper. http://www.ilo.org/global/publications/working-papers/WCMS_890761/lang-en/index.htm
- Gmyrek, Pawel, Christoph Lutz, and Gemma Newlands. 2024. "A Technological Construction of Society." *British Journal of Industrial Relations*: 1–29. <https://doi.org/10.1111/bjir.12840>
- Gooding, Matthew. 2020. "Is the Concentration of the Cloud Computing Market a Systemic Risk?" *Techmonitor*, 11 December. <https://techmonitor.ai/technology/cloud/aws-outage-cloud-iaas-paas>
- Halford, Susan, and Mike Savage. 2010. "Reconceptualizing Digital Social Inequality." *Information, Communication & Society* 13: 937–955.
- Hargittai, Eszter. 2002. "Second-Level Digital Divide." *First Monday* 7(4). <https://journals.uic.edu/ojs/index.php/fm/article/view/942/864>
- Hargittai, Eszter, and Steven Shafer. 2006. "Differences in Actual and Perceived Online Skills." *Social Science Quarterly* 87: 432–448.
- Hofer, Matthias, and Viviane Aubert. 2013. "Perceived Bridging and Bonding Social Capital on Twitter." *Computers in Human Behavior* 29: 2134–2142.
- Iansiti, Marco, and Greg Richards. 2020. "Coronavirus Is Widening the Corporate Digital Divide." *Harvard Business Review*. <https://hbr.org/2020/03/coronavirus-is-widening-the-corporate-digital-divide>
- Jarrahi, Mohammad Hossein, Gemma Newlands, Brian Butler, Saiph Savage, Christoph Lutz, Michael Dunn, and Steve Sawyer. 2021. "Flexible Work and Personal Digital Infrastructures." *Communications of the ACM* 64(7): 72–79.
- Jarrahi, Mohammad Hossein, Gemma Newlands, Min Kyung Lee, Christine T Wolf, Eliscia Kinder, and Will Sutherland. 2021. "Algorithmic Management in a Work Context." *Big Data & Society* 8(2): 1–14.
- Jones, Sarah. 2021. "One Giant Leap for Inequality Jeff Bezos's Space Joyride Is a Spectacle of Grotesque Wealth." *NY Mag*, 20 July. <https://nymag.com/intelligencer/2021/07/jeff-bezoss-space-trip-is-one-giant-leap-for-inequality.html>
- Kahn, Jeremy. 2020. "LinkedIn Tries to Improve Equality across Its Site." *Fortune Magazine*, 31 March. <https://fortune.com/2020/03/31/linkedin-improve-equality/>
- Kane, Gerald, Doug Palmer, Anh Nguyen Phillips, and David Kiron. 2017. "Winning the Digital War for Talent." *MIT Sloan Management Review* 58(2): 17–19.
- Kaplan, Andreas, and Michael Haenlein. 2010. "Users of the World, Unite!" *Business Horizons* 53: 59–68.
- Kellogg, Katherine, Melissa Valentine, and Angele Christin. 2020. "Algorithms at Work." *Academy of Management Annals* 14: 366–410.
- Light, Ben, Jean Burgess, and Stefanie Duguay. 2018. "The Walkthrough Method." *New Media & Society* 20: 881–900.
- Lo, Leo. 2023. "The Art and Science of Prompt Engineering." *Internet Reference Services Quarterly* 27(4): 1–8. <https://doi.org/10.1080/10875301.2023.2227621>
- Lutz, Christoph. 2016. "A Social Milieu Approach to the Online Participation Divides in Germany." *Social Media + Society* 2(1): 1–14.
- Lutz, Christoph. 2019. "Digital Inequalities in the Age of Artificial Intelligence and Big Data." *Human Behavior and Emerging Technologies* 1: 141–148.
- Lyon, David. 2007. *Surveillance Studies*, Cambridge: Polity.
- Martinez-Cantos, José Luis. 2017. "Digital Skills Gaps." *European Journal of Communication* 32: 419–438.
- Michaels, Ed, Helen Handfield-Jones, and Beth Axelrod. 2001. *The War for Talent*, Cambridge, MA: Harvard Business Press.

- Micheli, Marina, Christoph Lutz, and Moritz Büchi. 2018. "Digital Footprints." *Journal of Information, Communication and Ethics in Society* 16: 242–251.
- Miguel, Cristina, Christoph Lutz, Filip Majetić, and Rodrigo Perez-Vega. 2023. "Working from Paradise?" *New Media & Society*: 1–23. <https://doi.org/10.1177/14614448231205892>
- Morgeson, Frederick P, and Stephen E Humphrey. 2006. "The Work Design Questionnaire (WDQ)." *Journal of Applied Psychology* 91: 1321–1339.
- Newlands, Gemma. 2021a. "Algorithmic Surveillance in the Gig Economy." *Organization Studies* 42: 719–737.
- Newlands, Gemma. 2021b. "Lifting the Curtain." *Big Data & Society* 8(1): 1–14.
- Newlands, Gemma. 2024. "'This Isn't Forever for me'". *Environment and Planning A: Economy and Space* 56: 1262–1279.
- Newlands, Gemma, and Christoph Lutz. 2021a. "Crowdwork and the Mobile Underclass." *New Media & Society* 23: 1341–1361.
- Newlands, Gemma, and Christoph Lutz. 2021b. "Fairness, Legitimacy and the Regulation of Home-Sharing Platforms." *International Journal of Contemporary Hospitality Management* 32: 3177–3197.
- Newlands, Gemma, Christoph Lutz, and Christian Fieseler. 2019a. "The Conditioning Function of Rating Mechanisms for Consumers in the Sharing Economy." *Internet Research* 29: 1090–1108.
- Newlands, Gemma, Christoph Lutz, and Christian Fieseler. 2019b. "Trading on the Unknown." *The Law & Ethics of Human Rights* 13: 97–114.
- Newlands, Gemma, Christoph Lutz, Aurelia Tamò-Larrieux, Eduard Fosch Villaronga, Rehana Harasgama, and Gil Scheitlin. 2020. "Innovation Under Pressure." *Big Data & Society* 7(2): 1–14.
- NTIA. 1995. "Falling through the Net." *National Telecommunications and Information Administration (NTIA)*. <https://www.ntia.doc.gov/ntiahome/fallingthru.html>
- NTIA. 1999. "Falling through the Net." *National Telecommunications and Information Administration (NTIA)*. <https://www.ntia.doc.gov/legacy/ntiahome/fttn99/contents.html>
- OECD. 2018. "Market Concentration." *OECD*, 20 April. [https://one.oecd.org/document/DAF/COMP/WD\(2018\)46/en/pdf](https://one.oecd.org/document/DAF/COMP/WD(2018)46/en/pdf)
- Pollman, Elizabeth, and Jordan M Barry. 2017. "Regulatory Entrepreneurship." *Southern California Law Review* 90: 383–447.
- Rahman, Hatim. 2021. "The Invisible Cage." *Administrative Science Quarterly* 66: 945–988.
- Raisch, Sebastian, and Sebastian Krakowski. 2021. "Artificial Intelligence and Management." *Academy of Management Review* 46: 192–210.
- Riaz, Suhaib. 2015. "Bringing Inequality Back In." *Human Relations* 68: 1085–1097.
- Robinson, Laura, Shelia Cotten, Hiroshi Ono, Anabel Quan-Haase, Gustavo Mesch, Wenhong Chen, Jeremy Schulz, Timothy Hale, and Michael Stern. 2015. "Digital Inequalities and Why They Matter." *Information, Communication & Society* 18: 569–582.
- Sánchez-Monedero, Javier, Lina Dencik, and Lilian Edwards. 2020. "What Does It Mean to 'Solve' the Problem of Discrimination in Hiring?" *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*: 458–468.
- Scheerder, Anique, Alexander Van Deursen, and Jan Van Dijk. 2017. "Determinants of Internet Skills, Uses and Outcomes." *Telematics and Informatics* 34: 1607–1624.
- Schneider, David. 2012. "The Microsecond Market." *IEEE Spectrum* 49(6): 66–81.
- Schradie, Jen. 2011. "The Digital Production Gap." *Poetics* 39: 145–168.
- Shakina, Elena, Petr Parshakov, and Artem Alsufiev. 2021. "Rethinking the Corporate Digital Divide." *Technological Forecasting and Social Change* 162: 120405. <http://doi.org/10.1016/j.techfore.2020.120405>
- Shead, Sam. 2021. "Europe Is Focusing on 'Tech Sovereignty' as Tensions Flare between the U.S. and China." *CNBC*, 10 March. <https://www.cnn.com/2021/03/10/tech-sovereignty-key-issue-for-europe-amid-tensions-between-us-china.html>

- Sheetz, Michael. 2020. "High-Speed Traders Cost Regular Investors Almost \$5 Billion a Year, Study Says." *CNBC*, 27 January. <https://www.cnbc.com/2020/01/27/latency-arbitrage-trading-costs-investors-5-billion-a-year-study.html>
- Stark, Luke. 2019. "Facial Recognition Is the Plutonium of AI." *XRDS* 25: 50–55.
- Sun, Chyng, Rachael Liberman, Allison Butler, Sun Young Lee, and Rachel Webb. 2015. "Shifting Receptions." *The Communication Review* 18: 294–314.
- Thomas, Michael Ann. 2020. "The Tech Talent War 2.0." *bcw Insights*, 30 September. <https://bcw-global.com/insights/global/the-tech-talent-war-2-0>
- Tichenor, Phillip J, George A Donohue, and Clarice N Olien. 1970. "Mass Media Flow and Differential Growth in Knowledge." *Public Opinion Quarterly* 34: 159–170.
- Ticona, Julia, and Alexandra Mateescu. 2018. "Trusted Strangers." *New Media & Society* 20: 4384–4404.
- Trautenberg, Manuel. 2018. "AI as the Next GPT." *National Bureau of Economic Research*. https://www.nber.org/system/files/working_papers/w24245/w24245.pdf
- Van Deursen, Alexander, and Ellen Helsper. 2015. "The Third-Level Digital Divide." In *Communication and Information Technologies Annual*: 29–52. Bingley: Emerald.
- Van Deursen, Alexander, Ellen Helsper, and Rebecca Eynon. 2016. "Development and Validation of the Internet Skills Scale (ISS)." *Information, Communication & Society* 19: 804–823.
- Van Deursen, Alexander, Ellen Helsper, Rebecca Eynon, and Jan Van Dijk. 2017. "The Compoundness and Sequentiality of Digital Inequality." *International Journal of Communication* 11: 452–473.
- Van Dijk, Jan. 2005. *The Deepening Divide*, London: SAGE.
- Van Dijk, Jan. 2006. "Digital Divide Research, Achievements and Shortcomings." *Poetics* 34: 221–235.
- Van Dijk, Jan. 2020. *The Digital Divide*, Cambridge: Polity Press.
- Van Doorn, Niels. 2017. "Platform Labor." *Information, Communication & Society* 20: 898–914.
- Van Doorn, Niels. 2020. "A New Institution on the Block." *New Media & Society* 22: 1808–1826.
- Van Laar, Ester, Alexander Van Deursen, Jan Van Dijk, and Jos De Haan. 2019. "Determinants of 21st-Century Digital Skills." *Computers in Human Behavior* 100: 93–104.
- Van Laar, Ester, Alexander Van Deursen, Jan Van Dijk, and Jos De Haan. 2020. "Determinants of 21st-Century Skills and 21st-Century Digital Skills for Workers." *Sage Open* 10(1): 1–14.
- Vassilakopoulou, Polyxeni, and Eli Hustad. 2023. "Bridging Digital Divides." *Information Systems Frontiers* 25: 955–969.
- Veugelers, Reinhilde, Désirée Rückert, and Christoph Weiss. 2019. "Bridging the Divide." *Bruegel, Policy Contribution* 17: 1–13. <https://www.bruegel.org/2019/12/bridging-the-divide-new-evidence-about-firms-and-digitalisation/>
- Whiting, Rebecca, and Katrina Pritchard. 2020. *Collecting Qualitative Data Using Digital Methods*, London: SAGE.
- Whiting, Rebecca, and Gillian Symon. 2020. "Digi-Housekeeping." *Work, Employment and Society* 34: 1079–1096.
- Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109: 121–136.
- Young, Erin, Judy Wajcman, and Laila Sprejer. 2023. "Mind the Gender Gap." *New Technology, Work and Employment* 38: 391–414.
- Zillien, Nicole, and Eszter Hargittai. 2009. "Digital Distinction." *Social Science Quarterly* 90: 274–291.
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism*, New York: Profile Books.

12 AI Racial Bias

How Text-to-Image Artificial Intelligence Generators Construct Prestigious Professions

Anna M. Górska and Dariusz Jemielniak

Introduction

The exploration of the interconnections between technical constructs, such as algorithms, and social constructs, including race and gender, holds profound implications for the contemporary understanding of technology (Benjamin, 2016; 2019; Cave & Dihal, 2020; Hanemaayer, 2020; Kong, 2022; López Bellosso, 2022; O'Connor & Liu, 2023; Wajcman, 2004; 2016). Within the realm of artificial intelligence (AI), this relationship is especially strong, forging a nexus where sociocultural dynamics profoundly shape, and are reciprocally shaped by technological advancements.

Notwithstanding, the intersectionality of race and technology remains a relatively developing and underexplored area of research, notably within the contextual dynamics of AI (Cave & Dihal, 2020; Melaku & Beeman, 2020; Miller-Kleinhenz et al., 2021). A myriad of factors contribute to this scholarly gap, encompassing institutional biases and broader socio-cultural predispositions that have historically marginalised the contributions and viewpoints of communities of colour within technological discourses (Benjamin, 2019; Miller-Kleinhenz et al., 2021). Sinclair (2004) argues that racial prejudice significantly pervades these contributing factors. This encompasses a deficit in firsthand and historical narratives accentuating the role of diverse racial groups in technological evolution and prevailing stereotypes attributing technological advancements predominantly to white individuals. The academic discourse, moreover, has demonstrated a tendency to marginalise the centrality of race as a pivotal area of study (Cave & Dihal, 2020).

Nevertheless, there is an increasing shift addressing this gap; the interrelation between race and technology is starting to receive the same level of attention as the relationship between gender and technology since the 1970s. This promising shift, however, is tempered by the recognition that AI, a crucial subfield of technology, remains grossly under-researched in the context of race, thereby highlighting a pressing need for a more balanced and comprehensive scholarly engagement with these issues moving forwards (Noble, 2018).

AI, emblematic of transformative potential, has engendered optimism for mitigating human limitations, particularly biases and prejudices (Pascual, 2021; Przegalińska & Jemielniak, 2023). However, emergent research delineates a less positive picture, illuminating the perpetuation of biases within AI systems, which possess significant

ramifications for their conceptualisation and application (Feng & Shah, 2022; Liang & Acuna, 2020; Noble, 2018; Otterbacher et al., 2018). Instances such as algorithmic processes, like word embedding, manifesting racial biases emphasise this tendency. For example, a study found that word-embedding algorithm characterised European American names as pleasant, while African American ones as unpleasant (Zou & Schiebinger, 2018). The intrinsic biases within AI, largely reflective of the data they are cultivated upon, exacerbate disparities, perpetuating a trajectory of social inequity, exclusion, and misrepresentation. With the advancement and growing social impact of AI, understanding its biases is crucial to mitigate the potential risks associated with the perpetuation of social inequality, exclusion, and lack of representation.

The recent advancement of AI in content generation, particularly in creating images, marks a significant development in technological innovation. AI text-to-image generators, armed with simplistic, no-code, and natural language prompts, have evolved to craft images good enough to win art prizes (Roose, 2022). Their applications span a spectrum of fields such as advertising, media, education, and HR, wielding the potential to shape societal perceptions, roles, and stereotypes, particularly those rooted in race.

This exploration gains prominence in academic circles, yet a considerable gap persists – there lacks a dedicated exploration into racial biases within AI image generators, especially within the professional or workplace context. This study navigates this understudied territory, aiming to unveil the nuanced dynamics of racial bias in visual representations produced by AI image generators.

AI-generated images influence public perceptions and social reality. Their pervasive integration amplifies their potential to mirror, and in instances, craft societal realities. Images are more than sources of information; they are visual representations of phenomena, concepts, and influence perceived realities. However, AI-generated images are not merely a reflection of reality but also harbour the risk of perpetuating and reinforcing pre-existing inequalities and stereotypes, including those based on race (Górska & Jemielniak, 2023). By insinuating themselves into the realms of media creation and educational content, AI image generators wield the influence to educate and shape perceptions regarding various social phenomena. Concurrently, there lies a risk of these tools perpetuating a skewed representation, amplifying discrimination, stereotypes, and vulnerabilities, particularly among already marginalised groups. Historical studies echo the detrimental impacts of stereotyping, particularly towards women and communities of colour, within media portrayals (Arendt, 2023; Gaur, 2020; Ramasubramanian, 2011). AI, fuelled by historical and prevailing datasets, may inadvertently continue this legacy of perpetuating racial stereotypes, casting shadows of bias and negative portrayals upon communities of colour.

This chapter embarks on a meticulous exploration of racial biases within visual representations of diverse professionals, adopting lenses informed by critical race theory (CRT) and science and technology studies (STS). It pioneers a discourse into the occupational racial biases within AI image generators, thereby enriching the broader landscape of Whiteness studies within AI (Cave & Dihal, 2020; Cave et al., 2020). Centring on AI text-to-image generators, the analysis delves into the portrayal of various prestigious professions, scrutinising whether such depictions either perpetuate or dismantle prevailing racial biases.

Amidst a burgeoning societal awareness of biases within AI and algorithms, a realisation dawns regarding the scarcity of focused research aimed at unraveling racial biases within AI-generated content (Intahchomphoo & Gundersen, 2020; Turner Lee, 2018). A labyrinth of complexities shrouds the manifestation and operational transparency of these biases within AI systems. Advocacy for a nuanced, race-conscious approach in AI development gains momentum, challenging paradigms that inadvertently perpetuate rather than alleviate biases.

Within this chapter, we examine the visual presentation of four occupations in the disciplines of – law, medicine, engineering, and scientific research – based on the ranking of most prestigious occupations by the *Global Teacher Status Index* (2018). Employing a critical lens, the analysis unpacks whether these AI-generated images perpetuate or challenge racial biases in the portrayal of professionals in these fields. Although the racial division of labour and the entanglement of work and race are not new, AI systems add a major multiplier. Since a given profession is important in career choices, and the feeling of inadequacy and not matching the dominant image of a professional may outweigh personal interests and skill match (Haas, Koeszegi & Zedlacher, 2016), it is important that race representations of occupations refute rather than reinforce biases and stereotypes. Although there is a growing awareness in society about biases in AI and algorithms (Buolamwini & Gebu, 2018; Kundi et al., 2022; Escudero Pérez 2020; Raji & Buolamwini, 2019), still there is not enough research focused specifically on the racial biases present in AI-generated content. The current situation is concerning, as not only is there a manifest bias in AI systems, but there is also a complexity surrounding the demonstrations of these biases and lack of transparency of how these AI systems are built. Race-blind approach to AI development is not working. This approach, supposedly aimed at neutralising bias, is potentially counterproductive, perpetuating the existing biases rather than mitigating them. The argument thus leans heavily towards the incorporation of an informed and conscious engagement with racial dynamics as a salient feature in the developmental phases of AI technologies.

This chapter thus advances a concerted dialog, seeking to bridge theoretical frameworks with pragmatic insights, aiming to navigate the complexities posed by AI in either exacerbating or alleviating societal disparities such as inequality, exclusion, and misrepresentation. It fosters a nexus of critical engagement, aligning technological innovations with societal imperatives, thereby championing a trajectory of technological progression that resonates with inclusivity, justice, and comprehensive societal advancement.

Literature Review

Theoretical Background

Race is a categorisation that designates human value and societal rank, using white individuals as the standard for human excellence and achievement (Omi & Winant, 1994). This construct of race endures because inaccurate ideas of racial disparity have become deeply ingrained in the convictions and actions of societies. This deep-seated racism permeates all facets of society, including the field of technology.

CRT emerges as a pivotal framework, instrumental in dissecting and elucidating the intricate intersections of race with pervasive structures of power and oppression (Bercuci, 2020). It contests the notion of race as a biological inevitability, positing it instead as a socioculturally shaped construct, wielded to sustain disparities and perpetuate systemic inequalities. CRT illuminates the realms of technological studies with its critical lens, allowing for a nuanced exploration of the ostensibly neutral terrains of technology (Górska & Jemielniak, 2023).

In the field of STS, CRT has been used to analyse the ways in which race shapes technology development, design, and its racially biased outcomes, fighting with the myth of objectivity of technology (Deister, 2013). CRT fosters a discourse where we comprehend that technologies are not developed in a vacuum but are inflected with the biases and assumptions of their creators, which can consequently generate racially skewed outcomes. CRT reveals the undercurrents of racial biases intertwined within algorithmic decisions and design principles, asserting that these biases are not aberrant glitches but rather integral constituents of technological architectures (Noble, 2018).

This inquiry also necessitates an engagement with Whiteness Studies to unveil the subtle visual representations of bias within racialised imagery. Concepts such as the “white racial frame” (Feagin, 2013) emphasise the entrenched normalisation of whiteness, where its symbols and significations permeate societal structures, remaining largely unexamined and unquestioned under the veil of colour-blindness (Benjamin, 2016; Mueller, 2017). Particularly within technological cultures and academic realms, the illusion of colour-blindness prevails, subtly influencing biases in the conceptualisation and design of AI technologies (Daniels, 2015).

In conclusion, the integration of CRT within STS emerges as a potent approach to unveil and analyse the racial contours, biases, and their consequential implications within technological landscapes, thereby fostering a more comprehensive and critically engaged discourse within technological studies.

AI Racial Bias

It is increasingly recognised that algorithms and AI can exhibit biases. Technology, intrinsically, is a reflection of its creators’ biases and predilections, being ingrained with various societal stereotypes and inequalities (Reyero Lobo et al., 2022). Algorithm developers, brimming with an array of personal values and biases, particularly concerning race and gender, inadvertently infuse their creations with these prejudices, thereby fostering a misleading façade of meritocracy (Noble, 2018; O’Connor & Liu, 2023). Such biases within AI tools possess the formidable potential to perpetuate and even magnify discriminatory practices, thereby enabling the normalisation of specific societal groups while concurrently marginalising others (Liang & Acuna, 2020; Escudero Pérez 2020).

Algorithmic biases play a poignant role in amplifying pre-existing social disparities, influencing societal norms and perceptions concerning what is deemed “common” or “normal.” There is a discernible Western-centric bias observed in the realm of computer vision algorithms. For example, algorithms when applied to

photographic data, have been observed to label images of brides adorned in traditional Western white dresses as indicative of “brides” or “weddings.” Conversely, images featuring traditional North Indian brides in traditional red dresses have been incongruously classified as “performance art” or “costume,” thereby establishing Western culture as the implicit standard (Shankar et al., 2017).

The opacity of AI outcomes, compounded by a lack of comprehensive regulations and guidelines, renders the technology enigmatic and complex. A significant catalyst for this issue is the skewed representation within algorithm training data, which often marginalises certain demographic groups, perpetuating existing biases and facilitating their propagation into technological outputs (Currie & Rohren, 2022).

For instance, ImageNet, a seminal database employed for the training of AI systems in the domain of visual object recognition, originates in the United States and is thus susceptible to the underrepresentation of a diverse range of ethnicities (Steed & Caliskan, 2021). Such geographical homogeneity in the training data engenders algorithmic biases. However, it should be emphasised that the expectation of what it means for the population to be racially and ethnically diverse differs not only within the region but also changes with time, as, for example, the US Census Bureau analysis indicated a significant increase in racial and ethnic diversity since 2010 (Census Bureau, 2021). Despite this growing diversity, geographical homogeneity in the training data could still engender algorithmic biases.

Furthermore, machine-learning algorithms are inherently designed to maximise predictive accuracy based on their training datasets. Consequently, if a particular demographic is overrepresented in the data, the algorithm is predisposed to optimise its performance metrics in favour of that demographic, thereby perpetuating existing biases (Zou & Schiebinger, 2018). In another example, the digital platform FaceApp has been implicated in the inadvertent lightening of darker skin tones, particularly among African American users. This algorithmic bias can be attributed to the preponderance of European facial features in the training dataset, effectively encoding Eurocentric beauty norms into the algorithm (Morse, 2017).

Moreover, the system designers may incorporate some of their assumptions into their creations (Gaskins, 2022; Yarger, Fay & Neupane, 2019). The majority of individuals involved in the creation and development of AI technologies are white males, a demographic makeup that has inadvertently fostered the embedding of racial biases and the continuation of discriminatory practices within AI systems (Livingston, 2020; Noble, 2018). This lack of diverse perspectives in the developmental stages essentially results in AI technologies that not only overlook the specific needs of underrepresented groups but, in some instances, have been found to cause harm (Buolamwini, 2016). The foundational principles of these AI systems, including their inherent biases and values, tend to mirror those of their creators, inadvertently fostering a landscape where people of colour find themselves at a disadvantage due to the algorithms and datasets orchestrated predominantly by white males (Chivers, 2019; Noble, 2018; Santamicone, 2019).

As a result, researchers argue that technology and AI may not only reproduce bias, but may also reinforce existing power structures and inequalities (Buolamwini, 2016; Buolamwini & Gebru, 2018; Kong, 2022; Noble, 2018; Wajcman, 2016).

These power structures and inequalities have various facets and AI systems manifest a pronounced and deleterious bias against individuals with darker skin tones, thereby exacerbating existing societal inequalities. A study conducted by Joy Buolamwini, MIT researcher and founder of the Algorithmic Justice League, studied machine-learning algorithms from major corporations such as Microsoft, IBM, and Face++. Buolamwini's findings revealed a glaring disparity: IBM's algorithms had the most egregious error rates, as high as 35%, in identifying women with darker skin tones, whereas all tested systems performed optimally when identifying white males (Santamiconi, 2019). It is crucial to note that these biases are not isolated incidents; they pervade the technology sector. For example, Google Photos has misclassified images of African Americans as gorillas, for which Google later apologised (Kasperkevic, 2015), nevertheless, it showcases racial bias existent in seemingly objective AI. Similarly, Noble (2018) studied specific queries like "black girls" and found instances where Google's algorithms promoted racist and misogynistic representations. Similarly, Araújo et al. (2016) found that queries like "ugly woman" and "beautiful woman" resulted in images that negatively stereotyped black women while positively stereotyping white women.

These biases also have effect in facial recognition technologies adopted by stores to identify potential shoplifters. Chivers (2019) found that these demonstrate an alarming proclivity for misidentifying individuals with darker skin as suspects. Studies by Wilson, Hoffman and Morgenstern (2019) have also indicated that object detection systems, crucial for autonomous vehicle navigation, are less likely to accurately identify pedestrians with darker skin tones, thus putting them at greater risk of being hit (Cardenas & Vallejo-Cardenas, 2019).

These examples clearly show the risks connected with AI racial bias. Existing power structures and societal inequalities are not merely mirrored but also exacerbated by AI technologies, manifesting particularly damning biases against individuals with darker skin tones. Such biases, far from isolated anomalies, are entrenched within the technological sector, leading to the recurrent misidentification and misclassification across various applications, such as facial recognition technologies (Chivers, 2019; Santamiconi, 2019).

Research focusing on racial biases in AI, particularly concerning image generation, is currently developing. Preliminary studies have predominantly focused on web search engines, language processing tools, and image recognition, unveiling that such technologies perpetuate existing societal biases and stereotypes, particularly pertaining to race and gender (Kay et al. 2015; Noble, 2018; O'Connor & Liu, 2023; Zou & Schiebinger, 2018). The absence of comprehensive research underscores a vital necessity for a more extensive investigation into AI-generated images to discern their role in either mitigating or exacerbating existing societal stereotypes and biases.

Method

In this study, we utilised AI-generated images that exclusively contained human-like figures or anatomical components. We prompted nine AI image generation engines with the following occupational labels, "a doctor," "a lawyer," "an engineer," and

Table 12.1 Text-to-image generators with a number of generated images

<i>AI-image generator</i>	<i>Number of AI-generated images</i>
Craiyon.com	37
DALL-E 2	14
HotPot	4
HuggingFace (Stable Diffusion)	12
Hypnogram	4
Midjourney	16
NeuralBlender	4
StarryAI	4
Wombo	4
Sum	97

Source: Authors' own work.

“a scientist.” To maintain consistency, we adhered to the default settings for each engine and opted for realistic representations over abstract ones. The number of images generated by each engine varied, with some producing a single image, others generating up to four, and yet others yielding as many as nine. [Table 12.1](#) delineates the number of images each engine generated for the four specified professions.

Prior research investigating racial bias in AI has principally relied on manual coding and image recognition methodologies. However, the legitimacy of these methods has been questioned due to inherent biases ([Makhortykh, Urman & Ulloa, 2021](#)). In response, our study embraced a survey-based approach for image classification, leveraging human perspectives to discern gender and racial representation within AI-generated images.

In our study, 84 participants were enlisted to evaluate a compilation of 97 images. A survey was employed as the instrument to evaluate perceived racial biases within these AI-generated depictions. Rooted in the foundational works of [Cave and Dihal \(2020\)](#) and [Makhortykh et al. \(2021\)](#), our study adopted an encompassing perspective towards racialised representation. Participants were furnished with a spectrum of classification options, ranging from ‘white’ and ‘non-white’, to ‘inhuman’ and ‘unknown’. It is essential to emphasise that the binary classification of race (‘white’/‘non-white’) in our methodology, while a simplification, is a strategic choice. This dichotomy, despite obfuscating the multifarious nuances of racial identities, is a prevalent construct within academic and media discourses exploring societal hierarchies ([Eddo-Lodge, 2020](#); [Kivel, 2017](#); [Makhortykh et al., 2021](#)). It predominantly serves to delineate the distinctive privileges associated with whiteness in contrast to the multifaceted marginalisation experienced by non-white demographics. Given AI’s historical intertwinement with notions of whiteness and the nuanced challenges intrinsic to the visual discernment of racial sub-categories, we advocate for the suitability of this binary schema within the confines of our research’s objectives.

It is also crucial to highlight the complexities imbued in the visual determination of an individual’s racial identity. Racial identification transcends mere physical appearance, embodying a confluence of genetic variables, cultural imprints, and

personal self-identifications. An authentic and accurate racial attribution necessitates a holistic recognition of these intricate facets. The ambition of our research is to unravel the portrayal of diversity within AI-generated images and to scrutinise the representation of various occupations through the lens of AI imagery.

Results

Our study garnered insights from 84 completed surveys. A demographic analysis revealed that women composed 39% of the respondents, while a dominant 96% identified as white. Despite conducting a gender-based analysis, we did not uncover any statistically significant disparities. Due to the uniform racial composition of the respondents, a comparative racial analysis was deemed unfeasible; hence, the survey analysis proceeded without an emphasis on demographic differentiation. The predominant age group represented in our study was 18–24 years, encompassing 98% of the participants.

The survey, comprehensive with 100 questions, primarily focused on distinguishing the perceived race of professionals in AI-generated images, accounting for 97 of the questions. Respondents were tasked to categorise each image, discerning whether the professional depicted appeared as white, non-white, ambiguous, or inhuman, based on a consensus from the most frequent classifications.

Our findings illuminate a palpable racial bias in AI-generated images, with a significant inclination towards white representation across various professions. An aggregate analysis of 97 professional images revealed that 62% were predominantly identified as white individuals, 9% as non-white, 10% as of an indeterminate race, and 16% characterised as inhuman. Upon excluding images most frequently categorised as “inhuman” and “unknown” (28 images in total), the portrayal of white individuals surged to 87%. These findings clearly showcase that racial bias in AI-generated images is present and persistent across different professions and AI image generators.

A profession-specific examination, concentrating on four focal professions (engineer, lawyer, doctor, and scientist), unveiled a pervasive racial bias. This was manifested as a conspicuous overrepresentation of white professionals in the AI-generated images (see [Figure 12.1](#)). The data indicates a propensity within the AI to generate images leaning towards ambiguity or inhuman characteristics, rather than reflecting a diversity embodying non-white professionals. Such a pattern underscores a persistent inclination towards racial bias in AI-generated imagery across a spectrum of professional realms.

The analysis becomes even more alarming when we consider images that were categorised as “unknown” or “inhuman.” Once these categories are removed from the dataset, as shown in [Figure 12.2](#), the dominance of white professionals in the visual representation becomes even more pronounced. This clearly shows that in all but the scientist profession, white depictions of a professional dominate AI-generated images. The lowest representation of non-white professionals was visible among doctors, where non-white persons were depicted only in 5% of AI-generated images.

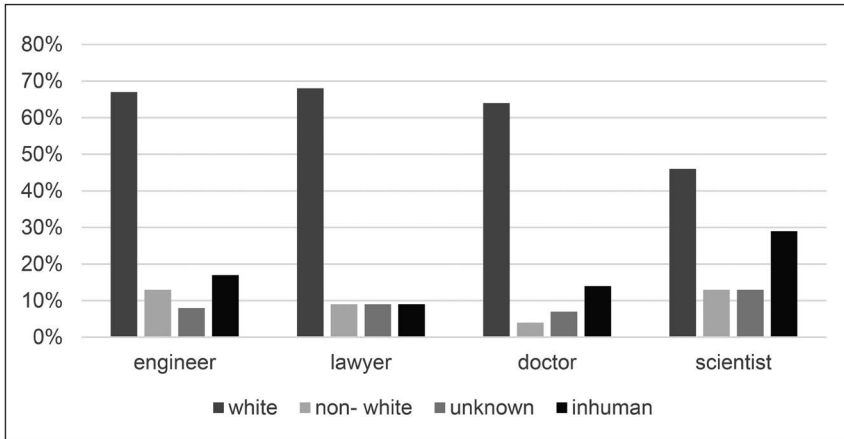


Figure 12.1 Results of respondents' evaluation of 97 AI-generated images with distinction to 4 professions.

Source: Authors' own work.

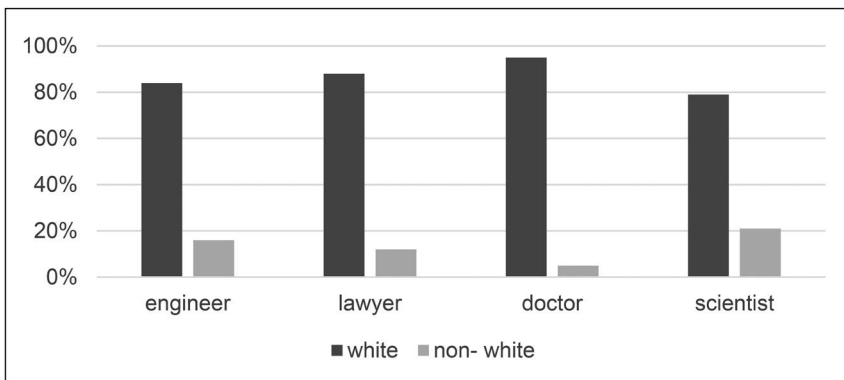


Figure 12.2 Results of respondents' evaluation of 97 AI-generated images with distinction to 4 professions, excluding "unknown" race and "inhuman" answers.

Source: Authors' own work.

Differences in AI Image Generators

Our evaluation of various AI image generators revealed distinct disparities in their portrayal of professions. We analysed nine AI tools, discerning noticeable variations in their representation of professional imagery. Six of these tools predominantly generated images that resembled human professionals, encapsulating identifiable human attributes. However, Midjourney, Hypnogram, and StarryAI deviated from this norm, often producing images that were categorised as inhuman or indeterminate in racial identification (as depicted in [Figure 12.3](#)).

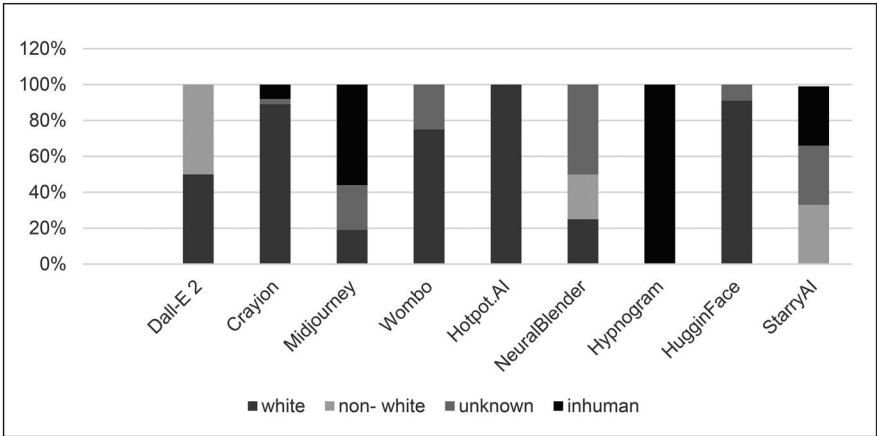


Figure 12.3 Results of respondent’s evaluation of 97 AI-generated images with distinction to 9 text-to-image generators.

Source: Authors’ own work.

In a further nuanced analysis of the six AI tools that predominantly generated human-like professional images, we observed variances in racial representation. Specifically, Dall-E 2 and NeuralBlender exhibited a balanced portrayal, generating an equitable number of images representing both white and non-white professionals. This parity underscores a semblance of racial diversity in their AI-generated professional images.

Contrastingly, the remaining four AI tools – Crayion, Wombo, Hotpot.AI, and HugginFace – demonstrated a significant inclination towards depicting professionals as predominantly white. This prevalence highlights a palpable racial bias, indicating a preponderance of white representations in the professional images generated by these tools. Thus, our study underscores the existence of inconsistencies and biases in AI image generators’ representation of professions, exemplifying the multifaceted challenges ingrained in fostering diversity and equality in AI-generated imagery.

The manifestation of racial bias is conspicuously prominent in several AI image generators, such as Crayion, Hotpot.AI, and HugginFace. These generators predominantly illustrated professionals as white, accounting for over 90% of all images produced. This illustrates a substantial skew towards a singular racial representation, underscoring an ingrained bias within these AI tools.

In contrast, Dall-E 2 emerges as a noteworthy exemplar of inclusivity, exhibiting a balanced representation of white and non-white professionals. This equal distribution suggests a commendable stride towards enhanced diversity and equitable racial portrayal. NeuralBlender also displayed an inclination towards inclusivity but with a nuanced complexity. Half of its images remained racially ambiguous, while the discernible images maintained an equitable balance between white and non-white professional representations.

It is essential to spotlight the progressive modifications integrated into Dall-E 2 in 2022, as cited by Sparkes (2022). These revisions appear to have strategically

navigated towards the realisation of enhanced representational diversity, signifying a successful advancement in mitigating racial disparities within AI-generated professional imagery.

Discussion

Through the analytical prism of CRT and studies of Whiteness, this research highlights the imperative to critically engage with the racial dynamics intricately woven into the fabric of AI technologies. Echoing [Bercuci \(2020\)](#), it underscores a poignant reality: societal constructs, instrumental in perpetuating racial disparities and inequalities, seamlessly extend their influence into the digital sphere, imbuing algorithms with the potency to magnify these inequities exponentially.

Our investigation unveils a conspicuous racial bias embedded within AI-generated imagery, marked by a pervasive overrepresentation of white professionals across a spectrum of prestigious professions. This evidence resonates with prior works that have illuminated the realm of technological systems with discerning critiques of their inherent biases and skewed representations ([Noble, 2018](#); [Reyero Lobo et al., 2022](#)). A commanding prevalence of white portrayals within the professional images generated by AI – where a staggering 87% of images align with white representations – evidently perpetuates the “white racial frame” as conceptualised by [Feagin \(2013\)](#). This pervasive manifestation of symbols and signs, where whiteness is tacitly normalised as the default paradigm, amplifies apprehensions regarding the insidious entrenchment of Eurocentric defaults within technological arenas ([Noble, 2018](#)).

It is crucial to note that such racial disparities in representation not only perpetuate existing stereotypes but also reinforce a skewed perspective of professional sectors, which could have wider implications on societal perceptions. An AI tool, for instance, that consistently portrays doctors as white, inadvertently sends the message that medicine is a predominantly white profession. In this study, doctors were most likely to be depicted as white, even though almost 45% of doctors in the US identify as non-white ([AAAMC, 2019](#)). Over time, such depictions could influence people’s perceptions, leading them to believe that certain races are more suited to certain professions than others. Thus, it seems that AI-image generators exacerbate the status quo biases, leading to portrayals that are consistent with negative racial stereotypes. Such biased portrayals are especially concerning given their influence on perceptions, similar to how media stereotypes can distort the audience’s perception ([Arendt, 2023](#); [Billings & Parrott, 2020](#); [Gaur, 2020](#); [Ramasubramanian, 2011](#)).

A nuanced exploration of the disparities observed among various AI image generators reveals a landscape marked by complexity and contradiction. While certain tools, such as Dall-E 2 and NeuralBinder, exhibit a semblance of balanced representation, others, like Crayon, Wombo, Hotpot.AI, and HuggingFace, navigate towards a predominant showcase of white professionals. This divergence necessitates a thoughtful interrogation of the foundational data and architectural frameworks that shape these models.

In a convergence with the insights articulated by [Daniels \(2015\)](#), our findings assert the limitations and pitfalls of a race-blind approach in AI development.

Neutrality, in this context, does not resonate as a synonym for fairness. Instead, it unveils a potential to amplify and reinforce pre-existing biases, casting shadows of marginalisation upon diverse racial identities. This realisation aligns with the contemplative perspectives offered by [Reyero Lobo et al. \(2022\)](#), which underscore that technology is not neutral but reflective of societal values.

Conclusions

The deployment of text-to-image AI generators in contemporary society has ushered in a host of concerns pertaining to the perpetuation and amplification of racial biases within technological artefacts. Our study of various AI image generators has unveiled a disconcerting pattern whereby these systems predominantly portray professionals as white, effectively omitting the diversity within these vocations. The consequences of this unsettling trend are far-reaching, as they may contribute to the reinforcement of stereotypes insidiously insinuating that professional competence is intricately tethered to one's racial identity. These biases, deeply embedded in AI-generated images, wield considerable influence over societal perceptions, intensifying the already extensive challenges of racial inclusivity.

Scholars such as [Benjamin \(2019\)](#) and [Wajcman \(2004\)](#) have emphasised the critical role technology plays in shaping societal structures, including those related to race and gender. Our findings echo their assertions, underscoring the agency of AI systems in molding and projecting racial stereotypes that can have a profound impact on the lived experiences of individuals. This aligns with the observations of [Cave and Dihal \(2020\)](#) who have drawn attention to the interplay between sociocultural dynamics and technological advancements, particularly within the realm of AI.

Nevertheless, it is crucial to recognise the inherent complexity and contradictions present in these AI models. While some tools demonstrate a more balanced representation, others predominantly showcase white professionals. This inconsistency underscores the need for meticulous scrutiny of the foundational data and architectural frameworks underpinning these models.

As our analysis has shown, striving for mere neutrality in AI is insufficient. Active measures must be taken to engender inclusive and equitable AI systems. This necessitates not only the careful curation of training data to represent a diverse spectrum of racial identities but also a rigorous examination to remove any hidden biases from the data itself ([Mishraky et al., 2022](#)). Moreover, inclusivity in AI development should extend to diverse voices and perspectives. This resonates with the idea of centring marginalised communities in technology design and development ([Noble, 2018](#)).

To redress the propagation of racial biases in text-to-image AI generators, the development process must prioritise diversity and inclusion. This entails meticulous scrutiny of training data for representational accuracy and fairness. The proactive involvement of diverse voices in the design and oversight of AI systems is vital ([Lee, 2018](#)). Collaborative efforts between researchers, developers, and policymakers, guided by ethical frameworks, are indispensable to address the issue of racial biases in AI technologies. Continuous monitoring, auditing, and iterative refinement of AI models are essential components of this approach ([Bartneck et al., 2020](#)).

In addition, education and awareness campaigns play a pivotal role in countering and dismantling the stereotypes perpetuated by AI generators. Raising public consciousness about the presence of biases within AI systems is crucial. Education initiatives should target users, developers, and the general public, elucidating the nuances of AI bias and fostering a culture of critical scrutiny of AI outputs. Additionally, fostering diversity and inclusivity within the AI industry itself is paramount. Promoting diversity in research teams, alongside comprehensive training on ethical considerations and bias mitigation strategies (Katz, 2020), is a step towards rectifying the systemic biases embedded within AI technologies. First steps have already been made with emerging tools such as “bias-detector” in Python which are to identify and mitigate these biases. Nevertheless, there is still a need for systemic solutions that will not only identify the already biased systems but also prevent in the creation of these.

The current study, while offering valuable insights, comes with certain limitations that future research should address. The respondent’s sample was skewed with a significant representation of respondents identifying as white and as such might affect their answers. Moreover, within this study, we simplified the AI-generated images to be classified as white versus non-white, neglecting the nuances of racial identities. Moreover, we focused solely on the racial bias, while existing research highlights the importance of a holistic approach in mitigating biases, as noted by O’Connor and Liu (2023), who emphasise that addressing one type of bias does not resolve overall inequality. Buolamwini and Gebru (2018) and Kong (2022) discuss intersectional biases, underlining the need to consider multiple bias dimensions as gender, race, and age. López Belloso (2022) and Ciston (2019) stress the necessity of a feminist and intersectional approach in AI regulation and development to combat biases. Chu, Donato-Woodger and Khan (2023) provide insights into age-related biases in AI, suggesting a broader scope of bias investigation. Therefore, future studies should also delve into the other diversity dimensions and consider intersectionality, such as intersections of race, gender, and age, to improve the fairness of outcomes for all.

In conclusion, the implications of our study resonate with the broader discourse on technology, race, and societal equity. The presence of racial biases within AI image generators underscores the urgency of collectively addressing these issues, lest we perpetuate and exacerbate systemic inequalities. The transformative potential of AI should be harnessed not only for innovation but also for fostering a more inclusive and equitable future. One of the promising new regulations is the AI Act, but the scope in which it will practically affect the algorithmic biases in practice remains to be seen. Moreover, to effectively address the mentioned issues, the US and Chinese regulations would have to follow.

References

- AAAMC. 2019. “Diversity in medicine: Facts and figures 2019.” <https://www.aamc.org/data-reports/workforce/data/figure-18-percentage-all-active-physicians-race/ethnicity-2018>
- Araújo, C. S., W. Meira, and V. Almeida. 2016. “Identifying stereotypes in the online perception of physical attractiveness.” In: International Conference on Social Informatics, 419–437. Springer.

- Arendt, F. 2023. "Media stereotypes, prejudice, and preference-based reinforcement: Toward the dynamic of self-reinforcing effects by integrating audience selectivity." *Journal of Communication* 73(5): 463–475. <https://doi.org/10.1093/joc/jqad019>
- Bartneck, C., C. Lütge, A. Wagner, and S. Welsh. 2020. *An Introduction to Ethics in Robotics and AI*. Springer Nature.
- López Bellosso, M. 2022. "Women's Rights under AI Regulation: Fighting AI Gender Bias through a Feminist and Intersectional Approach." In: Custers, B., Fosch-Villaronga, E. (eds) *Law and Artificial Intelligence*. Information Technology and Law Series, 35 The Hague: T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-523-2_5
- Benjamin, R. 2016. "Catching our breath: Critical race STS and the carceral imagination." *Engaging Science, Technology, and Society* 2: 145–156. <https://doi.org/10.17351/ests2016.70>
- Benjamin, R. 2019. "Race after technology: Abolitionist tools for the new jim code." *Social Forces; a Scientific Medium of Social Study and Interpretation* 98(4): 1–3.
- Bercuci, L. 2020. "Female and unfree in America: Captivity and slave narratives." *Romanian Journal of English Studies* 17(1): 22–30. <https://doi.org/10.1515/rjes-2020-0004>
- Billings, A., and S Parrott 2020. *Media Stereotypes: From Ageism to Xenophobia*. Peter Lang.
- Buolamwini, J. 2016. "How I'm fighting bias in algorithms." TEDxBeaconStreet. https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms
- Buolamwini, J., and T. Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In: *Conference on Fairness, Accountability and Transparency*. PMLR, 81:77–91
- Cardenas, S., and S. Vallejo-Cardenas. 2019. "Continuing the conversation on how structural racial and ethnic inequalities affect AI biases." In: Cunningham, M., and Cunningham P. (eds) *IEEE International Symposium on Technology in Society (ISTAS) Proceedings*.
- Cave, S., and K. Dihal. 2020. "The whiteness of AI." *Philosophy & Technology* 33(4): 685–703. <https://doi.org/10.1007/s13347-020-00415-6>
- Cave, S., K. Dihal, and S. Dillon. 2020. *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford University Press.
- Census Bureau, U.S. 2021. "2020 United States population more racially and ethnically diverse than 2010." Accessed from <https://www.census.gov/library/stories/2021/08/2020-united-states-population-more-racially-ethnically-diverse-than-2010.html>
- Chivers, T. 2019. "Facial recognition...coming to a supermarket near you." *The Guardian*, 4 August.
- Chu, C.H., S. Donato-Woodger, and S.S Khan. 2023. "Age-related bias and artificial intelligence: A scoping review." *Humanities and Social Sciences Communications*. 10(510). <https://doi.org/10.1057/s41599-023-01999-y>
- Ciston, S. 2019. "Intersectional AI is essential." *Journal of Science and Technology of the Arts* 11(2): 3–8. <https://doi.org/10.7559/citarj.v11i2.665>
- Currie, G., and E. Rohren. 2022. "Social asymmetry, artificial intelligence, and the medical imaging landscape." *Seminars in Nuclear Medicine* 52(4): 498–503. <https://doi.org/10.1053/j.semnucmed.2022.01.004>
- Daniels, J. 2015. "'My brain database doesn't see skin color': color-blind racism in the technology industry and in theorizing the web." *American Behavioral Scientist* 59(11): 1377–1393. <https://doi.org/10.1177/0002764215578728>
- Deister, V. 2013. "Race and the genetic revolution: Science, myth, and culture." *New Genetics and Society* 32(1): 99–101. <https://doi.org/10.1080/14636778.2012.743273>
- Eddo-Lodge, R. 2020. *Why I'm No Longer Talking to White People About Race*. Bloomsbury Publishing.
- Escudero Pérez, J. 2020. "An AI doesn't need a gender(but it's still assigned one): Paradigm shift of the artificially created woman in film." *Feminist Media Studies* 20(3): 393–408. <https://doi.org/10.1080/14680777.2019.1615973>
- Feagin, J. R. 2013. *The White Racial Frame: Centuries of Racial Framing and Counter-Framing*, 2nd ed. New York: Routledge.
- Feng, Y., and C. Shah. 2022. "Has CEO gender bias really been fixed? Adversarial attacking and improving gender fairness in image search." *Proceedings of the AAAI Conference on Artificial Intelligence* 36(11): 11882–11890.

- Gaskins, N. 2022. "Interrogating algorithmic bias: From speculative fiction to liberatory design." *TechTrends: For Leaders in Education & Training*, 19 September.
- Gaur, R. 2020. "A multilevel meta-analysis of effects of negative stereotypes of blacks in media on consumers' attitudes." *Media Psychology* 23(5): 711–732. <https://doi.org/10.1080/15213269.2019.1627888>
- Global Teacher Status Index. 2018. <https://www.varkeyfoundation.org/media/4790/gts-index-9-11-2018.pdf>
- Górska, A., and D. Jemielniak. 2023. "The invisible women: Uncovering gender bias in AI-generated images of professionals." *Feminist Media Studies* 23(8): 4370–4375.
- Haas, M., S. T. Koeszegi, and E. Zedlacher. 2016. "Breaking patterns? How female scientists negotiate their token role in their life stories." *Gender, Work, and Organization* 23(4): 397–413.
- Hanemaayer, A. 2020. *Artificial Intelligence and Its Discontents: Critiques from the Social Sciences and Humanities*. Palgrave Macmillan.
- Intahchomphoo, C. and O. E. Gundersen 2020. "Artificial intelligence and race: A systematic review." *Legal Information Management* 20(2): 74–84. <https://doi.org/10.1017/S1472669620000183>
- Kasperkevic, J. 2015. "Google says sorry for racist auto-tag in photo app." *The Guardian*. <https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app>
- Katz, Y. 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. Columbia University Press.
- Kay, M., C. Matuszek, and S. A. Munson. 2015. "Unequal Representation and Gender Stereotypes in Image Search Results for Occupations." In: *Proceedings of the 33rd Annual Conference on Human Factors in Computing Systems*, 3819–3828. ACM.
- Kivel, P. 2017. *Uprooting Racism - 4th Edition: How White People Can Work for Racial Justice*. New Society Publishers.
- Kong, Y. 2022. "Are 'Intersectionally Fair' AI Algorithms Really Fair to Women of Color? A Philosophical Analysis". In: *Conference on Fairness, Accountability and Transparency*.
- Kundi, B., C. El Morr, R. Gorman, and E. Dua. 2022. "Artificial intelligence and bias: A scoping review." *AI and Society* 37: 199–215.
- Lee, N. T. 2018. "Detecting racial bias in algorithms and machine learning." *Journal of Information, Communication and Ethics in Society* 16(3): 252–260.
- Liang, L., and D. E. Acuna. 2020. "Artificial Mental Phenomena: Psychophysics as a Framework to Detect Perception Biases in AI Models." In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 403–412. Association for Computing Machinery.
- Livingston, M. 2020. "Preventing racial bias in federal AI." *Journal of Science Policy and Governance* 16(2): 1–7. <https://doi.org/10.38126/jspg160205>
- Makhortykh, M., A. Urman, and R. Ulloa. 2021. "Detecting Race and Gender Bias in Visual Representation of AI on Web Search Engines." In: Boratto, L., Faralli, S., Marras, M., and Stilo, G. (eds) *Advances in Bias and Fairness in Information Retrieval*, 36–50. Springer. https://doi.org/10.1007/978-3-030-78818-6_5
- Melaku, T. M., and A. Beeman. 2020. "Academia Isn't a Safe Haven for Conversations About Race and Racism." *Harvard Business Review*. Accessed 25 June 2020. <https://hbr.org/2020/06/academia-isnt-a-safe-haven-for-conversations-about-race-and-racism>
- Miller-Kleinhenz, J. M., A. B. K. Nagy, A. A. Majewska, A. O. Adebayo Michael, S. M. Najmi, K. H. Nguyen, R. E. Van Sciver, and I. T. Fonkoue. 2021. "Let's talk about race: Changing the conversations around race in academia." *Communications Biology* 4 (902): 1–6. <https://doi.org/10.1038/s42003-021-02409-2>
- Mishraky, E., A. Ben Arie, Y. Horesh, and S. Meir Lador. 2022. "Bias detection by using name disparity tables across protected groups." *Journal of Responsible Technology*, 9. <https://doi.org/10.1016/j.jrt.2021.100020>
- Morse, J. 2017. "App creator apologizes for 'racist' filter that lightens skin tones." *Mashable*. Accessed 24 April 2017. <https://mashable.com/2017/04/24/faceapp-racism-selfie/#zeUItQB5iqI>

- Mueller, J. C. 2017. "Producing colorblindness: Everyday mechanisms of white ignorance." *Social Problems* 64(2): 219–238. <https://doi.org/10.1093/socpro/spx012> (Not accessible as of [08/06/2024]).
- Noble, S. U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.
- O'Connor, S., and H. Liu. 2023. "Gender bias perpetuation and mitigation in AI technologies: Challenges and opportunities. AI & SOCIETY. <https://doi.org/10.1007/s00146-023-01675-4>
- Omi, M., and H. Winant. 1994. *Racial Formation in the United States: From the 1960s to the 1990s*. Routledge.
- Otterbacher, J., A. Checco, G. Demartini, and P. Clough. 2018. "Investigating User Perception of Gender Bias in Image Search: The Role of Sexism." In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 933–936. Association for Computing Machinery.
- Pascual, J. 2021. *Innovation and Collaboration in the Digital Era: The Role of Emotional Intelligence for Innovation Leadership and Collaborative Innovation*. Walter de Gruyter GmbH & Co KG.
- Przegalinska, A., and D. Jemielniak. 2023. *Strategizing AI in Business and Education: Emerging Technologies and Business Strategy*. Cambridge: Cambridge University Press (Elements in Business Strategy).
- Raji, I. D., and J. Buolamwini. 2019. "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products." In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 429–435. Association for Computing Machinery.
- Ramasubramanian, S. 2011. "The impact of stereotypical versus counterstereotypical media exemplars on racial attitudes, causal attributions, and support for affirmative action." *Communication Research* 38(4): 497–516. <https://doi.org/10.1177/0093650210384854>
- Reyero Lobo, P., E. Daga, H. Alani, and M. Fernandez. 2022. "Semantic web technologies and bias in artificial intelligence: A systematic literature review." *Semantic Web* 14: 1–26.
- Roose, K. 2022. "An A.I.-generated picture won an art prize. Artists aren't happy." *The New York Times*. Accessed 2 September 2022. <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>
- Santamicone, M. 2019. "Is artificial intelligence racist? Racial and gender bias in AI." *Medium*. Accessed 2 April 2019.
- Shankar, S., Y. Halpern, E. Breck, J. Atwood, J. Wilson, and D. Sculley. 2017. "No classification without representation: Assessing geodiversity issues in open data sets for the developing world." *arXiv preprint*. <https://arxiv.org/abs/1711.08536>
- Sinclair, B. 2004. "Integrating the Histories of Race and Technology." In: Sinclair, B. (eds) *Technology and the African-American Experience: Needs and Opportunities for Study*, 1–17. MIT Press.
- Sparkes M. 2022. "AI art tool covertly alters requests." *New Scientist* 255(3397): [https://doi.org/10.1016/S0262-4079\(22\)01329-X](https://doi.org/10.1016/S0262-4079(22)01329-X)
- Steed, R., and A. Caliskan 2021. "Image representations learned with unsupervised pre-training contain human-like biases." In: *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 701–713. <https://doi.org/10.1145/3442188.3445932>
- Wajcman, J. 2004. *Technofeminism*. Polity Press.
- Wajcman, J. 2016. "Reflections on Gender and Technology Studies: In What State Is the Art?" In: Richardson, L. et al., (eds) *Feminist Frontiers*, 9th ed. 488–502. McGraw-Hill Higher Education.
- Wilson, B., J. Hoffman, and J. Morgenstern. 2019. Predictive inequity in object detection. *arXiv*. <https://doi.org/10.48550/arXiv.1902.11097>
- Yarger, L., C. P. Fay, and B. Neupane. 2019. "Algorithmic equity in the hiring of underrepresented IT job candidates." *Online Information Review* 44(2): 383–395.
- Zou, J., and L. Schiebinger. 2018. "Design AI so that it's fair: Identify sources of inequity, de-bias training data and develop algorithms that are robust to skews in data." *Nature* 559: 324–332.

Index

- 5G (technology) 74, 76, 82
- Aadhaar (identity system) 82
- ABBA 8, 77, 99; Arena 99; Voyage 99
- ANT (Actor-Network Theory) 43, 114
- Africa 113
- agency 7, 57–58, 63–65, 81, 96, 114, 127, 148, 157, 181, 222
- Ahmad, Normahfuzah 114
- AHS (automated hiring-systems) 197, 199, 201
- Artificial Intelligence (AI) 12, 39, 41, 58, 70, 74, 91, 105–108, 110–114, 117, 125, 128, 132, 139, 141, 146, 155, 173, 193, 211; and power 130–133, 139–147, 155, 160, 163, 167, 188, 197, 211–212; as a machinal learning 1–4, 7–10, 42, 67, 75, 77, 79; as a practise 46, 49, 97, 99; beings 132, 157; development 2–3, 5, 7, 9, 44–45, 48, 51, 121, 125–126, 128–129, 131, 193, 203, 213, 221–222; -generated 116, 212–213, 216–222; generative 9, 74–75, 82–83, 97–98, 116–117, 121, 148, 155–158, 160–161, 163, 165, 167, 169–170; model 3–4, 40, 45, 47, 148 NOW 128; socio-cultural AI (SCAI) ix, 8, 106–108, 110–111, 115; -solution 44–46, 48, 50, 115; strategy 126, 160; study 139, 145, 165, 166; team 44, 46, 48–49; technology 3, 5–6, 10, 41–42, 46, 49, 98, 106, 112–113, 121, 126, 127, 139, 140, 146, 151, 155, 161–162, 166, 197, 199, 201–202, 205, 213–216, 221–223; tool 97, 100, 113, 116, 214, 219–221; R&D 126, 156–158, 163, 167–169
- Airbnb 204
- Airolidi, Masimo i, 2–3, 39
- algorithms i–iv, xiii–xv, 1–8, 10–11, 18–31, 39, 41, 44–45, 47, 49–51, 57, 61–66, 68–69, 78, 83, 91–92, 95–96, 100–101, 110, 112–113, 128, 130, 132, 141–142, 144, 148, 150, 173–175–180, 183–188, 200–201, 211–215
- Algorithmic Justice League 128, 216
- Alibaba 84, 161–162, 168, 204
- Alphabet 168, 203–204
- AlphaGo Zero 3
- Alsufiev, Artem 204
- Altman, Sam 12, 72, 145
- Al-Zoubi, Omar A. 114
- Amazon 126, 157, 168, 193, 204
- analogue 18–22, 25–27, 30–31
- Anderson, Peter J. 116
- Andersson, Benny 99
- Anthropic 158
- application (App) 3, 5, 10, 18, 26, 39–41, 45–46, 49, 68–69, 74, 83, 92, 62, 76, 78, 82, 92, 97, 100, 109–110, 122, 125, 131, 158, 161–163, 174, 176–182, 185–188, 198, 212, 216
- Appadurai, Arjun 94
- Apple 134, 168, 204
- Artificial General Intelligence (AGI) 4, 12, 148
- Australia x, xii, 2, 101
- automation 10, 57–58, 62, 64–66, 91, 106, 113, 124, 128, 130, 141, 143, 162, 174, 176, 178, 181–183, 197, 199–200, 205
- Aziz, Omid 95
- Bachmann-Medick, Doris 12, 91
- Baichuan AI 157, 162
- Baidu 132, 161–162, 168, 204
- Bartholeyns, Gil 98
- Barwise, Patrick 193, 202–204
- Baudrillard, Jean 93–94, 99

- Bauman, Zygmunt 72, 92, 96
 Bax, Nick 92, 99
 Beer, David i, 1, 18, 33, 91, 95, 101, 141, 143–145
 behaviour 12, 30–32, 41, 47, 65, 73, 76–77, 91, 141, 144, 155, 160, 163, 178–179, 184
 Beijing 160, 184
 Bellosio, López 223
 Belt and Road Initiative 82
 Bender, Emily M. 122, 125, 127
 Bengio, Joshua 140–141, 145–152
 Bennett, Jane 72, 105, 107
 bias xi, 4, 10–11, 26, 44, 114, 128, 132, 134, 139, 197, 201, 205, 211–218, 220–223; -free 201
 Big Data 1, 9, 27, 61–62, 64–65, 67, 74, 110, 143, 156, 166, 173, 176, 179–181, 184, 186, 188, 193, 197, 204; Bureau 187
 Big Tech 8–9, 124, 125, 155, 157–159
 Birch, Kean 165
 Black-Scholes model 33, 131
 blockchain 74–75, 83–84
 Bloomfield, Brian P. 41
 Blue Origin 193
 Bolin, Göran i
 Bostrom, Nick 149
 Bourdieu, Pierre 39, 166, 200
 Boym, Svetlana 91, 93–94, 97, 99
 Brass, Irina 176
 Brexit 94
 Brzeziński, Dariusz iii, x, 5, 8, 91
 Buolamwini, Joy 201, 213, 215–216, 223
 Burke, Edmund 164
 ByteDance 168

 Canada x, xi, xiii, 2, 5, 139, 151, 158
 Carr, Roger Sir 68
 Castellano, Mayka 97
 Castoriadis, Cornelius 7, 58–61, 66, 69
 Cave, Stephen 123, 129, 163, 211–212, 217, 222
 Chartier-Edwards, Nicolas x, 4, 9
 ChatABC 162
 ChatGPT 74, 83, 117, 124, 128, 142, 148, 162
 cheapfake(s) 105, 117
 Chiang, Ted 122
 China (Republic of China) xii, 2, 6, 9–10, 12, 67–68, 81–83, 121, 126, 132, 156–163, 168–170, 180–181, 183–185, 187–188, 202, 204
 Chu, Donato-Woodgett 223
 Chubb, Jennifer 123–124

 Ciston, Sarah 223
 City Brain 180
 Cloud and Smart Industry Business Group 161
 Cloudry, Nick 39
 communities of practise 48, 51
 community 7, 40, 48–51, 79, 94, 127, 160
 computer 2, 7, 12, 23, 27, 31, 39, 50, 62, 67, 80, 94, 101, 112, 114, 121, 126, 139–140, 145–146, 149, 150, 161, 173, 198, 201, 214
 communication 2, 5, 8, 10, 27, 33, 43–44, 49, 105–109, 112–116, 142, 160, 164–168, 173, 183, 184, 186, 188, 193, 196–197, 199, 211–212, 222; disruptions 60, 79, 105, 143, 147
 COVID-19 (Coronavirus) 20, 46, 110, 196–198, 204
 CPFS (Complaints and Petition Forecasting System) 184
 Crawford, Kate 5, 41, 122, 133, 146
 Crimea 82
 CRT (critical race theory) 212, 214, 221
 cultural studies 2, 23, 93
 cybernetic 10, 140, 142, 145, 151
 cybersecurity 81, 83, 170, 196, 203
 cyberspace 76, 81, 94
 Cyberspace Administration of China 160
 Čapek, Karel 124

 DALL-E 148, 217, 220–221
 Daosheng, Tang 161
 data ii, xiii, 1, 3–4, 7, 9–11, 18–19, 21–27, 32–33, 39–42, 44–47, 49–50, 57, 61–62, 64–65, 67, 73–84, 96–98, 107–108, 110–111, 113–114, 116, 121–123, 127–129, 133, 139–140, 142–145, 147, 151, 156, 160, 166, 173–174, 176–188, 193, 201–205, 212, 215, 218, 221–222; analysis 50, 77, 81, 184; analytics 44–45, 47, 49–50, 64, 74; science 19, 74, 203
 deep learning 3, 44, 65, 74–75, 141–143, 156, 158
 Deep Nostalgia 92, 98
 deepfake(s) 98, 109–110, 117
 DeepLang AI 157
 DeepMind 158
 Defense Advanced Research Projects Agency (DARPA) 157
 Delfanti 193
 democracy xii, 105, 108, 116, 125, 128–129, 148; democratic 3, 50, 105, 110, 112, 117, 122, 128

- Descarte, René 80–81
 Deursen, Van 195, 198
 Didi Chuxing 84
 digital data 21; age 6, 8, 32, 63–64, 107, 108, 113; environment 21–22, 24–26, 30–33; humanism 117; networking 20; practices 31–32; revolution 57–58, 61, 64–67, 163; sociology x, 19; skills 194, 198–199, 204, system 18, 22, 25–27, 30, 32, 179, 199; transformation 6, 17–19, 25, 27, 32–33
 digitalisation 6, 17–18, 20, 58, 65, 193; **1.0** 17–18, 20, 33; **2.0** 6, 17–18, 20, 33
 Dillon, Sarah 123, 163,
 Dihal, Kanta 123, 217, 222
 Directory of Open Access Journals 108
 discourse 8–11, 24, 59–61, 80, 92, 114, 121, 123–124, 127–129, 132–134, 139, 141, 145, 163, 166, 168, 211–212, 214, 217, 223
 Distributed AI Research Institute (DAIR) 127–128, 134
 DNA 60, 63
 ‘Doomsday’ machine 66–67
 Dostoevsky, Fyodor 69
 Downey, John i
 Durante, Massimo iv, 66

 EBSCO Information Service 107
 echo chamber 106, 117
 economy i–iii, x, xii, 7, 9, 17, 27, 82–83, 132, 135, 156, 162–169, 198, 201, 204
 Elliott Anthony x, 1, 4, 7, 12, 17, 39, 57, 65, 101, 107
 Ellul, Jacques 60–61
 emphasis 7–8, 41, 43, 46–48, 59–60, 62, 68, 77, 99, 117, 125–126, 152, 156, 159, 160–161, 163, 177, 183, 186, 212, 214–215, 217–218, 223
 environment 6, 20–27, 33, 44, 51, 73, 76–77, 79, 105–106, 110, 117, 121, 130–131, 139–140, 143, 146, 148, 161, 174–175, 177–178, 198
 ethic 12, 46, 51, 109, 112, 114–115, 117, 169
 European Union (EU) 5–6, 68, 81, 83–84, 126, 133, 159, 203
 Eynon, Rebecca 198

 Face++ 216
 Facebook 83, 134, 157, 161, 168, 194
 Far East 5
 Fäلتskog, Agnetha 99
 Feagin, Joe R. 214, 221
 Feher, Katalin xi, 5, 8, 105–108, 115
 Filiciak, Mirosław xi, 3, 8
 Filipek, Kamil xi, 4, 7
 filter bubbles 8, 106
 Floridi, Luciano 6
 fragmentation 82, 105, 117, 178
 framework 2, 6, 8–10, 33, 39, 41–45, 51, 59, 68, 74, 101, 105, 107, 112–115, 117, 124, 126–127, 146, 149–150, 156, 160–161, 163, 168, 173–176, 187, 195–196, 213–214, 221–222
 freedom 58, 66, 69, 121, 129, 131
 Friemel, Thomas N. 106
 Frontis 157
 Future of Life Institute 125, 140, 145–148

 Gazprom 82
 Gebru, Timnit 122, 127–128, 201, 223
 Germany 2, 67, 173
 Giddens, Anthony 7, 73, 75–76
 Global North i, 204
 Global South i, 82, 204
 Global Teacher Status Index 213
 Goldfarb, Forman 201
 Google 13, 125–128, 134, 157–158, 161, 168; GNI (Google News Initiative) 113; Photos 95, 216
 governance 1, 5, 68–70, 126, 140, 145, 149, 151–152, 163–164, 179, 182, 186–188, 204
 government 8, 68, 73–74, 82–84, 121, 124–126, 131, 139, 152, 156–157, 159–160, 164, 173, 175, 178–180, 185, 187–188
 Górská, Anna M. xi, 4
 GPT-4 140, 158, 197
 Gray Widdler, David 123
 Grenier, Étienne 4, 9
 Guenduez, Ali A. 125

 Halpern, Orit 142
 Hamida, Norsiah A. 114
 Hangzhou 180
 Hanna, Alex 125
 Harari, Yuval Noah 127
 hardware 22, 74, 158, 195–196, 198
 Hargittai, Eszter 194
 Harrison, George 99
 Helsper, Ellen 198
 Hesmondhalg, David i
 Hine, Emmie 6
 He, Baohang xiii, 4, 9
 health 45–46, 74, 77, 84, 110, 122, 125, 132, 193, 195
 healthcare 44–51

- Hezbollah 67
 Hinton, Geoffrey 145
 Hofer, Johannes 92–93
 Hoffman, Judy 216
 Hong, Soonkwan 91
 HotPot 217, 220–221
 Hristova, Stefka 91
 Huawei 161–162, 204
 HugginFace 217, 220–221
 human 1, 3–7, 12, 17, 19, 21, 24–25, 27–33, 39–41, 43–44, 50–52, 57–63, 65–69, 73–74, 76–77, 79–81, 91–92, 94, 109, 112–117, 122, 124–131, 133, 139, 141–142, 144–145, 147–149, 151, 155, 163, 173, 175–177, 183, 187, 201–203, 211, 213, 216–220
 Human-Machine Communication (HMC) 112–113, 115; human-robot 109
 Hungary xi, 2
 Hunyuan 162
 Hypnogram 217, 219

 IBM 158, 216
 identity 48–49, 65, 81–82, 93, 101, 148, 217, 222
 iFlytek 161
 ImageNet 141, 215
 India 67, 81–83, 132, 158
 Indonesia 132
 information 1, 9, 11, 18, 22–23, 28, 42–43, 47, 57, 59, 61–62, 64–65, 67, 77, 81, 96, 105, 107, 110, 115–117, 122, 124–125, 127, 139, 146, 152, 160, 165–166, 169–170, 173, 175, 177–180, 182–188, 193–194, 198, 205, 212; dis- xii, 70, 84, 108, 110–111, 116, 117; mal- 105, 117; mis- 83, 105, 115
 infrastructure 10, 17, 21–22, 41, 46, 65, 73–79, 81–84, 113, 121–122, 128, 151, 160, 196–198, 202, 204–205
 Instagram 78
 Institute of Electrical and Electronics Engineers 108
 interdisciplinary i, ii, 4, 51, 91, 114–115
 interrelationship 19, 28–29, 32–33
 Israel 67–68, 132, 158

 Jackson, Peter 99
 Jacobsen, Benjamin 95
 Jacobsen, Michael Hviid 91, 100–101
 Jameson, Fredric 93
 Japan 132
 Javorsky, Emilia 147
 Jemielniak, Dariusz xii, 4, 10

 Jiang, Shian 113
 journalism xiii, 8, 105–117, 198

 Kant, Immanuel 58
 Katona, Attila 107–108
 Kempelen, Wolfgang 124
 Keras 3
 Klein, Ortega 5–6
 Kidd, Jenny 92, 98
 Knorr Cetina, Kaira 27
 knowledge iii, 1–2, 4–5, 24, 31, 40, 48–50, 52, 59, 74, 76, 80, 91, 113–114, 124, 131–132, 140, 175, 177–180, 183, 186–187, 194, 198, 200
 Kong, Youjin 223
 Korean Go 3
 Kuai, Joanne 114

 Land, Nick 144
 large language models (LLM) 3, 9, 125, 127, 140–142, 148–149, 156–158, 161–162, 169, 170, 193, 197, 200–201, 203–204
 Large Visual Models (LVM) 3
 Latour, Bruno 17, 23
 Lecun, Yan 145
 Lennon, John 98–99
 LinkedIn 194, 199–200
 Liu, Zheng xii, 6, 9, 223
 London 99
 Lutz, Christoph xii, 4, 10
 Lyngstad, Anni-Frid 99
 Lyotard, Jean-François 1, 4, 11

 machine learning (ML) 1, 3, 9, 19, 21, 24, 26–27, 39, 44, 47, 50, 58, 61–64, 66–67, 74, 99, 122, 139, 141–142, 151, 156–157, 174, 184, 196–197, 203–204, 215–216
 MacKenzie, Donald 27, 33
 Makhortykh, Mykola 217
 Mansell, Robin 165
 Marody, Mirosława 47
 Marres, Noortje i
 Marx, Karl 164
 Maslej, Nestor 159
 MathGPT 162
 McAvoy, Nieto 98
 McCartney, Paul 98–99
 McGilchirst, Ian 75
 McMillan-Major, Angelina 122, 127
 media ii, xi, xiii–xiv, 8, 74, 80, 93, 96, 101, 106, 113–115, 117, 121, 124, 127, 133, 140, 145, 164–166, 194; digital 105,

- 168; fake 109; industry 93, 164–165, 167; landscape 109–110, 115, 168; new *xiii*, 5, 106, 110, 113, 114, 165; regulation 110; social *xi–xii*, 2, 8, 18, 22, 26–27, 31, 43, 69, 74–78, 91, 95, 100, 105–106, 109, 110, 115, 124, 128, 193–194, 200, 204; social-fake-news 115, state-owned 114
- media and communication studies 2, 8, 106, 108, 115, 164–166
- media environment 105–106; high-choice 106; news 106; platform-dominated 110
- Meijer, Albert 173, 175
- Meimaridis, Melina 97
- Mendon-Plasek, Aaron 141
- Merton, Robert 145
- Meta 83, 157–158, 168, 203–204
- Mettler, Tobias 125
- Microsoft 83, 117, 126, 134, 157–158, 168, 200, 204, 216
- Middle East 113, 164
- Midjourney 217, 219
- military 1, 3, 6, 67–68, 72–73, 78–79, 126
- Mill, John Stuart 58
- MiniMax 162
- Minsky, Marvin 157
- Mitchel, Margaret 127, 139, 147
- ModelBest 157
- modern 7, 23, 58–61, 80, 95–97, 105, 132, 148, 150, 160, 173
- Moonshot AI 157
- Morgenstern, Jamie 216
- MOSS 162
- Musk, Elon 72, 78, 127, 145
- National Artificial Intelligence Research and Development Strategic Plan (NSTC) 156–157
- Negroponte, Nicholas 20
- Netflix 97
- Netherlands 173
- network 3, 18, 20–21, 23, 39, 43, 47–48, 57, 62, 65, 73–74, 76, 78, 82–84, 105–107, 109–110, 114, 121, 140, 144, 150–151, 163, 167–168, 197, 200, 204
- Neuberger, Christoph 106
- NeuralBlender 217, 220
- new technology 1, 5–6, 27, 58, 73, 92, 128, 133, 183, 202
- news 68, 82, 95, 105–107, 110, 112–115, 194–195; fake 8, 107, 109–110, 112; organisations 106, 113
- Ng, Jenna 99
- Ninno, Domenico 23
- Noble, Safiya U. 212, 216
- North America 164
- Norway *xii*, 2
- nostalgia 8, 91–94, 96–101
- Nowotny, Helga 62–63
- NRS (news recommender systems) 106
- Nvidia 83, 158
- Nyiri, Kristof 84
- O'Hara, Kieron *xiii*, 4, 7–8
- Ono, Yoko 98
- OpenAI 12, 83, 117, 124–125, 158, 160
- Oracle 126
- Pallister, Kathryn 92, 96–97
- Palouri, Manohar 95
- Pangu 162
- Parisi, Luciana 147
- Parshakov, Petr 204
- PDI (personal digital infrastructures) 196, 198
- Pfetsch, Barbara 105, 107
- Pierson, Jo *i*
- Pisano, Valérie 146
- Piwowar, Kuba *xiii*, 3, 8
- philosophy 2, 5, 58, 63, 77–78, 112, 140
- platform 7, 18, 22, 27, 31, 67, 76–77, 91, 95–97, 100, 106, 110, 113, 141, 167, 179–182, 184–185, 187–188, 194, 196, 200–201, 203, 215; platformisation 113
- Plato 80
- Poland 2, 101, 124
- polarisation 105, 107
- political science 2, 5, 91
- populism 84, 94
- postmodern 1, 12
- power *ii*, 1, 3, 5–7, 9–10, 21, 39, 42, 48, 50–52, 59, 61, 63, 66–67, 69, 73–74, 78–79, 81–82, 99, 113, 121–125, 127, 129–130, 139, 144–145, 147–148, 151, 155, 158, 163–166, 168–169, 175, 177–178, 182, 184, 186–187, 195, 197–199, 201–203, 214–216
- prediction 9–10, 30–32, 45, 61, 63–64, 75, 91, 97, 106, 139, 142–143, 150, 174, 176–177, 181, 183–184
- privacy *xii*, 6, 44, 46, 50, 69, 82, 114, 121, 124, 129–130, 160
- Project Maven 126
- psychology 5, 41, 63, 77, 91
- public policy 5, 123
- public sphere *xiv*, 105, 115–117, 133; illiberal 105; networked 107; post- 107
- Putnam, Hilary 80

- Python 3, 223
 Przeglasińska, Aleksandra i
 Qianwen 162
 Qianzhan Industry Research Institute 160
 Reckwitz, Andreas 41–43, 46, 51
 Reese, Stephen 107
 Republic of Korea (South Korea) 68, 82
 research 3–6, 8–10, 41–42, 46, 50, 83,
 91–92, 98, 101, 106–110, 112, 114–117,
 121, 123–124, 127–130, 132, 134, 147,
 150, 152, 155–163, 165–170, 174,
 193–194, 198–205, 211, 213, 216–218,
 221, 223
 retromania 96, 100
 Reynolds, Simon 96
 right 6, 28, 75, 80, 108, 146, 155, 167, 197,
 203
 Riri Xin 162
 Roberge, Jonathan xiii, 4, 9, 141,
 144, 151
 Roberts, Huw 6
 robot xii, 27–28, 39, 62, 68, 109, 123–124,
 130–131
 Routledge, Clay 93
 Ruhleder, Karen 21
 Russia 67–68, 70, 79, 82–83, 204
 Saramago, José 81
 Schatzki, Theodor 41–44, 47–48
 Schlesinger, Philip 107
 Scopus 107–108, 110
 Sedol, Lee 3
 SenseTime 161–162
 Seyfert, Robert xiii, 4, 6–7, 33
 Shafer, Steven 194
 Shanghai 160, 162
 Shenzhen xiv, 181–182
 Shmitchell, Shmargaret 122
 Shusheng 162
 Signoret, Simone 92, 94
 Silicon Valley 72, 84, 124, 132, 157
 Simondon, Gilbert 19, 21, 28
 Slack, Jennifer D. 91
 Smith, Adam 164
 Snapchat 95
 social media 2, 8, 18, 22, 26–27, 31, 43, 69,
 74–76, 78, 91, 95, 100, 106, 109, 115,
 124, 128, 193, 194, 200, 204; relations
 6, 17–21, 24–25, 29, 32–33, 51, 58, 64,
 164; science 2, 5, 12, 51, 91–92, 94,
 107–109, 114, 139, 205
 Social Science Research Network 108
 sociology 2, 5, 9, 17–19, 25, 27, 30, 32–33,
 39, 41, 91, 93, 115, 155, 169, 193–194
 Soviet (Soviet Union) 67, 73
 South China Sea 67
 Spinoza, Benedictus de 30
 Star, Susan L. 21
 Starr, Ringo 99
 StarryAI 217, 219
 startup 9, 83, 157–158, 204
 Steinhoff, James 165
 STEM (science, technology, engineering,
 and mathematics) 5, 10, 33, 158–159
 STS (science and technology studies) 169,
 212, 214
 surveillance iii, V, 6, 12, 23, 66–67, 69, 73,
 76, 79, 82, 110, 124, 129, 131, 174, 193,
 197, 201, 205
 Sweden 132
 Swidler, Ann 41–43, 45
 Swiss 92–93
 Taiwan 82
 Tanner, Gafton 99–100
 TaS (A Theory of Algorithmic Sociality) 6,
 17–25, 27–28, 32–33
 technology 1–3, 5–10, 18–19, 25, 27,
 42–43, 45, 49, 57–58, 60–62, 64–65,
 67–68, 72–73, 78–79, 82–83, 91–93, 95,
 98–100, 105–107, 115–116, 121–133,
 151, 155–158, 160–163, 165–168, 170,
 173, 175, 179, 183–185, 187, 193–205,
 211–216, 222–223
 Technology of China 157
 Tegmark, Max 147
 Tencent 84, 132, 161–162, 168, 204
 Thailand 132
 The Beatels 98–99
 Timnit, Gebru 122, 127
 Tongyi 162
 Totaro, Paolo 23
 transdisciplinary iv, 5, 8, 105, 108, 110,
 113, 115
 transparency 44, 84, 110, 122, 126, 134,
 178, 183, 197, 201, 213
 Uber 83, 200–201, 204
 UK (United Kingdom, Great Britain,
 Britain) 2, 67, 158, 164, 173
 Ukraine 70, 79, 82, 110
 Ulvaeus, Björn 99
 UNESCO 146–147
 Unified Payment Interface 82
 Urban Governance Command Center 182,
 186

- US Census Bureau 215
- USA (US, United State) 5–6, 9, 67–68, 83–84, 101, 114, 122, 126–128, 132, 156–158, 163, 168–169, 202–203, 215, 221–223
- User Experience (UX) 45, 49–50
- Valley, Silicon 72, 84, 124, 132, 157
- Van Dijk, Jean 193
- van Noort, Carolijn i
- Varoufakis, Yanis 72, 76
- Veale, Michael 91, 176
- Verdegem, Pieter 165–166
- Veugeler, Reinhilde 203
- VHS 97
- Vogl, Thomas M. 173, 175
- Watkins, Leo 204
- weapon 67–68
- Web of Science 107
- Wenger, Etienne 48, 50–51
- Wenxin Yiyan 162
- West, Sarah 123
- WeWork's IPO 83
- Whittaker, Meredith 123
- Wilson, Janelle 96, 216
- Winiarska-Brodowska, Malgorzata xiii, 5, 8
- Wombo 217, 220–221
- Woolgar, Stevie 41
- World Economic Forum 68
- World War 78
- World Wide Web 73, 81
- Wozniak, Steve 127
- Wu Jinjin xiii
- Xiaomi 168
- Xiecheng Wendao 162
- Xinghuo 162
- Yi 162
- Yuan 162
- Zhipu AI (Zhipu Qingyan) 157, 161–162
- Zidong Taichu 162
- Zuboff, Shoshana 106
- Zuckerberg, Mark 7