OXFORD

Imagining AI

How the World Sees Intelligent Machines

Edited by
STEPHEN CAVE
KANTA DIHAL

IMAGINING AI

IMAGINING AI

How the World Sees Intelligent Machines

Edited by

Stephen Cave
and
Kanta Dihal

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2023930758

ISBN 978–0–19–286536–6

DOI: 10.1093/oso/9780192865366.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

To Huw Price, who made all this possible.

Contents

Acknowledgements

This book originated in the Global AI Narratives (GAIN) project at the Leverhulme Centre for the Future of Intelligence, University of Cambridge. From 2018 to 2021, the GAIN team held twenty workshops to explore how AI is portrayed across cultural, geographical, regional, linguistic, and other boundaries and borders, and how these portrayals affect public perceptions of AI around the world. For helping us organize a wonderful series of events—in-person, virtual, and hybrid, at times in the face of political or social unrest and the pandemic—we thank our collaborators, including Nanyang Technological University and the NTU Institute for Science and Technology for Humanity (NISTH), Singapore; Waseda University, Tokyo; Teplitsa of Social Technologies, Russia; the Access to Knowledge for Development Center (A2K4D) at the American University in Cairo's School of Business; the Vā Moana Pacific Spaces research cluster and the University of Auckland, Aotearoa New Zealand; the Berggruen China Center at Peking University; the Department of Security Studies at Charles University, Prague; the Human Sciences Research Council of South Africa; the Pontificia Universidad Católica de Chile; the Present Futures Forum at the Technische Universität Berlin; and the School of Arts and Sciences at Ahmedabad University, India. We also thank all of the speakers, some of whose work is included in this book, as well as the many thousands of people who attended these workshops.

We acknowledge a deep debt of gratitude to the wonderful team of research assistants who worked on this project: Tonii Leach (who also assisted in the preparation of this volume), Elizabeth Seger, and Tomasz Hollanek.

We are immensely grateful to the many colleagues whose feedback has improved our understanding of this field in so many ways. In particular, for reviewing our own contributions to this volume, we thank Eleanor Drage, Kerry McInerney, Tomasz Hollanek, and Moritz Ingwersen.

The majority of contributions to this book are written by contributors who do not speak English as their first or primary language. We gratefully acknowledge the work of our translators Maya Feile Tomes, Jack Hargreaves, James Trapp, and James Hargreaves for their excellent translations from the Spanish, Chinese, and German.

We also express our gratitude to the team at Oxford University Press for once again having faith in us, and for their unfailingly professional and friendly support. In particular, we thank Sonke Adlung, Senior Commissioning Editor, and John Smallman and Giulia Lipparini, Senior Project Editors.

We acknowledge the generous support of the Templeton World Charity Foundation and DeepMind Ethics and Society for funding the Global AI

Narratives project; the Leverhulme Trust through their grant to the Leverhulme Centre for the Future of Intelligence; and the Cambridge–Africa ALBORADA Fund and the British High Commission Singapore for funding the Sub-Saharan Africa and Singapore workshops, respectively.

Finally, we thank Huw Price, former Bertrand Russell Professor of Philosophy at Cambridge and the founder of the Leverhulme Centre for the Future of Intelligence, for his unflagging support, and for making all this possible. It is to him that this book is dedicated.

Contributors

Rachel ADAMS, based in Cape Town, South Africa, is the Principal Researcher at Research ICT Africa. She is the Principal Investigator of the Global Index on Responsible AI, a new global instrument to measure country-level progress in advancing rights-respecting AI, and the African Observatory on Responsible AI. Rachel is the author of *Transparency: New Trajectories in Law* (Routledge, 2020), and the lead author of *Human Rights and the Fourth Industrial Revolution in South Africa* (HSRC Press, 2021). She is currently working on a monograph exploring the imperial forces at work in and through AI, and what is required to decolonize this novel technological superpower.

Macarena ARECO is a journalist and Full Professor of Latin American Literature at the Pontifical Catholic University of Chile, where she is also Director of Extension and Continuing Education at the UC Center for Chilean Literature Studies (CELICH). She is the author of the books *Cartografía de la novela chilena reciente* (2015), *Acuarios y fantasmas. Imaginarios de espacio y de sujeto en la narrativa argentina, chilena y mexicana reciente* (2017), and *Bolaño constelaciones: literatura, sujetos, territorios* (2020), and co-editor with Patricio Lizama of *Biografía y textualidades, naturaleza y subjetividad. Ensayos sobre la obra de María Luisa Bombal* (2015). She has published over a hundred articles and book chapters, and currently directs the research project 'Imaginarios sociales en la ciencia ficción latinoamericana reciente: espacio, sujeto-cuerpo y tecnología'.

Noelani ARISTA ('Ōiwi) is Director of the Indigenous Studies Program at McGill University, Montreal, and Associate Professor in History and Classical Studies. Her research focuses on Hawaiian governance and law, indigenous language textual archives and traditional knowledge organization systems. Her work seeks to support indigenous communities in the ethical transmission of knowledge and to develop methods that can be applied in multiple indigenous contexts. Her next book project focuses on the first Hawaiian constitutional period 1839–1845. She is a co-author of the award-winning essay *Making Kin with the Machines,* and co-organizer of the Indigenous AI workshops.

Abeba BIRHANE is Senior Fellow in Trustworthy AI at the Mozilla Foundation. Her interdisciplinary research explores various broad themes in cognitive science, AI, complexity science, and theories of decoloniality. Birhane examines the challenges and pitfalls of computational models (and datasets) from a conceptual, empirical, and critical perspective.

Stephen CAVE is Director of the Leverhulme Centre for the Future of Intelligence at the University of Cambridge. He is author of a *New Scientist* Book of

the Year, *Immortality* (Penguin Random House, 2012), and *Should We Want to Live Forever* (Routledge, 2023), as well as co-editor of *AI Narratives: A History of Imagining Intelligent Machines* (Oxford University Press, 2020) and *Feminist AI* (Oxford University Press, 2023). He has written extensively for a broader public, including for the *New York Times* and *Atlantic*, and his work is featured in media around the world. He regularly advises governmental and international bodies on the ethics of AI.

Madeleine CHALMERS is Teaching Fellow in French at Durham University, UK. Her current book project *Unruly Technics* traces a genealogy of thinking and writing about technics that germinates in French avant-garde writings of the late nineteenth and early twentieth centuries, takes root in the mid-twentieth century in the work of major French philosophers such as Gilbert Simondon and Gilles Deleuze, and continues to bloom in the contemporary 'nonhuman turn' in Anglo–American theory that draws inspiration from those philosophers. Her work on technology and its intersections with French literature and thought has appeared in journals including *French Studies*, *Nottingham French Studies*, and *Dix-Neuf*.

Raul CRUZ is an artist and illustrator born in Mexico City. His work is a fusion of futuristic and sci-fi aesthetics with the art of Mesoamerican cultures such as the Maya and Aztecs, blending past and present with unpredictable futures. He has exhibited at universities, galleries, and cultural spaces in Mexico, New York, San Diego, Chicago, Los Angeles, Texas, Cuba, Italy, and Spain. His work appears in publications such as New York Illustrators Society, *Heavy Metal* magazine, SPECTRUM Fantastic Art, *The Future of Erotic Fantasy Art I* and *II*; has been compiled in a volume edited by the Universidad Autónoma Metropolitana de México and Editorial Milenio España; and features in two books on Mexican Fantastic Art. He participated in the Third Spectrum exhibition, gallery of The Society of Illustrators, in New York, and was selected as one of the 20 best artists in manual techniques of Fantastic Art by the book Fantasy + 3.

Kanta DIHAL is Lecturer in Science Communication at Imperial College London and Associate Fellow of the Leverhulme Centre for the Future of Intelligence, University of Cambridge. Her research focuses on science narratives, particularly those that emerge from conflict. She was Principal Investigator on the project 'Global AI Narratives' from 2018–2022. She is co-editor of *AI Narratives: A History of Imagining Intelligent Machines* (Oxford University Press, 2020) and has advised the World Economic Forum, the UK House of Lords, and the United Nations. She holds a DPhil from the University of Oxford on the communication of quantum physics.

Hans ESSELBORN is Emeritus Professor of the University of Cologne and specializes in the Enlightenment, Jean Paul, Early Twentieth Century German

Literature, Interculturality, and Science Fiction. He has been a visiting professor at Lawrence (Kansas), Nancy, Paris, Lyon, and Krakow. He is the author of *Georg Trakl* (PhD 1981) and *Das Universum der Bilder. Die Naturwissenschaft in den Schriften Jean Pauls* (habilitation 1989), *Die literarische Science Fiction* (2000), *Die Erfindung der Zukunft in der Literatur. Vom technisch-utopischen Zukunftsroman zur deutschen Science Fiction* (2019). He is the editor of *Utopie, Antiutopie und Science Fiction* (2003), *Ordnung und Kontingenz. Das kybernetische Model in den Künsten* (2009), and *Herbert W. Frankes SF-Werke* (since 2014).

Tomasz HOLLANEK is a design and technology ethics researcher with a background in cultural studies, philosophy, UX design, and communications. Currently, he is a Postdoctoral Research Fellow at the Leverhulme Centre for the Future of Intelligence at the University of Cambridge. His ongoing research explores the possibility of reconciling human-centric technology design principles with the goals of sustainable and restorative design for a more-than-human world. Previously, Tomasz was a Vice-Chancellor's PhD Scholar at Cambridge and a Visiting Research Fellow at the École normale supérieure in Paris. He has contributed to numerous research projects, including the Global AI Narratives Project at LCFI and the Ethics of Digitalization research program at the Berkman Klein Center for Internet and Society at Harvard.

Nik HYNEK is Professor specializing in Security Studies and affiliated with Metropolitan University Prague. He received his research doctorate from the University of Bradford and was previously affiliated with Saltzman Institute of War and Peace Studies (SIWPS) at Columbia University, the London School of Economics and Political Science (LSE), the Australian National University (ANU), Carleton University, and Ritsumeikan University. His latest monograph, co-authored with Anzhelika Solovyeva, is *Militarizing Artificial Intelligence: Theory, Technology, and Regulation* (Routledge, 2022).

Hirofumi KATSUNO is Associate Professor of Anthropology and Media Studies in the Faculty of Social Studies at Doshisha University, Kyoto. His primary research interest is the socio-cultural impact of new media technologies, particularly focusing on the formation of imagination, agency, and presence in technologically mediated environments in the age of AI and robotics. He is a co-organizer of the ethnographic research project Model Emotion (www.modelemotion.org) and a co-investigator of the international joint research project *Rule of Law in the Age of AI: Distributive Principles of Legal Liability for Multi-Species Societies*, funded by UKRI (UK) and JST (Japan).

So Young KIM is Professor and former head of the Graduate School of Science and Technology Policy and the director of the Korea Policy Center for the Fourth Industrial Revolution at KAIST. Her research deals with high-stake issues at the interface of science and technology (S&T) and public policy such

as research and development funding and evaluation, S&T workforce policy, science-based official development assistance, and the governance of emerging technologies. Her scholarly work includes publications in *International Organization*, *Journal of Asian Survey*, and *Science and Public Policy*, as well as the co-edited volumes of *Science and Technology Policy: Theories and Issues*, *The Specter of the Fourth Industrial Revolution*, and *A Return of the Future: COVID-19 and the Fourth Industrial Revolution* [in Korean]. As a public intellectual, she has served on numerous committees of government ministries and international entities, including the World Economic Forum's Global Future Councils.

Edward KING is Associate Professor in Portuguese and Latin American Studies at the University of Bristol. His research explores contemporary cultures of connectivity, with a particular focus on the use of speculative fiction and multimedia texts to expose and contest the shifting power dynamics of the digital age. His latest monograph *Twins and Recursion in Digital, Literary and Visual Cultures* (Bloomsbury Academic, 2022) traces the ways in which twins have been used in different fields of knowledge and representational practices to map the increasingly intricate entanglements of human subjects and their technological and natural environments. His co-authored book *Posthumanism and the Graphic Novel in Latin America* (UCL Press, 2017; with Joanna Page) studies the emergence of the graphic novel in Latin America as a uniquely powerful medium through which to explore the nature of twenty-first-century subjectivity and especially forms of embodiment or mediatization that bind humans to their nonhuman environment.

Bogna KONIOR is Assistant Professor at the Interactive Media Arts department at New York University Shanghai, where she teaches classes on emerging technologies, philosophy, humanities, and the arts. She also co-directs the university's Artificial Intelligence and Culture Research Centre. www.bognamk.com

Cheryl Julia LEE is Assistant Professor with the English department at Nanyang Technological University, Singapore, and critical editor of *prose.sg*. Her research interests include Singaporean and Southeast Asian literature. She has published articles in *Textual Practice* and *Asiatic*.

Jason Edward LEWIS is a digital media theorist, poet, and software designer. His research/creation projects explore computation as a creative and cultural material, and he is interested in developing intriguing new forms of expression by working on conceptual, critical, creative, and technical levels simultaneously. He is the University Research Chair in Computational Media and the Indigenous Future Imaginary as well Professor of Computation Arts at Concordia University, Montreal, where he directs the Initiative for Indigenous Futures, and co-directs the Indigenous Futures Research Centre and

the Aboriginal Territories in Cyberspace research network. He is Hawaiian and Samoan, born and raised in northern California.

Eleonora LIMA is Research Fellow in Digital Humanities at Trinity College Dublin, where she is part of the EU-funded project Knowledge Technologies for Democracy (KT4D). She is currently working on the first comprehensive history of computing in Italian literature. She has published on the interconnections between literature, science, and technology, as well as on Italian cinema and visual arts. Eleonora is a member of the Ethically Aligned Design for the Arts Committee, part of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. She holds a PhD in Italian and Media Studies (University of Wisconsin at Madison, 2015). She was previously an EU Marie Skłodowska-Curie Postdoctoral Fellow (2018–2020) and a Postdoctoral Fellow at the University of Toronto (2017–2018).

Yang LIU 劉 洋 is Senior Research Fellow in the Faculty of Philosophy and the Leverhulme Centre for the Future of Intelligence. His research interests include logic, foundations of probability, mathematical and philosophical decision theory, and philosophy of artificial intelligence (AI). His current research is supported by the Leverhulme Trust and the Isaac Newton Trust. Before moving to Cambridge, he received his PhD in Philosophy from Columbia University. He is co-founder and co-organizer of a seminar series on logic, probability, and games, as well as a conference series on decision theory and AI.

Graham MATTHEWS is Associate Professor in Contemporary Literature at Nanyang Technological University, Singapore. His research interests include the literary representation of science, medicine, and technology, and perceptions of risk. He is the author of *Will Self and Contemporary British Society* and has contributed to journals including *Modern Fiction Studies*, *Textual Practice*, *Literature & Medicine*, *Journal of Modern Literature*, *English Studies*, *Pedagogies*, and *Critique*.

Upamanyu Pablo MUKHERJEE is Professor of English and Comparative Literary Studies at Warwick University. He is the author of *Crime and Empire: The Colony in Nineteenth-Century Fictions of Crime* (Oxford University Press, 2003), *Natural Disasters and Victorian Empire: Famines, Fevers and the Literary Cultures of South Asia* (Palgrave Macmillan, 2013), and *Final Frontiers: Science Fiction and Techno-Science in Non-Aligned India* (Liverpool University Press, 2020).

Anton Ivanovich PERVUSHIN (Антон Иванович ПЕРВУШИН) holds an MSc in Engineering and is an independent scholar and science fiction writer. He is a member of the St. Petersburg Writers' Union, the St. Petersburg Union of Scientists, and the Russian Cosmonautics Federation. He researches the history of Soviet cosmonautics, science fiction, and astroculture. Among his

nonfiction books, the two most notable are *12 Myths about Soviet Science Fiction* (*12 мифов о советской фантастике*), and *Space Mythology: From Atlanteans on Mars to Lunar Conspiracy Theory* (*Космическая мифология: от марсианских атлантов до лунного заговора*).

Nagla RIZK is Professor of Economics and Founding Director of the Access to Knowledge for Development Center (A2K4D) at the American University in Cairo's School of Business. Her research area is the economics of knowledge, technology, and innovation with emphasis on human development in the digital economy in the Middle East and Africa. Her recent work focuses on the economics of data, AI, and inclusion. She authored the chapter 'Artificial Intelligence and Inequality in the Middle East' in *The Oxford Handbook of Ethics of Artificial Intelligence* (Markus Dubber et al. (eds.), Oxford University Press, 2020). She is a member of the Steering Committee of the Open African Innovation Research Partnership and is Faculty Associate at Harvard's Berkman Klein Center for Internet and Society.

Anzhelika SOLOVYEVA is Assistant Professor at the Department of Security Studies, Faculty of Social Sciences at Charles University in Prague. She has been affiliated with the Charles University Research Centre of Excellence dedicated to the topic of 'Human-Machine Nexus and the Implications for the International Order'. Her latest monograph, co-authored with Nik Hynek, is *Militarizing Artificial Intelligence: Theory, Technology, and Regulation* (Routledge, 2022).

SONG Bing is Senior Vice President of the Berggruen Institute and the Director of the Institute's China Center. She is responsible for the strategy and overall development of the Institute's China-based research, fellowship programmes, and public engagement. Prior to joining the Berggruen Institute, Bing came from a longstanding legal and banking career. Her research interests lie in the intersection of Chinese philosophy, frontier technologies, and governance. Recently she edited a volume on AI and Chinese philosophy, published by CITIC Press and Springer.

Apolline TAILLANDIER is a Postdoctoral Research Fellow at the Leverhulme Centre for the Future of Intelligence at the University of Cambridge and at the Center for Science and Thought at the University of Bonn. She also holds affiliated positions with the Department of Politics and International Studies at Cambridge and the Center for European Studies at Sciences Po in Paris. Previously, she was a Doctoral Fellow at the Max Planck Sciences Po Center on Coping with Instability in Market Societies, and held visiting fellowships at the University of California, Berkeley, and Cambridge. She is currently writing a monograph on the intellectual history of postwar transhumanism. In her new project, she traces feminist ideas in the history of AI and computer science in the US, the UK, and France.

Miao TIAN is Professor and department head at the Institute for the History of Natural Sciences, Chinese Academy of Sciences; her main research area is the History of Science and Technology in China, and she specializes in the history of mathematics and mechanics, the history of knowledge dissemination, and comparative studies of mathematics in Chinese and European history.

Nadine WEHEBA is Associate Director for Research at the Access to Knowledge for Development Center (A2K4D) at the School of Business of the American University in Cairo. Weheba's research interests are the political economy of information communication technologies, innovation, and development. She studies new business models in the digital economy, new technologies, such as artificial intelligence, and inclusion in the context of Egypt, Africa, and the region. She received her MSc in Development Studies from the School of Oriental and African Studies (SOAS) at the University of London and holds a BA in Economics from The American University in Cairo.

Daniel WHITE is Senior Research Associate in the Department of Social Anthropology at the University of Cambridge. He co-organizes an ethnographic research project called Model Emotion, in which he works across disciplines with anthropologists, psychologists, computer scientists, and robotics engineers to trace and critique how theoretical models of emotion are built into machines with emotional capacities for care and well-being. His publications, podcasts, and other projects document how different cultural imaginaries, largely from Japan, are shaping diverse futures for artificial intelligence and can be found at www.modelemotion.org.

Yan WU is Professor at the Center for Humanities at Southern University of Science and Technology in Shenzhen. He is also a science fiction writer, winner of the Chinese Galaxy and Nebula Awards, and the 2020 Clareson Award of the Science Fiction Research Association (SFRA).

Feng ZHANG is a science fiction researcher. He is visiting researcher at the Humanities Center of Southern University of Science and Technology in Shenzhen, chief researcher of Shenzhen Science & Fantasy Growth Foundation, honorary Assistant Professor of the University of Hong Kong, Secretary-General of the World Chinese Science Fiction Association, and editor-in-chief of *Nebula Science Fiction Review*. His research covers the history of Chinese science fiction, development of the science fiction industry, science fiction and urban development, and science fiction and technological innovation.

INTRODUCTION

1

How the World Sees Intelligent Machines

Introduction

Stephen Cave and Kanta Dihal

1.1 Myths and Realities

Artificial intelligence (AI) was a cultural phenomenon long before it was a technological one. In some cultures, visions of intelligent machines go back centuries, even millennia (Liveley and Thomas, 2020; Zhang and Tian, Chapter 22 this volume). Such visions spread with industrialization until, by the twentieth century, a future with such machines was being richly imagined around the world. When the term 'artificial intelligence' was coined in the US in 1956, it was not to name a new invention, but rather to express a determination to realize a long-standing fantasy (Cave et al., Chapter 2 this volume).

Some would say that AI is still a cultural phenomenon and not a technological one: that for all the innovations in computing, and all the hype in industry and policy, no existing systems deserve to be called truly intelligent (e.g. Simkoff and Mahdavi, 2019; Taulli, 2019). Others, however, argue that we are surrounded by AI: that it pervades our daily lives—through our smartphones, the online services we use, and the hidden systems that govern us (e.g. Smith, 2021). It is hard to think of another technology in history about which such a debate could be had—a debate about whether it is everywhere, or nowhere at all. That it can be held about AI is a testament to its mythic quality.

Of course, innovation in digital technology is real and rapid, and it is continually in interplay with this long-standing mythology of intelligent machines. This cultural backdrop shapes what motivates funders and engineers, how products are designed, whether and by whom technologies are taken up,

how they are regulated, and so on. The nodes of production—academia and industry, media and policy—are interwoven and mutually influential.

Take, for example, the 2014 Hollywood film *Transcendence*, which tells the story of AI expert Will Caster (Johnny Depp), who has his mind 'uploaded' onto an AI system by his wife Evelyn (Rebecca Hall) and his colleague Max Waters (Paul Bettany) (Pfister, 2014). During the film's opening weekend, three well-known scientists, Stephen Hawking, Max Tegmark, and Stuart Russell, published a *Huffington Post* article titled 'Transcending Complacency on Superintelligent Machines' (Hawking et al., 2014). Later republished with the name of Nobel Prize winner Frank Wilczek added, the article argues that:

> As the Hollywood blockbuster *Transcendence* debuts this weekend with Johnny Depp, Morgan Freeman and clashing visions for the future of humanity, it's tempting to dismiss the notion of highly intelligent machines as mere science fiction. But this would be a mistake, and potentially our worst mistake ever. (Hawking et al., 2014)

The article is intended to convince policy makers and other publics of the importance of AI. That same month, Tegmark founded the Future of Life Institute (FLI), which aims 'to ensure that tomorrow's most powerful technologies are beneficial for humanity'. It is partly funded by Elon Musk, who is a real-life technology magnate and pioneer of AI-driven cars. Musk also appears in a cameo in *Transcendence*, as an audience member of a lecture on AI. In addition, Morgan Freeman, who plays one of the film's heroes, sits on FLI's Board, alongside Musk. The film therefore perfectly reflects what we call the 'Californian feedback loop': the multiple entanglements of Hollywood and the broader culture it reflects, of academic research and Silicon Valley industrial production, of narrative and the billionaire-funded fight to shape the future. Stories such as *Transcendence* co-construct what AI is understood to be, embedding or disputing existing attitudes and approaches.

But crucially, these attitudes and approaches are not the same around the world. They are shaped by the particular histories, philosophies, ideologies, religions, narrative traditions, and economic structures of different countries, cultures, and peoples. *Transcendence* is a product of the US, and trades in well-worn Hollywood tropes, such as AI as the ultimate technology (and therefore the ultimate solution to all problems, i.e. pollution, the energy crisis, disease), yet at the same time the ultimate threat to humanity; the reduction of the individual to data and computation, and therefore the possibility of digital immortality; and the lone male genius scientist and the subordinate female who must in some way sacrifice herself. Individually, these tropes are not unique to Hollywood—each can be found elsewhere—but collectively they form the distinctive mythology of AI in America.

In our 2020 volume (with Sarah Dillon) *AI Narratives: a History of Imaginative Thinking about Intelligent Machines*, we surveyed the predominant themes in anglophone Western portrayals of AI (Cave et al., 2020a). In this volume, *Imagining AI: How the World Sees Intelligent Machines*, we look beyond the mainstream traditions of the US and UK to how other cultures have conceived of this technology. There are a number of motivations for this.

First, AI is now a global phenomenon. The term originated in the US, and much of the technology continues to be developed there. But those systems are being taken up around the world, and other countries are scrambling to develop their own AI industries. Each will do so informed by their own mythologies of AI—the concatenation of different stories and ideologies that shape their expectations and anxieties around what this technology can be. Understanding how AI will develop requires, therefore, an understanding of the many sites in which its story is unfolding.

Second, the debate about how AI is developed responsibly and governed has been dominated by anglophone actors. This is starting to change, as more countries develop their own AI strategies. But they are entering a space shaped by anglophone Western assumptions. There is a risk that efforts to regulate AI will fail as these assumptions are insensitive to different cultural contexts, or that solutions imposed will unwittingly prejudice some traditions. It is imperative, therefore, to develop a better understanding of the diversity of views about what AI should be.

Third, we hope this comparative approach will shed new light for scholars, whether of the anglophone tradition or others. Seeing where cultures share or differ in their approaches to AI gives insight into the forces that shape these traditions. For example, we could look at what narratives are common to capitalist countries, or colonizing ones. Those who live and work outside the anglophone tradition may discover in this collection narratives from elsewhere in the world that resemble their own perspectives and concerns much more closely. For instance, we have included chapters on anti-colonial or decolonial AI narratives from Latin America, South Asia, Sub-Saharan Africa, and Indigenous North American nations. They show that this resistance can take many shapes, but shared themes resonate across continents: the platforming of non-Western knowledge and forms of knowing with respect to AI 'to critically reflect on such designations as "advanced" and "backward"' (Mukherjee, Chapter 15 this volume); the re-appropriation of technologies from the anglophone West for purposes and art forms their original creators did not envision or even explicitly excluded; and the deployment of fictional and non-fictional narratives of AI for the explicit purpose of post-colonial nation building. At the same time, we see narratives about 'catching up' to other countries, particularly the US and UK, from South Korea, India, and Russia alike. Our contributors also show how narratives have historically resisted and

supported a wide range of ideologies, from communism in mid-twentieth-century China, the Soviet Union, and Italy, to neoliberalism in Chile, and technocracy in Singapore.

Fourth, each cultural perspective is limited and particular, privileging some within that culture and prejudicing others. Certainly this is true of the anglophone Western tradition, which, as we and others have noted elsewhere, is inflected by ideologies of racial, gender, and class hierarchy, by polarizing dichotomies, and a fixation on domination and control (Cave, 2020; Cave and Dihal, 2020, 2019). There have therefore been many recent calls for new imaginaries of technology. Race scholar Ruha Benjamin quotes producer and activist Kamal Sinclair, 'story and narrative are the code for humanity's operating system', before calling to 'reimagine science and technology for liberatory ends' (Benjamin, 2019, pp. 193–5). For any given culture, in attempting to imagine how the future can be different from what its mainstream prescribes, we hope it will be illuminating to consider how other cultures imagine AI otherwise.

Of course, the limitations of Western narratives will not be solved simply by adding a Chinese work to one's reading list, consulting an Indigenous person, or mentioning Ubuntu ethics at a workshop on AI regulation. Tokenistic approaches to the diversification, de-localization, and decolonization of AI have been attempted and criticized repeatedly in the past decade (Birhane and Guest, 2020; Snell, 2020). Nonetheless, as Priyamvada Gopal points out in her discussion of curriculum decolonization, 'diversity is, in fact, important both for its own sake and for pedagogical and intellectual reasons—a largely white or largely male curriculum is not politically incorrect, as is often believed, but intellectually unsound' (Gopal, 2021, p. 877). *Imagining AI* addresses this issue of diversity on two levels. First, at the level of the narratives themselves: this collection presents a wealth of novels, films, comics, visual art, and other media from outside the anglophone West, many—though not all—of which are largely unknown outside their region. Second, through intellectual engagement with these sources: all of our authors engage in analysis and contextualization of the narratives they bring to the table, bringing out motifs and arguments that introduce new themes to the now-growing field of AI narratives, or shed new light on existing themes.

It is with these four motivations in mind that we started the Global AI Narratives research project at the University of Cambridge's Leverhulme Centre for the Future of Intelligence, in collaboration with nine partner institutions on six continents. The project aimed to understand and analyse how different cultures and regions perceive the risks and benefits of AI, and the influences that are shaping those perceptions. We convened a series of twenty workshops across the globe between 2018 and 2021 and built an international network of experts on portrayals and perceptions of AI beyond the English-speaking West,

many relating these diverse visions to pressing questions of AI ethics and governance. *Imagining AI* is the product of these workshops, at which many of our contributors first shared the work collected here.

Our highly interdisciplinary group of contributors consists of leading experts from academia and the arts, selected for their expertise on a given region or culture. As in *AI Narratives*, the discourses they analyse range across myth and legend; literature and film, including science fiction; and nonfiction such as policy documents and government propaganda. We do not draw a clear distinction between fiction and nonfiction in examining AI narratives: as the chapters show in detail, and as our previous work has argued, fictional and nonfictional AI narratives together form 'sociotechnical imaginaries' (Jasanoff, 2015) that shape public perceptions of AI on the one hand, and the direction technology development takes on the other (Cave et al., 2020b, 2019; Cave and Dihal, 2020, 2019).

It is of course not possible to achieve a complete survey of AI imaginaries in all the many regions and cultures of the world. While we have aimed to be broadly geographically representative, we encountered the problem that some regions are much more intensely studied than others. We were therefore able to find a large number of excellent scholars writing on, for example, Continental Europe or China. But it proved much harder to find contributors on, for example, the imaginaries of Sub-Saharan Africa or India. For this edited collection, this problem was exacerbated by the fact that the most underrepresented cultures and regions were also those worst hit by the Covid-19 pandemic, which forced some of our initial contributors to instead attend to more pressing personal circumstances. It is a source of great regret to us that we cannot better represent these regions in this volume, and we hope that it will perhaps inspire more work on such important but understudied areas. Despite these lacunae, we hope that readers will enjoy the unprecedented wealth of perspectives on AI that our contributors share in the following chapters.

1.2 Overview of the Book

The chapters of this book are clustered geographically.[1] We have chosen this arrangement, rather than a focus on themes or historical periods, to allow us to group chapters from similar linguistic or cultural backgrounds. For example, we have four short chapters on China that each cover a different historical period, together highlighting a range of aspects of a deep narrative history. Of course, different regions or countries are not pristine islands of cultural uniqueness: each is shaped by new immigrants or its own diasporas, by the arrival of new religions and ideologies from elsewhere, by layers of conquest and

settlement, by absorption of or reaction to the cultural exports of more powerful actors. Many of our contributors have aimed to tease out the uniqueness of the narratives they discuss, taking account of the diverse influences that have shaped them, and the diversity within the traditions they represent.

We open this collection with a comparative chapter focusing not on narratives themselves, but rather on the words narrators have at their disposal for expressing the concepts covered in this book. Chapter 2 investigates the terms used to denote intelligent machines in five language groups—Germanic, Slavonic, Romance, Chinese, Japanese—to elucidate the value-laden character of the terminology as well as explore its influence on the perceptions of contemporary AI technologies around the globe. The chapter traces the implied meanings behind the technology's original name—its links to specific conceptions of (human) intelligence and reason in the West, as well as the cultural history of artificiality in Europe—and investigates what happens to those connotations when the term is translated and popularized in non-anglophone and non-Western cultures. It examines how, in some cases, other cultural-linguistic groups have reproduced these associations, while in others, the English term is widely used, but with wholly different connotations.

Part I encompasses Europe, including Russia. While most of the chapters focus on the twentieth century, the section opens with Chapter 3, which looks at late-nineteenth-century French culture and its explorations of the birth of the technologies that underpin our contemporary life. Madeleine Chalmers proposes that there is a uniquely French mode of approaching what we would now recognize as AI: a mode that is distinct from the fragmentation of Anglo–American modernism or the fascist sex-and-speed-machines of Italian futurism. The French touch lies in its meta-reflexive narrative play, which allows it to test out technological and political scenarios. Chalmers asks how looking back to the imaginings of the past can help us to find critical distance, as debates about the existential and ethical risks of developing technologies permeate mainstream culture faster than we can conceptualize their ramifications.

Chapter 4 looks at Italy and its 1980s counterculture, whose new political sensitivity found expression through images of androids and cyborgs that embodied the new rebellious subject of the 1980s: fluid, highly technological, urban, more at ease with pop culture references than with Marxist ones. Eleonora Lima introduces its most prominent example, the comic strip *Ranxerox* (1978–1986), whose protagonist is a violent, uninhibited, and amoral android, created by a group of dissident college students by assembling pieces from a photocopy machine—a Rank Xerox—as a weapon against the police. The third of our Western European chapters is written by Hans Esselborn, who in Chapter 5 investigates two phases in thinking about AI in science fiction written in German. First, the 1960s cybernetics revolution that inspired Austrian author H.

W. Franke to write several innovative texts on supercomputers governing and controlling societies. Second, in the twenty-first century, public discussions on AI and its practical achievements have encouraged several German science fiction authors to write novels about the prospects and dangers of self-learning programs, including their awakening and achieving world supremacy through networking and computing power.

Part I then moves towards Eastern Europe. In Chapter 6, Bogna Konior explores the works of Stanisław Lem, particularly his *Summa Technologiae* (1964) in which he reveals his preoccupation with the long-term techno-biological evolution of humanity. Focusing on artificial life and 'the breeding of information' as examples of existential, rather than utilitarian, technologies, Konior's chapter explores Lem's unique reading of cybernetics, and reflects on its relationship to the dominant militaristic model of cybernetics during the Cold War. Simultaneously, the chapter reflects on the place of the *Summa* within Lem's oeuvre, and within the context of the 1960s Communist bloc and communism's own tautological ideas of political evolution of civilizations.

Chapter 7 zooms out to identify trends concerning AI in science fiction of the USSR (1922–1991). Anton Pervushin identifies three trends: the use of intelligent machines as a cheap workforce, thus making them the new proletariat of a future world; intelligent machines as an instrument, used by bourgeois society to suppress and further exploit the proletariat; and society compelling intelligent machines to acquire emotions and understand the taboos and laws that are considered a norm in this society. Pervushin's chapter concludes that these three trends were eventually consolidated into two main stereotypes of the intelligent machine in Soviet science fiction: the evil robot and the funny robot. Moving beyond the Soviet period, Chapter 8 by Anzhelika Solovyeva and Nik Hynek traces a hundred-year history of imagining intelligent machines in Russia to the present day. Contextualized by references to nascent representations of automata in the Russian Empire, they distinguish three formative phases based on transformations in science, politics, literature, and visual arts. The earliest attempts to conceptualize the human–machine nexus originated in the Bolshevik Revolution and the Russian avant-garde. In the second phase (mid- to late twentieth century), the Soviets' progress in computer engineering, cybernetics, and AI paved the way for machine automation and facilitated fantasies about intelligent machines. In the third phase (post-Soviet Union and especially in Putin's Russia) both popular culture and research envision a future with lifelike machines.

From Europe we move across the Atlantic for Part II, which focuses on the Americas and Pacific. Our own Chapter 9 opens this section and is the only chapter in this book whose focus is mainstream anglophone Western conceptions of AI. Surveying Hollywood films and science fiction literature,

we argue that American AI narratives tend to be extreme: either utopian or (more often) dystopian. We illustrate the utopian strand with an analysis of Isaac Asimov's 'The Last Question' (1956), situating it within an American ideology of technology that associates technological mastery with ontological superiority, the legitimation of settler-colonialism and the pursuit of a 'second creation' on earth. We then examine the inherent instability of this vision, which explains why it can tip so readily into the dystopian. This includes fears of decadence and the slave-machine uprising, and revolves around loss of control. We then examine three degrees of loss of control: from willingly giving up autonomy to the machine in the 2008 film *WALL-E*, to value misalignment in Martin Caidin's *The God Machine* (1968), to utter extermination in the *Terminator* franchise.

Chapter 10 concerns Brazil, and Edward King focuses on the use of Afrofuturist aesthetics to produce what Ruha Benjamin (2019) describes as 'subversive countercodings' of the dominant practices of racialization. Artists working in various media—including filmmaker Adirley Quierós and visual artists Vitória Cribb and the Afrobapho Collective—have adapted a science fiction aesthetic developed in the 1960s and 1970s US by musicians such as Sun Ra and George Clinton to challenge the normalized disparity between Black culture and science and technology. Although varied in their approaches, these practitioners are united in their use of Afrofuturism to contest what André Brock, Jr. (2020) identifies as the conflation of online identity with whiteness, 'even as whiteness is itself signified as a universal, raceless, technocultural identity'. In the process, they propose alternative conceptions of the human as intimately imbricated with computer systems that do not repeat pseudo-universal versions of modernity that are predicated on anti-Blackness.

Equally focused on subversion and resistance is the selection of artworks by Mexican artist Raúl Cruz in Chapter 11 and his explanation about how his work fuses science fiction tropes of AI with the art of Mesoamerican cultures, particularly the Maya and Aztecs. He argues that, although new technologies can cause global homogenization, they will not erase the traditions, myths, and customs of Latin America. The cultural legacy of the peoples of this continent will merge with technological advances, giving rise to distinctive visions of intelligent machines, including robots with Aztec aesthetics, cybernetic catrinas, and Quetzalcoatl ships. Cruz's works are distinctively Latin American visions of intelligent machines that merge the fantastic, the literary, the mythic, and the mechanical. In Chapter 12, Macarena Areco moves to an analysis of recent Chilean science fiction featuring AI. She examines how the imagined spaces, technologies, and subjectivities of the neoliberal era are condensed in representations of AI in the works of Chilean author Jorge Baradit.

The final two chapters of Part II offer Indigenous perspectives from North America and the Pacific/Moananuiākea. In Chapter 13, Jason Edward Lewis

describes the work done at the 2019 Indigenous Protocol and AI Workshops and the ways in which this work led to imagining what one's relationship to AI should be from an Indigenous perspective, and what Indigenous AI might look like. Then, in Chapter 14, Noelani Arista introduces the concept of Maoli Intelligence, or that corpus of ancestral knowledge where the collective *'ike* (knowledge, traditional knowledge) of Kānaka Maoli, Hawaiian people, continues to be accessed. She presents a compelling view of Indigenous data sovereignty and provides examples from her *lāhui* (Nation, people) built upon a robust oral to textual, filmic, and material culture 'archive'. Her chapter concludes with her work with an all-Indigenous team of data and computer scientists to build a prototype of an image recognition application forecasting what can be done with Maoli Intelligence that can be applied on a broader scale.

Part III covers Africa, the Middle East, and South Asia. In Chapter 15, Upamanyu Pablo Mukherjee discusses AI in the fiction of the Indian writer, filmmaker, musician, and designer Satyajit Ray, who frequently used science fiction to interrogate some of the key premises of European 'Enlightenment'. In his robot stories in particular, Ray complicated the presumed circuit between reason and intelligence by introducing emotions as a key but variable component. Mukherjee examines Ray's science fiction to wonder whether a different account of AI can be sketched from the post-colonial moment of the mid-twentieth century.

Abeba Birhane's Chapter 16 concerns what she terms 'the AI invasion of Africa': the ways in which Western tech monopolies, with their desire to dominate, control, and influence social, political, and cultural discourse, share common characteristics with traditional colonialism. While traditional colonialism used brute force domination, colonialism in the age of AI takes the form of 'state-of-the-art algorithms' and 'AI solutions' to social problems. Birhane argues that not only is Western-developed AI unfit for African problems, the West's algorithmic invasion simultaneously impoverishes development of local products while also leaving the continent dependent on Western software and infrastructure. Then, in Chapter 17, Rachel Adams investigates what it means and requires to decolonize AI and explores an alternate cultural perspective on nonhuman intelligence from that portrayed in the more well-known canon of the Western imaginary. Her enquiry focuses on the transgendered *ogbanje* of Nigerian Yoruba and Igbo cultural traditions—a changeling child or reincarnated spirit. Through engagement with the works of Chinwe and Chinua Achebe and Akwaeke Emezi, she explores the roles anthropomorphism, representation, and gender play in making intelligence culturally identifiable; whether this offers an alternative imaginary for transcending the normative binaries that AI fortifies; and what kind of politics this requires.

To close this section, Chapter 18 is a collaboration between the Access to Knowledge for Development Center at the American University in Cairo and the Leverhulme Centre for the Future of Intelligence at Cambridge and covers the Middle East and North Africa (MENA) region. While people in

MENA have been imagining intelligent machines since the Islamic Golden Age (9th–14th centuries), Western perceptions of success, development, progress, and industrialization influence the hopes and dreams for the future of technology today. In this chapter, we analyse the various factors that make the MENA region a unique environment for imagining futures with intelligent machines, mapping local visions of technological progress onto the region's complicated past, as well as contemporary economic and political struggles.

Part IV focuses on East and South East Asia, regions frequently portrayed as radically different from the anglophone West in its portrayals and perceptions of AI. Chapter 19 by Hirofumi Katsuno and Daniel White investigates the political dimensions of this idea with regard to Japan. According to this view, whereas in the Western robotic imaginary intelligent machines signify a threat to humanity, in the Japanese imaginary machines are partners to humans, offering their technological skills to address problems of shared human–robot concern. Based on ethnographic fieldwork in Japan, their chapter analyses the imaginaries of animism, animation, and animacy among Japanese roboticists building emotionally intelligent companion robots. The chapter argues that while emphases on emotionality in Japanese AI narratives challenge distinctions between reason and emotion in anglophone AI research, even more importantly, observations on the cultural politics of these distinctions diversifies the notion of 'culture' itself.

Chapter 20 focuses on South Korean narratives, particularly policy discourse on AI after the 2016 AlphaGo match. So Young Kim argues that, held right after the 2016 Davos Forum that popularized the term 'the Fourth Industrial Revolution', the match rendered the esoteric technology a household name, with no day going by without AI featuring in national news media. The following four years saw a huge outpouring of news, events, studies, artworks, policies, and more on AI—including the creation of the Presidential Committee on the Fourth Industrial Revolution (2017), the declaration of the National AI Strategy (2019), and the announcement of the AI-driven Digital New Deal (2020). Kim's chapter explores the central features of South Korean policy discourse on AI, revealing the imprints of developmental state legacies that permeate every corner of Korean society.

The next four short chapters each address a different aspect of AI narratives in China. First, in Chapter 21 Bing Song explains which philosophical ideas and practices may have shaped Chinese thinking towards the development of frontier technologies and the approach to human–machine relationships, focusing on the three dominant schools: Confucianism, Daoism, and Buddhism. Next, in Chapter 22 Baichun Zhang and Miao Tian discuss the attitudes towards new technologies in pre-Qin Dynasty China (pre-221 BCE) and look at a range of attitudes towards both imaginary technologies that we would now

call AI, and real innovations in battlefield technologies. They argue that pragmatic motivations tended to trump philosophical concerns. In Chapter 23, Yan Wu fast-forwards to twentieth-century China and discusses science fiction from 1949 to 1983. While advanced computers and robots featured early in Chinese science fiction, in these stories the AI characters are largely human-like assistants, chiefly collecting data or doing manual labour. The reform and opening-up of 1978 caused the number and quality of robot-themed works to balloon. Important themes in these new works included the idea that the growth of AI requires a suitable environment, stable family relationship, and social adaptation.

Finally, in Chapter 24 Feng Zhang explores the newest generation of AI-themed science fiction from China. On the one hand, contemporary authors are aware that the 'soul' of current AI, to a large extent, comes from machine-learning algorithms. As a result, their works often highlight the existence and implementation of algorithms, bringing manoeuvrability and credibility to the portrayal of AI. On the other hand, the authors prefer to focus on the conflicts and contradictions in emotions, ethics, and morality caused by AI that penetrate human life. If earlier AI-themed science fiction is like a distant robot fable, recent narratives of AI assume contemporary and practical significance.

Our final chapter 25, is the most comprehensive of the surveys of AI narratives corpora discussed in this collection. Cheryl Julia Lee and Graham Matthews present an evaluation of the entire history of AI narratives in Singapore. They argue that fictional portrayals of AI in Singapore problematize the Smart Nation initiative of Prime Minister Lee Hsien Loong, which simultaneously valorizes and objectifies the human. The narratives typically eschew the tropes of existential risk—annihilation and enslavement—and present instead visions of coexistence and mutual dependency between humans and machines. The authors question whether narratives of AI in Singapore promote visions of an AI-driven future that we should accept or be wary of surrendering to—and the extent to which these conclusions depend on a dichotomy between 'Western technology' and 'traditional Eastern values'.

Together, these essays show that the imaginary of intelligent machines is an incredibly rich site for exploring the varied influences—historical, political, and artistic—that shape how different cultures reconcile the demands of being human in a technological age.

Endnote

1. For consistency, this overview gives all author names in first name–family name order.

References

Benjamin, R. (2019) *Race after technology: Abolitionist tools for the new Jim code*. Medford, MA: Polity Press.

Birhane, A. and Guest, O. (2020) 'Towards decolonising computational sciences', *ArXiv* [online], 29 September. Available at: https://arxiv.org/abs/2009.14258

Brock, Jr., A. (2020) *Distributed Blackness: African–American cybercultures*. New York: New York University Press.

Cave, S. (2020) 'The problem with intelligence: Its value-laden history and the future of AI', in *Proceedings of the 2020 AAAI/ACM conference on AI, ethics, and society, AIES '20*. New York: ACM, pp. 29–35. https://doi.org/10.1145/3375627.3375813

Cave, S., Coughlan, K., and Dihal, K. (2019) '"Scary robots": Examining public responses to AI', in *Proceedings of the 2019 AAAI/ACM conference on AI, ethics, and society, AIES '19*. New York: ACM, pp. 331–7. https://doi.org/10.1145/3306618.3314232

Cave, S. and Dihal, K. (2019) 'Hopes and fears for intelligent machines in fiction and reality', *Nature Machine Intelligence*, 1, pp. 74–8. Available at: https://doi.org/10.1038/s42256-019-0020-9

Cave, S. and Dihal, K. (2020) 'The whiteness of AI', *Philosophy & Technology*, 33, pp. 685–703. Available at: https://doi.org/10.1007/s13347-020-00415-6

Cave, S., Dihal, K., and Dillon, S. (eds.) (2020a) *AI narratives: A history of imaginative thinking about intelligent machines*. Oxford: Oxford University Press.

Cave, S., Dihal, K., and Dillon, S. (2020b) 'Introduction: Imagining AI', in Cave, S., Dihal, K., and Dillon, S. (eds.) *AI narratives: A history of imaginative thinking about intelligent machines*. Oxford: Oxford University Press, pp. 1–21.

Gopal, P. (2021) 'On Decolonisation and the University', *Textual Practice*, 35(6), pp. 873–99. Available at: https://doi.org/10.1080/0950236X.2021.1929561

Hawking, S., et al. (2014) 'Transcending complacency on superintelligent machines', *Huffington Post* [online]. Available at: https://www.huffingtonpost.com/stephen-hawking/artificial-intelligence_b_5174265.html (Accessed: 21 June 2019).

Jasanoff, S. (2015) 'Future imperfect: Science, technology, and the imaginations of modernity', in Jasanoff, S. and Kim, S.-H. (eds.) *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*. Chicago: University of Chicago Press, pp. 1–33.

Liveley, G. and Thomas, S. (2020) 'Homer's intelligent machines: AI in Antiquity', in *AI narratives: A history of imaginative thinking about intelligent machines*. Oxford: Oxford University Press, pp. 25–48.

Transcendence. (2014) Directed by Wally Pfister [Film]. Warner Bros. Pictures.

Simkoff, M. and Mahdavi, A. (2019) 'AI doesn't actually exist yet', *Scientific American* [online], 12 November. Available at: https://blogs.scientificamerican.com/observations/ai-doesnt-actually-exist-yet/ (Accessed 22 April 2022).

Smith, C. S. (2021) 'A.I. here, there, everywhere', *The New York Times* [online], 9 March. Available at: https://www.nytimes.com/2021/02/23/technology/ai-innovation-privacy-seniors-education.html

Snell, T. (2020) 'Tokenism is not "progress"', *POCIT People Color Tech* [online]. Available at: https://peopleofcolorintech.com/articles/tokenism-is-not-progress/ (Accessed 22 April 2022).

Taulli, T. (2019) 'Splunk CEO: Artificial intelligence does not exist today', *Forbes* [online], 25 October Available at: https://www.forbes.com/sites/tomtaulli/2019/10/25/splunk-ceo—artificial-intelligence-does-not-exist-today/ (Accessed 22 April 2022).

2

The Meanings of AI

A Cross-Cultural Comparison

Stephen Cave, Kanta Dihal, Tomasz Hollanek, Hirofumi Katsuno, Yang Liu, Apolline Taillandier, and Daniel White

2.1 Introduction

'Too flashy', 'kind of phony', 'attendees balked at the term': since its coining in 1955, the term 'artificial intelligence' (AI) has been contentious for the connotations of both its elements—'artificial' and 'intelligence' (McCorduck, 2004, pp. 114–16). Invented as an eye-catching, attention-grabbing term for a new scientific field, it certainly can be considered successful, with AI—or its equivalent in other languages—dominating headlines around the world. But what terms are used to refer to AI in those other languages? Are they equally contentious? Are their connotations different, and if so, what does this mean for the hopes, fears, and expectations surrounding AI?

This chapter traces the history of the technology's original name in English and the implied meanings behind it—its links to specific conceptions of (human) intelligence and reason in the West, as well as the cultural history of artificiality in Europe—and investigates what happens to those connotations when the term is translated and popularized in non-Anglophone and non-Western cultures. It examines how in some cases other cultural-linguistic groups have reproduced these associations (for example, the relation between art and artifice is preserved in many European languages, such as the German *Kunst* [art]—*künstliche Intelligenz*); while in others, the English acronym is widely used alongside a term with quite different connotations (as in Japan, where 'AI' is used as a buzzword alongside *jinkō chinō*).

The chapter investigates the terms used to denote intelligent machines in five languages and language groups—Germanic (first English, then Continental Germanic languages), Romance, Slavonic, Japanese, and Chinese. These have been chosen because they are or contain languages that are both among the most spoken globally and which are spoken in countries that are major

sites of AI development. In what follows we explore these terms in each linguistic context in turn, elucidating the ways in which they are value-laden, and how they might reflect and shape attitudes to AI in the countries where they are spoken.

2.2 Origins of the term 'artificial intelligence'

Famously, the term 'artificial intelligence', as a description of the use of machines to emulate some aspect of human thought, was coined by John McCarthy and a handful of colleagues in 1955. It was the term McCarthy chose to convene a group of leading figures in this nascent field at an event that came to be called 'The Dartmouth Summer Research Project on Artificial Intelligence', held at Dartmouth College, New Hampshire, USA, from June to August 1956. In her history of AI based on interviews with many of the protagonists, Pamela McCorduck relates:

> many attendees balked at that term, invented by McCarthy. 'I won't swear that I hadn't seen it before', he recalls, 'but artificial intelligence wasn't a prominent phrase particularly. Someone may have used it in a paper or a conversation or something like that, but there were many other words that were current at the time. The Dartmouth Conference made that phrase dominate the others'.
>
> (McCorduck, 2004, pp. 114–15)

It was far from inevitable that this term would label the field for the coming decades. As McCarthy mentioned, there were many plausible alternatives. According to historian of AI Jonnie Penn, these included:

> 'engineering psychology', 'applied epistemology', 'neural cybernetics', 'non-numerical computing', 'neuraldynamics', 'advanced automatic programming', 'automatic coding', 'fully automatic programming', 'hypothetical automata', and 'machine intelligence'.
>
> (Penn, 2020, p. 206 fn)

It is noteworthy that Alan Turing had used the term 'intelligence' in the title of his now-famous 1950 paper 'Computing Machinery and Intelligence'. But he does not use the term AI, and indeed, apart from in the title, does not refer to 'intelligence' at all, but focuses on thought and thinking. McCarthy later said that Turing was not a great influence on those at Dartmouth (McCarthy, 1988, p. 7).

John McCarthy has made a number of comments on why he chose the term 'artificial intelligence'. First, he had at the time a collaboration with Claude Shannon, already well known for his information theory, which included

jointly editing a book that was published as *Automata Studies* (Shannon and McCarthy, 1956). McCarthy had wanted a broader title, but Shannon had dismissed the alternatives as 'too flashy' (McCorduck, 2004, p. 115). McCarthy was, however, disappointed with the narrowness of the papers that the 'Automata Studies' title attracted, and was determined not to repeat that mistake, as he saw it, for the Dartmouth gathering.

Second, writing in 1988, McCarthy explained why he avoided another plausible alternative:

> As for myself, one of the reasons for inventing the term 'artificial intelligence' was to escape association with 'cybernetics'. Its concentration on analog feedback seemed misguided, and I wished to avoid having either to accept Norbert . . . Wiener as a guru or having to argue with him.
>
> (McCarthy, 1988, p. 6)

Bringing together these two motivations, we can see that McCarthy recognized that the term 'artificial intelligence' had a number of advantages: it was attention grabbing—perhaps even 'flashy'; and even while it drew on traditions such as cybernetics, it was not tied to any one of them, but could encompass a wide range of approaches to the emulation of human thought.

McCarthy managed to persuade his collaborators Marvin Minsky, Nathaniel Rochester, and Shannon of the utility of this term, and in 1955 they together submitted to the Rockefeller Foundation a request for funds for the Dartmouth event. They summarized their intentions thus:

> We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.
>
> (McCarthy et al., 1955)

As McCorduck records, some of the participants at the Dartmouth event did not take to the new appellation: 'Neither [Allen] Newell nor [Herbert] Simon liked the phrase, and called their own work complex information processing for years thereafter' (McCorduck, 2004, p. 115). Later in 1956, Minsky, who went on to found MIT's AI Lab, addressed some of these critics in the report *Heuristic Aspects of the Artificial Intelligence Problem* for the US Department of Defense. He points out that some people regard intelligence as 'a kind of gift which can not be performed by a machine even in principle' (Minsky, 1956, p. iii). He

even notes a phenomenon that plagued the field of AI for decades thereafter—that we consider a certain behaviour intelligent only until it is accomplished by a machine, at which point the goalposts are promptly shifted. Five years later, in 1961, Minsky still felt the need to preface his paper 'Steps toward Artificial Intelligence' with a defence of the term 'artificial intelligence', which he places in scare quotes. But as Penn notes: 'Only a decade later . . . the MIT Artificial Intelligence Laboratory—scare quote free—was fully operational with a budget purported to be in the millions' (Penn, 2020, p. 206).

The path to the present prevalence of AI was not smooth: while the term's popularity rose through the 1960s, it fell in the 1970s, in particular in the UK, where a 1973 Parliament-commissioned assessment, the Lighthill Report, condemned a lack of progress in the field, and prompted the first 'AI winter'. In the 1980s, interest rose again with the deployment of programs known as 'expert systems' but fell into the second AI winter when these, too, disappointed (Russell and Norvig, 2010, p. 24). However, by the start of this century, the term had entered common parlance—as evidenced and perhaps aided by the title of Steven Spielberg's 2001 blockbuster film *A.I. Artificial Intelligence* (2001). It now features daily in newspaper headlines and corporate and government strategies alike.

Despite the intense interest today, the term AI remains controversial. There is no widely accepted definition of it—one of the leading textbooks, *Artificial Intelligence: A Modern Approach*, offers four groups of definitions (Russell and Norvig, 2010). This controversy and ambiguity can be traced back to McCarthy's original intentions to coin a phrase that would be bold, broad, attention grabbing, and unconstrained by any particular scientific field or body of knowledge. The term is therefore only partly a reference to a group of technologies, and more an evocation of possibilities; less a field of study, and more a bold, ambitious, and rather ill-defined goal. As one of those aforementioned policy initiatives, that of the French Parliament led by French mathematician and MP Cédric Villani, astutely notes:

> It is probably this relationship between fictional projections and scientific research which constitutes the essence of what is known as AI. Fantasies—often ethnocentric and based on underlying political ideologies—thus play a major role, albeit frequently disregarded, in the direction this discipline is evolving in.
>
> (Villani et al., 2018, p. 4)

In the rest of this chapter, we examine those underlying ideologies and fantasies, 'often ethnocentric', that shape the meaning and connotations of the term AI in English and as it has entered discourse in other language groups.

2.3 The meanings of 'AI' in English: art, trickery, and hierarchies of the human

It is easy to see how the phrase 'AI' fulfilled McCarthy's need for a name that was attention grabbing and free from ties to any particular approach to computing and the emulation of thought. The term 'artificial' broadly means 'made by humans', and so is not restricted to any of the many ways in which humans were at the time attempting to make computing devices, such as cybernetics or automatic programming. 'Intelligence', too, is a broad term that encompasses many of the capacities that specific projects aimed to recreate, such as learning, language processing, or decision making. But untying this agenda from any more narrow, technical tradition also came with costs that have troubled AI ever since: vagueness in its methods and aims, inflated expectations (and fears), and the importation of ideological connotations, in particular with the term 'intelligence', as we explore next.

2.3.1 Artificial

The primary meaning of 'artificial' is 'made or constructed by human skill, esp. in imitation of, or as a substitute for, something which is made or occurs naturally' ('artificial, adj. and n.', n.d.). It came to English from Latin via Old French. In Latin, *artificialis* means 'of or belonging to art', and *artificium* is 'a work of art; but also a skill; theory, or system'. The Latin *ars* corresponds to the Greek *techne.* An *artifex* is both 'a craftsman or artist' but also a schemer or a mastermind. From its origins, the word therefore interweaves art and artifice, technology and trickery (see section 2.1 in Hollanek, 2020).

These connotations persist today and underlie one of the main anxieties around AI: that it is fooling us into thinking it is something it is not. At one level, this fear is simply of being gulled, as when the Automaton Chess Player of 1770, popularly known as the Mechanical Turk, fooled audiences for decades until it was revealed to have a diminutive man inside (see chapter 2 in Wood, 2002). This implication of fraudulence troubled some of the participants at the Dartmouth event. According to Arthur Samuel, '[t]he word artificial makes you think there's something kind of phony about this, or else it sounds like it's all artificial and there's nothing real about this work at all' (in: McCorduck, 2004, p. 115).

But these anxieties about deception implied by the term 'artificial' also take a deeper and more sinister form: that an AI could live among us, passing for human. This is a persistent theme in fiction: in 1816, while the Mechanical Turk was still touring, the Prussian author E. T. A. Hoffmann wrote 'The Sandman', in which the protagonist Nathanael is bewitched by the beauty of a woman

called Olimpia—until the discovery that she is an automaton drives him to suicide (Cave and Dihal, 2018; Hoffmann, 1816). Two centuries later, blockbuster film franchises like *Blade Runner* play on the same fear that the people around us are not what they seem, or that they might even, as in the *Terminator* films, be out to destroy us (Dihal, 2020, p. 207).

2.3.2 *Intelligence*

While the term 'artificial' is crucial to the meaning of 'AI', it is arguably 'intelligence' that is doing the real work, both in setting the research agenda and evoking its grandiose potential. It is also both highly contested in meaning and laden with ideological baggage (Cave, 2020; Legg and Hutter, 2007). Given the central role that the concept of intelligence now plays in many discourses, not just AI, it is perhaps surprising that until the twentieth century it had 'remained largely in the backwaters of English-language discourse' (Carson, 2006, p. 79). What is also surprising to many who use the term today is that, when it began to rise to widespread usage, it was closely tied to eugenics and ideologies of white supremacy, colonialism, classism, and patriarchy.

The idea that the most intelligent people should rule over others has a long history, reaching back to Ancient Greece. But it only attained widespread acceptance in the West when it proved useful in supporting colonialism. Aristotle's argument that some people, because of their superior intellect, were born to rule and others to be ruled over provided a justification for the conquest and enslavement of non-Western peoples. The concept of race served as the scaffold for this narrative: the non-white races were intellectually inferior—in the words of Rudyard Kipling, 'Half-devil and half-child'—and so it was a right, or even a duty, for white peoples to rule over them and their lands (Cave, 2020, p. 30).

Over the course of the nineteenth century, the Western scientific establishment engaged ever more systematically in attempts to evidence these claims. The key figure in this was the English scientist Sir Francis Galton, who was the first to develop a battery of tests to measure intellectual ability, also coined the term 'eugenics', as he put it: 'the science of improving stock . . . to give to the more suitable races or strains of blood a better chance of prevailing speedily over the less suitable' (Galton, 1869). The two went hand-in-hand: as the eugenics agenda spread around the world, intelligence was considered the key variable in determining which peoples should flourish and which were 'less suitable' (Levine, 2017, p. 25).

The role of intelligence—in particular in the form of intelligence testing—in the political history of the twentieth century is vast and complex. It was widely used to justify not only racist and imperialist ideologies, but also

patriarchy and classism. While the most egregious examples of the exploitation of those considered less intelligent are confined to the first half of the century, intelligence testing has remained important to this day. At the same time, so do the associations between degrees of innate intelligence and different races, genders, and classes (Saini, 2019).

As Cave (2020) notes, there is good reason to think that these associations are shaping expectations for AI. They might, for example, be driving concern towards middle class white professionals, in the form of worries about lawyers or doctors losing their jobs, when, in fact, it is the poor and marginalized who are more likely to be impacted negatively (Eubanks, 2017). At the same time, the association of this technology with a term—intelligence—long claimed by white men might be harming the prospects of women and people of colour in this important technology sector, so exacerbating a cycle of injustice (Cave, 2020, pp. 33–4).

The English term 'AI' has therefore fulfilled McCarthy's hopes of commanding attention, stimulating the imagination, and encompassing a wide range of technical approaches. But the realization of his ambitions has come at the cost of awakening fears—whether of deception and disaster or displacing the middle classes—that detract from other pressing ethical and political concerns arising from digital technology, such as the regulation of technology giants or mitigating the impact of automation on the most vulnerable.

2.4 'AI' in the other Germanic languages

2.4.1 KI or AI?

Although etymologically different from the Romance-rooted 'artificial', the term used in most Germanic languages has a very similar meaning. German (*künstliche Intelligenz*), Dutch (*kunstmatige intelligentie*), Danish, and Norwegian bokmål and nynorsk (*kunstig intelligens* in all three) use a term derived from the Middle Low German *kunst* ('knowledge' or 'ability'). This term itself derives from the reconstructed Proto-Germanic verb ∗*kunnaną*, denoting 'to know'.[1] Just as 'artificial' interweaves 'art' and 'artifice', so does *kunst*: while rooted in words denoting skill, the term currently means 'art' in all of the aforementioned languages.

As the most widely spoken language of the family, English has had a particularly strong influence on the other Germanic languages. At the same time, perhaps because of the similarity in meaning between 'art' and *kunst*, the prefix 'art-' has been present in several of these languages. Dutch and German have respectively *artificiële intelligentie* and *artifizielle Intelligenz* as alternative terms for AI: while *artificieel* and *artifiziell* are less commonly used words in either language, this version benefits from both the AI acronym and its similarity to the English

term. Dutch is particularly ambivalent about KI or AI, to the extent that a newspaper article may use one in its headline and the other in the standfirst (Sheikh, 2021). In the Van Dale dictionary, the 'intelligence' lemma refers to both. The agreed-upon scientific terminology did not help the cause: although Dutch AI scientists in the twentieth century agreed upon the term *kunstmatige intelligentie*, they used the English 'AI' as its abbreviation (van den Herik, 1990; Visser, 1995). Danish, similarly, allows the English abbreviation AI as an alternative to KI ('kunstig intelligens', n.d.). The aforementioned role of the film *AI: Artificial Intelligence* may have influenced Dutch more so than the other Germanic languages, as it is the only language in which the title of the film was left untranslated.[2]

Swedish, meanwhile, is the only Germanic language other than English to have adopted only the 'art' root in its version of the phrase 'AI': it uses *artificiell intelligens*, despite Swedish having the word *kunst* for 'art'. Swedish is also the only Germanic language to have the term *maskinintelligens* (machine intelligence) as a synonym for AI.

One Germanic language does not use either AI or KI. Icelandic has *gervigreind*, from *gervi* meaning 'imitation, artificial, or pseudo-' and *greind* meaning 'intelligence'. If 'artificial' already suggests that 'there's something kind of phony about this', then the Icelandic term makes this phoniness explicit.

2.4.2 Being 'intelligent'

The 'intelligence' part of 'artificial intelligence' is less contested: all Germanic languages except Icelandic use the same term derived from the Latin *intelligentia* via the French *intelligence*. This means, however, that the aforementioned ideological baggage attached to the concept of intelligence has carried across into the understanding of AI in all of these languages. This ideological baggage does not come from English alone: it has been co-constructed through intelligence research in countries including the US, England (the aforementioned Galton), France, and Germany in the early twentieth century.

One of the most famous means of classifying intelligence comes from Germany, where the term *Intelligenz* is still considered a technical term from psychology ('Intelligenz', 2021). The term *Intelligenzquotient*, abbreviated to IQ, was invented in 1912 by the Jewish German William Stern, building on the work of the French psychologist Alfred Binet (Section 2.5) (Stern, 1912). This body of work was abused by eugenics movements both in Stern's homeland when the Nazi regime took power, and in the US, to which he fled from this regime. In Nazi Germany, intelligence tests were administered for the purpose of identifying and exterminating 'congenital feeble-mindedness' (*angeborener Schwachsinn*), though the tests were heavily skewed to disadvantage anyone who disagreed with Nazi ideology (Büüsker, 2015).

While these Nazi practices made worldwide eugenics movements significantly less popular, the association between race and intelligence—particularly that white, Western European peoples are more intelligent than others—remains. For example, stereotypes about ethnic minorities being less intelligent than white secondary school pupils have been proven to persist in schoolteachers in the Netherlands (van den Bergh et al., 2010) and Germany (Bonefeld and Dickhäuser, 2018). It is therefore likely that the stereotyped preconceptions about intelligence that underlie the history of AI are equally fundamental to conceptions of AI in northwestern Europe.

2.5 'AI' in the Romance languages

In the mid-1950s, IBM started commercializing in France a new electronic machine named *ordinateur*. The term, an ancient religious word to describe God as organizer of the world, was meant to avoid the literal translation of the English 'computer' as *calculateur*, and their evocation of early basic calculators rather than new machines with powerful capacities (Centenaire IBM France, 2014). By contrast, the translation of 'artificial intelligence' into French closely follows the English version: AI in French is called *intelligence artificielle* and both the term and its corresponding acronym 'IA' have become buzzwords in the general media since 2015. While specific usage patterns vary across contexts, the term appears to share a common etymology across the main Romance languages. Thus, IA stands for *inteligencia artificial* in Spanish, *intelligenza artificiale* in Italian, *inteligência artificial* in Portuguese, and *inteligență artificială* in Romanian.

This proximity with English terminology is both linguistic and cultural. In French academic, policy-oriented, and popular scientific discourse, AI is generally described as a research program with ill-defined boundaries but clear Anglo–American origins, associated most prominently with the works of Alan Turing, John McCarthy, Marvin Minsky, or Herbert Simon (see, e.g. Cardon et al, 2018; LeCun, 2016). Just as in English-speaking contexts, the term also conveys science fiction imaginaries and fantasies. For instance, the French computer scientist and Oulipo member Paul Braffort introduced the 1968 book *L'intelligence artificielle* by noting how, for the non-specialist, AI inevitably brought up fantastic images and representations. As he explained, with *intelligence artificielle* 'arise the key words "robot", "electronic brain", "thinking machine" in a science fiction atmosphere, a clinking of gears, a smell of ozone: Frankenstein is not far away!' (Braffort, 1968, p. 3).

Six decades later, the aforementioned Villani report claims that although AI's development 'has always been accompanied by the wildest, most alarming and far-fetched fantasies', a new dominant and Californian narrative underlies its most recent developments (Villani et al., 2018, pp. 4–5). According to

the authors, the ethnocentrism and determinism of dominant AI fantasies today is most apparent in transhumanist ideas of the Singularity, and requires a strong French and European industrial response for ensuring what they call meaningful AI.[3] As an echo of mid-1990s critics of the 'Californian ideology' (Barbrook and Cameron, 1995), a recurring theme in French public and media discourse around AI in recent years has been to point out the dangers of dominant AI imaginaries associated with Silicon Valley.

As mentioned, both words 'artificial' and 'intelligence' get their meaning from Latin. However, the notion of intelligence conveys specific connotations in the French context. John Carson has demonstrated how *intelligence*, originally a synonym for knowledge, came to suggest 'an ability existing in degrees' for mid-nineteenth-century French anthropologists (Carson, 2006, p. 79). Intelligence was then a specialized anthropological term that played a key role in explaining or justifying classifications and hierarchies in natural science taxonomies of both animal and human life. In large part, intelligence referred to mental powers, which were expected to occur in different degrees across human groups.

While attempts to systematically measure intelligence through cranial volume or brain mass largely failed, the notion that intelligence was an objective faculty remained in place. When Binet and fellow psychologist Théodule Ribot introduced the first psychological scales of intelligence in 1905, it was based on the idea that intelligence was a material and measurable characteristic. Beyond the walls of the Collège de France and the Sorbonne, intelligence measures were seen as consistent with the principles of the Third Republic. When compared with Galton's eugenics or nineteenth-century racial science, the political agenda of experimental psychology was distinctively progressive: the aim of measuring 'mental age' (*âge mental*) was to help trace children's development and to improve the education of the 'abnormal' or 'retarded' (Malabou, 2019, p. 18). Intelligence scales met the positivist orientation of early twentieth-century science and helped justify the existence of a new administrative elite without resorting to essentialist criteria.

At the same time, the modern concept of intelligence implied by psychological tests was also strongly resisted. As the philosopher Catherine Malabou highlights, measurable intelligence stood in sharp contrast with romantic notions of *esprit* (a word signifying at once 'mind' and 'spirit'), Enlightenment notions of universal reason, or antique notions of intellect or *entendement* (commonly equated with a symbolic capacity of understanding), which all presumed an active participation of the thinking subject. Philosophers, such as Henri Bergson, and novelists, such as Marcel Proust, deemed the psychological concept of intelligence a dangerous one, bearing the threat of normalization and standardization against the autonomy of 'intuition' (Malabou, 2019, pp. 6–7, 39). From such perspectives, intelligence was not on the side of

thought, but on the side of the mechanical and the biological; not on the side of creative adaptation, but on that of automatisms and repetition.

The notion of an irreducible tension between philosophical and symbolic realms of intelligence on the one hand, and between biological and psychological knowledge on the other, is still at play in many public and philosophical discussions of AI, in which artificiality and creativity are commonly opposed. French psychologists' work on intelligence was also much more influential and visible in the United States: in France, explorations of intelligence focused on clinical studies of individual variations, rather than mass IQ testing. These disjunctive histories of intelligence continue to play out in the continuous reference to US-dominated AI imaginaries in France, and related attempts to foster alternative meanings of AI.

2.6 'AI' in the Slavic languages

We have suggested that Spielberg's 2001 blockbuster *A.I. Artificial Intelligence* might have contributed to the popularity of McCarthy's term in the Anglophone West. In Central and Eastern Europe the term 'artificial intelligence', as a translation of the English original into Russian, Polish, or Czech, appeared in academic papers as early as the 1960s, and then in science fiction and popular science writing throughout the 1970s, 1980s, and 1990s. But it was, arguably, Spielberg's film that helped 'artificial intelligence' migrate from specialist journals and relatively niche science fiction publications into common parlance. Translated into Russian as *Иску́сственный ра́зум* [*Iskússtvennyy rázum*], into Czech as *A.I. Umělá inteligence,* and into Polish as *A.I. Sztuczna inteligencja,* the film's title is a good indication of how the Slavic versions of McCarthy's 'artificial intelligence' differ from the original term—and from one another.

If the Polish and Czech translations of the title retain the English abbreviation, 'AI', this corresponds to how the acronym functions in the vernacular of some Central and Eastern European countries, such as Poland or the Czech Republic, where the local equivalents have never truly caught on. In Poland, for example, the Polish term *sztuczna inteligencja* (the direct translation of 'artificial intelligence') often features in brackets as an expansion of the English abbreviation 'AI'—which is prevalent in industry reports, media coverage, and technology companies' promotional material. We could relate the contemporary tendency to supplant the local *SI* with the original 'AI' to the dominance of English in the age of globalization, as observed in other countries and linguistic areas (and as we discuss in this chapter). However, it is worth noting that the use of English loanwords has a particular history in Central and Eastern European, where, during the communist period, the original names of some consumer goods used to convey a sense of potential and freedom coded 'Western'.