



Open Research Online

Citation

Ali, Mustafa; Dick, Stephanie; Dillon, Sarah; Jones, Matthew L.; Penn, Jonnie and Staley, Richard (2023). Histories of artificial intelligence: a genealogy of power. BJHS Themese, 8 pp. 1–18.

URL

<https://oro.open.ac.uk/94940/>

License

(CC-BY-NC-ND 4.0) Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from [Open Research Online \(ORO\) Policies](#)

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding



INTRODUCTION

Histories of artificial intelligence: a genealogy of power

Syed Mustafa Ali¹, Stephanie Dick², Sarah Dillon³, Matthew L. Jones⁴, Jonnie Penn⁵
and Richard Staley⁶

¹School of Computing and Communications, the Open University, UK, ²School of Communication, Simon Fraser University, Burnaby, BC, Canada, ³Faculty of English, University of Cambridge, UK, ⁴History Department, Princeton University, NJ, USA, ⁵Centre for the Future of Intelligence, University of Cambridge, UK and ⁶Department of History and Philosophy of Science, University of Cambridge, UK and Department of Science Education, University of Copenhagen, Denmark
Corresponding author: Richard Staley; Email: raws1@cam.ac.uk

An introduction to the history of AI: genealogies of power in the management age

Like the polar bear beleaguered by global warming, artificial intelligence (AI) serves as the charismatic megafauna of an entangled set of local and global histories of science, technology and economics. This *Themes* issue develops a new perspective on AI that moves beyond conventional origin myths – AI was invented at Dartmouth in the summer of 1956, or by Alan Turing in 1950 – and reframes contemporary critique by establishing plural genealogies that situate AI within deeper histories and broader geographies. ChatGPT and art produced by AI are described as generative but are better understood as forms of pastiche based upon the use of existing infrastructures, often in ways that reflect stereotypes. The power of these tools is predicated on the fact that the Internet was first imagined and framed as a ‘commons’ when actually it has created a stockpile for centralized control over (or the extraction and exploitation of) recursive, iterative and creative work. As with most computer technologies, the ‘freedom’ and ‘flexibility’ that these tools promise also depends on a loss of agency, control and freedom for many, in this case the artists, writers and researchers who have made their work accessible in this way. Thus, rather than fixate on the latest promissory technology or focus on a relatively small set of elite academic pursuits born out of a marriage between logic, statistics and modern digital computing, we explore AI as a diffuse set of technologies and systems of epistemic and political power that participate in broader historical trajectories than are traditionally offered, expanding the scope of what ‘history of AI’ is a history of.

‘AI’ is everywhere and nowhere, and so is its history. On one hand, despite a growing body of critique, it is only recently that historians have begun to devote sustained scholarly attention to the subject. This issue maps and consolidates new and ongoing research in this area, and brings academic historical perspectives to bear on other disciplinary approaches to the study and critique of AI, and vice versa. On the other hand, centering on AI as a specific set of technical systems with its developers and varied uses risks obscuring the fact that artificial intelligence participates in and concretizes many broader

© The Author(s), 2023. Published by Cambridge University Press on behalf of British Society for the History of Science. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<https://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

logics and histories – of industrialization, militarism, colonialism, social science, capitalism and, of course, management – all subject to long-standing historical investigation. In this sense, histories of AI have been written, even if they are as yet unrecognized in their pertinence and multiplicity.

What is presented here, then, is perhaps less a history, as that might be traditionally understood, than a genealogy, in the Foucauldian sense. Less a search for origins than a multiple tracing of interconnected, interlayered events and phenomena, informed by the recognition that there is “something altogether different” behind things: not a timeless and essential secret, but the secret that they have no essence or that their essence was fabricated in a piecemeal fashion from alien forms.¹ Foucault asserts that what genealogy finds ‘at the beginning of things is not the inviolable identity of their origins; it is the dissonance of other things. It is disparity’.²

The genealogy of AI presented here pays attention to this disparity. Contributors from around the globe consider the ‘vicissitudes of history’, from the palimpsest of ‘details and accidents that accompany every beginning’ to ‘the subtle, singular, and subindividual marks that might possibly intersect’ on the palimpsestuous history of AI that forms ‘a network that is difficult to unravel’.³ Throughout this work, our guiding question is not ‘what are the origins of this thing called AI?’ but rather, ‘what are the histories within which it makes sense to bracket a host of political, technical and epistemic systems under this umbrella?’

To answer this question requires the skills of historians, but also of other scholars from across the humanities – the heterogeneity of the subject matter requires a corresponding pluralism of method. Like the seminar which informs it, this issue constitutes, then, a work of synchronic interdisciplinarity in which multiple disciplines bring their different methods to bear on a common object.⁴ In ‘From work to text’, Roland Barthes presents a productive definition of this form of interdisciplinarity when he observes that ‘what is new ... comes not necessarily from the internal recasting of each of these disciplines, but rather from their encounter in relation to an object which traditionally is the province of none of them’.⁵ AI is just such an object. It references a range of technologies gathered loosely under a single banner because they, collectively, (re)produce behaviour presumed to be ‘intelligent’. It also references a list of other things: a sociotechnical phenomenon, an invitation for speculative rhetoric, a manufacturing philosophy, a claim to the limits of consciousness and an extension of managerial authority. To encompass such diverse sources and endeavours, our contributors draw on a range of methodologies and disciplinary perspectives. In doing so, they provide critical and comparative research on the historical character of AI technologies, including their entanglements in systems of politics, profit, power and control.

Recurring themes in the histories of AI

The articles in this issue all, more or less explicitly, address, interrogate and evolve four thematic threads that animated the work of the HoAI Seminar: *hidden labour*, *encoded behaviour*, *disingenuous rhetoric* and *cognitive injustice*.⁶ The intersection of time periods, geographical locales and these themes enables a rich and novel picture of recurrences

1 Michel Foucault, ‘Nietzsche, genealogy, history’, in Foucault, *Language, Counter-memory, Practice* (ed. Donald Bouchard), Ithaca, NY: Cornell University Press, 1980, pp. 139–64, 142.

2 Foucault, *op. cit.* (1), p. 142.

3 Foucault, *op. cit.* (1), pp. 144–5.

4 Sarah Dillon, ‘The deictic humanities: towards a taxonomy of interdisciplinarity’, forthcoming.

5 Roland Barthes, ‘From work to text’, in Barthes, *Image, Music, Text* (tr. Richard Howard), New York, Hill & Wang, 1977, pp. 155–64, 155.

6 For an account of the themes from the original HoAI Seminar proposal (2019), see www.ai.hps.cam.ac.uk/about-0/themes.

in AI's histories. Addressing diverse aspects of artificial intelligence as 'applied epistemology' (a synonym of 'AI' for John McCarthy, who coined the term), these themes manifest both in its formation and in its propagation, as well as in recent critiques of AI.⁷ These critiques reveal the social and power relations materialized within AI systems, and reconfigured by its use.⁸

Hidden labour aims to shed light on the unacknowledged human work required to make AI-powered systems practical (such as data creation, data labelling, data structuring, content moderation, mineral mining and infrastructure maintenance). The introduction of automated systems tends not to replace human labour but rather to require or enact the reconfiguration and redistribution of work, authority and responsibility within broader political economies.⁹ The second and third themes, *encoded behaviour* and *disingenuous rhetoric*, direct attention to the ways in which users, citizens and commercial audiences have engaged with AI systems in unanticipated ways. As critics have long argued, there is a distinction between how those systems have been imagined, described, taught, advertised or sold (as well as the manner in which they are defended after a crisis or failure) and their actual uses and effects.¹⁰ 'AI' systems, via reinforcement learning, have disciplined and continue to discipline and/or frame the behaviour of those who encounter them, a process that occurs in tension with the open, transparent connectivity that such technologies are said to offer.¹¹

7 For details on McCarthy's alternative name see Jonnie Penn, 'Inventing intelligence: on the history of complex information processing and artificial intelligence in the United States in the mid-twentieth century', PhD dissertation, University of Cambridge, 2020, pp. 116–58. For recent critiques of AI see, for example, Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, 1st edn, New York: Crown, 2016; Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, 1st edn, New York: St Martin's Press, 2017; Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World*, Cambridge, MA: MIT Press, 2018; Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York: New York University Press, 2018; Ruha Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code*, Medford, MA: Polity, 2019; Neda Atanasoski and Kalindi Vora, *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*, Durham, NC: Duke University Press, 2019; Yarden Katz, *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*, New York: Columbia University Press, 2020; Shunryu Colin Garvey, 'Unsavory medicine for technological civilization: introducing "Artificial Intelligence & Its Discontents"', *Interdisciplinary Science Reviews* (3 April 2021) 46(1–2), pp. 1–18; Wendy Hui Kyong Chun and Alex Barnett, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*, Cambridge, MA: MIT Press, 2021; Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven: Yale University Press, 2021; Justin Joque, *Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism*, London and New York: Verso, 2022; Dan McQuillan, *Resisting AI: An Anti-fascist Approach to Artificial Intelligence*, Bristol: Bristol University Press, 2022.

8 For articulation of computers as sites where social relations are materialized and are therefore easier to explore and understand historically, see Janet Abbate and Stephanie Dick (eds.), *Abstractions and Embodiments: New Histories of Computing and Society*, Baltimore: Johns Hopkins University Press, 2022.

9 Mary L. Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*, Boston, MA: Houghton Mifflin Harcourt, 2019; Lilly Irani, 'The hidden faces of automation', *XRDS: Crossroads, the ACM Magazine for Students* (2016) 23(2), pp. 34–37.

10 Sarah T. Hamid, 'Abolishing carceral technologies', *Logic Magazine*, 31 August 2020, at <https://logicmag.io/care/community-defense-sarah-t-hamid-on-abolishing-carceral-technologies>; Wendy Hui Kyong Chun, *Updating to Remain the Same: Habitual New Media*, first MIT Press new paperback edn, Cambridge, MA and London, England: MIT Press, 2017.

11 Fred Turner, *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*, 1st paperback edn, Chicago: The University of Chicago Press, 2008; Turner, *The Democratic Surround: Multimedia & American Liberalism from World War II to the Psychedelic Sixties*, Chicago and London: The University of Chicago Press, 2015; Marc Aidinoff, 'Centrists against the center: the Jeffersonian politics of a decentralized internet', in Abbate and Dick, op. cit. (8), pp. 40–59; Christopher M. Kelty, *Two Bits: The Cultural Significance of Free Software*, Durham, NC: Duke University Press, 2008; Wendy Hui Kyong Chun, *Control and Freedom: Power and*

At the same time, the history of imaginative thinking around AI, in fact and fiction, influences how AI is produced, perceived and regulated, and the rhetorical framing of 'AI', past and present, by scientists, technologists, governments, corporations, activists and the media, performatively creates and shapes the very phenomenon purportedly under analysis.¹² The final theme, *cognitive injustice*, points to epistemic and ontological injustices that are entangled with AI in its prehistory and its development, examining the ways in which the definitions and protocols of 'intelligence' that it deploys appear to narrow and delimit knowledge, particularly for marginalized groups.¹³ This is pervasive throughout the operations of AI and its histories.

Historiographies of AI and management

The articles in this issue situate AI not only within the customary histories of computing and information, but also within histories of management and control, including those born of industry, statecraft and coloniality. With a frequency that we had not anticipated given the breadth of periods and locales they explore, our contributors emphasize the centrality of managerial techniques and concerns – over and against logical and technological features. Their writing analyses that metonymic status to reveal how contemporary discourse about AI mistakes it as an emancipatory upshot of the 'Information Age' (from the 1950s to today) rather than as an extension of, and euphemism for, what we call the 'Management Age' (from the 1500s to today).

The verb 'manage' was introduced in the mid-sixteenth century to describe actions that 'control or direct by administrative ability'.¹⁴ By our account, the Management Age conditions the Information Age. To show this, the articles gathered here largely situate

Paranoia in the Age of Fiber Optics, Cambridge, MA and London: MIT Press, 2006; Alexander R. Galloway, *Protocol: How Control Exists after Decentralization*, Cambridge, MA: MIT Press, 2004.

12 See, for example, Lynette Hunter, 'Rhetoric and artificial intelligence', *Rhetorica* (November 1991) 9(4) pp. 317–40; Stephen Cave, Kanta Dihal and Sarah Dillon (eds.), *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford: Oxford University Press, 2020; The Royal Society, *Portrayals and Perceptions of AI and Why They Matter* (2018), at <https://royalsociety.org/-/media/policy/projects/ai-narratives/AI-narratives-workshop-findings.pdf>; Sarah Dillon, 'The Eliza effect and its dangers: from demystification to gender critique', *Journal for Cultural Research* (2020) 24(1), pp. 1–15; Sarah Dillon and Jennifer Schaffer Goddard, 'What AI researchers read: the role of literature in artificial intelligence research', *Interdisciplinary Science Reviews* (2023) 48(1), pp. 15–42; Jascha Bareis and Christian Katzenbach, 'Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics', *Science, Technology, & Human Values* (2022) 47(5), pp. 855–81; Ali A. Guenduez and Tobias Mettler, 'Strategically constructed narratives on artificial intelligence: what stories are told in governmental artificial intelligence policies?', *Government Information Quarterly* (2023) 40(1), pp. 1–13; Carolijn van Noort, 'On the use of pride, hope and fear in China's international artificial intelligence narratives on CGTN', *AI & Society* (2022), at <https://doi.org/10.1007/s00146-022-01393-3>; Ching-Hua Chuan, Wan-Hsiu Sunny Tsai and Su Yeon Cho, 'Framing artificial intelligence in American newspapers', *AIES*, January 2019, pp. 339–44; Alexa Robertson and Max Maccarone, 'AI narratives and unequal condition: analyzing the discourse of liminal expert voices in discursive communicative spaces', *Telecommunications Policy* (2023) 47(5), 102462; Astra Taylor, 'The automation charade', *Logic(s)*, 1 August 2018, at <https://logicmag.io/failure/the-automation-charade>.

13 Alison Adam, *Artificial Knowing: Gender and the Thinking Machine*, London and New York: Routledge, 1998; Katz, *op. cit.* (7); Sanjay Seth, *Beyond Reason: Postcolonial Theory and the Social Sciences*, New York: Oxford University Press, 2021; Stephanie Dick, 'The Marxist in the machine', *Osiris* (1 July 2023) 38, pp. 61–81; John Carson, 'The culture of intelligence', in Theodore M. Porter and Dorothy Ross (eds.), *The Cambridge History of Science*, 1st edn, Cambridge: Cambridge University Press, 2003, pp. 635–48.

14 Douglas Harper, 'Etymology of manage', *Etymology Online* (13 October 2021), at www.etymonline.com/word/manage (accessed 1 October 2022). On related tensions between the histories of science, technology and knowledge see Lorraine Daston, 'The history of science and the history of knowledge', *KNOW: A Journal on the Formation of Knowledge* (2017) 1(1), pp. 131–54; Sebastian Felten and Christine von Oertzen, 'Bureaucracy as knowledge', *Journal for the History of Knowledge* (2020) 1(1), pp. 1–16.

AI, as a flagship of the Information Age, within longer histories of population management, biometrics, racial capitalism and mass media. We investigate efforts to digitize social practices alongside efforts to organize bodies, naturalize state and corporate power, and valorize archival and actuarial epistemes. The essays in the issue show the intersections of digital decision tools with notions of scientific and political authority in the US from Truman and Eisenhower to George H.W. Bush; in Soviet Russia from Stalin to Gorbachev; from nineteenth-century Argentina to the present day; and across comparable periods in India, Australia and Brazil. The result is a multi-decade, multinational, interdisciplinary picture of the history of AI that is contingent upon and responsive to local conditions, yet also operates – across time and place – as a means to consolidate power.

Early histories of AI were offered by reflective practitioners sharing their informed perspectives on the intellectual history of their field and by anthropologists, sociologists and critical practitioners who questioned the coherence and consequences of AI's dominant methods and claims.¹⁵ Many focused on the small (yet assertive) elite Western academic communities engaged in technical research on what would eventually be called AI and robotics. The work of Lucy Suchman, Diana Forsythe, Harry Collins and other early social studies of AI aimed in part to articulate the theories of 'knowledge', 'reasoning' and 'intelligence' taking shape as technologists sought to reproduce these in the machine. They recognized that, in one sense, anthropologists, historians of science and AI researchers were interested in *doing the same thing*: each offered definitions, accounts, theories and models of 'intelligence' and 'knowledge'. But Collins, Forsythe and others highlighted just *how different* were the theories of 'intelligence' taking shape (almost in parallel) within AI and within the social studies of science. The former sought to reduce intelligence to a formalism, or to the product of formal and data-driven processing.¹⁶ The latter insisted that knowledge is unavoidably social and embodied, requiring experiences and capacities that computers would always lack. This evolving debate spilled into the public arena in 1994 as Forsythe and James Fleck traded barbs in the *Social Studies of*

15 Practitioner histories include Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge and New York: Cambridge University Press, 2010; Stuart J. Russell, Peter Norvig and Ernest Davis, *Artificial Intelligence: A Modern Approach*, 3rd edn, Upper Saddle River, NJ: Prentice Hall, 2010; Allen Newell, 'Intellectual issues in the history of artificial intelligence', in Ronald Chrisley and Sander Begeer (eds.) *Artificial Intelligence: Critical Concepts*, New York: Routledge, 2000, pp. 187–227; John McCarthy, 'Reminiscences on the theory of time-sharing', *Professor John McCarthy*, at <http://jmc.stanford.edu/computing-science/timesharing.html> (accessed 10 September 2022); Margaret A. Boden, *Mind as Machine: A History of Cognitive Science*, 2 vols., Oxford and New York: Clarendon Press, 2006. Critical histories include Diana Forsythe and David J. Hess, *Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence*, Stanford, CA: Stanford University Press, 2001; Lucille Alice Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd edn, Cambridge and New York: Cambridge University Press, 2007; Philip E. Agre, *Computation and Human Experience*, Cambridge and New York: Cambridge University Press, 1997; Philip E. Agre, 'Toward a critical technical practice: lessons learned in trying to reform AI', in Geoffrey C. Bowker, Les Gasser, S.L. Star and Bill Turner (eds.), *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*, Mahwah, NJ: Lawrence Erlbaum Associates, 1997, pp. 131–57; Hubert L. Dreyfus, *What Computers Can't Do: A Critique of Artificial Reason*, 1st edn, New York: Harper & Row, 1972; Hubert L. Dreyfus, *What Computers Still Can't Do: A Critique of Artificial Reason*, Cambridge, MA: MIT Press, 1992; James Fleck, 'Development and establishment in artificial intelligence', in Norbert Elias, Herminio Martins and Richard Whitley (eds.), *Scientific Establishments and Hierarchies*, Dordrecht: Springer Netherlands, 1982, pp. 169–217; H.M. Collins, *Artificial Intelligence: Against Humanity's Surrender to Computers*, Medford, MA: Polity Press, 2018.

16 Hunter Heyck, 'Defining the computer: Herbert Simon and the bureaucratic mind – Part 1', *IEEE Annals of the History of Computing* (April 2008) 30(2), pp. 42–51; Hunter Heyck, 'Defining the computer: Herbert Simon and the bureaucratic mind – Part 2', *IEEE Annals of the History of Computing* (April 2008) 30(2), pp. 52–63; Stephanie Dick, 'Of models and machines: implementing bounded rationality', *Isis* (2015) 106(3), pp. 623–34; Paul Erickson, Judy L. Klein, Lorraine Daston, Rebecca M. Lemov, Thomas Sturm and Michael D. Gordin, *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*, Chicago and London: The University of Chicago Press, 2013.

Science over whether anthropology or knowledge engineering was more inclined to positivism.¹⁷

On the one hand, social and computational theories of intelligence could not be more different, as these early histories reveal. Social historians and historical epistemologists have a unique and essential skill set for making genealogical sense of AI and the social and technical logics materialized within it because they have a robust set of alternative theories of ‘intelligence’ to work with. But on the other hand, historians do not occupy a privileged position and in fact share contexts and culture with AI. They often reach for concepts that are also central to AI – such as ‘network’ and ‘systems’ used to characterize the social character of knowledge. Speaking in Fordist terms of ‘knowledge production’, ‘knowledge consumption’ and ‘knowledge circulation’ highlights that our own fields have drawn from many of the same conceptual and cultural resources as AI when framing intelligence in industrial, bureaucratic and cybernetic ways. Given these overlaps, it is also not inconceivable that AI research makes use of our conceptual and cultural resources in the years ahead.

With this in mind, we consider the history of AI and social studies of science as related projects, and seek to open lines of inquiry into the mutual concerns, historical entanglements and shared paradigms that have informed competing Western accounts of ‘intelligence’ in the twentieth and twenty-first centuries, including those most familiar to the readership of *BJHS Themes*. In order to do so, we believe that robust histories of AI ought to be contextualized and historicized beyond Western frameworks. In *Human-Machine Reconfigurations*, Suchman argues that precisely what it means to be human is both revealed and reconfigured everywhere machines are said to mimic human behaviour, revealing that AI and robotics raise questions about the character not only of knowledge and intelligence, but of humanness itself. The history of AI, accordingly, partakes in histories of colonial power which always centred on hegemonic control over the definition of what is ‘human’ and what cognitive capacities reveal one’s humanness.¹⁸

Recently, several scholarly studies have traced the broader socio-technical-colonial contexts within which AI emerged and which it served to mobilize, both during and after the field’s formal inception and naming in Cold War-era American defence establishment research.¹⁹ Histories of related information technologies have revealed how powerful actors such as the British Civil Service, the US military, defence contractors like Palantir and Axon and corporations like Google Inc. have leveraged such tools to accomplish hidden economic, ideological and political aims.²⁰

17 Diana E. Forsythe, ‘Engineering knowledge: the construction of knowledge in artificial intelligence’, *Social Studies of Science* (1993) 23(3), pp. 445–77; James Fleck, ‘Knowing engineers? A response to Forsythe’, *Social Studies of Science* (1994) 24(1), pp. 105–13; Diana E. Forsythe, ‘STS (re)constructs anthropology: a reply to Fleck’, *Social Studies of Science* (1994) 24(1), pp. 113–23.

18 Suchman, *op. cit.* (15). On the history of ‘humanness’ within AI see Atanasoski and Vora, *op. cit.* (7); Benjamin, *op. cit.* (7); Katz, *op. cit.* (7); Dick, *op. cit.* (13); Jenny Carla Moran, ‘Loveability: a critical theory for understanding love, humanness, and futurity in the age of the sex robot’, PhD dissertation, University of Cambridge, 2023; Stephanie Dick, Wendy Chun and Matt Canute, ‘Verifying generative AI: adversarial mimicry’, *Issues in Science and Technology*, forthcoming; Alan Blackwell, ‘Moral codes: designing alternatives to AI’, advance public release of a book to be published by MIT Press in 2024, at <https://moralcodes.pubpub.org> (accessed 2 November 2023).

19 Garvey, *op. cit.* (7); Penn, *op. cit.* (7); Theodora Dryer, ‘Settler computing: water algorithms and the equitable apportionment doctrine on the Colorado River, 1950–1990’, *Osiris* (2023) 38, pp. 265–85; Katz, *op. cit.* (7); Atanasoski and Vora, *op. cit.* (7); Seth, *op. cit.* (13); Joque, *op. cit.* (7); Chun and Barnett, *op. cit.* (7); Benjamin, *op. cit.* (7); Jason Edward Lewis, Noe Arista, Archer Pechawis and Suzanne Kite, ‘Making kin with the machines’, *Journal of Design and Science* (2018) 3(5), at <https://jods.mitpress.mit.edu/pub/lewis-arista-pechawis-kitem> (accessed 1 July 2020).

20 Jon Agar, *The Government Machine: A Revolutionary History of the Computer*, Cambridge, MA: MIT Press, 2003; Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America*, Cambridge, MA and London: MIT Press, 1997; Noble, *op. cit.* (7).

In providing novel historical explorations of AI, this issue likewise casts new light on the period in which it emerged: the Cold War. AI is clearly a product of the Cold War. But *which* Cold War? Decades of sustained historiography have delineated the multiple conflicts, scales and logics across the second half of the twentieth century.²¹ Some are well studied: the Cold War of the bomb, of scientific diplomacy, of McCarthyism, of large-scale computing, of the arms race and the space race pitting the US against the USSR, and of the Cold War university.²² Recent work on the history of social science paints another picture, less reductive than earlier studies.²³ This literature helps us to recognize how research in AI, like cybernetics, brought together multiple registers of Cold War epistemology, politics and practice, and developed as much within Cold War social science as within computing or cognitive science. At first glance, post-war social science asserted aspirational characteristics similar to those of early AI: neutral objectivity, universal applicability, overconfidence in scientific maturity and faith in systematized rationalization through professionalization.²⁴ Both areas received substantial funding from the US defence establishment in the post-war period, often through nominally independent subsidiaries of federal agencies like RAND. Situating AI in this manifold requires us to account for its kaleidoscopic touchpoints with various disciplinary practices and patronage networks, from statistical and computer-engineering techniques to funding aims for operations research. Some of these touchpoints originated in the Cold War; others did not.²⁵

These historiographical nuances speak to the surprising continuities that emerge from a sustained historical treatment of what has been called, at various junctures since the 1950s, 'AI'. Early symbolic manipulation (1950s–1970s), expert-systems databases (1960s–1980s), and the now dominant approaches of data-driven machine learning (1940s–) have, on the face of it, very little in common. The first modelled human reasoning as heuristic symbolic information processors.²⁶ The second encoded human knowledge as databases of 'if-then' rules.²⁷ The third trains algorithms, especially artificial neural networks today, to compute patterns and correlations in order to make forecasts

21 Hunter Heyck and David Kaiser, 'Introduction', *Isis* (2010) 101(2), pp. 362–66.

22 Daniel Volmar, 'The computer in the garbage can: air-defense systems in the organization of US nuclear command and control, 1940–1960', PhD dissertation, Harvard University, 2018; Edwards, op. cit. (20).

23 Joy Rohde, *Armed with Expertise: The Militarization of American Social Research during the Cold War*, Ithaca, NY: Cornell University Press, 2013; Hunter Crowther-Heyck, *Age of System: Understanding the Development of Modern Social Science*, Baltimore: Johns Hopkins University Press, 2015; Egle Rindzeviciute, *The Power of Systems: How Policy Sciences Opened Up the Cold War World*, Ithaca, NY: Cornell University Press, 2016; Jamie Cohen-Cole, *The Open Mind: Cold War Politics and the Sciences of Human Nature*, Chicago and London: The University of Chicago Press, 2014; Mark Solovey and Hamilton Cravens (eds.), *Cold War Social Science: Knowledge Production, Liberal Democracy, and Human Nature*, New York: Palgrave Macmillan US, 2012; Paul Erickson, *The World the Game Theorists Made*, Chicago and London: The University of Chicago Press, 2015.

24 Solovey and Cravens, op. cit. (23).

25 Richard Barbrook, *Imaginary Futures: From Thinking Machines to the Global Village*, London: Pluto, 2007; Thomas Borstelmann, *The Cold War and the Color Line: American Race Relations in the Global Arena*, Cambridge, MA: Harvard University Press, 2009.

26 Hunter Crowther-Heyck, *Herbert A. Simon: The Bounds of Reason in Modern America*, Baltimore: Johns Hopkins University Press, 2005; Dick, op. cit. (16); Ekaterina Babintseva, 'Engineering the lay mind: Lev Landa's algorithmic theory and artificial intelligence', in Abbate and Dick, op. cit. (8), pp. 319–40; Penn, op. cit. (7); Erickson *et al.*, op. cit. (16); Roberto Cordeschi, *The Discovery of the Artificial: Behavior, Mind, and Machines before and beyond Cybernetics*, Dordrecht and Boston, MA: Kluwer Academic Publishers, 2002.

27 Collins, op. cit. (15); Philip Shiman and Alex Roland, *Strategic Computing: DARPA and the Quest for Machine Intelligence, 1983–1993*, Cambridge, MA: MIT Press, 2002; David C. Brock, 'Learning from artificial intelligence's previous awakenings: the history of expert systems', *AI Magazine* (28 September 2018) 39(3), pp. 3–15; Stephanie A. Dick, 'Coded conduct: making MACSYMA users and the automation of mathematics', *BJHS Themes* (2020) 5, pp. 205–24; Forsythe and Hess, op. cit. (15).

based on large databases.²⁸ Looking past these purported differences shows that shared logics – especially managerial, military, industrial and computational – cut across them, often in ways that reinforce oppressive racial and gender hierarchies.²⁹ This *Themes* issue engages distinctively Cold War elements of this story (Mendon-Plasek, Babintseva, Schirvar, Kirtchik, Powell), but also proposes continuities with intellectual structures and research practices that pre-date or traverse that period (Sahoo, Penn, Hamid, Moreschi), or explores quite different contexts (Stark; Lysen; Law; Taylor; Hagerty, Aranda and Jemio). The refinement of formal abstraction in AI intersected with efforts at social control across scales. Historians of automata, automation, axiomatization and biometrics have implied as much on disciplinary, professional, national and international scales.³⁰

28 R.C. Eberhart and R.W. Dobbins, 'Early neural network development history: the age of Camelot', *IEEE Engineering in Medicine and Biology Magazine* (September 1990) 9(3), pp. 15–18; M. Olazarán, 'A historical sociology of neural network research', (University of Edinburgh, 1991), at <http://hdl.handle.net/1842/20075>; David A. Medler, 'A brief history of connectionism', *Neural Computing Surveys* (1998) 1, pp. 18–72; James A. Anderson and Edward Rosenfeld (eds.), *Talking Nets: An Oral History of Neural Networks*, Cambridge, MA: MIT Press, 1998; Tara H. Abraham, '(Physio)logical circuits: the intellectual origins of the McCulloch–Pitts neural networks', *Journal of the History of the Behavioral Sciences* (2002) 38(1), pp. 3–25; Aaron Plasek, 'On the cruelty of really writing a history of machine learning', *IEEE Annals of the History of Computing* (October 2016) 38(4), pp. 6–8; Dominique Cardon, Jean-Philippe Cointet and Antoine Mazières, 'La revanche des neurones: L'invention des machines inductives et la controverse de l'intelligence artificielle', *Réseaux* (2 November 2018) 211(5), 173–220; Matthew L. Jones, 'How we became instrumentalists (again): data positivism since World War II', *Historical Studies in the Natural Sciences* (November 2018) 48(5), 673–84; Chun and Barnett, op. cit. (7); Aaron Mendon-Plasek, 'Mechanized significance and machine learning: why it became thinkable and preferable to teach machines to judge the world', in Jonathan Roberge and Michael Castelle (eds.) *The Cultural Life of Machine Learning*, Cham: Springer International Publishing, 2021, pp. 31–78; Xiaochang Li, "'There's no data like more data": automatic speech recognition and the making of algorithmic culture', *Osiris* (1 July 2023) 38, pp. 165–82; Chris H. Wiggins and Matthew L. Jones, *How Data Happened: A History from the Age of Reason to the Age of Algorithms*, New York: W.W. Norton & Company, 2023; Ranjodh Singh Dhaliwal, Lucy Suchman and Théo Lepage-Richer, *Neural Networks*, Minneapolis: University of Minnesota Press, 2024.

29 Adam, op. cit. (13); D.D. Mahendran, 'Race and computation: an existential phenomenological inquiry concerning man, mind, and the body', PhD dissertation (unpublished), University of California, 2011; Penn, op. cit. (7); Devin Kennedy, 'Virtual capital: computers and the making of modern finance, 1929–1975', PhD dissertation, Harvard University, Graduate School of Arts & Sciences, 2019.

30 Disciplinary scales are addressed by Alma Steingart, *Axiomatics: Mathematical Thought and High Modernism*, Chicago and London: The University of Chicago Press, 2023; Bernard Dionysius Geoghegan, *Code: From Information Theory to French Theory*, Durham, NC: Duke University Press, 2023. Professional scales are addressed by Lorraine Daston, 'Enlightenment calculations', *Critical Inquiry* (1994) 21(1), pp. 182–202; Daston, *Rules: A Short History of What We Live By*, Princeton, NJ: Princeton University Press, 2022; Simon Schaffer, 'Babbage's intelligence: calculating engines and the factory system', *Critical Inquiry* (1994) 21(1), pp. 203–27; Schaffer, 'OK computer', in Michael Hagner (ed.), *Ecce Cortex: Beiträge zur Geschichte des modernen Gehirn*, Göttingen: Wellenstein, 1999, pp. 254–85; Schaffer, 'Ideas embodied in metal: Babbage's engines dismembered and remembered', in Joshua Nall, Liba Taub and Frances Willmoth (eds.), *The Whipple Museum of the History of Science*, Cambridge: Cambridge University Press, 2019, pp. 119–58; Joanne Yates, *Control through Communication: The Rise of System in American Management*, Baltimore: Johns Hopkins University Press, 1989; Crowther-Heyck, op. cit. (23). National scales are addressed by Cohen-Cole, op. cit. (23); Kelly Gates, *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*, New York: New York University Press, 2011. International scales are addressed by Projit Bihari Mukharji, 'Profiling the profiloscope: facialization of race technologies and the rise of biometric nationalism in inter-war British India', *History and Technology* (2015) 31(4), pp. 376–96; Iris Clever, 'Biometry against fascism: Geoffrey Morant, race, and anti-racism in twentieth-century physical anthropology', *Isis* (2023) 114(1), pp. 25–49; Audra J. Wolfe, *Freedom's Laboratory: The Cold War Struggle for the Soul of Science*, Baltimore: Johns Hopkins University Press, 2018; Arjun Appadurai, 'Number in the colonial imagination', in Carol Appadurai Breckenridge and Peter van der Veer (eds.), *Orientalism and the Postcolonial Predicament: Perspectives on South Asia*, Philadelphia: University of Pennsylvania Press, 1993, pp. 314–39. Pablo F Gómez, *The*

Structure of the issue

The structure of this *Themes* issue provides a basis for our consideration of several key, if unexpected, genealogies in the development of AI. Section 1, ‘Origins? Intelligence, capture, discovery’, considers general historical, historiographic and epistemological perspectives. Section 2, ‘Creativity, economy, and human–machine distinctiveness’, analyses researchers developing machine learning and AI technologies within the US, the UK and the Soviet Union from the 1950s to the 1980s. Section 3, ‘Seeing through computer vision, historically’, examines diverse elements of the means by which AI techniques have been incorporated in visual work from post-war radiography to large-scale visual data sets. Section 4, ‘“The social implications of machine intelligence” in the biometric state’ highlights the ways surprisingly long-term state, corporate and academic entanglements have shaped both early practitioners’ concerns with the social implications of AI and current bureaucratic programmes implementing AI in citizen identification and health policy in India and Argentina.

Section 1: Origins? Intelligence, capture, discovery

The articles in Section 1 explore central, often unspoken, paradigms, knowledge forms and justificatory frameworks within AI, and emphasize in particular the ways in which AI participates in the extractive and racist legacies of colonialism, race science and capitalism. A growing body of scholarship reveals that AI is predicated on the reproduction of colonial supply chains, systems of extraction and structures of power. For example, rare-earth minerals like coltan, often mined in abhorrent conditions in places like the Democratic Republic of the Congo, power the microelectronics that allow for the vast extraction and centralization of data in the hands of corporate, state and military actors, largely in the global North and West, who in turn use that data to develop AI systems that increase their profits and consolidate their power.³¹ Others have explored how AI reproduces epistemic forms of colonialism. For example, Chun and Barnett have shown how the ‘homophilic’ logic that ‘like belongs with like’ – which was central to eugenics and other race sciences and, with them, social order – is also central to the data-driven systems of classification and ‘prediction’ that constitute AI today.³² Still others have explored how imperial states enrol contemporary AI systems in order to preserve and maintain their power – for example, Theodora Dryer’s recent work articulating how AI supports settler control of natural resources.³³ The articles in this section take up the underlying logics of AI and its entanglements with colonial ways of knowing and engaging with the world, but from previously unexplored vantage points. Historian of science Jonnie Penn proposes a parallel between AI’s orientation to the mind and settler colonial orientations to the land; critical media scholar Luke Stark explores the forms of inference at work in machine learning and connects them to histories of phrenology and other forms of race science; and abolitionist Sarah T. Hamid identifies logics of capture and erasure of violence that are at work both in AI and in the histories that seek to ground its critique.

Experiential Caribbean: Creating Knowledge and Healing in the Early Modern Atlantic, Chapel Hill: University of North Carolina Press, 2017.

31 Nathan Ensmenger, ‘The environmental history of computing’, *Technology and Culture* (2018) 59(4), pp. S7–S33; Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, CT: Yale University Press, 2022; James H. Smith, ‘Tantalus in the digital age: coltan ore, temporal dispossession, and “movement” in the eastern Democratic Republic of the Congo’, *American Ethnologist* (2011) 38(1), pp. 17–35.

32 Chun and Barnett, op. cit. (7).

33 Dryer, op. cit. (19).

In ‘*Animo nullius*’, Penn advances a new concept to parallel the settler colonial notion of *terra nullius* – land that was said to belong to no one. Colonizers developed legal frameworks based on Western theories of private property, ‘civilization’ and statehood that allowed them to claim that the lands they colonized belonged to no one because they were not being cultivated and claimed according to Western logics.³⁴ Penn suggests that, similarly, AI sets up ‘intelligence’, or more generally ‘the mind’ (and some of its products), as unclaimed territory, spaces to be owned and structured according to the prescriptions of Western formalization – ignoring, erasing and discrediting the many cultures of knowledge and wisdom that are already there. Penn sets out first to explore how scientific appeal to neurophysiology has been mobilized by early and contemporary proponents of AI and obscures the latter’s entanglement with capitalist bureaucracy. Second, he makes the case for thinking about the genealogy of AI by drawing analogies with historical colonialism and contemporary discourse about data colonialism. Following Jon Agar’s suggestion to read the histories of information technologies and modern state formation together, Penn offers *animo nullius* for ‘no persons’ mind’, as a ‘heuristic’ to draw attention to seizure (and, in the context of capitalism, forms of enclosure) as elements vital to the economic logic of AI. Penn suggests transcending the constructed dichotomy between symbolic and connectionist approaches to AI in favour of focusing on what they have in common, namely mathematical formalization as a way of ‘claiming’ cognition for computing. Penn connects this history to the onset of private cloud infrastructure and the corporate capture of big data; crucially, he maintains that tech conglomerates ‘indulged AI’s sociotechnical imaginary to veil acts of seizure as acts of novel transformation or discovery’.

In ‘Artificial intelligence and the conjectural sciences’, Stark draws attention to the role of correlation as against causation (associated with modelling and theory) in the mobilization of abductive logic within AI – more specifically, machine learning. Building on prior work in the history and philosophy of statistics, including his own recent co-authored exploration of physiognomic AI, he explores how contemporary machine learning generates ‘automated conjectures’ based on concepts associated with the discredited conjectural pseudosciences of physiognomy and phrenology, central to nineteenth-century race science and eugenics.³⁵ Stark explores this phenomenon through the lens of Italian historian Carlo Ginzburg’s idea of ‘empirical science’, defined in terms of regularity and repeatability (rather than contingency), which he suggests is more properly understood in terms of a move from inductive or statistical to deductive inference based on probability rather than universal law. Stark’s overarching concern is to thwart the perpetuation and extension of societal injustice by ‘restricting the use of automated conjecture’. Commenting on the divinatory affordances of machine learning, Stark maintains that it ‘performs a double dance: abductive claims become deductive ones, and a contingent narrative about the past becomes a necessary one about the future’, thereby pointing to what others have referred to as the colonization of futures in the automated ‘prediction’ of the future.

In ‘History as capture’, Hamid interrogates the presumption that history can – and should – be mobilized as a means to critique what she refers to as ‘the cultural hegemony’ of computing. Hamid turns critiques of AI and its logics back onto historical scholarship itself, highlighting entanglements between AI and the many fields that now claim to

34 Yogi Hale Hendlin, ‘From terra nullius to terra communis’, *Environmental Philosophy* (2014) 11(2), pp. 141–74; Penn’s approach extends Nick Couldry and Ulises A. Mejias, ‘Data colonialism: rethinking big data’s relation to the contemporary subject’, *Television & New Media* (2019) 20(4), pp. 336–49.

35 Luke Stark and Jevan Hutson, ‘Physiognomic artificial intelligence’, *Fordham Intellectual Property, Media, and Entertainment Law Journal* (2022) 4, pp. 922–78.

critique it or to hold it to account. Most centrally, she proposes that both AI and traditional history of computing treat violence and oppression as non-normative, exceptional and peripheral to computing rather than constitutive and pervasive. Hamid identifies logics and practices of ‘capture’ at work in the history of computing that closely parallel those increasingly identified in AI: senses of ‘history’ and ‘development’ at work in the International Congress of Mathematics; recent historiographies of computing (and mathematics, science and technology) offered by so-called ‘guild historians’; and hegemonic discourses that academic historians establish and maintain for disciplines such as mathematics and AI, even while critiquing them. Rather than abandon history, she invites the reader to do it ‘differently’, attending to histories that have been ‘displaced and banalized’ via something akin to a weaponization of critical reflexivity. In this connection she points to ‘a line of continuity through carceral geographies: the Middle Passage, the plantation, the reservation, the prison, the housing project, the refugee camp, the detention centre, the border, and so on.’ Hamid proposes that these are not sites where Western ideas and technologies were poorly applied or badly wielded – they are the sites in which central technologies and concepts were conceived to solve problems and maintain control precisely there.

Taken together, these three articles expand the sense in which we understand AI’s entanglement with the logics of colonialism and race science. Each explores Western forms of knowledge and power as they are manifested in AI’s core organizing logics and frameworks. Together they reveal how coloniality structures and delimits our relationships with knowledge, as well as with land and people. Such logics aim not to be seen, but rather to be taken for granted as perfectly natural, inevitable and unchangeable. This opening section works against that obfuscatory project to articulate epistemic facets of colonial violence as they structure AI and its histories.

Section 2: Creativity, economy, and human–machine distinctiveness

A central facet of the longer history of AI is a recurring insistence on its very impossibility. In the middle of the nineteenth century, while reflecting upon Charles Babbage’s proposed analytical engine, Ada Lovelace denied that machinery could originate anything new.³⁶ In decades of reflection on the limits of machines versus human beings, Harry Collins came to focus on the resolutely social qualities of key facets of human reasoning. ‘The Western technical intelligentsia’, the Marxist philosopher Evald Ilyenkov wrote, is ‘entangled in the problem of “man–machine” because they don’t know how to formulate it properly; that is, as a social problem, as a problem of the relationship between man and man, mediated by the material body of civilization, including the modern machine technology of production.’³⁷ To see machines as intelligent was to forget the fundamental lessons of nothing less than commodity fetishism itself: to mistake technological developments for their underlying social foundation.

Rather than assuming any teleologies about the trajectories of human and machine intelligence, in their respective essays historians of science Ekaterina Babinsteva, Aaron Mendon-Plasek and Sam Schirvar, and historian and sociologist Olessia Kirtchik, each underscore how generative it has been to seek the divide between what machines and

³⁶ ‘The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform.’ L.F. Menabrea, ‘Sketch of the analytical engine, invented by Charles Babbage’ (tr. Ada Augusta, Countess of Lovelace), 82, Bibliothèque universelle de Genève, 1942, at www.fourmilab.ch/babbage/sketch.html (accessed October 2020).

³⁷ Evald Ilyenkov, Aleksandr Arsen’ev and Vassily Davidov, ‘Mashina i chelovek: kibernetika i filosofiya’ (Machine and human: cybernetics and philosophy), in Fedor Konstantinov (ed.), *Leninskaya Teoriya Otrazheniya i Sovremennaya Nauka*, Moscow: Nauka, 1966, pp. 265–83, quoted in Kirtchik in this issue.

human can do, and what they can each do best, in both the USSR and the West. Thinking about machine intelligence in their cases involves thinking about the powers and limits of human intelligence, and not necessarily supplanting it. Creativity figured prominently as a philosophical, a pedagogical and, fundamentally, an economic concern. In accentuating the centrality of dramatic technological transformations to develop and upend economies and labour markets, leaders and researchers in the US and the USSR alike sought to understand creativity and to produce practices and tools to enhance and support it, and to reflect upon how transformations in creativity would have deep impacts for the nature of future labourers. However much historians of technology rightly urge the rejection of histories of technological development focused upon innovation, beliefs about the political and economic necessity for innovation loosened the resources to undertake long-term research programmes on humans and machines. While decidedly grounded in military and economic support, the results do not simply map onto a Cold War logic; they reveal diverse possible Cold War programmes investigating and studying human creativity.

Bringing together the history of Xerox with the accounts of machines under dialectical materialism highlights the historical contingency of what gets classified as AI and what does not. What was deemed AI in the Soviet Union in the 1970s came to be branded primarily as ‘human-computer interaction’ in the United States, with dramatically different disciplinary developments – and historiographies and critical commentary. In his famous manifesto ‘Man-computer symbiosis’, the US psychologist J.C.R. Licklider sharply contrasted technologies for human-machine collaboration with fully machinic intelligence that he was certain would come.³⁸ As the articles here show, the partition was different in the USSR, and could be again.

In the Soviet context, dialectical materialism precluded strong claims that machines might achieve human intelligence. Far from limiting researchers, this approach amplified programmes to seek coordination between humans and machines, or, more often, between hierarchies of humans and hierarchies of machines. While much of the research of the first few decades of AI work focused heavily upon efforts to formalize aspects of reasoning and human action, other researchers centred on just those facets least amenable to formalization. Kirtchik and Babintseva reveal the diverse ways in which Soviet researchers developed research agendas presuming that human reasoning was not fully formalizable. Fundamental questions of control in a socialist economy rested on a superior account of human and machine capacities – and limits.

Tracing the roots of Lev Landa’s algo-heuristic theory (AHT), Babintseva’s article, ‘Rules of creative thinking’, explores how and why the Soviet Union came heavily to support research on human creativity in the late 1960s. The workforce of the future would be less about physical work in factories than about creative work using automatic control and digital technologies. A revitalized Soviet cybernetics sought to replace qualitative pedagogical approaches to stimulating creativity with a powerful quantitative approach to psychology. Finding famous American projects to automate theorem proving to have empirically and theoretically inadequate accounts of the mind at work, Veniamin Pushkin studied problem solving in action by documenting, for example, the eye movements of chess players. Babintseva explains that Pushkin found that ‘Simon and Newell’s neat decision trees had little to do with the actual messiness of human cognition’. Improving automation – and improving the humans involved with automation – required grasping how machine capacities differed from human minds.

³⁸ J.C.R. Licklider, ‘Man-computer symbiosis’, *IRE Transactions on Human Factors in Electronics* (1960) HFE-1, pp. 4–11.

In ‘The Soviet scientific programme on AI’, Kirtchik charts how Soviet scientists and engineers came to view machines as ‘tools to think with’ rather than ‘thinking machines’. Focusing on the former general and researcher Dmitry Pospelov, Kirtchik shows how the term ‘AI’ was redefined, from the 1970s onward, to refer to ‘a control system dealing with complex and weakly formalised domains and problems, not with deterministic and numerical methods, and simulating the way humans think and operate’. The distinctive conception of AI that Pospelov skilfully delineated enabled an entire research ecosystem to emerge. Rather than assess problems of optimization or statistical induction, this version of AI sought to provide more qualitatively robust forms of planning and control. Kirtchik argues that in this era Soviet AI ‘lies precisely at the blurred boundary where cybernetic *control* of machines becomes *management* of human societies’.

While much of this Soviet effort remained largely theoretical, the empirical study of the texture of control and management was at the heart of the Applied Information-Processing Psychology Project (AIP) at the Xerox Corporation. In ‘Machinery for managers’, Schirvar tracks the dramatic shifts in the assumptions about the humans involved in human-computer interaction: gendered assumptions about labour, creativity and skill. Early advocates for improved human-computer practices like Licklider and Douglas Englebart envisioned their ideal users as knowledge workers like themselves, autonomous, creative, buried in paperwork – and almost exclusively male. Working within the commercial imperative of Xerox, researchers deployed their psychological methods to understand the so-called ‘naive’ user – the secretary, gendered female, involved in routine yet skilled tasks, above all in typing. And yet the coming – and successful marketing – of the personal computer by the early 1980s displayed the default user to be a neutered, but implicitly male, manager and thinker. In these empirical studies, researchers underscored the distinctiveness of human behaviour while using computing machines, ultimately justifying the founding of a distinctive discipline of human-computer interaction.

In his ‘Irreducible worlds of inexhaustible meaning’, Mendon-Plasek offers three case studies of researchers in the early 1950s who envisioned learning by machines as ‘the capacity to respond appropriately to unexpected or contradictory new data by generating interpretations that might complement, surprise or challenge human interpretations’. These advocates insisted not on the objectivity of computerized approaches but rather on their subjectivity and creativity, notably when confronted with complex empirical data. Precisely because a computer could create new categories in learning, they were simultaneously important to computing and philosophy. Rather than merely reproducing current scientific and social classifications, these forms of learning seemed to offer the possibility of radically reworking existing ways of dividing up the world. Machines might act and see otherwise than their human creators. In his approach, Mendon-Plasek seeks to understand how machine learning’s emphasis on subjectivity can serve as a kind of social relation generator.

These four grounded histories dramatically reshape traditional accounts of the development of AI through their expansion of the actors considered, the salience of questions of philosophy and labour throughout, and their fundamental resistance to easy political, ethical and technological teleologies. All four articles suggest how intelligence and its ramifications might be envisioned and institutionalized in diverse ways.

Section 3: Seeing through computer vision, historically

Although with the release of ChatGPT, Google Bard and other generative AI chatbots textual manipulation has recently overshadowed visual exemplifications of AI capabilities, over the previous decade many of the most powerful and most problematic

exemplifications of AI have come in pattern recognition and machine-learning systems that link visual recognition with purportedly cognitive abilities. At border controls travellers with e-passports are checked against facial-recognition systems; self-driving cars rely on continuously assessing and updating their models of the environment; ‘visual capabilities’ enable drones to deliver surveillance, firepower and medical goods; and Microsoft PowerPoint offers design suggestions as you incorporate different media in a slide presentation. As Anthony McCosker and Roman Wilken note in their sociological study *Automating Vision*, the 2018 demonstration of the Chinese government’s capacity to link citizen information with facial-recognition systems and prosecute people jaywalking at a crowded intersection can stand as a defining image of this manifestation of AI.³⁹ Yet the questions raised never concern solely how computers process visual information; they also require a careful understanding of how and for whom they do visual work. That the government and people of China have integrated these technologies in a form of ‘social contract’ speaks to the ambivalences of automatic vision, which McCosker and Wilken regard as amounting, in the present surveillance society, to a new age of ‘camera consciousness’. Without in any way diminishing recognition of the scale and significance of present image work, the articles in this section instead underline important developmental continuities across time, as expressed both in the self-understanding of AI researchers and in the databases on which publicly facing image technologies have been trained.

Complementing and extending Orit Halpern’s investigation of the nexus between reason and vision in cybernetics and urban planning, as well as Jacob Gaboury’s study of computer graphics, and recent concerns with facial recognition, these articles collectively go several steps toward providing an unusually comprehensive and probing investigation of significant features in the development and uses of ‘computer vision’, from the 1950s to the present.⁴⁰ Although their centres of gravity vary, each contributes significant insight into the relations between research communities and the public and commercial environments in which computer-aided vision has been developed, in very different contexts.

In ‘Errors and fallibility in radiology’, historian of science and media Flora Lysen draws on the concerns of science and technology studies and historical epistemology to study the medical detection of lung disease from the 1950s onwards, showing that current arguments for AI systems deploy similar strategies to much earlier work integrating computerized records into image reading. Similarly, historian Harry Law’s tightly focused account of critical work in optical character recognition in the 1990s shows researchers deploying new versions of brain metaphors with roots in AI research in the 1950s and 1960s, a point that strongly reinforces Penn’s *animus nullius* heuristic. Simon Michael Taylor, a scholar of biometrics, governance and digital technologies, and Bruno Moreschi, academic researcher, artist and filmmaker, focus instead on the broadly based systems in which animal bodies or photographic images have been incorporated

³⁹ Anthony McCosker and Roman Wilken, *Automating Vision: The Social Impact of the New Camera Consciousness*, New York: Routledge, 2020, p. 13 and Chapters 1–3 on ‘Interrogating seeing machines’, ‘Camera consciousness’ and ‘Face values’.

⁴⁰ Orit Halpern, *Beautiful Data: A History of Vision and Reason since 1945*, Durham, NC: Duke University Press, 2014. Recently Jacob Gaboury has explored the virtual reproduction of images and environments in computer graphics and the underlying algorithmic and data-driven techniques of representation in *Image Objects: An Archeology of Computer Graphics*, Cambridge, MA: MIT Press, 2021. The history of ‘facial recognition’ – one of the largest application domains of computer vision – has also begun to receive sustained attention. See, for example, Eden Medina, ‘Forensic identification in the aftermath of human rights crimes in Chile: a decentered computer history’, *Technology and Culture* (2018) 59(4), pp. S100–S133; Gates, *op. cit.* (30); Stephanie Dick, ‘The standard head’, in Jeffrey Yost and Gerardo Con Diaz (eds.), *Just Code*, Baltimore: Johns Hopkins University Press, 2024.

into AI technologies. In ‘Species *ex machina*’, Taylor shows that real-time estimates of cattle meat and fat proportions in use today draw on heterogeneous sources shaped strongly by product surveillance regimes responding to mad cow disease from the 1980s. Moreschi’s article, ‘Five experimentations in computer vision’, demonstrates how the large visual databases currently being used to train AI systems to recognize elements of everyday life are problematically marked by the limitations of current Western populations – even as they are labelled by precarious labourers working in the global South as well as in the US.

While rigorous efforts to manage error emerge as a major issue in Lysen’s study of radiology and Law’s account of refinements in machine capabilities in reading numerals, the capacious looseness of categorical work and the absence of critical scrutiny mark equally Taylor’s study of cattle crushes and the ready transposition of techniques across animal and human environments, as well as Moreschi’s experimental work examining how database image labels are applied and might be developed more responsibly. Collectively, these authors remind us that training AI to ‘see’ relies on hidden human labour in its production (likely this is without exception), but often also on blinding human vision in its implementation (or, at least, on promoting conceptual myopias). Yet their detailed studies show that examining expert communities and the subjects of their work can offer ways of understanding and reworking the implicit power structures deployed in computer vision systems and parsing some of the more or less subtle senses in which they involve redefining what it means to ‘see’.

Lysen’s account of radiographers’ work to improve diagnostic success in reading X-ray photographs in the 1950s shows that this exposed a troublingly high error rate, and differences even between the same observers. Their focus on the fallibility of human judgement, Lysen shows, prepared the ground for the development of computer programs to collate collective experience as well as work to formalize judgement procedures and render them accessible to statistical measure and improvement. It is an extremely important point that proposals for computerized decision making and vision have often relied on an argument for the fallibility of human judgement. Histories of AI, therefore, ‘are also histories of imaginaries of human (in)competences’; but, as Lysen shows here, these technologies have aided, without in any way escaping, the necessity of human judgement in the expert community of radiographers. Radiographers have periodically renewed engagement with these difficulties of interpretation over the past seventy years and economic and sociocultural considerations have shaped both their conception and proposed solutions.

Moreschi’s account engages both work practices and the images used in large visual databases – and discloses major limitations in their comprehensiveness and the ways categories deployed in labelling have been uncritically derived from textual databases. Moreschi shows the hidden work through which images stripped of context are then reconstituted for computer systems, and demonstrates unambiguously that the product reflects the legacies of colonial power structures. For example, a subfolder’s fish images are cradled in white hands, often as trophies – our largest visual databases reflect fishing as recreation rather than the diversity of fishing work and fish throughout the world. Moreschi’s article is methodologically innovative: first, in yielding historical insight through experimentations; second, in taking the distinctive further step of drawing on artistic practices developed in 1960s resistance to South American dictatorships to show how visual databases can be curated to disclose rather than obscure the power structures of communal vision.

These articles also indicate how important tight control of their subject matter has been to commercial success in the development of AI techniques. This is true of the Bell Labs research group primarily responsible for the implementation of techniques of

back propagation, convolutional neural networks and statistical weight management that have drastically improved computational speed. It is also evident in the mix of agricultural and computer science expertise that has taken off-the-shelf video-gaming devices to yield under-the-skin surveillance and assessment of animal flesh. In ‘Bell Labs and the “neural” network, 1986–1996’, Law shows that managing what counts as a handwriting sample enabled Bell Labs researchers to present error rates in their favour, through careful curation of the training database and test procedures that they deployed to develop optical character recognition of numerals. Combined with an imaginative use of brain metaphors only loosely connected with the techniques they described, this helped depict machine reading techniques as autonomously cognitive – with the statistical weight management described as ‘optimal brain damage’ – when their relative failure to match human skill might easily have seemed as significant. Examining the more distributed research environment of industrial agriculture and showing how important it is for our studies of AI to incorporate work with animals, Taylor discloses an ambivalence in the total control asserted over animal bodies on the path to slaughterhouses that helps researchers shift AI and digital surveillance techniques easily between different fields of commercial operation. Taylor’s examples owe part of their origins to food safety regulations, and on the other hand might escape ethical scrutiny because they concern (in the present moment) non-human animals, not humans. This is just one of the many instances in which our authors’ analytic work has used historical research to heighten moral conscience as well as more conscious command of the diverse ways we use AI.

Section 4: ‘The social implications of machine intelligence’ in the biometric state

As occurs throughout this issue, the three articles that comprise its final section revisit themes of state–corporate–academic entanglement, ambivalence and/or disingenuousness over technologists’ role(s) in social engineering, and, perhaps most surprisingly, the pronounced historical continuities or conservatism (rather than liberatory rupture via technology) that ‘AI’ and statistical tools have helped to sustain across varied locales and time periods. First, in ‘The “artificial intelligentsia” and its discontents’, historian Rosamund Powell chronicles AI researchers’ efforts in the 1970s to speculate on – yet, conspicuously, do almost nothing to address – the societal impacts wrought by their craft. In ‘Biometric data’s colonial imaginaries continue in Aadhaar’s minimal data’, media studies scholar Sananda Sahoo considers the prehistory of biometrics in India, from research by Thomas Nelson Annandale and P.C. Mahalanobis in the early twentieth century to the 2010 launch of Aadhaar – the largest biometric system in the world. Lastly, in ‘Predictive puericulture in Argentina’, anthropologist Alexa Hagerty, researcher–activist Florencia Aranda and researcher–journalist Diego Jemio connect an AI-enabled ‘predictive’ platform for adolescent pregnancy deployed in Salta Argentina in the 2010s to long-standing forms of biopolitical governance in that context – another instance of the world made old, not new, by AI.

State–corporate–academic entanglements underlie these papers in different measures. Sahoo’s account probes overlaps between efforts ‘to develop statistical methods on the one hand and aid governance on the other’. Powell, similarly, calls the organizers of the 1972 Social Implications of Machine Intelligence Research conference the ‘Serbelloni group’ after the villa along Lake Como, Italy, where they convened. Passed from a Catholic archbishop to the Duke of San Gabrio, the villa was by then operated by the Rockefeller Foundation, a philanthropy at the heart of the US foreign-policy establishment that had served as a primary funder of the Dartmouth Summer Research Project on Artificial Intelligence two decades earlier. Powell shows that the accuracy of AI researchers’ predictions had less to do with their area of expertise than with their

standing within the dominant military-industrial-university nexus of their time. 'Following the 1970s', she writes, 'the symbolic AI approach was largely abandoned in favour of neural networks, and gradually the very harms predicted by the Serbelloni group came to pass because of new methods which they had not considered.'

One sees in these papers the initial contours of an AI history that takes stock of its evolving patrons, partners and benefactors – in this instance, the Indian state, an Argentinian province, an American parasocial philanthropy named after a nineteenth-century corporate titan, and civil society organizations with strong ties to the Catholic Church. Each entanglement is contextual and distinct, if linked by a commitment to managerialism as an ideal. The Serbelloni group aspired to map out and plan for AI's 'social implications' even if they did not want to address them. Mahalanobis is well known as the progenitor of the Mahalanobis distance, a technique still popular for cluster analysis and classification. Sahoo captures how statistical imaginaries about state planning that stemmed from his 1920s work – and relied upon sampling populations – by the 2010s and 2020s had given way to a mutated state-corporate form of biometrics. This regime now imposes upon individual citizens' access to private services like banking and telecoms, an expansion of biometric management to include industry. In Argentina, similarly, the notion of pregnancy 'prediction' emerged as a collaboration between the government of Salta and Microsoft in the mid-2010s, capitalizing on discourse about underpopulation that dates back to the nineteenth century.

As is true of other articles in this issue, each of the final contributions speaks to the present from the perspective of history. In *Rules: A Short History of What We Live By* Lorraine Daston celebrates this move: 'One of the uses of history', she argues, 'especially history pursued on a longer time scale, is to unsettle present certainties and thereby enlarge our sense of the thinkable'.⁴¹ Each author moves historical certainties – about the morality of pregnancy, or risks of biometrics, or the self-appointed indemnity of technical genius – from black and white to grey. Powell works between the 1970s and the period since the 2010s to illustrate, for example, that the binary between AI's champions and discontents, which Garvey has most recently brought to light, was not that clear cut, and that contemporary hopes for remedies like algorithmic auditing are, by now, half a century old.⁴² In sum, this set of accounts brings to the fore how techniques treated by some as innovative figured in longer efforts to oppose innovation. As Daston suggests, to revisit these histories is to challenge what could be new.

Conclusion

The articles of this issue aim to deepen our understanding of AI, its genealogy and its historical character: as an intellectual project, a science, an industrial art, a management tool, a promise. Historical understanding can be a powerful tool for breaking with long-taken-for-granted paradigms and assumptions about language, norms and possibility. At its best, history salvages the complexities of past decisions, and decision makers, to populate one's imagination about potential new social practices. The histories offered in this issue aim not to critique AI for the purpose of its betterment, but rather to develop a clearer picture of what and where AI is, what and where it *might be*, and what and where it perhaps *should not be*. Abolitionist frameworks and a renewed interest in the Luddites signal a growing politics of refusal in response.⁴³ By asking after the empirics behind

⁴¹ Daston, *Rules*, op. cit. (30), p. 22.

⁴² Garvey, op. cit. (7)

⁴³ Ruha Benjamin, 'Informed refusal: toward a justice-based bioethics', *Science, Technology, and Human Values* (2016) 41(6), pp. 967–90; Joanna Radin, 'Digital natives: how medical and indigenous histories matter for big

origin myths that treat AI as a natural marriage of logic and computing in the mid-twentieth century, these articles situate AI within longer – and more socially contingent – histories of industry, statecraft, epistemology and control.

In doing so, this issue reveals how AI figures in a long and steady expansion and naturalization of managerial power, one that extends even beyond the significant powers afforded to management and bureaucracy in the context of nineteenth-century industrialization. Managerial techniques from actuarial sciences, office culture, population sampling, livestock handling and elsewhere are often directly imported into and encoded within AI systems. As historians of computing Daniel Volmar, Thomas Haigh and Nathan Ensmenger have explored, modern digital computers reinforced and expanded the scope of that managerial power, even while reconfiguring it within the context of corporate offices, government agencies, and the American defence establishment.⁴⁴

AI, in its various historical manifestations, represents a further expansion of managerial forms of control, managerial epistemologies and the philosophy of management across sites and scales. The articles gathered here create space to consider what is being managed by AI systems – whether it is populations, students’ minds, natural resources, images, livestock, stories, office work, diagnosis or discourse – and according to what techniques – from abductive reasoning and biometric surveillance to record-keeping practices and technocratic institutionalization. Together they signal the value of an expansive history of AI that allows us to appreciate how contemporary technologies concretize epistemes, ideologies and genealogies far beyond what dominant origin myths and traditional computer histories can reveal.

Supplementary material. The supplementary material for this article, its preface, can be found at <https://doi.org/10.1017/bjt.2023.15>

Acknowledgements. We are enormously grateful for the role that Susie Gates played as the HoAI Mellon Sawyer Seminar events coordinator, and for the support of administrators in the Department of History and Philosophy of Science in different aspects of the formation and operation of the seminar, including especially Tamara Hug, Louisa Russell and Aga Launucha, as well as Kay Daines and Yaritza Bennett at the Research Office of the University of Cambridge. The generosity of the Mellon Foundation was exemplified by all its staff, and we thank in particular Yoona Hong and Martha Sullivan. This *Themes* issue was focused through a call for papers and our Winter Symposium – we thank all participants for the way they shaped the project and our collective scholarship as it emerged. In particular, Bo An, Michael Castelle, Fernando Delgado, Shunryu Garvey, Andrew Meade McGee, Paola Ricarte Quijano and Youjung Shin were important in our thinking and goals for this *Themes* issue, and we are grateful for all the work they put towards it too. This issue has benefited greatly from an unusually diverse group of referees and from the patient care with which Rohan Deb Roy and Trish Hatton at *BJHS Themes* have helped us refine its contributions.

data’, *Osiris* (2017) 32, pp. 43–64; McQuillan, op. cit. (7); Gavin Mueller, *Breaking Things at Work: The Luddites Were Right about Why You Hate Your Job*, New York: Verso, 2021; Syed Mustafa Ali, ‘A brief introduction to decolonial computing’, *XRDS: Crossroads, The ACM Magazine for Students* (2016) 22(4) pp. 16–21; Chelsea Barabas, ‘To build a better future, designers need to start saying “no”’, *OneZero*, 13 October 2020, at <https://onezero.medium.com/refusal-a-beginning-that-starts-with-an-end-2b055bfc14be> (accessed 28 November 2023); Jonnie Penn, ‘Algorithmic silence: a call to decomputerize’, *Journal of Social Computing* (December 2021) 2(4), pp. 337–56, <https://doi.org/10.23919/JSC.2021.0023>.

⁴⁴ Volmar, op. cit. (22); Thomas Haigh, ‘Technology, information and power: managerial technicians in corporate America, 1917–2000’, PhD dissertation, History and Sociology of Science, University of Pennsylvania, 2003; Thomas Haigh, ‘Inventing information systems: the systems men and the computer, 1950–1968’ *Business History Review* (2001) 75(1), pp. 15–61; Nathan Ensmenger, *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*, Cambridge, MA: MIT Press, 2012.

Cite this article: Ali SM, Dick S, Dillon S, Jones ML, Penn J, Staley R (2023). Histories of artificial intelligence: a genealogy of power. *BJHS Themes* 8, 1–18. <https://doi.org/10.1017/bjt.2023.15>