Harvard Data Science Review • 1.1

Artificial Intelligence

Stephanie Dick

Published on: Jun 23, 2019

Updated on: Oct 04, 2019

DOI: [10.1162/99608f92.92fe150c](https://doi.org/10.1162/99608f92.92fe150c)

There is a plaque at Dartmouth College that reads: “In this building during the summer of 1956 John McCarthy (Dartmouth College), Marvin L. Minsky (MIT), Nathaniel Rochester (IBM), and Claude Shannon (Bell Laboratories) conducted the Dartmouth Summer Research Project on Artificial Intelligence. First use of the term ‘Artificial Intelligence.’ Founding of Artificial Intelligence as a research discipline ‘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.’” The plaque was hung in 2006, in conjunction with a conference commemorating the 50th anniversary of the Summer Research Project, and it enshrines the standard account of the history of Artificial Intelligence—that it was born in 1955 when these veterans of early military computing applied to the Rockefeller Foundation for a summer grant to fund the workshop that in turn shaped the field. The plaque also cites the core conjecture of their proposal: that intelligent human behavior consisted in processes that could be formalized and reproduced in a machine (McCarthy, Minsky, Rochester, & Shannon, 1955).

Grounded in the postwar traditions of systems engineering and cybernetics, and drawing from the longer history of mathematical logic and philosophy aimed at formal descriptions of human thinking, they held that cognitive faculties could be abstracted from the supporting physical operations of the brain. Thus the former could, in principle, be reproduced in different material substrata so long as the formal rules could be executed there (Kline, 2011, 2015). Two of the attendees, Herbert Simon and Allen Newell, influentially proposed more specifically that human minds and modern digital computers were ‘species of the same genus,’ namely *symbolic information processing systems*; both take symbolic information as input, manipulate it according to a set of formal rules, and in so doing can solve problems, formulate judgments, and make decisions (Crowther-Heyck, 2008; Heyck, 2005; Newell & Simon, 1972). After the 1956 workshop, this became the dominant approach, and artificial intelligence researchers accordingly set out to identify the formal processes that constituted intelligent human behavior in medical diagnosis, chess, mathematics, language processing, and so on, in hopes of reproducing that behavior by automated means.

Overwhelmingly, however, artificial intelligence today resembles this symbolic approach in name only. Most key commitments and approaches were abandoned over the course of the twentieth century. Perhaps most notably, *human intelligence* was the central exemplar around which early automation attempts were oriented. The goal was to reproduce intelligent human behavior in machines by uncovering the processes at work in our own intelligence such that they could be automated. Today, however, most researchers want to design automated systems that perform well in complex problem domains by *any* means, rather than by *human-like* means (Floridi, 2016). In fact, many powerful approaches today set out intentionally to bypass human behavior, as in the case of automated game-playing systems that develop impressive strategies entirely by playing only against themselves, keeping track of what moves are more likely to produce a win, rather than by deploying human-

inspired heuristics or training through play with human experts (Pollack & Blair, 1997; Tesauro, 1995). That the core project could have changed so dramatically highlights the fact that what counts as *intelligence* is a moving target in the history of artificial intelligence.

That research communities picked out different behaviors and processes as constitutive of intelligence is actually also highlighted in the early history itself. Standard historical accounts of artificial intelligence often overstate the significance of the Dartmouth workshop and the symbolic approach associated with it. Indeed, even according to the participants themselves, the workshop was something of a disappointment. McCarthy recollected that “anybody who was there was pretty stubborn about pursuing the ideas that he had before he came, nor was there, as far as I could see, any real exchange of ideas” (McCorduck, 2004, p. 114). McCarthy’s lamentation also hints at the fact that approaches to artificial intelligence research were more multifaceted than accounts of “good old-fashioned AI” (as symbolic artificial intelligence was dubbed in the 1980s) might suggest.

For example, proponents of a field called ‘expert systems’ rejected the premise that human intelligence was grounded in rule-bound reasoning alone. They believed, in part because of the consistent disappointment attendant to that approach, that human intelligence depended on what experts *know* and not just how they think (Brock, 2018; Collins, 1990; Feigenbaum, 1977; Forsythe, 2002). Edward Feigenbaum (1977), the Stanford-based computer scientist who named this field, proposed that:

We must hypothesize from our experience to date that the problem-solving power exhibited in an intelligent agent’s performance is primarily a consequence of the specialist’s *knowledge* employed by the agent, and only very secondarily related to the generality and power of the inference method employed. Our agents must be knowledge-rich, even if they are methods-poor. (p. 3, emphasis added)

In this approach, ‘knowledge engineers’ would interview human experts, observe their problem-solving practices, and so on, in hopes of eliciting and making explicit what they knew such that it could be encoded for automated use (Feigenbaum, 1977, p. 4). Expert systems offered a different explanation of human intelligence, and their own theory of knowledge, revealing that both were moving targets in this early research.

Still others, many of whom were interested in automated pattern recognition, focused on attempts not to simulate the human *mind* but to artificially reproduce the synapses of the *brain* in ‘artificial neural networks.’ Neural networks themselves date from the 1940s and 50s, and were originally meant to simulate brain synapses by digital means (Jones, 2018). These neural networks, now largely stripped of all but the most cursory relationship to human brains, are at work in many of today’s powerful

machine learning systems, emphasizing yet again the protean character of ‘intelligent behavior’ in this history.

The history of artificial intelligence is, therefore, not just the history of mechanical attempts to replicate or replace some static notion of human intelligence, but also a changing account of how we think about intelligence itself. In that respect, artificial intelligence wasn’t born at Dartmouth in 1955, as the standard account would have us believe, but rather participates in much longer histories of what counts as intelligence and what counts as artificial. For example, essays in Phil Husbands, Owen Holland, and Michael Wheeler’s *The Mechanical Mind in History* (2008) situate symbolic information processing at the *end* of a long history of mechanical theories of mind.

Artificial intelligence belongs in the history of *human intelligence*, in ways that complicate teleological accounts in which symbolic AI emerge naturally and inevitably from attempts across centuries to reduce human reasoning to a logical formalism. Human cognitive faculties have been theorized, partitioned, valued, and devalued in different ways at different times. For example, in the 18th and early 19th centuries, complex arithmetic calculations were associated with virtuosic mathematical genius (Daston, 1994). Famous mathematicians the likes of Gauss and Laplace were celebrated in their time for their “lightning arithmetic.” But by the mid-nineteenth century, calculation had largely lost its association with genius, and was “drifting from the neighborhood of intelligence to that of something very like its opposite” (Daston, 1994, p. 186). Calculation was demoted to the “merely mechanical” or the “automatic,” requiring no real presence of mind or cognitive attention. Historian Lorraine Daston tracks this shift within the efforts of French mathematician Gaspard de Prony, who applied Adam Smith’s principles of the division of labor to mathematical calculation, breaking down complex calculations into small and simple arithmetic operations that could be executed (indeed executed best) by uneducated people. The “deskilling” of calculation and the demotion of this cognitive faculty precipitated a shift in the demography of calculators; once the purview of *philosophes* and mathematicians, calculation became in the 19th century the domain of unskilled laborers and of women. In large numbers, these “human computers” (indeed, for most of its history, the work ‘computer’ referred to a person rather than to a machine), could carry out complex calculations by aggregating their arithmetic labor. Once demoted to the realm of the “merely mechanical,” calculation was also ripe for machine automation, and mathematicians like Charles Babbage and Charles Thomas de Colmar set out to design machines that could do for mental labor what factory automation did for physical labor (Schaffer, 1994). Daston (1994) uses this historical transformation to argue that “what intelligence meant and- still more telling- who had it shifted in tandem with the meanings and subjects of calculation” (p. 186). Human, and especially women, computers were employed well into the 20th century, especially by the British and American governments, and many also became programmers of the early digital computers that would eventually replace them, remaking yet again the economy of labor and intelligence (Abbate, 2012; Ensmenger, 2010; Grier, 2007; Hicks, 2018; Light, 1999).

This history also points to the fact that attempts to produce intelligent behavior in machines often run parallel to attempts to *make human behavior more machine-like*. From the disciplining of 19th-century factory workers' bodies by the metronome to the automatic and unthinking execution of arithmetic that De Prony sought in his human computers, automation efforts often parallel the disciplining of human minds and bodies for the efficient execution of tasks. Indeed, Harry Collins has argued that machines can only appear to be intelligent in domains where people have already disciplined themselves to be sufficiently machine-like, as in the case of numerical calculation (Collins, 1992). This historical perspective invites a reconsideration of 21st century artificial intelligence as well. As anthropologist Lucy Suchman (2006) has proposed, artificial intelligence “works as a powerful disclosing agent for assumptions about the human” (p. 226). What behaviors are selected as exemplars of intelligence to be replicated by machinery? Whose cognitive labor is valued and devalued, displaced or replaced by the new economies of intelligence that surround modern digital computers or powerful machine learning systems?

In fact, the most powerful and profitable artificial intelligences we have produced—those of today's machine learning—exhibit a rather limited range of intelligent behavior. Overwhelmingly, machine learning systems are oriented towards one specific task: to make accurate predictions. Drawing on statistical techniques that date back to the mid-20th century, machine learning theorists aim to develop algorithms that take a huge amount of data as input to a neural network, and output a prediction rule or a classifier for the relevant domain in polynomial time (Valiant, 1984; Wald, 1947). Given this specific focus and particular history, many machine learning theorists would not situate their work in artificial intelligence in the traditional sense of the term at all. However, perhaps ironically, machine learning systems are billed as artificial intelligence by many public facing institutions, developers, and corporations, even though the original goal—to simulate human intelligent behavior—has not been a part of machine learning since neural networks were stripped of all but the barest reference to human brains in the 1960s (Jones, 2018). The name has stuck, in spite of the quite different definitions of ‘intelligence’ and ‘artificial’ at work there.

Though machine learning systems have proven to be extremely powerful prediction engines, the new “artificial intelligence” has also been controversial. In 1997, John McCarthy penned a critique of DeepBlue in the wake of its victory over Kasparov. DeepBlue did not make use of neural networks, but it anticipated some debates that surround machine learning by deploying a brute force approach that would be unavailable to human players. McCarthy lamented that the human exemplar had not grounded its design—he didn't want a program that could play chess well by any means, he wanted the recreation of human intelligence by automated means (1997). The critiques against inhuman artificial intelligences continue today. Some, like Noam Chomsky, have found the emphasis on predictive accuracy in machine learning epistemologically problematic and unsatisfying (Norvig, 2011). He doesn't believe scientists should forfeit their concern with understanding and explanation in exchange for

powerful data-driven prediction. Others, like Virginia Eubanks, Julia Angwin, Safiya Noble, and Cathy O’Neil, have noted how machine learning predictions can introduce ethically and socially problematic decision-making practices and reproduce or entrench bias and inequality (Angwin, 2016; Eubanks, 2017; Noble, 2018; O’Neil, 2016). Both critiques draw in part from the fact that, especially in the case of non-interpretable systems, machine learning models can predict *that* something is the case, but do not offer any explanation of *why* it is the case, restricting the forms of knowledge, understanding, and accountability they afford. Some worry that today’s artificial intelligence doesn’t begin with human intelligence as a model, but that, in a sense, it may not *end* with human intelligence either.

There isn’t a straightforward narrative of artificial intelligence from the 1950s until today. One important arc, however, is that the human exemplar, once the guide and the motivation for artificial intelligence research in its many postwar forms, has largely been displaced from the field, and so too have certain perspectives on what we are meant to *know* and *do* with intelligent machines.

References

- Abbate, J. (2017). *Recoding Gender: Women’s Changing Participation in Computing*. Cambridge, MA: MIT Press.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *Pro Publica*.
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Brock, D. (2018). Learning from Artificial Intelligence’s Previous Awakenings: The History of Expert Systems. *AI Magazine* 39(3), 3-15.
- Collins, H. (1992). *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge, MA: MIT Press.
- Crowther-Heyck, H. (2008). Defining the Computer: Herbert Simon and the Bureaucratic Mind - Part 1. *IEEE Annals of the History of Computing* 30(2), 42-51.
- Daston, L. (1994). Enlightenment Calculations. *Critical Inquiry* 21(1), 182-202.
- Ensmenger, N. (2010). *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*. Cambridge, MA: MIT Press.
- Eubanks, V. (2017). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: Macmillan.

- Feigenbaum, E. (1977). *The Art of Artificial Intelligence: Themes and Case Studies of Knowledge Engineering*. Stanford Heuristics Programming Project Memo HPP-77-25.
- Floridi, L. (2016). *The 4th revolution: how the info sphere is reshaping human reality*. Oxford: Oxford University Press.
- Forsythe, D. (2002). *Studying those Who Study Us: An Anthropologist in the World of Artificial Intelligence*. Stanford, CA: Stanford University Press.
- Grier, D. (2007). *When Computers Were Human*. Princeton, NJ: Princeton University Press.
- Heyck, H. (2005). *Herbert Simon: The Bounds of Reason in Modern America*. Baltimore, MD: Johns Hopkins University Press.
- Hicks, Marie. (2017). *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. Cambridge, MA: MIT Press.
- Husbands, P., Holland, O., Wheeler, M., eds. (2008). *The Mechanical Mind in History*. Cambridge, MA: MIT Press.
- Jones, M. (2018). How We Became Instrumentalists (Again): Data Positivism since World War II. *Historical Studies of the Natural Sciences* 48(5), 673-684.
- Kline, R. (2011). Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence. *IEEE Annals of the History of Computing* 33(4), 5-16.
- Kline, R. (2015). *The Cybernetics Moment: Or Why We Call Our Age the Information Age*. Baltimore, MD: Johns Hopkins University Press.
- Light, J. 1999. When Computers were Women. *Technology and Culture* 40(3), 455-483.
- McCarthy, J., Minsky, M., Shannon, C. E., Rochester, N., & Dartmouth College. (1955). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence.
- McCarthy, J. (1997). AI as Sport. *Science* 276(5318), 1518-1519.
- McCorduck, P. *Machines Who Think*. Natick, MA: A K Peters.
- Newell, A., & Simon, H. (1972). *Human Problem Solving*. Oxford, England: Prentice-Hall.
- Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: New York University Press.

Norvig, P. (2011). On Chomsky and the Two Cultures of Statistical Learning.
<http://norvig.com/chomsky.html>.

O’Niel, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Broadway Books.

Schaffer, S. (1994). Babbage’s Intelligence: Calculating Engines and the Factory System. *Critical Inquiry* 21(1), 203-227.

Suchman, L. (2006). *Human-Machine Reconfigurations*. Cambridge, UK: University of Cambridge Press.

Tesauro, G. (1995). Temporal Difference Learning and TD-Gammon. *Communications of the ACM* 38(3), 58-68.

Valiant, L. (1984). A Theory of the Learnable. *Artificial Intelligence and Language Processing* 27(11): 1134-1142.

Wald, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons.

This article is © 2019 by Stephanie Dick. The article is licensed under a Creative Commons Attribution (CC BY 4.0) International license (<https://creativecommons.org/licenses/by/4.0/legalcode>), except where otherwise indicated with respect to particular material included in the article. The article should be attributed to the author identified above.