Manufacturing an Artificial Intelligence Revolution

Manufacturing an Artificial Intelligence Revolution

Yarden Katz[1,2]

[1]Dept. of Systems Biology, Harvard Medical School, Boston, MA, USA
[2]Berkman Klein Center for Internet & Society, Harvard University, Cambridge, MA, USA

November 27, 2017

Scientific controversy is not a matter of "naked power" versus the "naked truth." Scientific truth itself is clothed in technical capacity and institutional power.

—Jan Sapp, *Where The Truth Lies* (1990)

Thus is the problem of rich and poor to be solved. The laws of accumulation will be left free; the laws of distribution free. Individualism will continue, but the millionaire will be but a trustee for the poor; entrusted for a season with a great part of the increased wealth of the community, but administering it for the community far better than it could or would have done for itself.

—Andrew Carnegie, "Wealth" (1889)

Abstract

While the term "Artificial Intelligence" (AI) was coined in the 1950s, in recent years AI has become a focus of attention in mainstream media. Yet the forces behind AI's revival have been unclear. I argue here that the "AI" label has been rebranded to promote a contested vision of world governance through big data. Major tech companies have played a key role in the rebranding, partly by hiring academics that work on big data (which has been effectively relabeled "AI") and helping to create the sense that super-human AI is imminent. However, I argue that the latest AI systems are premised on an old behaviorist view of intelligence that's far from encompassing human thought. In practice, the confusion around AI's capacities serves as a pretext for imposing more metrics upon human endeavors and advancing traditional neoliberal policies. The revived AI, like its predecessors, seeks intelligence with a "view from nowhere" (disregarding race, gender and class)—which can also be used to mask institutional power in visions of AI-based governance. Ultimately, AI's rebranding showcases how corporate interests can rapidly reconfigure academic fields. It also brings to light how a nebulous technical term (AI) may be exploited for political gain.

1 The revolution is here

Artificial Intelligence (AI) is supposedly remaking the world. Although the term has been around since the 1950s, AI, we are told, is finally poised to reshape society, from the workplace to the court system—and if it isn't now, it will be soon. *Wired* magazine acknowledges that "Silicon Valley is not exactly averse to hyperbole," but assures us that "in the field of AI, the change is real" (Metz, 2016). It's been widely accepted that "AI" describes something coherent (and revolutionary) that we must contend with. But less attention has been paid to what the term "AI" means and the conditions that have made it prominent in recent years. Why is AI so fashionable now?

As the standard story goes, there was too much optimism about AI in the 1960s, which later caused the field to crash. A 1966 memo by MIT's Artificial Intelligence Group, for example, gave the impression that "significant parts of a [artificial] visual system," including the ability to recognize objects, could be developed over the course of a summer (Papert, 1966). By the late 1980s, these hopes for progress were shattered and the grand vision of AI—to understand human intelligence and instantiate it in machines—was set aside. That was the start of the so-called "AI winter," when researchers instead turned to making smarter software without the appeal to human intelligence. Things are different today, apparently because of groundbreaking advances in computing. AI is back and *The New York Times* runs lengthy articles on "The Great A.I. Awakening" (Lewis-Kraus, 2016).

This narrative is misleading because it doesn't capture the malleability of the AI label, or the fact that the agenda of AI researchers can change. The label "AI" has in fact recently undergone a rebranding. Corporations have helped manufacture an "AI revolution" in which AI stands for a confused mix of terms—such as "big data," "machine learning," or "deep learning"—whose common denominator is the use of expensive computing power to analyze massive centralized data. AI has essentially become a convenient redressing of a stale vision long promoted by Silicon Valley entrepreneurs. It's a vision in which truth emerges from big data, where more metrics always need to be imposed upon human endeavors, and where inexorable progress in technology can "solve" humanity's problems. Powerful companies have played a crucial role in the rebranding by hiring academics working on statistical analysis of big data (a term now interchangeable with AI), intervening more aggressively in academic research, and dominating mainstream discourse on AI.

By uncritically accepting that the "age of AI" is upon us, many discussions have basically adopted the framing set by these major corporations. The manufactured AI revolution has created the false impression that current systems have surpassed human abilities to the point where many areas of life, from scientific inquiry to the court system, might be best run by machines. However, these claims are predicated on a narrow and radically empiricist view of human intelligence. It's a view that lends itself to solving profitable data analysis tasks, but leaves no place for the politics of race, gender, or class. Meanwhile, the confusion over AI's capabilities serves to dilute critiques of institutional power. If AI runs society, then grievances with society's institutions can get reframed as questions of "algorithmic accountability." This move paves the way for AI experts and entrepreneurs to present themselves as the architects of society.

My aim here is to spell out some of the conditions that enabled the latest AI boom and the political projects embedded in it. I first trace the popularity of the term "AI" since the 1980s, focusing on U.S. media, and sketch some of the corporate maneuvers that helped rebrand AI as the new big data dream (Section 2). Using artificial vision as a case study, I then critically review some of the claims made about the capabilities of current AI systems and the ideology about intelligence that's packed into them (Sections 4-5). In Section 6, I argue that in practice, AI is being used as a vehicle for advancing familiar neoliberal political projects and economic policies. Finally, using algorithmic sentencing as an example, I argue in Section 7 that the confusion over AI can

be exploited to divert attention from structural forces of oppression and inequality, and to reduce these to the inadequate language of algorithmic accountability and "bias."

2 Rebranding "Artificial Intelligence"

When some American computer scientists began using the term "Artificial Intelligence" in the 1950s[1] (McCarthy et al., 2006), they were optimistic about instantiating intelligence in machines. This excitement was not limited to academia; a 1958 magazine article described an IBM computer as a "giant brain" that manages to "perform miracles that touch the lives of all of us," such as translating scientific publications into different languages or answering questions about "major historical events" (Strother, 1958).

Artificial Intelligence continued to generate excitement in the press through the 1980s. A *New York Times Magazine* piece in 1980 was sanguine about the efforts to create "machines that can think" (Stockton, 1980). The article discussed the prospect of developing machines that "understand" language, learn from experience and even feel emotion. Such intelligent machines may be able to scour databases to "amass knowledge about worldwide terrorist activities" and answer questions about "how terrorists in the Middle East, for example, are different from those in Italy or El Salvador." They could function as personal assistants with the equivalent of a college education. Some academics fueled these expectations: an M.I.T. professor was quoted in the *Times* saying that there are "excellent chances" of having artificial intelligence by the end of the 20th century (Shenker, 1977).

Of all this AI hype, the philosopher John Searle said:

> ...there's a lot of nonsense that comes out about AI, like the idea that computers are a deep threat to human beings and that computer achievement will destroy our sense of human dignity. That's *crap!* I have a pocket calculator that can beat any mathematician in the world, but that's no threat to anybody's dignity. (Rose, 1985)

Human dignity, and job loss to machines, has indeed been a constant theme in this coverage, along with speculation on whether machines can "think." But what's changed through the years is what the label "Artificial Intelligence" actually refers to. The term has always been somewhat nebulous, particularly when used in popular media. In terms of methodology, symbolic and logic-based approaches to AI were the focus in the 1970s, while in the 1980s, the connectionist program (which is centered around neural networks and uses statistical tools) received much of the media attention. It seems the term "AI" can be made to fit nearly any cutting-edge computation offered by computer scientists. AI's scope changes too, hovering between being "just a tool" for solving a particular and difficult task (e.g., helping scientists analyze data or playing a good game of backgammon) to a sci-fi creation that threatens human dignity. AI is bound to be slippery and contentious, partly because any claim about AI implies something about what counts as human intelligence (a question I'll return to in Section 4).

Despite the malleability of the term, it is only very recently that the AI label has become hugely popular. A CEO of a technology company, writing in 2014, described the rapid shift: "A few months ago, the phrase 'artificial intelligence' suddenly started being tossed around presentations, blogs, headlines, seminars—even a Facebook earnings meeting—as if it were the most benign concept in the world" (Silver, 2014). In the late 2000s,

[1] It's worth noting that there's been a lack of critical histories of AI. As the organizers of a panel at the 2017 Society for the History of Technology observed, many of the accounts the field's past have been a kind of "Whig history" written by practitioners.

the phrase "Artificial Intelligence" surged in academic journals and media discussions, and it increased most sharply in 2013 (Figure 1). AI started to peak after terms like "big data" (or "machine learning") had already become popular. In U.S. mainstream media, the phrase "Artificial Intelligence" started rising again only in the late 2000s (Figure 1, where the media attention on AI in the mid-1980s is also clear).[2]

In this latest surge of AI, the label is being used mostly synonymously with familiar catch phrases from the Silicon Valley orbit, such as "big data" or "machine learning." The relatively recent addition to the mix has been "deep learning," which refers to the training of multi-layered neural networks (from the connectionist tradition) on big data (LeCun et al., 2015). Academics who work in these areas of statistical data analysis started using the AI label more frequently and liberally. For instance, the computer science conference NIPS, which is traditionally focused on machine learning (as well as computational neuroscience), started publishing more papers that contain the term "Artificial Intelligence" (Figure 2). The sharpest increase in AI mentions started in the 2000s, and the fraction of conference papers that mention AI roughly doubled between 2005 and 2015. Prior to this peak, NIPS was already funded by corporations like Facebook, Twitter and Google, who naturally wanted to use the tools developed by this community of researchers. Although AI is a flexible term, many areas that traditionally fall under that academic discipline, such as knowledge representation, are not the focus of the NIPS community's flavor of AI (and aren't of great interest to its corporate sponsors). Essentially, the latest usage of AI became synonymous with the kind of statistical analysis of big data that Silicon Valley companies have been promoting for years.

As part of this rebranding campaign, major Silicon Valley companies started their own AI research laboratories. These companies hired academics, many from the NIPS community, to head their new AI labs (see Table 1) and started funding PhD fellowships in AI (Shead, 2017). In the media, this has been portrayed as a battle about dominance in a burgeoning field, with headlines such as "Apple Lags Behind Google and Facebook on AI" or "Facebook's Race to Dominate AI."

Table 1: Subset of researchers recruited from academia to head AI labs at major corporations and the number of co-authored publications these researchers had in the conference NIPS (1987-2016). Number of NIPS publications is used here as a crude indicator for having shared research interests with the NIPS community. Note that at the time of writing, the majority of individuals listed have retained an affiliation or part-time position with their academic institute.

Name	**Academic institute**	**Corporate position**	**No. NIPS publications**
Ryan Adams	Harvard University	Google/Twitter	18
Andrew Ng	Stanford University	Google	41
Sebastian Thrun	Stanford University	Google	21
Geoffrey Hinton	University of Toronto	Google	58
Craig Boutilier	University of Toronto	Google	5
Nando de Freitas	University of British Columbia, Oxford University	Google	15
Raquel Urtasun	University of Toronto	Uber	14
Zoubin Ghahramani	University of Cambridge	Uber	50
Ruslan Salakhutdinov	University of Toronto, Carnegie Mellon University	Apple	33
Yann LeCun	New York University	Facebook	17
Alex Smola	Carnegie Mellon University	Amazon	48

But why has the AI label become so attractive in recent years, enough to have companies fight over it? It's hard to find a clear answer, partly because AI is confounded, in both popular and academic writing, with so

[2]Here I take *The New York Times* and *The Wall Street Journal* as representatives of mainstream press and business press, respectively.

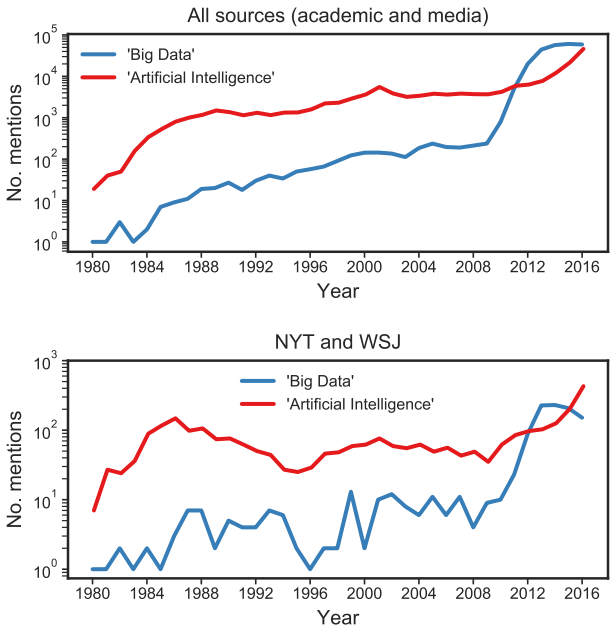

Figure 1: Top: Number of mentions of the terms "Big Data" or "Artificial Intelligence" in academic and media sources, 1980-2016. Bottom: Number of mentions in *The New York Times* and *The Wall Street Journal*, used as proxies for U.S. mainstream media and business media. Note logarithmic y-axis scale. Source: Factiva.

many other loosely defined terms. One explanation that is sometimes given is that new advances in machine learning, notably in deep learning, were what enabled companies like Google to build AI systems that surpassed human performance. According to *Wired* magazine, these breakthroughs were made possible by "post-paucity computing" (Weinberger, 2017), meaning that computing power isn't scarce anymore. Computing has hardly reached "post-paucity" (though it might feel that way for those at places like Google), but the claim is that having more computing power is what enabled the AI revolution.

The argument about increased computing power was actually used in an earlier attempt to promote the connectionist program during the 1980s. Seymour Papert described this in a reflection on the enthusiasm about connectionism (what he called "Snow White's awakening") and the stories that were used to explain how the field suddenly overcame the limitations of certain connectionist architectures (which he and Marvin Minsky had pointed out in their 1969 book *Perceptrons*):

> A purely technical account of Snow White's awakening goes something like this: In the olden days of Minsky and Papert, neural networking models were hopelessly limited by the puniness of the computers available at the time and by the lack of ideas about how to make any but the simplest networks learn. Now things have changed. Powerful, massively parallel computers can implement very large nets, and new charming algorithms can make them learn. No romantic Prince Charming is needed for the story. (Papert, 1988)

Papert argued that while increased computing power has clearly had some role to play in the excitement over connectionism in the 1980s, a "sociological explanation" was also needed.[3] His own explanation was that

[3] According to Marvin Minsky, some of the changes that took places in AI in the 1980s are also explained by corporate interventions: "This kind of progress of trying new experiments with computers kept happening in the 1960s and '70s and part of the '80s, but then things tightened up. The great laboratories somehow disappeared, economies became tighter, and companies had to make a profit—they couldn't start projects that would take 10 years to pay off."

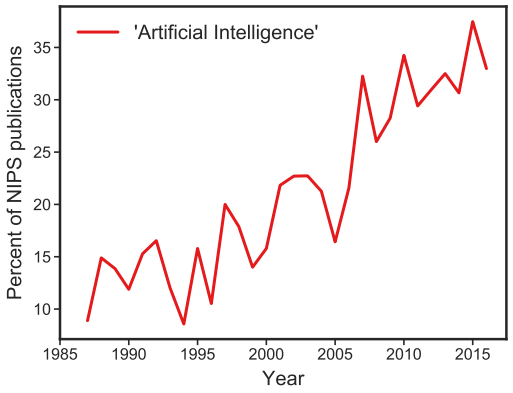

Figure 2: Percentage of papers in the NIPS conference (1987-2016) that mention the phrase "Artificial Intelligence" in the abstract or main text.

connectionism had a kind of "cultural resonance": those who are compelled by behaviorism—the view that all behavior can be shaped through reinforcement, and that behavior is what counts[4]—see the appeal of the connectionist approach. In both approaches, the focus is on an input-output relationship that's learned by tuning the model using error signals in data, without worrying about any internal states that govern that relationship. Papert recognized that behaviorism resonated with researchers, and the neural networks of the connectionist tradition, which were argued to resemble biological neurons, even gave the behaviorist framework a biological feel.

In the present rebranding of AI, connectionism and behaviorism have also been linked, and here too the phenomenon merits an explanation beyond increased computing power. The connectionist program has been revived in the form of deep learning, which has been the focus of attention. But another area that's being celebrated is reinforcement learning, a field that in many ways embodies a behaviorist approach, where computational agents learn to perform tasks using reinforcement signals from the environment. Although the most popular textbook on reinforcement learning was published in the late 1990s (Sutton and Barto, 1998), *MIT Technology Review* nonetheless included "reinforcement learning" in its 2014 list of "Breakthrough Technologies." Similarly, Google's DeepMind attributed the success of its Atari playing system to the combination of deep learning and reinforcement learning (Silver et al., 2016). The behaviorist core of this system was even made explicit by *Nature* magazine, which published DeepMind's paper. A major challenge for systems like the Atari playing AI, according to the magazine, is "avoiding 'superstitious behavior' in which statistical associations may be misinterpreted as causal" (Schölkopf, 2015). "Superstitious behavior" here is a reference to B. F. Skinner's famous 1948 paper in which he claimed that pigeons can be fooled into "superstitiously" believing that their behavior won them food (reinforcement), even when there's no causal link between the two[5] (Skinner, 1948). As a field, reinforcement learning offers multiple modeling approaches, including ones where agents form representations that a behaviorist such as Skinner might reject as "mentalistic." Nonetheless, the presentation of

[4]I won't get here into the several forms of behaviorism that have been defined over the years, because the behaviorist features I point to in the latest AI systems are so overt (and in some cases explicitly identified with B. F. Skinner's famous formulation). It's important to note that contrary to many accounts, behaviorism (in all of its forms) was not swept away from psychology, or even cognitive science, by the so-called "cognitive revolution," even if "behaviorism" became a kind of slur in some circles. The cultural resonance of behaviorism that Papert describes is, in that sense, not so deviant.

[5]Skinner's "superstition" interpretation has been disputed, including by psychologists who view themselves as "behaviorists." See (Staddon, 1992) for a critical review.

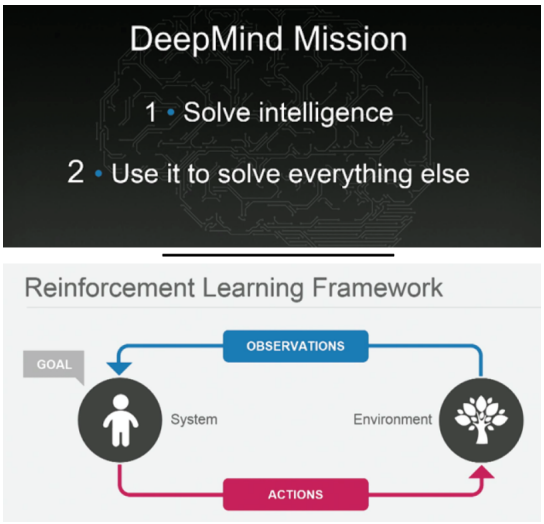

Figure 3: Google's DeepMind presentation slides. Mission statement (top) and framework for achieving it (bottom).

reinforcement learning as panacea signals the belief that all intelligent behavior can be learned by reinforcement. A behaviorist view of human intelligence is imprinted in these latest AI systems, but it has been marketed as a novel technology and clothed in computer science jargon.

3 Portrait of an AI universe

What does AI hold for the future, according to promulgators of the revolution? The promises about AI's effects are best described as "technophoric cyberdrool," to borrow Judith Squires's phrase. According to *Wired* magazine, AI may discover the Higgs Boson through its indecipherable "alien knowledge," but humans won't know it (Weinberger, 2017). Journalism will be reshaped too: the Nieman Foundation predicted that in the year 2017, "Robots will analyze complex editorial content of all lengths, and provide feedback to the humans sitting behind the keyboard." Andrew Ng, a prominent AI researcher, announced on Twitter in 2016 that "Pretty much anything that a normal person can do in less than one second, we can now automate with AI."[6] Less modestly, Google's company DeepMind declared that their mission is to first "solve" intelligence, which would then enable them to "solve everything." Their "solution" to intelligence is apparently the behaviorist framework of reinforcement learning (Figure 3), so when Google's Eric Schmidt says "we can use technology to build our dream society," it's more than reminiscent of Skinner's vision of sculpting a desirable society using a "technology of behavior" (Skinner, 1971).

A dream society can supposedly be made by growing the economy using AI. In his book *Life 3.0: Being Human in the Age of Artificial Intelligence*, physicist Max Tegmark explains that in ancient Athens, citizens "had lives of leisure where they could enjoy democracy" because "they had slaves to do much of the work"—and that "AI-powered robots" could provide the equivalent slave labor for all people today. Such an AI-driven economy would "not only eliminate stress and drudgery and produce an abundance of everything we want today, but it would also supply a bounty of wonderful new products and services that today's consumers haven't yet realized they want" (Tegmark, 2017).

[6]Ng later walked back his claim, writing in a piece for *Harvard Business Review* that "If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future."

AI's effects won't be limited to earth's economy, however, and imperial metaphors—even of the intergalatic variety—aren't hard to find in mainstream discussions of AI. Jürgen Schmidhuber, a noted AI researcher, claims that it's a matter of decades until "human-level" intelligence is implemented artificially. From there, it'll be only a small step to create a super-human AI that, in order to efficiently utilize the vast "resources" of space, will "spread out slowly through the milky way," and a couple of millions years later "establish a network of senders and receivers all over the galaxy."[7] By Tegmark's account, these super-civilizations spawned by AI may then encounter one another and apparently recapitulate colonialist dynamics. He writes that while "Europeans were able to conquer Africa and the Americas because they had superior technology," one super-intelligent civilization may not so easily "conquer" another. But since "assimilating your neighbors is a faster expansion strategy than settlement," one super-human civilization may persuade the other based on the "superiority" of its ideas, thereby "leaving the assimilated better off." In summary, AI can "make us the masters of our own destiny not only in our Solar System or the Milky Way Galaxy, but also in the cosmos" (Tegmark, 2017).

4 Seeing like a deep network

Figure 4: Figure reproduced with permission from (Lake et al., 2016). Captions generated by a deep network (Karpathy and Fei-Fei, 2015). Image credits: Gabriel Villena Fernández (left), TVBS Taiwan / Agence France-Presse (middle), AP Photo / Dave Martin (right).

In spite of these claims about human-level AI (and beyond), the systems that are now called "AI" aren't in danger of approaching human thought. These claims are based on a narrowly empiricist notion of intelligence that ignores, as the field of AI has traditionally, the historical context of human life.

It's helpful to examine some of the recurring success stories of the rejuvenated AI to see these limitations. One is DeepMind's system that learns to play Atari computer games and supposedly achieves "human-level" performance using deep learning and reinforcement learning. This system exemplifies the radically empiricist epistemology described earlier. It receives as input only images of the game and learns to play based purely on reinforcement signals (how many points it scored in the game). According to its developers, the system achieves "human-level" performance and even exceeds it in some cases. Similar claims have been made about the AlphaGo system, which has beat human champions in the game of Go.

[7] Schmidhuber's company, "Nnaiscent," is working on creating the human-level AI to set off this process.

Figure 5: Captions generated by Google's "Show and Tell" deep network. Image credits: Ammar Awad / Reuters (left), U.S. Department of Justice (middle), Reuters (right).

There are numerous problems with the claims about what these systems have supposedly learned from data, as well as with their comparison to humans. These claims have been criticized by cognitive scientists and AI researchers, but in mainstream discussions, the critical evaluation has been largely eclipsed by hype. Recently, (Lake et al., 2016) critically evaluated deep learning-based systems in particular and it's worth dwelling on some of their objections. First, the Atari-playing system received the equivalent of roughly 38 days' worth of play time (Lake et al., 2016). This extensive training allows the deep learning network to perform well, particularly when the games are amenable to strategies that don't require long-term planning. However, a human player who receives only two hours of training can beat the deep learning system in games that require longer-term goals to be completed (Lake et al., 2016). Second, and more importantly, these systems do not acquire the same knowledge about the games that people do. These AI systems are imprinted with particulars of the training data that prevent the kind of generalization to other contexts that humans do effortlessly. As (Lake et al., 2016) point out, the trained deep networks are "rather inflexible to changes in its inputs and goals: changing the color or appearance of objects or changing the goals of the network would have devastating consequences on performance if the network is not retrained." For instance, the game playing system was trained with the goal of maximizing its score—an objective to which it is locked. People, on the other hand, can flexibly adopt different goals. If asked to play with a different objective, such as losing as quickly as possible (the opposite of maximization), or getting to the next level but just barely, people have no problem using what they've learned to do just that (Lake et al., 2016).

The AlphaGo system suffers from the same limitations. It's highly tuned to the configuration of the Go game on which it was trained. If you change the size of the game's board, for instance, there's no reason to expect the trained AlphaGo model to do well. AlphaGo also reveals that these deep learning systems are not as radically empiricist as advertised. The rules of Go are built into AlphaGo, a fact that's typically mentioned in passing. This is hard-coded symbolic knowledge, not the "tabula rasa" that DeepMind declares (Silver et al., 2017).

Vision has been another domain where deep learning is celebrated. Here, the gap between what's marketed and what's possible is perhaps starkest. It's been claimed that deep learning systems developed by Microsoft and other companies outperform people in classifying images or recognizing objects (He et al., 2015; Wu et al., 2015). But what does the world look like from the perspective of a deep network? How does a deep network "see"?

To a deep network, an image of people escaping a flood may look like "people on a beach," while an airplane

crashing down looks like "an airplane on a tarmac" (Figure 4). The knowledge of objects and their relations, or of human emotions and intentions (easily observed in the faces of those photographed) is missing. But there is a more fundamental gap between human thought and these AI systems, which has to do with the historical conditions of human life. The social and political context of the present affects how people perceive their world. Yet aspects of gender, racism, class (or having a body) are systematically ignored or dismissed by AI researchers. In presuming a universal intelligence, AI researchers have often ignored the fact that thought, action and perception are embodied—both literally (in a body) and in a historical context. The power dynamics among people can be read in body language, yet these systems are blind to it.

To illustrate this, I obtained Google's deep learning-based image captioning system (called "Show and Tell" (Vinyals et al., 2015)), trained on hundreds of thousands of images, and used it to analyze a series of photographs. These photographs were chosen to show how historical context shapes the interpretation of scenes. The images I chose are rather different from the images that Google showcases when presenting this system, which typically lack obvious historical significance, and are selected to be instances where the deep network produces impressive captions.[8]

Consider a photograph of Palestinians arriving at a checkpoint controlled by Israeli soldiers (Figure 5, left). A Palestinian lifts his shirt to show the soldier, who is motioning to him from the top of a small hill, that he is unarmed. Google's deep network gave the image the caption "a group of people standing on top of a snow covered slope." For a statistical pattern recognizer, the outline of the hill and the light dirt might look like a snow covered slope—but the sun, the clothing, the relationship among those photographed make that an absurd description. Similarly, a 1960 photograph of Ruby Bridges, a six-year old African American girl being accompanied to a desegregated school by U.S. marshalls, is registered as "a group of men standing next to each other" (Figure 5, middle).

Figure 6: Captions generated by Google's "Show and Tell" deep network. Image credits: Ahmad Gharabli / AFP (left), unknown (right).

There are many more complex relations among the photographed that are missed. Consider the scene of an Israeli soldier, holding down a young Palestinian boy, while the boy's family try to remove the soldier[9] (Figure 5, right). Google's deep network produces the caption "people sitting on top of a bench together" (the "bench" perhaps being the boy). The motives, goals, and intentions of the actors are entirely lost.

[8] See for instance Google's blog post on "Show and Tell."

[9] The incident occurred during a protest in the village of Nabi Saleh in the West Bank in the summer of 2015. See "A Perfect Picture of the Occupation" in *Haaretz* newspaper, Aug. 31, 2015.

Figure 7: Images involving gender politics. Captions generated by Google's deep network ("Show and Tell"). A sexist ad for cigars from the 1960s (left) and an image used to depict unequal division of household work between men and women (right). Image credits: Cigar Institute of America, Inc. (left), CORBIS (right).

Apart from interpreting body gestures or violence, there are interactions among groups that people can pick out instantly, but that visual pattern recognition systems would miss. An image of Palestinian Muslims praying outside the mosque in protest, with the Dome of the Rock in the background, is registered by Google's deep learning system as "a crowd of people standing around a parking lot filled with kites," probably because of the colored shirts of the men in prostration (Figure 6, left). Similarly, a 1960 photograph from South African apartheid, in which black men line up to receive passbooks from a panel of all-white officials, is captioned as "a black and white photograph of a group of people" (Figure 6, right). A sexist ad from the 1960s where a woman is used as ashtray support for a man smoking a cigar is captioned as "a black and white photo of a woman wearing a tie" (Figure 7, left). Another scene that was staged to illustrate unequal division of household work (where the woman is a fleeting figure in the background), and that can be grasped instantly by people, is registered by the deep network as "a woman sitting on the couch with a laptop" (Figure 7, right).

One response to these examples might be that with more training data, these systems would be able to "understand" even these complex images. The presumption is that mappings from images to labels are sufficient; that the "information" is there, and it's only a matter of finding the right model to decode it. But more training on labeled images won't prepare the system for something like Ruby Bridges's photograph. This image isn't an instance of a "type" of visual configuration that can be extracted from an arsenal of captioned images (the world of photographs simply can't be meaningfully parceled into ever finer categories such as "six-year old African American girls being protected from violence on first day of school in the United States of the 1960s").

Putting aside whether this specific behaviorist conception of scene understanding is tenable, claims about AI's limits aren't likely to be persuasive, particularly in the current climate. As the philosopher Hubert Dreyfus observed, "Artificial Intelligence workers who feel that some concrete results are better than none, and that we should not abandon work on artificial intelligence until the day we are in a position to construct such artificial men, cannot be refuted" (Dreyfus, 1992). Dreyfus's observation is all the more relevant when there's great inducement and funds for building AI-based futures.

5 A view from nowhere

I don't bring up these absurd outputs from AI systems in order to provoke the long-standing philosophical debate on whether human thought can be reduced to a computational process, or the extent to which such processes can be recreated in a computer (without, say, having a body, as discussed in (Dreyfus, 1992; Clark, 1998)). Nor is the intention here to rehash the flaws of a connectionist research program of one form or another, which were intensely debated in the late 1980s (see Fodor and Pylyshyn (1988); Smolensky (1988); Fodor and McLaughlin (1990)). It is also not to deny that people are capable of performing sophisticated computational feats, or that some aspects of thought can be usefully modeled by formal computational frameworks.

The point, rather, is that mainstream AI researchers have been mostly disinterested in the social and political conditions of human life. Examples from state of the art systems merely help to unpack the narrow view of "intelligence" that AI researchers have focused on. In her book *Artificial Knowing: Gender and the thinking machine*, Alison Adam showed how that narrow notion of intelligence pervades AI's research traditions (Adam, 2006).[10] AI researchers generally seek "a view from nowhere," as Adam put it. The identity of the thinking subject is rarely mentioned, although subjects are nearly universally assumed to be motivated by the rational pursuit of goals (and analytic problem solving is taken to be an adequate exemplar of intelligence) (Adam, 2006).

Yet in spite of efforts to erase the subject, AI research definitely constitutes "a view from somewhere." As one of several examples, Adam analyzes the "somewhere" implicit in Herbert Simon and Allen Newell's work on a "unified" theory of cognition (which is decidedly in the symbolic tradition of AI, and has little in common, mathematically, with connectionist models such as deep networks). While Simon and Newell generally neglected to mention aspects like gender when discussing their pool of subjects for psychological experiments, Adam was able to piece together from the records that the subjects used for this "general" theory of cognition were all male undergraduates at the prestigious university where Newell and Simon worked. In the AI tradition, the benchmark for intelligence is typically "based on the behavior of a few, technically educated, young, male and probably middle-class, probably white, college students working on a set of rather unnatural tasks in a US university" (Adam, 2006).

The aspiration to a "view from nowhere" that masks the identity of human subjects also manifests in the training and evaluation of the systems we have discussed. To take one example, Google's image captioning system, "Show and Tell," was trained on hundreds of thousands of captioned images—but who provided the captions? Some image datasets that were used for training the system, such as Microsoft's COCO (Lin et al., 2014), were captioned by workers on Amazon's Mechanical Turk platform (AMT). Other image captions were scraped from individual accounts on the image sharing platform Flickr (Ordonez et al., 2011). To evaluate their model, Google researchers also used AMT workers to score model-generated captions. From the universal intelligence perspective, what matters is that captions were produced or validated by *some* human. The identity of the viewer isn't considered relevant, even though it clearly matters. A photograph of a checkpoint in occupied Palestinian territories may well be perceived differently by a viewer in Ramallah compared with a viewer in London. AMT does in fact allow the employer to select workers based on country, but apparently Amazon has recently limited the worker pool to the U.S. (Irani, 2015). As Lilly Irani argued, the predominantly U.S.-based employers prefer U.S. workers because, among other things, "they are likelier to be culturally fluent in the kinds of linguistic and categorization tasks" that are delegated to AMT (Irani, 2015). In spite of the likely restriction to

[10] As Adam shows, AI researchers largely embraced the dominant epistemology of the analytic Anglo-American philosophical tradition.

the U.S., the authors of Google's "Show and Tell" system reported that there was only 65% agreement among AMT workers regarding the validity of model-generated captions. In cases of disagreement, the developers averaged the scores (Vinyals et al., 2015). This would blur out contextual differences, which makes sense if one is seeking the "view from nowhere" that Adam described.

Within the AI tradition, there have been some notable rejections of the view from nowhere, such as Terry Winograd and Fernando Flores's 1986 book *Understanding Computers and Cognition*. While Winograd and Flores do not address the politics of race, gender or class, they embrace the idea that "every representation is an interpretation" and that individuals do not act alone but in reference to a group and a tradition (whose assumptions they cannot make explicit, let alone reason about in a detached manner when acting in the world) (Winograd and Flores, 1986).[11] But as Winograd and Flores themselves acknowledge, their approach is a radical deviation from most work in AI.

Even if it doesn't capture the richness of human thought and action, building systems that are premised on the universalist perspective that Winograd and Flores, or Adam, reject is undeniably profitable. Statistical pattern recognizers are suited for average-case performance, where small increases can be lucrative. As Microsoft researchers pointed out, in a click prediction task, for example, "even 0.1% of accuracy improvement would yield greater earnings in the hundreds of millions of dollars" (Ling et al., 2017). Claims of super-human AI can be a pretext for applying such pattern recognizers to new areas.

6 Governance by the numbers

The rebranding of "AI" comes at a time when there's increasing concern about the influence of major tech corporations and the data they collect. The NSA files leaked by Edward Snowden in 2013 helped make the surveillance dimension of big data tangible. The affair also drew attention to collusions between governments and tech companies. The wave of AI hype dilutes these critical looks at big data. By injecting confusion about wild futures, the manufactured AI revolution diverts attention from surveillance and manipulation (what Shoshana Zuboff identified as "surveillance capitalism" (Zuboff, 2015)) that the big data vision has come to be associated with.

At the core of the big data vision is what legal scholar Alain Supiot called "governance by the numbers" (Supiot, 2012). Governance by the numbers works by first defining quantitative metrics and imposing them upon the world, and then using these metrics to "program" behavior through rankings and benchmarks. The aspiration here is to reduce the "diversity of beings and things" so as to create "a total market, which seeks to encompass all of human kind and all the products of the planet" (Supiot, 2012).

The talk of super-human AI is a boon to governance by the numbers. The reason is simple: the AI systems that are being promoted can't work without metrics. It's easy to see this in the domain of science, where there has been a long-running effort to commodify research and evaluate it in economic terms (Mirowski, 2011). If supposedly super-human AI systems can guide or replace scientists, then metrics for what counts as "good" science will be needed (otherwise what would AI systems "maximize"?). The business of companies like Google is to collect these metrics, such as citation counts (e.g., through Google Scholar). This requires access to vast and fragmented online resources along with immense computing power that companies like Google can

[11] Winograd and Flores criticize what they call the "rationalistic" approach that dominates most AI research (Winograd and Flores, 1986), and like Adam, they trace it to the analytic philosophy tradition.

afford.[12] Google is not alone in this metrics for science game. The Chan-Zuckerberg Initiative (CZI), co-founded by Facebook's Mark Zuckerberg, recently acquired a startup company whose AI software supposedly "helps scientists read, understand and prioritize millions of scientific papers" by analyzing such metrics.

As companies like Google try to gain a foothold in science, the metrics that they collect are being held up as rankings of scientific worth. *MIT Technology Review*, for instance, wrote that Google was having an "annus mirabilis" and that its "surging investment in machine learning" launched the company "into the scientific stratosphere" (Regalado, 2017). The evidence for this was that Google published over two hundred papers in 2016 alone, including some in prestigious venues such as *Nature*, as well as in machine learning conferences. Science, machine learning, and AI all get muddled up, but what comes through clearly is the notion that scientific insight can be ranked and optimized.

If scientific inquiry is to be ruled by metrics, the logical next step is to reframe science's problems around Silicon Valley's tools. CZI's foray into biology does exactly that. CZI announced its plan to "end all disease" with a three billion dollar investment in research—roughly one-tenth of the annual budget of the National Institutes of Health (the major public sponsor of biomedical research in the U.S.)—which raised some questions about how CZI will achieve this feat. Cori Bargmann, a biologist from Rockefeller University who heads the initiative, explained that Silicon Valley could advance biology, since in her lab, "everyone now writes code; thats a bit like everyone making their own soap"—and that instead, "we should be finding ways of doing this that are general and powerful, that allow us to interact and share our knowledge" (Hayden, 2016). The problems facing biologists are thus reframed around what companies like Facebook have to offer: programming pipelines, computing power and big data analysis. AI is a pretext for promoting these tools.

The rebranded AI label also provides an opportunity to claim areas of knowledge through patenting, for those who can afford to do so. Google has already filed patents on common algorithmic techniques related to training neural networks, as well as combining neural networks with reinforcement learning, while Microsoft has filed a broader patent in the area of "active" machine learning (Gillula and Nazer, 2017). AI is nebulous enough a term to allow for even broader patents. A MasterCard-owned company filed for a patent on a "Method for Providing Data Science, Artificial Intelligence and Machine Learning As-A-Service." AI made for ill-defined patents in the past, too: in 1985, a patent simply titled "Artificial intelligence system" claimed "an artificial intelligence system for accepting a statement, understanding the statement and making a response to the statement based upon at least a partial understanding of the statement" (U.S. Patent No. 4,670,848). It's not far off from a 2017 patent for an "AI learning method and system" that "may transmit a question to users through a messaging service and may acquire learning data for the AI through reactions of the user to the transmitted question."

The vagueness of AI helps refuel existing patent arms races, and some investment sites have been keeping score. Microsoft is leading the current AI patent spree, according to one investment site, having filed more than 200 "AI-related" patents since 2009, while companies such as Apple are "widely criticized for being slow" in filing (CB Insights, 2017). So success at the patent office is reinstated as a metric of innovation.

But beyond promoting such metrics, AI has also been a pretext for peddling more conventional policies. It's perhaps most transparent in discussions of AI and the economy. In the book *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence*, Jerry Kaplan discusses strategies for dealing with the job loss that AI is assumed to bring about. As in many discussions in this literature, job loss is framed here as

[12]Not surprisingly, Google has been extremely protective over its Google Scholar data, which cannot be systematically downloaded or queried. Some researchers have in the past recreated citation data for specific fields (e.g., publications in various fields of physics), but this requires downloading a massive number of publications and computing their citation relationships.

an unpleasant but inevitable byproduct of technological progress. Like Max Tegmark, Kaplan believes that AI will enable the economy to expand and that this will cure social maladies. He assures us that "we don't need to take from the wealthy and give to the less fortunate," since "our economy is not standing still; its continually expanding, and this growth is likely to quicken. So all we need to do is distribute the benefits of future growth more widely, and the problem will slowly melt away" (Kaplan, 2015). What would the new distribution of benefits look like? Kaplan draws on the thinking of Milton Friedman to offer "free-market solutions" to this question. While Kaplan grants that economic inequality is an issue, his proposed solution is to cut taxes for corporations using a new "objective" metric and to restructure Social Security (which he refers to as a "monolithic and opaque centralized system of investment" that limits individual choice) (Kaplan, 2015).

To remain competitive in the age of AI, restructuring the economy won't do, though; education reform is also needed. Christof Koch, a neuroscientist at the Allen Brain Institute and major proponent of AI, wrote that looming super-intelligent AI will require people to augment their brains to stay relevant (Koch, 2017). While some may opt for "education" to keep afloat, Koch argues that "training (and retraining) people takes time" and that not everybody may be able to switch from driving trucks (a vanishing profession) to one of the few jobs that AI experts predict to be secure (but not for too long). Likewise, in his widely discussed book *Superintelligence: Paths, Dangers, Strategies*, the philosopher Nick Bostrom discounts education as "probably subject to diminishing returns," and considers it an ineffective means of acquiring the "superintelligence" that the AI age demands (Bostrom, 2016).

When the AI shell is stripped away, these reflections on the economy and education quickly reduce to old tropes of individuals needing to compete in ever-changing markets. AI is a vehicle for promoting this thinking and imposing governance by the numbers.

7 Epistemology in the service of power

Governance by the numbers, enacted through computation on big data, offers many opportunities for manipulation and control. Even before the recent explosion of AI, there were calls for "algorithmic accountability" as a way of coping with these issues. At the very least, the push for algorithmic accountability is a recognition that the technical features of computational data-driven systems—such as the selection of data and the particulars of the algorithms used to process it—encode politics. Algorithmic accountability can be a sleight of hand, though. The notion that "algorithms exercise their power over us" is misleading (Diakopoulos, 2014). It can obscure the fact that algorithms don't do things in the world, people do. Algorithmic accountability may mask the power structures that enact violent decisions and create the conditions for the decisions to be made at all (whether prescribed by human or computer). It can also give the illusion that a technological solution is within reach; that one can "de-bias" the black box. The latest wave of AI hype exacerbates these problems with algorithmic accountability. If there's super-human AI out there, after all, then why not use it over the artisanal judgments of people?

The prospect of AI-based "algorithmic sentencing" in the U.S. highlights, for one, how algorithmic accountability can render institutional forces invisible. Some believe that when it comes to sentencing, "algorithms are also seen as a way to dispense justice in a more efficient way that relies more on numerical evidence than personal judgments" (Smith, 2016). Similarly, others yearn for "robojudges" who, using AI, will be unbiased and replace or aid human judges (Tegmark, 2017). From the perspective of algorithmic accountability, this raised the worry that the sentencing software made by private companies will be biased against particular populations.

It has in fact been shown that software used in several states systematically assigns higher "risk assessment" scores for African Americans (these scores then factor into judges' decisions). But it's here that the violent institutions that carry out the sentencing—and that enable a private company to even enter into a critical role in the process—recede into the background.[13]

As Kimberlé Crenshaw argued, when discussions of the prison system are framed around "at risk" populations, it leads to "subtle erasure of the structural and institutional dimensions of social justice politics" (Crenshaw, 2011). Looking for "at risk" populations inside a computational system can have the same effect. An analysis of bias at the sentencing level might hide, for instance, the growth incentives for the mass incarceration system, or the structural forces that work to incarcerate and criminalize specific populations (Crenshaw, 2011). Crenshaw surveys examples of these structural forces, including how policies related to immigration, housing and child welfare can together result in certain groups being "overpoliced and underprotected" based on race, class, gender and legal status (Crenshaw, 2011). The structural phenomena Crenshaw describes are not adequately captured in terms of "bias" at the moment of sentencing. To speak of an "unbiased" AI judge is to presume that the harm is localized to one decision point (of an otherwise fine system) that can be corrected by technology.

There's also a technical aspect to the latest AI systems that helps mask the structural forces that Crenshaw discusses. Deep networks, which have been most popular in recent years, are thought to be opaque to humans (our "puny" human brains cannot comprehend their "alien knowledge," as one writer put it (Weinberger, 2017)). There's a small element of truth to this: when a neural network is trained on a data set, there won't necessarily be an easily interpretable rule that comes out. But this blanket acceptance of indecipherability is a gift to systems of power. If AI systems outperform us (and hence must be used), yet are indecipherable, then who can be held accountable?

In dealing with the indecipherability of computational systems, (Ananny and Crawford, 2016) are right to say that simply calling for transparency can "come at the cost of a deeper engagement with the material and ideological realities of contemporary computation." Deeper engagement with the computation, as they argue, would mean analyzing the relations between the computational system and the individuals that operate it or are affected by it. The computation of the sort used for prison sentencing, however, could not exist without institutional violence that builds, enacts, and profits from it. There are limits to what can be understood by analyzing the interrelations of individuals with a technical system, without also considering the institutional forces that enable such systems to exist in the first place.

When our gaze turns to the technical intricacies of these computational systems, it sets the stage for technical experts to become the architects of society. Indeed, this has been a theme of the discussions on the effects of AI on society. It seems that the researchers that are deeply embedded in the corporate world are expected to both anticipate and correct the undesirable effects of the revolution that they're invested in. Conferences such as NIPS, for example, host symposia on the AI and society with panels made up entirely of prominent researchers who lead AI research groups at places like Facebook and Google.[14] Other high-profile attempts to plan the AI-driven society of the future have had a similar composition.[15] The mainstream press has largely played a cheerleading role in these AI debates, as is often the case in coverage of science and technology (Katz, 2016). AI's academic

[13]This might explain the ease with which public conversations are had about "algorithmic accountability," rather than the overlapping racist and disciminatory institutions that result in mass incarcerations of certain groups in the U.S., for example.

[14]See the "Algorithms Among Us" symposium at the NIPS conference.

[15]See the "Beneficial AI 2017" conference organized by the Future of Life Institute, which is largely sponsored by Elon Musk. Many of the participants were AI entrepreneurs. According to the conference co-organizer, journalists were "banned" from attending the meeting because of the way the press has covered Elon Musk's remarks on AI at a talk he gave at MIT (Tegmark, 2017).

entrepreneurs have been embraced by the media as the protagonists of a revolution (without questioning its terms or timing).

The rejuvenated AI field is therefore in lock step with corporate interests. It's perhaps not surprising that a field that seeks universal "intelligence" with no regard to social context (and bent on reproducing it in machines) would be amenable to occasional rebranding campaigns driven by industry. The pragmatic interest on the part of industry is natural, since the behaviorist approach that has appealed to many AI researchers aligns with the profit motives of surveillance capitalism. But apart from being useful, the behaviorist design of the latest AI systems also says something about how human nature is viewed. The vision elaborated by companies like Google, as Shoshana Zuboff observed, shares with behaviorist psychology the idea that "human autonomy is irrelevant and the lived experience of psychological self-determination is a cruel illusion" (Zuboff, 2015). The antidote to this cruel illusion, at least for the masses, is governance by the numbers.

Governance by the numbers is aided by the confusion over what AI is and whether its inner logic can be deciphered. AI might be the perfectly nebulous term to use if the task is to convey a sense of technological disruption that licenses sweeping political change (especially when many are already willing to believe in the unbounded power of big data). This is epistemology in the service of power. It's obviously not the first time that a techno-scientific field's promise to bring about utopia (or dystopia) has been exploited. Given the behaviorist core of today's celebrated AI systems, it's worth revisiting the 20th century debates on behaviorism-based visions of a future society. In a critique of B. F. Skinner's promises that human behavior can be reshaped to produce a desirable society using the scientific methods of reinforcement, Noam Chomsky wrote: "One waits in vain for psychologists to make clear to the general public the actual limits of what is known. Given the prestige of science and technology, this is a most unfortunate situation" (Chomsky, 1972). The assessment is remarkably accurate for AI today, particularly when AI researchers are being bought to sell Silicon Valley dreams.

8 Thoughtlessness

Like behaviorist visions, the AI-based vision of future society is totalizing. AI would need to enter all of society, much like the principle of reinforcement would need to transform every social institution.[16] But the project of AI goes even further than reshaping human society on earth. It isn't only about finding a view from nowhere when it comes to intelligence or cognitive ability, but about "escaping the human condition" altogether, as Hannah Arendt put it (Arendt, 2013). Though Arendt didn't use the term AI, the starting point of *The Human Condition*, which opens with a discussion of how science and technology may radically change human life, and the possibility of using machine automation to do away with labor, couldn't be more relevant to AI in the present.

Arendt might as well have been talking about AI when writing that the creation of "artificial life," enabled by the emerging space technology, seeks to sever "man's connection to the children of nature." For Arendt, the highly abstract character of scientific theories had already distanced scientists, who aimed to view the earth as if from an Archimedean point set in space, from everyday earthly experience. The conquest of space, through opaque techno-scientific means, may be the final act of distancing.

AI presents an even purer form of distancing: a project with a view from nowhere that promises to launch artificial civilizations into space that were never bound to earth. One can hear a celebration of this severing of ties with the human condition in Schmidhuber's forecast of AI:

[16] E.g., as described by Skinner in (Skinner, 1953).

> We are currently witnessing the beginning of something that is huge. This is not just another industrial revolution. This is more than all of civilization. This is a step, a new step, on the path of the universe towards higher and higher complexity...This goes beyond human kind, this transcends human kind, and it's a privilege to be part of that and witness the beginning of that...We shouldn't think of us versus them...But view all of us including human civilization and these future beings as part of one grand scheme that allows the universe to go from simple states towards more complex states and it's great to be a part of that.

Arendt diagnosed that this type of pursuit would make society "thoughtless." Our techno-science would be so opaque that it wouldn't be expressible in ordinary speech, so we'd "need artificial machines to do our thinking." In trying to escape the human condition, we would be "helpless slaves" not of the machines we make, but of our ability to produce technology. Humans would become "thoughtless creatures at the mercy of every gadget which is technically possible, no matter how murderous it is" (Arendt, 2013).

Thoughtlessness may be the defining feature of the so-called age of AI. What is governance by the numbers, or the reduction of structural injustice to the language of bias and algorithmic accountability, if not ways to limit thought by piling metrics upon metrics?

The AI revolution shows that powerful techno-scientific projects don't need to produce technologies that live up to their promises in order to induce thoughtlessness. Arendt thought that space technology would eventually permit human life away from earth, but it's not clear that the current vision of AI could ever be realized.[17] It's the pursuit alone that can be destructive. And if opacity, or distance from common human experience, is a driver of thoughtlessness, then AI shows that opacity doesn't necessarily have to originate in the content of scientific theories. AI is a propagandistic term that sends the media (itself an opaque computational system of sorts) to a frenzy. Even if the science of AI wasn't opaque to begin with, it may become so in public discourse once churned through the media.

Arendt wrote that when scientists live in opaque worlds, their political judgment (as scientists) can't be trusted. Yet the pressing question, whether science should be used to escape the human condition, is a "political one of the first order." It has no technocratic answers. As with politics, it's up to the "agreement of the many."

Acknowledgments

Thanks to Ariella Azoulay, Tristan Foster, Ulrich Matter, Grif Peterson and Lauren Surface for helpful discussions and comments on early drafts of this work.

[17] Arendt didn't seem to think that something like AI, as described nowadays, could be realized either: "It is highly unlikely that we, who can know, determine, and define the natural essences of all things surrounding us, which we are not, should ever be able to do the same for ourselves—this would be like jumping over our own shadows" (Arendt, 2013).

References

Adam, A. (2006). *Artificial knowing: Gender and the thinking machine*. Routledge.

Ananny, M. and Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society*.

Arendt, H. (2013). *The Human Condition*. University of Chicago Press.

Bostrom, N. (2016). *Superintelligence: Paths, dangers, strategies*. OUP Oxford.

CB Insights (2017). Microsoft, Google Lead In AI Patent Activity, While Apple Lags Behind.

Chomsky, N. (1972). Psychology and ideology. *Cognition*, 1(1):11–46.

Clark, A. (1998). *Being there: Putting brain, body, and world together again*. MIT press.

Crenshaw, K. W. (2011). From private violence to mass incarceration: Thinking intersectionally about women, race, and social control. *UCLA L. Rev.*, 59:1418.

Diakopoulos, N. (2014). Algorithmic accountability reporting: On the investigation of black boxes. *Tow Center for Digital Journalism, Columbia University*.

Dreyfus, H. L. (1992). *What computers still can't do: a critique of artificial reason*. MIT press.

Fodor, J. and McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, 35(2):183–204.

Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1):3–71.

Gillula, J. and Nazer, D. (2017). Stupid Patent of the Month: Will Patents Slow Artificial Intelligence? *Electronic Frontier Foundation*.

Hayden, E. C. (2016). Facebook couple commits $3 billion to cure disease. *Nature*, 537(7622):595–595.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034.

Irani, L. (2015). Difference and dependence among digital workers: The case of Amazon Mechanical Turk. *South Atlantic Quarterly*, 114(1):225–234.

Kaplan, J. (2015). *Humans need not apply: A guide to wealth and work in the age of artificial intelligence*. Yale University Press.

Karpathy, A. and Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3128–3137.

Katz, Y. (2016). Cheerleading with an agenda: how the press covers science. *3:AM Magazine*.

Koch, C. (2017). We'll need bigger brains. *The Wall Street Journal*, page C1.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2016). Building machines that learn and think like people. *Behavioral and Brain Sciences*, pages 1–101.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lewis-Kraus, G. (2016). The Great A.I. Awakening. *The New York Times Magazine*.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer.

Ling, X., Deng, W., Gu, C., Zhou, H., Li, C., and Sun, F. (2017). Model ensemble for click prediction in bing search ads. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 689–698. International World Wide Web Conferences Steering Committee.

McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on Artificial Intelligence, August 31, 1955. *AI magazine*, 27(4):12.

Metz, C. (2016). Inside OpenAI, Elon Musk's Wild Plan To Set Artificial Intelligence Free. *Wired*.

Mirowski, P. (2011). *Science-Mart: Privatizing American science*. Harvard University Press.

Ordonez, V., Kulkarni, G., and Berg, T. L. (2011). Im2text: Describing images using 1 million captioned photographs. In *Advances in Neural Information Processing Systems*, pages 1143–1151.

Papert, S. (1966). The Summer Vision Project. *MIT Artificial Intelligence Group*.

Papert, S. (1988). One AI or Many? In Graubard, S., editor, *The Artificial Intelligence Debate: False Starts, Real Foundations*, pages 241–267. MIT Press, Cambridge.

Regalado, A. (2017). Google's AI Explosion in One Chart. *MIT Technology Review*.

Rose, F. (1985). *Into the Heart of the Mind: An American Quest for Artificial Intelligence*. Vintage Books.

Schölkopf, B. (2015). Artificial intelligence: Learning to see and act. *Nature*, 518(7540):486–487.

Shead, S. (2017). DeepMind has started paying to put PhD students through Oxford. *Business Insider*.

Shenker, I. (1977). Thinking machines are getting smarter. *The New York Times*, (10):8.

Silver, C. (2014). Artificial Intelligence Is No Longer a Four-Letter Word - And Could Even Win an Oscar. *Wired*.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. (2017). Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359.

Skinner, B. (1971). *Beyond Freedom and Dignity*. A Bantam/Vintage book. Hackett Publishing.

Skinner, B. F. (1948). "Superstition" in the pigeon. *J Exp Psychol*, 38(2):168–172.

Skinner, B. F. (1953). *Science and Human Behavior*. Simon and Schuster.

Smith, M. (2016). In Wisconsin, a Backlash Against Using Data to Foretell Defendants' Futures. *The New York Times*.

Smolensky, P. (1988). The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn. *The Southern Journal of Philosophy*, 26(S1):137–161.

Staddon, J. E. (1992). The 'superstition' experiment: a reversible figure. *J Exp Psychol Gen*, 121(3):270–272.

Stockton, W. (1980). Creating computers that think. *The New York Times Magazine*, pages 48–54.

Strother, R. (1958). Thinking machines are getting smarter. *Mechanix Illustrated*, (10):159–167.

Supiot, A. (2012). *The spirit of Philadelphia: Social justice vs. the total market*. Verso.

Sutton, R. and Barto, A. (1998). *Reinforcement Learning: An Introduction*. Bradford Book.

Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf.

Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2015). Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164.

Weinberger, D. (2017). Our machines now have knowledge we'll never understand. *Wired*.

Winograd, T. and Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Intellect Books.

Wu, R., Yan, S., Shan, Y., Dang, Q., and Sun, G. (2015). Deep image: Scaling up image recognition. *arXiv preprint arXiv:1501.02876*, 7(8).

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1):75–89.