REVIEW ESSAY

Artificial Intelligence and the Problem of Judgment

Zeynep Pamuk

The Promise of Artificial Intelligence: Reckoning and Judgment, Brian Cantwell Smith (Cambridge, Mass.: MIT Press, 2019), 184 pp., cloth $24.95, eBook $14.99.

The Political Philosophy of AI: An Introduction, Mark Coeckelbergh (Cambridge, U.K.: Polity, 2022), 176 pp., cloth $64.95, paperback $22.95, eBook $18.

Will existing forms of artificial intelligence (AI) lead to genuine intelligence? Typing this question into the AI chatbot ChatGPT produces the following answer: "While AI systems can be incredibly sophisticated and can perform many tasks that once required human intelligence, they lack the ability to truly understand the world and make independent decisions in the same way that humans do. They do not have consciousness, emotions, or subjective experiences, and they are not capable of introspection or self-awareness."[1]

This is not a bad answer, though Brian Cantwell Smith's new book *The Promise of Artificial Intelligence: Reckoning and Judgment* offers a more interesting and nuanced one. While ChatGPT and most literature on the topic of artificial intelligence focus on the abilities that AI lacks, Smith argues that its deficiencies lie in its stance toward the world—or rather, its lack of a proper stance. Human-level intelligence requires commitment to and engagement with the world *as world*

Zeynep Pamuk, London School of Economics and Political Science, London, United Kingdom (z.pamuk@lse.ac.uk)

Ethics & International Affairs, 37, no. 2 (2023), pp. 232–243.

doi:10.1017/S089267942300014X

and an awareness of there being genuine stakes. Only an entity that fulfills these conditions can live up to the normative ideal of judgment, which Smith views as a remarkable achievement of humanity forged over thousands of years and in diverse cultures. Though AI's abilities are becoming ever more impressive, they lie in calculation or reckoning, not judgment (hence the book's title). Smith maintains that neither deep learning nor other forms of second-wave AI (nor any proposal currently advanced for third-wave AI) will lead to anything remotely close to genuine intelligence.

Smith does not intend to put forward a new or controversial account of judgment, but the book's ontological reflections are anchored in an existentialist understanding of the concept. Although his argument falls mainly within more abstract domains of philosophy, such as the philosophy of mind and metaphysics, the way he conceives of judgment and the distinction between reckoning and judgment have important political implications. This essay intends to draw out these implications by bringing Smith's technical insights into dialogue with the political challenges discussed in Mark Coeckelbergh's *The Political Philosophy of AI: An Introduction*, focusing on the role of judgment in both works.

Coeckelbergh's book is an introduction to the political philosophy of AI, which uses fundamental political concepts such as freedom, equality, democracy, justice, and power to analyze the social and political issues raised by AI. If Smith is interested in how much AI is like us or can become so, Coeckelbergh can be seen as exploring the opposite question: How is AI changing us—our society, institutions, norms, and concepts—and what dilemmas do these changes create for us? One of the book's central aims is to push back against the simplistic view that technology is merely a tool and articulate a more complex vision of how it shapes humans. As such, when Coeckelbergh takes up the problem of judgment, his concern is less about whether AI has or can have judgment and more about what happens to our very concept of judgment in societies that are increasingly reliant on AI.

In considering the problems of judgment in AI raised by these two books, this essay will make three main points. First, I will argue that the existentialist conception of judgment that Smith defends is highly idealized and demanding, even for human beings. Holding it up as a standard of intelligence may be appropriate, but its implications for when and how AI should be deployed are not as clear as Smith suggests, especially given rival conceptions of judgment that are less demanding. Second, the displacement of judgment that Smith is concerned with is not unique to AI and long preceded it. The relationship between AI technologies and the

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

longer-term erosion of judgment raises stimulating questions that could help situate developments in AI within the context of broader structural changes. Yet these developments are only intelligible if we move beyond ontology and metaphysics and into political philosophy. Finally, I argue that Coeckelbergh's distinctly political conception of judgment might offer a solution to an important boundary-drawing problem between tasks requiring judgment and those requiring reckoning, thus filling a gap in Smith's argument and clarifying its political stakes.

Reckoning and Judgment

Smith starts with a brief history of AI, focusing first on its failures in the 1960s and 1970s and then its spectacular recent success. He points out that first- and second-wave AI were built on fundamentally different philosophical views, concerning not only the nature of intelligence but the world itself. The first wave subscribed to a Cartesian understanding of intelligence as rational thought, shown most clearly in logical inference. Its corresponding ontology assumed that the world was made up of discrete formal concepts that relate unambiguously to one another. The challenge was to teach AI these formal relationships. According to Smith, this approach failed because its underlying ontology was deeply flawed. The world itself is "unutterably rich" (p. 95). Intelligence can only emerge from a tacit background of knowledge and sense-making that cannot be captured in the rigid and limited formal ontology fed into AI by humans.

The ongoing second wave of AI—a suite of technologies called "machine learning"—eschewed logic entirely and was developed instead on the principles of statistics. It operates by recognizing patterns and making predictions based on large amounts of data. There have also been important architectural/neurological developments in the second wave, leading AI to be modeled more closely after the brain's architecture, which involves parallel neural networks working in tandem. These changes shifted the focus from clear relationships between a small number of discrete concepts to a large number of weak correlations and mappings between inputs and outputs. Smith argues that the impressive success of second-wave AI in tasks such as pattern recognition and classification can be traced to the fact that it operates beneath the narrow conceptual categories of human beings and can take into account vastly more detail. Problems arise only when we try to fit these patterns back into a small number of conceptual forms, thus forcing AI to adopt the limits of our own categories.

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

Despite these recent advances, Smith claims that AI still falls short of genuine intelligence. The main reason, according to Smith, is epistemological; AI systems simply do not know what they are talking about. They cannot engage with or defer to the world itself. We interpret their claims as being about the world, but this is different in an important way than the system itself understanding that there is a world. To be able to do this, Smith argues, the system would need to be existentially committed. He lists several requirements for this: The system must be intentionally directed toward reality; it must be able to distinguish between objects and their representations; it must have a sense of things mattering or there being stakes; and it must be able to distinguish between the actual, the possible, and the impossible. Not only must it be able to make these distinctions but it must also be able to care about them. This requires an orientation to the world backed by a web of normative commitments.

Without such authentic engagement with the world, AI will not be able to exercise judgment, which Smith defines as "dispassionate deliberative thought, grounded in ethical commitment and responsible action, appropriate to the situation in which it is deployed" (p. XV). Judgment is the normative ideal to which human intelligence aspires, and Smith distinguishes it from the ever more impressive feats of reckoning that AI performs. He defines "reckoning" as the ability to manipulate symbolic representations without a commitment to whether the world is the way it is represented. A reckoning system can be extraordinarily powerful in making calculations—as AI already is—but incapable of understanding what those calculations are about. Smith is careful not to claim that judgment is a distinctly human attribute, but he does not see how an AI system would develop it. He argues that current approaches are far from grappling with this problem of judgment.

Smith's existentialist characterization of intelligence and judgment are elegant and moving, especially when he emphasizes the uniqueness of this achievement and describes it as a mark of "the sacred, the beautiful, and the humane," if not of the human (p. 146). This may well be a good characterization of genuine intelligence, although I must admit I am not sure what it would take to insist that one characteristic rather than another is required for intelligence. Smith does not mention opponents who are likely to disagree with him or defend his view against possible rivals, so it is difficult to know what the criteria are for evaluating a given definition of intelligence. Either way, judgment as existential commitment seems rather demanding. People have long asked what intelligence is, and as AI

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

meets more of the conditions that used to be seen as hallmarks of intelligence, definitions of intelligence seem to evolve in ways that diminish the significance of what AI can do. In the eighteenth and nineteenth centuries, brilliant mathematicians and astronomers were distinguished by their ability to manipulate large sets of numbers.[2] Their feats of mental arithmetic—which Smith would describe as reckoning—were regarded as signs of genius. The Marquis de Condorcet claimed that calculation was the foundation of all intellectual operations, including the formation of ideas, judgment, and reasoning.[3] As calculation became mechanized, however, it became commonplace to dismiss it as a mindless task, inferior to abstraction, originality, and understanding.[4] According to one study from the 1920s, fourteen experts defined intelligence as involving abstract thinking, problem solving, adaptability to new environments, and the capacity to learn.[5] As AI today displays each of these abilities to an impressive degree, we have new definitions of intelligence that emphasize capacities AI does not yet possess.

Perhaps a more useful way to think about the problem of defining intelligence is to ask what exactly is at stake. Here, Smith has a clear response: unless AI is genuinely intelligent, we ought to use it only when we are prepared to take responsibility for "every registration scheme, every inferential step, and every 'piece of data' that they use along the way" (p. 80). The problem, however, is that, as Smith admits, humans do not always live up to this demanding ideal of judgment, either. It is an aspirational standard, not one that can be met in every cognitive act by a human. Smith himself notes that standards of public discourse today appear to fall short of his ideal of dispassionate judgment, but he points out that this very criticism is a sign that we have not yet forgotten norms of judgment, even if they are under threat.

This is hardly reassuring. If humans regularly violate norms of judgment and the state of public discourse is evidence that things are in fact getting worse, then the argument for always keeping a human in the loop is weakened. To be persuaded of this conclusion, we need something more than the claim that humans simply have the capacity for judgment. We need to know if and how often humans exercise this capacity well and what happens when they fail to exercise it or exercise it poorly. We might also want to ascertain whether the consequences are worse when humans fail or when a decision requiring judgment is entrusted to AI.

Since AI does not seem to be willfully subversive, cruel, manipulative, or self-interested (for now), there might well be cases where we are better off removing

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

the human in the loop, even for decisions that appear to require judgment. The distinction between human- and AI-appropriate tasks does not depend categorically on having or aspiring to judgment. Rather, it requires the ability to weigh the relative likelihood of error when it comes to certain types of decisions (such as those that Smith classifies as requiring judgment) and the consequences of making a mistake. Contra Smith, we may not need AI to be actively committed to our survival if its reckonings are sufficiently good at ensuring it.

Of course, this framing of the issue as a weighing problem, with the assumption that we may be able to predict expected consequences and compare their pros and cons, may be evidence that I am already hopelessly in the grip of algorithmic thinking—the very possibility that Smith fears. In fact, two of the most thought-provoking possibilities in the book appear as fears: Smith is "terrified" that, first, we will assign AI tasks that require judgment rather than reckoning, and second, that we will be so impressed by reckoning that we will change our expectations of human intelligence in this direction (p. XIX), prioritizing feats of calculation over judgment.

Regarding his first concern, the distinction between tasks that require judgment and those that require only reckoning raises the question of how we would distinguish between the two. Smith does not specify. In many areas of life, and certainly in many kinds of jobs, the kind of reasoning required of humans will in fact be more like reckoning. Or it may be that more tasks begin to appear that way only after AI begins to perform them rather well. It would be helpful to have some criteria to tell these apart. I suspect this task will not be easy, especially considering ever-changing norms and expectations about what AI can and should be entrusted with. A discussion at this level would take us outside ontology—and thus beyond the book—into the realm of ethics and politics.

Smith's own argument presupposes a particular ethical and political view—an existentialist one—that grounds his emphasis on authentic engagement with the world as a condition of intelligence, although he never outright defends this. From within other ethical or political traditions, however, AI's failings in judgment may not appear as shortcomings at all. For a utilitarian, right action requires solving a maximization problem under constraints. One does not have to have a sense of the world as the world to solve for the morally correct action; one just needs data. Given the simplicity of the maxim and the daunting calculations required to apply it, AI may be much better at it than humans, not least since

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

humans are often prone to biases in reasoning precisely because we are in the world.

Furthermore, a utilitarian would not at all subscribe to the division between problems requiring judgment and problems requiring reckoning. Nor, for that matter, would Hobbes, who famously defined reason as "nothing but reckoning."[6] We might even enlist some Rawlsians if we consider the possibility that the difference principle could be more successfully applied by a sophisticated algorithm. There are theories of ethics and politics that require less than authentic engagement with the world; existentialism may be unique in being defined by this commitment.

Smith's second worry—that humanity will be so impressed by AI that we will shift our expectations of intelligence in a reckoning direction—is strikingly similar to fears about technocracy, which revolve around the displacement of judgment and its replacement with technical calculation in modern politics. In Habermasian language, we could express this as a worry about the increasing dominance of a certain mode of instrumental rationality, which has taken over (or "colonized") spheres that are not appropriately governed by its rules.

The displacement of judgment and the rise of instrumental rationality, however, were well underway before second-wave AI. It would be a mistake, then, to explain these changes only through technological advancements.

The elective affinity between these two complaints is revealing because it suggests one way in which technical developments in AI can be tied to more fundamental structural changes wrought by modernity. There is clearly a relationship between the rise and tremendous success of a certain understanding of calculative rationality and the successes of second-wave AI, marked by its mode of intelligence as reckoning. These observations hold out clues for understanding how technological visions of intelligence are connected to the historical and political context in which they are embedded, though, of course, causality is difficult to establish. If AI did not cause the trend away from judgment toward reckoning, is it a product of it? Could it be that AI can reckon because reckoning has long triumphed over judgment as a governing rationality?[7] Or could both changes be driven by more fundamental economic or social structures? Answering these questions is beyond the scope of this essay, but it is clear that we must turn to political philosophy to begin our search.

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

Political Philosophy and AI

Although Smith and Coeckelbergh cover very different grounds, they are united by the belief that we need to explore more than just ethics to understand the challenges that AI poses for society. Smith argues that ethics fails to uncover the deepest philosophical dilemmas raised by AI because it skips over foundational questions about the nature of intelligence, and he views such questions as prerequisites for meaningful discussions on the topic. Coeckelbergh agrees that "ethics of AI" falls short, but instead of turning toward technical discussions of mind, world, and intelligence, he maintains that we must pay more attention to politics. While Smith does not mention politics even once in his book, Coeckelbergh insists that the tools and concepts of political philosophy are indispensable for thinking through the normative issues raised by AI.

Coeckelbergh charts the different ways in which political philosophy sheds light on and is in turn transformed by AI. First, AI raises important challenges for political values such as freedom, equality, democracy, and power. Political philosophy can help us diagnose problems by pointing out how different understandings of these concepts will highlight different issues. Second, AI has both intended and unintended effects on long-standing social and political problems, and important insight can be gained by reexamining these issues in light of new technologies. Finally, and most interestingly, Coeckelbergh argues that AI calls into question our very understanding of the concepts we use in political philosophy. While the first point requires showing how historical and contemporary views in political philosophy apply to and elucidate the effects of new technologies, the second and third focus on how AI leads us to rethink the traditional problems and concepts of political philosophy.

The Political Philosophy of AI is organized around particular concepts; freedom, equality, democracy, power, and nonhumans each get their own chapter. The chapters go through the concepts in question and show how they illuminate issues around AI. The chapter on freedom, for instance, starts with a discussion of negative freedom and how AI technologies are used to facilitate state surveillance and interference. It then turns to freedom as autonomy and considers the compatibility of AI-driven nudges with ideals of autonomous choice. Next up is freedom as emancipation, focusing on exploitation and alienation in an automated workplace. The chapter ends with the Arendtian view of freedom as political participation and a discussion of social media and the manipulation of voters.

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

As this brief summary may indicate, the book covers a lot of ground, both in political philosophy and AI, without exploring the issues in depth. It is an accessible introduction that spans the history of political thought and contemporary political philosophy without committing to any approach or position in the field. Coeckelbergh does not develop a clear thesis in the book or try to resolve the problems he highlights but raises many thought-provoking issues that should leave the good student of political philosophy eager to explore further.

He does, however, emphasize a few general lessons throughout the book. They are not particularly controversial, though still worth repeating: Technology is not neutral; it is inherently political. AI does not lend itself intrinsically to good or bad ends; its effects are always traceable to decisions made by humans, who must be held accountable. While the conjoining of these two claims might sound paradoxical, Coeckelbergh argues that AI can exert power by shaping decisions and decision environments. These, however, are ultimately traceable to particular design and deployment decisions.

The effects of AI may be intended or unintended, and design decisions may be intentional or the result of tech workers' unthinking obedience to rules. In fact, one of the interesting suggestions Coeckelbergh makes, following Dan McQuillan, is that AI systems themselves may encourage uncritical rule following.[8] Drawing on Hannah Arendt's *Eichmann in Jerusalem*, Coeckelbergh defines this thoughtlessness as the inability to critique instructions, the lack of reflection on potential consequences, and a commitment to the belief that a correct order is being carried out.[9] These can also be defined in terms of lacking judgment, thus putting Coeckelbergh on the same page as Smith. However, the causal connections the two authors draw between AI and the erosion of judgment are different. While Smith suggests that we might be impressed by AI and want to emulate its style of intelligence, Coeckelbergh traces the displacement of judgment to the opacity and seemingly incontrovertible statistical authority of AI recommendations. This leaves humans with no option but to accept and follow.

While there is little overlap in the issues that Coeckelbergh and Smith cover and the literatures they engage with, there is a continuity in their reflections on judgment. I have already mentioned Smith's two major fears: that AI will be used in areas where judgment is needed, and that humans will shift expectations of intelligence away from judgment and toward reckoning. We can read Coeckelbergh as taking up these concerns where Smith left off.

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

The most critical intervention Coeckelbergh makes on the topic of judgment is in introducing a focus on *political* judgment. While Smith manages to discuss judgment without ever mentioning politics, Coeckelbergh evokes a long tradition in political philosophy, stretching from Aristotle to Arendt, that maintains that judgment can be developed only through political interactions among citizens. Where Smith emphasizes the importance of engaging with the world, Coeckelbergh cites Arendt to emphasize the necessity of engaging with a *shared* world. This involves engaging imaginatively with others' viewpoints, which makes judgment possible.

Adopting a political conception of judgment has important implications. First, it makes the attainment of judgment even more demanding for AI, in that it is not enough for AI to recognize that there is a world; it must also recognize the world as shared and be able to understand the perspective of others in deliberative engagement. This is an incredibly tall order.

Coeckelbergh himself points out that the Arendtian notion of political judgment is rare even among humans. For many citizens, political activity will simply be a narrow calculation of self-interest based on incomplete information. While a conception of judgment that gives political activity a central role may be more attractive as an ideal, it may also turn out to have an irreducibly elitist dimension.[10] As I pointed out in the previous section, if our conception of judgment is highly idealized and excludes many instances of ordinary human decision-making, it will be an unhelpful standard for judging AI and determining which decisions we should entrust it with.

At the same time, Coeckelbergh's turn to political judgment may offer one way in which we could carve out the space between decisions that need judgment and those that need merely reckoning. Although he does not take this step himself, one conclusion we can derive from his discussion is that if a decision is political, it will require judgment; if it is purely technical, reckoning will do. If we follow Coeckelbergh's claims that issues with AI are almost always political rather than merely technical, we can conclude that the importance of judgment extends to most, if not all, decisions involving AI. This offers one way of delineating the scope for judgment vs. reckoning and thus potentially filling in the gap in Smith's argument.

Coeckelbergh turns to the issue of judgment once more in his chapter on power, where he explores how AI shapes our understanding of self and subjectivity. He introduces this idea through Ray Kurzweil's transhumanist

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

fantasy of resurrection and immortality. Kurzweil claims that machine learning will soon be able to reconstruct a digital version of his dead father and make it possible to have conversations with his avatar. He then goes on to claim that the avatar would be more like his father than his actual father—an idealized self that is more coherent, consistent, and rational than any actual self could be.

Coeckelbergh interprets this as an illustration of how AI molds our conceptions of self in ways that encourage abstracting away from the motivations, intentions, and desires that constitute the reflexive, imperfect human self. Fascination with AI leads to conceptions of subjectivity that abstract from all the things that Smith thinks make judgment possible, including deliberative engagement with the world, intersubjective awareness, and ethical commitment. The resurrected avatar can be more real than the actual father only if authentic engagement with the world and with others is seen as inessential, while data and information uploaded from the brain to a computer are viewed as the essence of a self. This fantasy, of course, would be Smith's nightmare.

Conclusion

As AI displays ever more astonishing reckoning powers, both books discussed here will be invaluable guides for understanding the dilemmas this raises. Reading the two together, we are led to pay attention to the relationship between the abstract philosophical question of whether AI will achieve genuine intelligence and political questions about how AI is changing our world, concepts, and sense of self. This juxtaposition deepens our understanding of judgment in the context of AI, but also exposes the difficulties of moving from the ontological to the practical and political. An existentialist conception of judgment is attractive as the basis of a definition of intelligence, but when it comes to determinations about the deployment of AI, considerations of accuracy, bias, and consequences—and the relative reckoning prowess of AI and humans—cannot be easily bracketed in favor of a capacity for judgment. A fuller understanding of the relationship between reckoning and judgment would situate the technological changes addressed in these books within the broader social and political trend of the displacement of judgment and the triumph of calculative rationality over a long period that predates the rise of AI.

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press

Notes

[1] OpenAI's ChatGPT AI language model, response to question from author, February 3, 2023.

[2] Lorraine Daston, *Rules: A Short History of What We Live By* (Princeton, N.J.: Princeton University Press, 2022)

[3] Jean-Antoine-Nicolas de Caritat, *Moyens d'apprendre à compter sûrement et avec facilité* (Paris: Moutardier, 1804; repr. in *Enfance* 42 [1989], pp. 61–62).

[4] Daston, *Rules*.

[5] Rolf Pfeifer and Christian Scheier, *Understanding Intelligence* (Cambridge, Mass.: MIT Press, 2001), p. 6.

[6] Thomas Hobbes, *Leviathan*, ed. Richard Tuck (New York: Cambridge University Press, 1996), Ch V.

[7] See, for example, Wendy Brown, *Undoing the Demos: Neoliberalism's Stealth Revolution* (Cambridge, Mass.: MIT Press, 2015).

[8] Dan McQuillan, "The Political Affinities of AI," in Andreas Sudmann (ed.), *The Democratization of Artificial Intelligence: Net Politics in the Era of Learning Algorithms* (Bielefeld, Germany: transcript, 2019), pp. 163–73.

[9] Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (New York: Penguin, 2006).

[10] Dana Villa, *Arendt and Heidegger: The Fate of the Political* (Princeton, N.J.: Princeton University Press, 1995), p. 35.

Abstract: Will existing forms of artificial intelligence (AI) lead to genuine intelligence? How is AI changing our society and politics? This essay examines the answers to these questions in Brian Cantwell Smith's *The Promise of Artificial Intelligence* and Mark Coeckelbergh's *The Political Philosophy of AI* with a focus on their central concern with judgment—whether AI can possess judgment and how developments in AI are affecting human judgment. First, I argue that the existentialist conception of judgment that Smith defends is highly idealized. While it may be an appropriate standard for intelligence, its implications for when and how AI should be deployed are not as clear as Smith suggests. Second, I point out that the concern with the displacement of judgment in favor of "reckoning" (or calculation) predates the rise of AI. The meaning and implications of this trend will become clearer if we move beyond ontology and metaphysics and into political philosophy, situating technological changes in their social context. Finally, I suggest that Coeckelbergh's distinctly political conception of judgment might offer a solution to the important boundary-drawing problem between tasks requiring judgment and those requiring reckoning, thus filling a gap in Smith's argument and clarifying its political stakes.

Keywords: artificial intelligence, judgment, technology, rationality, transhumanism, calculation, politics, ontology

https://doi.org/10.1017/S089267942300014X Published online by Cambridge University Press