

Book Review

# Humans are the Danger, Not AI: A Review of Shannon Vallor's *The AI Mirror*

Shannon Vallor, *The AI Mirror: How to reclaim our humanity in an Age of Machine Thinking*. (Oxford University Press, 2024).

Blessing Adewuyi <sup>1,\*</sup>

<sup>1</sup> Religion Department, University of Georgia, Athens, GA. [blessing.emmanuel@uga.edu](mailto:blessing.emmanuel@uga.edu)

\* Correspondence: [blessing.emmanuel@uga.edu](mailto:blessing.emmanuel@uga.edu)

**Abstract:** This review essay critically engages with Shannon Vallor's *The AI Mirror: How to reclaim our humanity in an Age of Machine* which is an astute exposition to the threats that Artificial Intelligence (AI) possess to our humanity; a threat the author argues from a philosophical framework is distinguishable from exaggerated motifs and media representations of AI's existential danger to human survival and of Artificial General Intelligence (AGI) which is still yet to come. Drawing from her experience as a consultant in Silicon Valley for one of the leading AI companies and her scholarship in philosophy and ethics, the author boldly presents an unpopular position that whatever dangers or threats AI possess to human survival came from within humanity; AI itself isn't the danger but humans! This is a bold assertion that offers an alternative perspective to mainstream positions in AI ethics, and the most obvious rejoinder to such a position is how are we (humans) the danger? Aren't humans the potential 'victims' of AI's existential threats?

**Citation:** Adewuyi, Blessing. 2026. Humans are the Danger, Not AI: A Review of Shannon Vallor's *The AI Mirror*. *Journal of Ethics and Emerging Technologies* 36: 1. <https://doi.org/10.55613/j eet.v36i1.225>

**Keywords:** Artificial Intelligence; Humanity; Mirror; Philosophy; Ethics; Virtue.

## 1. Introduction

The last few decades of human society have experienced an avalanche of innovative and unprecedented technological inventions that have no doubt reshaped and redefined what it means to be human in the world. The *AI Mirror* uniquely crafts a narrative that engages the overarching concerns of the book which is to redirect our focus as a society to the source of the dangers posed by AI. The Author argues that "We (humans) are the source of the danger to ourselves from AI." (p7) The book advances that although AI is celebrated as a powerful tool for opening new futures, it has undoubtedly and repeatedly pointed humanity to the past; presenting AI as a digital mirror and a reflection of humanity that points backwards. In seven capacious chapters contained in the book, Vallor makes a case for what AI could be, in helping humanity reclaim her full potential for growth rather than being swept away by a reprojection of the past. But to achieve this, humanity must recreate itself by rethinking what AI is, what it can be and most importantly, what we (humans) want to be with it. The author tacitly discusses the intricacies and complexities heralded by the interdependence between humans and AI using simple terms and relatable themes accessible to interested readers within and outside academic scholarship.

## 2. AI as a Mirror of Humanity

Received: 10/03/2026  
Accepted: 24/03/2026  
Published: 24/03/2026

**Publisher's Note:** IEET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

In the first two chapters, the author introduces the overall objective and discusses the rationale for employing the mirror metaphor in stating the thesis that Artificial Intelligence (AI) is a mirror of humanity, “not as we ought to be or could be, but as we already are and have long been” and that whatever characteristics we see portrayed by AI is but a reflection of what exists in and with humanity (p10). This statement is impressively exemplified by AI’s capacity for extracting and analyzing large amounts of information “statistical patterns from past human-generated data and projecting these patterns forward into optimized predictions, selections, classifications and compositions” something AI does exceptionally well! (p38) But that is just one side of the coin in the argument. The second (and perhaps more definitive) side of the coin to the argument is that, unlike AI, humanity has the capacity to break free from the algorithmic predictions and patterns that have engulfed our understanding of ourselves and our world – a world in which no AI system is an active/sentient participant. By this, it is implied that the absence of embodiment and biological mind – which the author took a couple of pages to explain – separates AI systems from experiencing the world in the same way a human being would experience it.

While this may be a statement of fact, it does not take away the reality of AI’s engagement with human society through different pre-programmed and learned artificial neural frameworks, which are increasingly being developed with Large Language Models (LLMs) and allied algorithms. It is important to foreground a critic of the parallels drawn between the functioning of AI and the human mind in computational terms. Vallor more specifically takes a stand against the computational theory of the mind, as limited and often incapable of delineating the capaciousness of the embodied human mind in a human world; a crucial stand many philosophers and ethicists will align with. (p40) Going beyond the human mind, it can be argued that a computational understanding of the human person may result in a reductive understanding what it means to be human, in terms of lived experiences and personal narratives that go beyond a mechanical and statistical framework.

### **3. Virtues and Vices**

In her third chapter, the author draws from her philosophical engagements with virtues to draw a distinction between moral virtues (such as courage, honesty and others) and intellectual virtues (such as wisdom and curiosity). Both types are essential paradigms of human character formation, and intellectual virtues have played significant roles in the development of technology, robotics and AI. As humanity increasingly interacts with AI, the author boldly asserts that “AI has the potential to change our human character; to make it better or worse” on the one hand (p67). On the other hand, AI also tends to obscure the most important virtues of humanity, thereby making it difficult for us to recognize these virtues in ourselves and in one another especially as human interactions and relationships are increasingly mediated through technological channels that have little to no inherent semblances of human virtues. The author invites the reader into the world of an intellectual virtue called imagination. Although a critical assessment may cursorily rule out imagination as insufficient to be categorized as a virtue, especially since not that many scholars and philosophers would have thought of imagination as a virtue. However, as the author increasingly unravels the benefits of human imagination for human society, it becomes obvious and perhaps convincing to tentatively accept imagination as an important quality to possess, even if it might not be as important or placed on the same pedestal as other virtues, especially those necessary for human survival. Her aim in presenting imagination as a virtue is to herald a call for humanity to both recognize the ways in which technology and AI can strengthen or undermine human virtues, on one hand, as well as the need for humanity to integrate our virtues and values into the development of technology and AI, on the other.

However, the author is silent about imagination as a vice, which has no doubt led to unthinkable evil unleashed on human society. Arguably, this should distinguish imagination from other virtues that are fundamentally benign, but this delineation is not clearly presented in the discussion. While it can be argued that any virtue can become a vice in extreme contexts, the essential presentation of imagination as a virtue without recourse to its potential as a vice is somewhat imbalanced. So, while imagination may be tentatively acceptable as a virtue, a counterbalance must be established that imagination is both a virtue and a vice.

#### **4. Moral Deskilling**

Chapters four and five of the book explore important critiques of technology related to the attenuation of moral reasoning and responsibility in human beings since AI makes the process of decision making easier and faster, and because of the opacity of such processes. When humans become increasingly dependent on algorithms to make crucial decisions regarding important aspects of human life, the author argues we accentuate 'moral deskilling' – describing a situation in which humans lose the capacity for moral reasonings and the spaces for moral deliberations (p117). Since it is implausible to ascribe moral responsibility to AI algorithms that make decisions, how can society apportion praise or blame? On top of that, the propensity for AI to manufacture probable responses which are not truths or lies but 'bullshit' narratives complicate the importance of a society built on trust. Furthermore, algorithms make decisions drawing from dominant patterns that were shaped by only a subset of society which become encoded as data created from a deeply flawed algorithmic architecture. The author gives an example of how any large language model trained to extract information from digitized texts that were published over a few decades ago will capture predominantly only male voices and perspectives. In this sense, AI algorithms will intrinsically accentuate historical perspectives backed by predominant voices in power.

#### **5. Project Bootstrapping**

In the final chapters, the author calls for a reflection on renewed paradigms for building a sustainable world based on virtues, values and moral ideals that are derived from the exercise of moral imagination as a virtue. More importantly, the author argues for the ways in which AI, rather than erode these virtues, can actually participate in the human project of creative-self renewal; what the author calls auto fabrication and "boot-strapping challenge for humanity – and for the future design and use of AI" (p163). In these final pages, the author advances the need for human society to rethink our dominant values, habits, moral virtues and epistemic postures, much like a call for humanity to reinvent herself as an ecumenical project. Of course, many challenges threaten the feasibility of such project, including the moral inadequacy of current human society to engineer a change that ensures the future sustainability of the planet. Furthermore, the author argues that AI presents both a danger and an opportunity for humans. One the hand, AI is a danger because it accelerates the transfer of wealth and power into the hands of a few people and because by humanity's overreliance on AI, we gradually and increasingly lose confidence in ourselves and others; we become convinced that AI and machines are better and we lose "faith in our capacity to be any better than our mirrors" (p200). On the other hand, the opportunity rests in humanity's unique ability to remake – to create. Humans have the capacity to remake ourselves and our society into "something that never existed in that shape before, and thereby bring into the world a new value system; a new image of the good..." (p201) This call to recreate and reinvent ourselves and our values gives humanity the opportunity to foreground the virtues and values that will build a sustainable future, and for intentionally recrafting what AI can be for a world we will all be proud of.

## 5. Conclusions

Summarily, The AI mirror calls out humanity for its past inadequacies and culpability in AI's existential threat to humanity and offers a democratic project of re-creation with emphasis on understated virtues of wisdom, excellence and prudence in helping humanity recreate itself and leverage AI as a tool in that project. According to the author, this project will only be successful based on "coordinated and co-constructed democratic efforts of creative moral improvisation...guided by practical wisdom" as a virtue related to prudence (p165). She argues that coordinated and widespread human excellence (as a virtue) in centering the sustainability of our planet and a divine rescue are the only two options that can protect human society from existential threats. This may signal a pessimistic view of human capacity to turn from the bad towards the good; and while there may be valid reasons for such positions given the untoward and disastrous acts of humanity with irreversible effects on the planet, I argue that it might be too early to give up on humanity just yet, even if the kind of change needed will be "hard to get going" (p169). The joint belief in humanity to remake itself presents an opportunity, despite all odds, and necessitates a reinvention that can safeguard the future. AI isn't all bad and humanity isn't all doomed. Human society can indeed reinvent itself to prioritize the values and virtues that truly matter and AI can be remade for a humane future and as a tool to serve the purposeful ends of sustainability and preservation, especially if humans choose and demand that AI is built to serve those purposes.

**Funding:** This research received no external funding

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

(Vallor 2024) Vallor, Shannon. (2024) *The AI Mirror: How to reclaim our humanity in an Age of Machine Thinking*. New York: Oxford University Press. pp. 1-228.