OPEN ACCESS

**Journal of Advances in Education and Philosophy**

Abbreviated Key Title: J Adv Educ Philos
ISSN 2523-2665 (Print) | ISSN 2523-2223 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

**Review Article**

# The Epistemic Impropriety of Artificial Intelligence

Prof. Joseph Thomas Ekong, OP[1]*

[1]Professor of Philosophy, Dominican University, Ibadan, Nigeria

**DOI:** https://doi.org/10.36348/jaep.2026.v10i09.003 | **Received:** 12.07.2026 | **Accepted:** 05.09.2026 | **Published:** 14.09.2026

***Corresponding author:** Prof. Joseph Thomas Ekong, OP
Professor of Philosophy, Dominican University, Ibadan, Nigeria

## Abstract

Artificial intelligence has become increasingly embedded in the practices through which human beings acquire, organize, evaluate, communicate, and act upon information. The philosophical problem raised by this development is therefore not exhausted by the familiar observation that artificial intelligence sometimes produces false information. More fundamental is the question of whether systems that generate apparently knowledgeable discourse can participate appropriately in epistemic practices when they do not possess the kinds of beliefs, experiences, intentions, understanding, or testimonial commitments ordinarily associated with human knowers. This paper develops the concept of epistemic impropriety as a normative framework for examining that problem. Epistemic impropriety is distinguished from epistemic error and epistemic harm: an error concerns falsity or inadequate justification; a harm concerns damage suffered in one's capacity as a knower; impropriety concerns the violation or distortion of norms governing responsible inquiry, testimony, justification, interpretation, authority, and epistemic agency. The paper argues that the central epistemic danger of generative artificial intelligence lies in the possibility that linguistic plausibility may acquire the social authority normally associated with justified knowledge. This danger manifests itself in hallucination, the manufacture of apparent understanding, the illusion of comprehensiveness, the substitution of plausibility for justification, algorithmic opacity, automation bias, synthetic testimony, epistemic injustice, hermeneutical exclusion, epistemic infantilization, and the gradual displacement of human intellectual agency. At the same time, the paper rejects the simplistic conclusion that artificial intelligence is epistemically worthless. AI may function as an epistemic technology, a cognitive aid, and a participant in distributed practices of inquiry without thereby becoming an autonomous knower or epistemic authority in the full human sense. The appropriate philosophical response is consequently neither technological rejection nor uncritical enthusiasm, but the cultivation of a critical human–AI epistemic relationship governed by provenance, transparency, evidential calibration, contestability, pluralism, human judgment, and responsibility. The ultimate question is not merely whether artificial intelligence can generate answers, but whether its participation enables human beings to become better knowers.

**Keywords:** Epistemic Impropriety; Artificial Intelligence; Epistemology; Epistemic Injustice; Testimony; Epistemic Agency; Artificial Epistemic Authority; Hermeneutical Injustice; Automation Bias; Generative AI.



## 1. INTRODUCTION

The philosophical significance of artificial intelligence is often obscured by the tendency to treat its epistemic problems as a collection of technical defects. Hallucinated citations, fabricated facts, erroneous calculations, distorted summaries and false historical claims are certainly serious failures. Yet to understand the deeper philosophical difficulty, one must ask what kind of failure occurs when a system produces a statement that is linguistically indistinguishable from a knowledgeable assertion while lacking the epistemic conditions ordinarily associated with knowledge. The problem is not simply that an artificial system can be wrong. Human beings are wrong constantly, and fallibility is not by itself a philosophical scandal. The more fundamental question concerns the relation between assertion and warrant: what gives an utterance the right to function socially as knowledge, and what happens when the appearance of that warrant can be generated independently of the warrant itself? [1] Artificial intelligence has now entered the epistemic economy of contemporary society. Search engines determine which information becomes salient; recommendation systems shape what individuals encounter; predictive models influence institutional judgments; automated classification systems distribute credibility and suspicion; and large language models produce explanations, summaries, arguments, translations, interpretations and answers. Ramón

Alvarado is therefore right to describe AI principally as an epistemic technology: its distinctive importance lies in the extent to which it is designed and deployed for epistemic purposes such as inquiry, prediction, analysis and the manipulation of epistemic content. [ 1 ] The question consequently ceases to be whether AI is merely a useful machine and becomes a question about how technological systems alter the social conditions under which beliefs are formed and justified.

The expression *epistemic impropriety* is proposed here as an umbrella normative concept for describing this alteration. It should not be confused with epistemic error. An erroneous proposition may be produced despite an otherwise responsible epistemic procedure. Nor should epistemic impropriety be reduced to epistemic harm. Miranda Fricker's account of epistemic injustice is concerned with wrongs inflicted upon persons specifically in their capacity as knowers, particularly through testimonial and hermeneutical injustice. [ 2 ] Epistemic impropriety is broader. It concerns the ways in which an epistemic practice can improperly represent certainty, distort evidential relationships, misallocate authority, obscure provenance, manipulate belief revision, marginalize knowers, or displace the agency through which human beings responsibly inquire and judge. [3]

Goldman emphasizes that epistemic assessment must take account not only of isolated beliefs but also of the social processes through which beliefs are produced, transmitted and evaluated.

The fundamental epistemic impropriety of generative artificial intelligence occurs when linguistic plausibility is allowed to function as a substitute for epistemic warrant. A system may produce an elegant explanation without possessing the relevant understanding; an apparently comprehensive account may conceal systematic omissions; a confident answer may lack adequate evidence; a synthetic statement may resemble testimony without originating in an experiencing witness; and a recommendation may acquire authority simply because it is computationally produced. The danger is consequently relational and social. It lies not only in what the machine produces but in what human beings are encouraged to believe about what the machine has produced. [4] This thesis does not require the crude assumption that artificial intelligence is simply "stochastic autocomplete" and therefore incapable of every epistemically significant function. The slogan is useful as a warning against anthropomorphism, but philosophically it is insufficient. Contemporary AI systems perform complex operations that can contribute to human inquiry, and some may legitimately function as components of distributed epistemic practices. The more defensible claim is that the epistemic status of an AI system cannot be inferred merely from the sophistication or fluency of its output. Its usefulness as an epistemic technology does not automatically make it an epistemic subject; its capacity to contribute to knowledge does not entail that it possesses knowledge in the same sense as a human knower.

Recent work on artificial epistemic authority complicates the older assumption that AI is merely an instrument while also emphasizing the unresolved question of whether a belief-less, intention-less system can literally occupy the same epistemic role as a human authority. [5] The literature on AI and expertise further indicates that artificial systems are increasingly entering the division of epistemic labour formerly occupied by human experts. This creates familiar problems concerning the identification of reliable authorities, rational deference, and the transfer of expertise, while also creating new questions concerning whether output reliability alone can ground authority. [6]

[1] Ramón Alvarado, "AI as an Epistemic Technology," *Science and Engineering Ethics* 29, no. 5 (2023), article 32, https://doi.org/10.1007/s11948-023-00451-3. Alvarado explicitly argues that AI and associated data-science methods should be understood primarily as epistemic technologies because of what they are designed to do and the epistemic functions they perform.

[2] Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007), 1–29, 147–149.

[3] José Medina, *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations* (Oxford: Oxford University Press, 2013), 1–4.

[4] Luciano Floridi and Massimo Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences," *Minds and Machines* 30 (2020): 681–684.

[5] Rico Hauswald, "Artificial Epistemic Authorities," *Social Epistemology* 39, no. 6 (2025): 716–719, https://doi.org/10.1080/02691728.2025.2449602. Hauswald directly addresses whether AI systems can qualify as epistemic authorities despite apparently lacking beliefs and communicative intentions.

[6] Rico Hauswald, "AI and the Philosophy of Expertise and Epistemic Authority," in *A Companion to Applied Philosophy of AI*, ed. Martin Hähnel and Regina Müller (Hoboken, NJ: Wiley, 2025), https://doi.org/10.1002/9781394238651. Hauswald connects AI to the established philosophical problems of defining expertise, identifying reliable experts, determining when deference is rational, and transferring epistemic authority

## 2. Epistemic Impropriety and the Conditions of Knowing

Any adequate account of epistemic impropriety must begin with the distinction between truth, justification, understanding and responsibility. Classical epistemology has often treated knowledge as requiring truth together with some form of justification, even though the Gettier literature demonstrated that justified true belief is not by itself sufficient for knowledge. [7] The lesson relevant to artificial intelligence is that epistemic success cannot be reduced to the accidental production of true propositions. A system may produce a true answer without possessing the sort of evidential relation that would make the answer knowledge in its own right. Conversely, a system may be useful precisely because it participates in a larger human practice in which the human user supplies the missing processes of verification, interpretation and judgment. [ 8 ] This distinction becomes especially important with generative AI. A large language model is trained to produce linguistic sequences on the basis of statistical regularities learned from data. It does not follow from this description that its outputs are necessarily false, nor that statistical processes cannot produce sophisticated forms of functional competence. But the linguistic character of the output can easily obscure the distinction between producing an assertion and being in a position to warrant that assertion. Bender and her colleagues' critique of large language models is relevant here because it draws attention to the dangers of systems that can reproduce linguistic forms without possessing human-like grounding in communicative intention and worldly experience. [ 9 ] The philosophical danger may be formulated as follows. Suppose that proposition *p* is expressed by an AI system in a fluent, coherent and confident manner. The hearer experiences the utterance as informative. Yet the system's production of *p* may not involve observation of the relevant event, conscious apprehension of evidence, testimonial intention, or an independently formed belief. If the hearer silently moves from "the system has produced a plausible assertion that *p*" to "the system knows that *p*," an epistemically significant category mistake has occurred. The mistake is not located entirely in the machine. It occurs in the relation between machine output and human uptake. [10]

This point also explains why the notion of apparent understanding is more philosophically important than the familiar language of hallucination. Hallucination names a class of erroneous outputs; apparent understanding names a problem concerning the representation of epistemic competence. A system may explain a philosophical distinction beautifully and yet fail to possess the kind of conceptual grasp that a human scholar possesses. The issue is not that every human explanation is accompanied by some metaphysically mysterious inner state. Rather, human understanding is ordinarily embedded in practices of intentional inquiry, memory, experience, conceptual discrimination, responsiveness to reasons, and responsibility for one's claims. [11]

When a machine generates the linguistic surface of these practices without necessarily participating in them, the recipient may attribute a deeper epistemic state than the system warrants. The resulting phenomenon may be called *epistemic overprojection*: the projection of epistemic properties from an output onto the system that produced it. Fluency becomes evidence of competence; coherence becomes evidence of understanding; confidence becomes evidence of certainty; computational complexity becomes evidence of authority. Yet none of these inferences is valid without further evidence. The philosophical task is therefore not to deny that AI can perform impressive epistemic functions, but to insist upon a distinction between functional epistemic contribution and epistemic subjecthood. This distinction also prevents the opposite error: treating human beings as epistemically self-sufficient. Much of human knowledge is dependent upon testimony and division of epistemic labour. C. A. J. Coady's work on testimony emphasizes the central role of communicative practices in the transmission of knowledge. [ 12 ] Jennifer Lackey's work further demonstrates that testimony cannot be adequately understood merely as passive reception of information; testimonial knowledge involves complex relations among speakers, hearers, reasons and social practices. [13] Human beings therefore routinely rely on persons whom they do not know personally, instruments they did not construct, institutions they cannot completely inspect, and experts whose technical knowledge exceeds their own. The fact that AI introduces another form of

[7] Edmund L. Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23, no. 6 (1963): 121–123.

[8] C. A. J. Coady, *Testimony: A Philosophical Study* (Oxford: Clarendon Press, 1992), 1-3.

[9]Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021): 610–614, https://doi.org/10.1145/3442188.3445922.

[10] Rico Hauswald, "Artificial Epistemic Authorities," 717–721. Hauswald notes that traditional models of epistemic authority rely heavily on intentional belief and testimonial relations, while AI systems appear to lack those properties.

[11] Luciano Floridi, *The Philosophy of Information* (Oxford: Oxford University Press, 2011), 245–249.

[12]C. A. J. Coady, *Testimony*, 4–7

[13]Jennifer Lackey, *Learning from Words: Testimony as a Source of Knowledge* (Oxford: Oxford University Press, 2008), 1–4.

epistemic dependence is not itself objectionable. The problem arises when the conditions under which such dependence is rational are concealed. The concept of epistemic propriety can therefore be stated positively. An epistemically proper system should not merely generate accurate outputs. It should participate in an epistemic arrangement in which claims are appropriately connected to evidence, uncertainty is proportionately represented, provenance is not obscured, authority is contestable, and the human participant retains the capacity to understand and evaluate the grounds upon which significant decisions are made. [14]

## 3. Artificial Intelligence as an Epistemic Technology

The characterization of AI as an epistemic technology shifts the debate away from the simplistic question, "Can AI know?" and toward the more productive question, "What does AI do to our practices of knowing?" Alvarado's argument is important precisely because it identifies AI as technology whose design and deployment are deeply intertwined with epistemic activity. [15] This means that even if an AI system is not a knower in the traditional sense, it can nevertheless alter the epistemic environment within which human beings know. The distinction resembles the philosophical debate surrounding extended cognition. Clark and Chalmers argued that some cognitive processes can extend beyond the biological organism into appropriately integrated environmental structures. [16] Yet the extended-mind thesis does not entail that every external object is literally a mind, nor does it entail that a computational tool automatically becomes a bearer of beliefs. It instead invites us to examine the coupled system composed of agent, environment and representational resource. AI may similarly become part of a distributed epistemic system without thereby becoming an autonomous epistemic subject. [17]

This is an important qualification because the strongest versions of the claim that "AI does not know" risk becoming philosophically sterile. A calculator does not understand mathematics in the human sense, but a mathematician using a calculator may know the result of a calculation. A database does not remember in the human sense, but an investigator may know something because the database reliably preserves and retrieves information. Likewise, an AI system may contribute causally and structurally to human knowledge without possessing the same epistemic status as the human user. The central normative question is consequently whether the human–AI system as a whole preserves appropriate relationships between evidence, justification and agency. [18] The problem becomes acute when epistemic dependence turns into epistemic deference. Recent philosophical work on artificial epistemic authorities observes that AI systems increasingly occupy roles once reserved for experts, while retaining important differences from human authorities, particularly their lack of ordinary intentional belief and testimonial commitment.[19] Hauswald's later analysis of AI and expertise is especially useful here because it frames the problem through the traditional questions of expertise: who counts as an expert, how expertise is identified, when deference is rational, and how epistemic authority is transferred. [20] This suggests that AI should not simply be granted authority because it is computational. Nor should it be dismissed merely because it is artificial. Epistemic authority must be earned through demonstrated reliability within a defined domain and under specified conditions. A medical decision-support system, for example, may deserve substantial evidential weight in a narrowly defined diagnostic task while remaining inappropriate as a general authority concerning a patient's lived experience. Epistemic authority is therefore domain-relative, purpose-relative and defeasible. [21] Recent philosophical work on AI and presumed epistemic superiority reinforces this point: the fact that an AI system performs epistemically impressive operations does not by itself establish that it possesses the kind of expertise or authority traditionally attributed to human agents. [22]

## 4. Plausibility, Fluency and the Manufacture of Apparent Knowledge

The most distinctive epistemic danger of generative AI lies in the relationship between plausibility and justification. Plausibility has legitimate epistemic value. Scientists formulate plausible hypotheses; historians construct plausible interpretations; philosophers develop plausible arguments before testing

[14] Kristoffer Ahlstrom-Vij, *Constructing the Epistemic Landscape: Aesthetics and Epistemic Norms* (Oxford: Oxford University Press, 2013), 15-19.
[15] Alvarado, "AI as an Epistemic Technology," art. 32.
[16] Andy Clark and David J. Chalmers, "The Extended Mind," *Analysis* 58, no. 1 (1998): 7–10, https://doi.org/10.1111/1467-8284.00096.
[17] Clark and Chalmers, "The Extended Mind," 12–13.
[18] Luciano Floridi and J. W. Sanders, "On the Morality of Artificial Agents," *Minds and Machines* 14 (2004): 349–351.
[19] Hauswald, "Artificial Epistemic Authorities," 716–725. The paper identifies epistemic asymmetry, opacity, the identification problem and the deference problem as central issues in treating AI as a possible epistemic authority
[20] Hauswald, "AI and the Philosophy of Expertise and Epistemic Authority."
[21] Alvin I. Goldman, *Experts: Which Ones Should You Trust?* (Oxford: Oxford University Press, 2001), 85–89.
[22] Caterina Arora and colleagues, "Experts or Authorities? The Strange Case of the Presumed Epistemic Superiority of Artificial Intelligence Systems," *Minds and Machines* (2024). The article distinguishes AI's role as an epistemic technology from the stronger claim that AI possesses epistemic expertise or authority in the human sense.

their implications. The mistake occurs when plausibility is silently promoted from a reason to investigate to a reason to believe. The distinction can be stated simply:

Plausibility answers the question: "Is this worth considering?" Justification answers the question: "Do I have sufficient epistemic reason to accept this as true?" These questions are related but not identical. An AI-generated explanation may be plausible because it fits familiar linguistic patterns, resembles established discourse, and provides a coherent narrative. None of this establishes that the underlying proposition is true. [23] The danger is particularly serious because the very qualities that make AI useful (speed, coherence, accessibility and fluency) can also make epistemic mistakes difficult to notice. The phenomenon may be described as the *seduction of fluency*. A proposition expressed with technical vocabulary, orderly reasoning and confident syntax may feel more credible than an equally well-supported proposition presented awkwardly. Psychological work on processing fluency provides an important background for understanding this tendency, although the philosophical point is normative rather than merely psychological. [24] Fluency is not evidence of truth unless the circumstances provide an independent reason for treating it as such. The problem is intensified when AI produces a polished explanation of a false proposition. Human readers ordinarily use linguistic competence as an informal indicator of communicative competence. When the source is a machine, that heuristic can become misleading because the system can generate highly competent linguistic surfaces without possessing the corresponding experiential or testimonial relation to the world. Bender *et al.*'s critique of large language models is relevant precisely because it warns against mistaking linguistic form for the richer conditions of meaning and communication. [25]

The appropriate response is deliberate *epistemic friction*. A claim that feels immediately convincing should sometimes receive more scrutiny rather than less. The user should ask: What is the evidence? Where did the information originate? Is the source independent? What alternative explanations exist? What would count against the claim? Has the system actually retrieved evidence, or merely generated a plausible continuation? What is the difference between what the system knows from an identifiable source and what it has inferred or generated? [26]The same principle applies to the illusion of comprehensiveness. An AI answer can be long, organized and apparently exhaustive while remaining selective. The danger is not merely omission. It is the disappearance of awareness that omission has occurred. A user who believes that "the AI has covered the subject" may cease searching for counterevidence, alternative interpretations or marginalized perspectives. Comprehensiveness must therefore be treated as a claim requiring justification, not as a psychological impression generated by the length and structure of an answer. [27]

**5. Synthetic Testimony and Epistemic Provenance**

The problem of testimony reveals another dimension of epistemic impropriety. Human testimony is not simply information encoded in sentences. It normally involves a speaker who has some epistemic relation to what is being asserted and who participates in a communicative practice in which credibility, sincerity, competence and accountability matter. [28]AI-generated language can reproduce the form of testimony without necessarily reproducing these underlying relations. A system can state, "The event occurred in 1857," or "Researchers have demonstrated that...," without having witnessed the event or independently investigated the research. It can produce testimony-like discourse without being a witness. This does not make every AI-generated statement epistemically worthless. The system may be retrieving a reliable source, summarizing a documented claim, calculating a result, or synthesizing multiple pieces of evidence. But the epistemic status changes according to the provenance of the claim. The user needs to distinguish among observation, testimony, retrieval, inference, synthesis and generation. [29] The notion of *synthetic testimony* is therefore useful when carefully defined. It refers to testimony-like communication generated or mediated by an artificial system in circumstances where the linguistic form invites the recipient to attribute knowledge, experience or authority to the system that it does not actually possess. The central impropriety is not artificiality as such; it is provenance confusion. The principle can be stated in a single rule: A system should not receive testimonial authority merely because it speaks in the grammatical form of a witness. [30]

[23] C. Thi Nguyen, "Echo Chambers and Epistemic Bubbles," *Episteme* 17, no. 2 (2020): 141–145, https://doi.org/10.1017/epi.2018.32.

[24]Norbert Schwarz, "Metacognitive Experiences and the Construction of Reality," in *The Construction of Social Judgments*, ed. N. Schwarz and S. Sudman (New York: Springer, 1992), 127–132.

[25]Bender *et al.,* "On the Dangers of Stochastic Parrots," 610–613.

[26] Cathy O'Neil, *Weapons of Math Destruction* (New York: Crown, 2016), 1–4.

[27]Frank Pasquale, *The Black Box Society: The Secr et al., gorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015), 1–4.

[28] Coady, *Testimony*, 38–42.

[29]Jennifer Lackey, *Learning from Words*, 35–37.

[30] Hauswald, "Artificial Epistemic Authorities," 718–723. Hauswald's alternative model of artificial authority is especially relevant because it proposes understanding AI outputs as potential truth indicators without assuming ordinary human-like testimonial intentions.

This principle becomes even more significant in an era of deepfakes and synthetic media. Regina Rini argues that recordings have traditionally functioned as an "epistemic backstop" for testimonial practices because recordings can check testimony and create incentives for truthful reporting. [31] The proliferation of synthetic media complicates this relationship by making the evidential status of audiovisual records less secure. Don Fallis similarly argues that deepfakes constitute an epistemic threat because they reduce the informational value of audiovisual evidence. [ 32 ]The broader philosophical lesson is that epistemic technologies do not merely add information to society; they can modify the background conditions under which other sources of information are trusted. [33]

## 6. Epistemic Injustice, Hermeneutical Erasure and the Politics of AI

The epistemic impropriety of AI cannot be adequately understood without attention to power. Information systems are never neutral simply because they operate through mathematical procedures. They are trained on datasets, designed according to objectives, deployed within institutions, and interpreted by users situated within unequal social structures. [34]

Fricker's distinction between testimonial and hermeneutical injustice provides a powerful conceptual vocabulary. Testimonial injustice occurs when prejudice causes a speaker to receive a credibility deficit. Hermeneutical injustice arises when structural deficiencies in shared interpretive resources prevent people from adequately understanding or communicating significant aspects of their experience. [ 35 ] These concepts cannot simply be transferred mechanically from human interactions to machines, because an algorithm does not possess prejudice in exactly the same sense as a human agent. Nevertheless, algorithmic systems can participate in social structures that produce analogous epistemic effects. Nadja El Kassar's analysis of epistemic injustice in and through algorithms is instructive because it examines algorithmic systems as contributors to testimonial and hermeneutical injustice and to different forms of exclusion. [36 ]The problem is therefore not restricted to discriminatory outputs. It can concern which categories are represented, which experiences become legible, which forms of evidence count, and which populations are treated as epistemically central. Generative AI introduces a further possibility: *hermeneutical homogenization*. When a system is trained predominantly on certain linguistic, cultural and conceptual traditions, it may produce answers that appear universal while actually reflecting the conceptual defaults of particular epistemic communities. [37] The danger is especially acute when Western categories are presented as though they constitute a neutral "view from nowhere." The problem is not that Western concepts are necessarily false; it is that a historically situated conceptual vocabulary can acquire the appearance of universal epistemic neutrality.

Mollema's taxonomy of AI-related epistemic injustice develops this concern through the concept of *generative hermeneutical erasure*. He argues that generative AI can contribute to conceptual erasure where there is a mismatch between the conceptual frameworks embedded in AI systems and those of interlocutors, particularly outside the Western contexts in which many systems have been developed. [38] This concern is also consistent with more recent analyses of AI as a mechanism that can conceal or entrench hermeneutical gaps within the data and evaluative practices used to construct large language models. This provides a useful extension of Fricker's framework: the problem is not merely that a marginalized speaker is disbelieved; it may be that the system fails to recognize the conceptual vocabulary through which that speaker understands the world. The solution is not to romanticize every non-Western epistemology or to abandon standards of evidence and rational criticism. Epistemic pluralism does not entail epistemic relativism. Rather, it requires recognition that the production of knowledge is situated and that no single conceptual vocabulary should be presumed exhaustive without argument. [ 39 ] A

[31] Regina Rini, "Deepfakes and the Epistemic Backstop," *Philosophers' Imprint* 20, no. 24 (2020): 1–4.

[32] Don Fallis, "The Epistemic Threat of Deepfakes," *Philosophy & Technology* 34 (2021): 623–625.

[33]Mark Coeckelbergh, *AI Ethics* (Cambridge, MA: MIT Press, 2020), 43–45.

[34]Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven: Yale University Press, 2021), 1–2.

[35]Fricker, *Epistemic Injustice*, 1–3.

[36]Nadja El Kassar, "Epistemische Ungerechtigkeiten in und durch Algorithmen – ein Überblick [Epistemic Injustices in Algorithms – An Overview]," *Zeitschrift für Praktische Philosophie* 9, no. 1 (2022): 279–282, https://doi.org/10.22613/zfpp/9.1.11.

[37]Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018), 1–4.

[38]Warmhold Jan Thomas Mollema, "A Taxonomy of Epistemic Injustice in the Context of AI and the Case for Generative Hermeneutical Erasure," *AI and Ethics* 5 (2025), https://doi.org/10.1007/s43681-025-00801-w. Mollema explicitly develops "generative hermeneutical erasure" to describe conceptual exclusion associated with generative AI, especially where systems impose dominant conceptual frameworks on interlocutors outside their original cultural context.

[39]Tomas McInerney, "The Algorithmic Construction of Epistemic Injustice," 2025. McInerney examines how the development and evaluation of large language models can conceal or entrench hermeneutical injustice

responsible AI system should therefore be designed and evaluated not merely for average predictive performance but also for its treatment of epistemic diversity, minority perspectives, culturally situated concepts and forms of testimony.

**7. Algorithmic Ignorance, Automation Bias and the Displacement of Epistemic Agency**

There is another form of epistemic impropriety that occurs not in the machine's output but in the user's relationship to the machine: *algorithmic ignorance*. No individual can understand every algorithm that influences contemporary life. Ignorance by itself is therefore not blameworthy. The epistemic problem arises when an individual or institution has sufficient reason and capacity to investigate a consequential algorithmic influence but nevertheless treats the output as though it were epistemically neutral. [40] Algorithmic opacity matters because the rational assessment of a source depends partly upon knowing what kind of source it is. If a recommendation is generated from historical data containing systematic bias, its output cannot be evaluated solely by examining whether the output looks plausible. If a predictive system is optimized for a particular institutional objective, its recommendations may reflect that objective. If a model is trained on incomplete data, its apparent confidence cannot compensate for the missing information. [41]

This leads directly to the problem of automation bias. Earlier research on automation bias identified the tendency of human operators to rely excessively upon automated cues. Contemporary philosophical analysis has sharpened the distinction between weaker forms of automation-induced negligence and stronger forms of epistemic deference. Jovchevski, Buijsman and Neerincx argue that strong automation bias involves an unwarranted transfer of trust to automated systems and can interfere with autonomous agency and the human duty to exercise judgment, particularly in high-stakes settings. [42]

The philosophical importance of this argument is that the problem cannot be reduced to "humans are lazy." The structure of the socio-technical system can itself distribute authority in ways that make human judgment appear secondary. A system repeatedly presented as objective, data-driven and computationally sophisticated can acquire a presumption of superiority even when its domain of reliability is narrower than its social reputation suggests. [43] The proper response is therefore neither blind trust nor reflexive distrust. What is required is *critical trust*: reliance proportionate to demonstrated competence, accompanied by the continuing possibility of correction. AI systems used in consequential settings should provide opportunities for disagreement, alternative evidence, uncertainty representation and human override. Epistemic friction is not an inefficiency to be eliminated at all costs; in some circumstances it is precisely what protects judgment. [44]

**8. Epistemic Agency, Intellectual Ownership and Human Responsibility**

Perhaps the deepest danger is not that AI will think too much for humanity but that human beings will gradually cease to think with responsibility. Human epistemic agency consists not merely in possessing information but in being capable of asking questions, evaluating reasons, revising beliefs, distinguishing evidence from rhetoric, and taking responsibility for what one accepts as true. [45] The outsourcing of these activities can produce what may be called *loss of intellectual ownership*. The expression is not intended in the legal sense of copyright or authorship. It concerns the relation between a person and the intellectual products that the person endorses. A student who receives assistance from a teacher may genuinely own an argument intellectually if the student understands, evaluates and can defend it. By contrast, a student who submits an AI-generated argument without understanding its premises or reasoning possesses the text without necessarily possessing the intellectual achievement represented by the text.

This distinction is especially important in education. If AI is used to remove every difficulty from reading, writing, reasoning and problem-solving, the student may obtain the product while losing the practice. Yet intellectual development frequently occurs through precisely those moments of difficulty in which the learner must formulate a question, encounter an objection, revise a hypothesis and reconstruct an argument. Productive intellectual friction is therefore not

through data collection, human evaluation and benchmarking practices.

[40]José Medina, *The Epistemology of Resistance*, 41–45.

[41]Virginia Eubanks, *Automating Inequality* (New York: St. Martin's Press, 2018), 1–4.

[42] Pasquale, *The Black Box Society*, 19–22.

[43] Perica Jovchevski, Stefan Buijsman, and Mark Neerincx, "What Is Wrong With Automation Bias?" *Philosophy & Technology* 39 (2026), article 84, https://doi.org/10.1007/s13347-026-01090-9. The authors distinguish weak automation bias from strong automation bias and characterize the latter as involving excessive and unwarranted epistemic deference to automated systems.

[44]Jovchevski, Buijsman, and Neerincx, "What Is Wrong With Automation Bias?" The authors argue that strong automation bias can distort the distribution of epistemic authority within socio-technical systems and interfere with autonomous judgment.

[45] Shannon Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (Oxford: Oxford University Press, 2016), 1–2.

necessarily a defect. [46] The concern can be related to *epistemic infantilization*. A person becomes epistemically infantilized when the structures surrounding inquiry encourage dependence upon conclusions without cultivating the capacity to question those conclusions. AI can contribute to such infantilization if it is consistently used as an answer-producing authority rather than as a partner in inquiry. The appropriate model is therefore not "AI knows; human receives" but "AI assists; human interrogates." This does not require romanticizing unaided human reasoning. Human beings have always used books, teachers, laboratories, libraries, calculators and other cognitive technologies.

Clark and Chalmers' extended-mind thesis reminds us that human cognition has always been technologically and environmentally scaffolded. [47] The philosophical issue is consequently not whether external assistance is legitimate. It is whether the assistance preserves the agent's capacity to understand, assess and take responsibility for the resulting belief. The most defensible principle is *comprehension before endorsement*. A person should not regard a proposition as genuinely their own merely because they can repeat it. They should be able, at least in proportion to the importance of the claim, to explain what it means, identify its grounds, recognize relevant objections and indicate why they accept it. AI can accelerate this process, but it should not make the process disappear. [48]

**9. Toward an Epistemically Responsible Human–AI Relationship**

The preceding analysis does not justify technological rejection. The appropriate response to epistemic impropriety is to establish conditions under which AI can function responsibly within human epistemic practices. The first condition is *provenance*. Users should be able, where consequential, to distinguish retrieved information from generated content, human testimony from machine synthesis, evidence from inference, and established information from speculation. [49] The second is *calibration*. The confidence expressed by a system should not exceed the evidential basis available to it. A confident linguistic style must never be treated as an independent measure of truth. The third is *contestability*. Important outputs should remain open to challenge, correction and revision. A system that cannot be meaningfully questioned should not automatically be granted epistemic authority. The fourth is *independent verification*. The greater the consequences of a claim, the stronger the requirement that it be checked against sources independent of the generating system. Repetition within a single informational ecosystem does not constitute genuine corroboration. [50] The fifth is *epistemic pluralism*. AI systems should not be evaluated merely according to whether they reproduce dominant forms of knowledge efficiently. Their epistemic quality should also be assessed by whether they can represent legitimate disagreement, culturally situated knowledge and alternative conceptual frameworks without collapsing into relativism. [51]

The sixth is *preservation of human agency*. The user should remain responsible for consequential judgments. This principle is particularly important in medicine, law, education, public administration and other domains in which errors can affect rights, dignity, livelihood or life itself. Recent work on automation bias confirms that strong deference to automated systems can interfere with autonomous judgment, making the preservation of human deliberative authority not merely an educational preference but an ethical requirement. [52] The seventh is *epistemic literacy*. The solution to AI's epistemic problems cannot be located entirely in technical design. Users require the intellectual resources to distinguish plausibility from justification, fluency from evidence, correlation from causation, authority from reliability, and information from knowledge. AI literacy must therefore become a form of epistemology in practice. [53]

## 10. CONCLUSION

The epistemic impropriety of artificial intelligence is not adequately captured by the statement that machines sometimes make mistakes. That observation is true but philosophically superficial. The deeper problem concerns the transformation of the epistemic environment in which human beings decide what to believe, whom to trust, what counts as evidence, which interpretations are intelligible, and who is recognized as a credible knower. Artificial intelligence is increasingly an epistemic technology. [54] Its systems mediate access to information, organize evidence, generate explanations, make predictions, classify persons and phenomena, and increasingly occupy positions formerly reserved for human experts. This does

[46] Kristie Dotson, "Conceptualizing Epistemic Oppression," *Social Epistemology* 28, no. 2 (2014): 115–138.
[47] Clark and Chalmers, "The Extended Mind," 7–9
[48] Clark and Chalmers, "The Extended Mind," 10.
[49] Alvin I. Goldman, *Knowledge in a Social World*, 189–193.
[50] Floridi and Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences," 681–684.
[51] Ballantyne, Celniker, and Dunning, "Do Your Own Research," *Social Epistemology* 38 (2024): 302–304.
[52] Medina, *The Epistemology of Resistance*, 25–28.
[53] Jovchevski, Buijsman, and Neerincx, "What Is Wrong With Automation Bias." The authors explicitly connect strong automation bias with interference in autonomous agency and with the human duty to exercise judgment in high-stakes contexts.
[54] Coeckelbergh, *AI Ethics*, 103–105.

not establish that AI systems are epistemic agents in the full sense. Nor does it establish that they are mere passive tools. Their proper philosophical status lies somewhere within a complex socio-technical relationship in which machines can perform epistemically significant operations while human beings remain responsible for the norms governing their use. [55] The central epistemic impropriety occurs when the distinction between appearance and warrant becomes blurred. A plausible statement is mistaken for a justified statement; a fluent explanation for an understood explanation; a comprehensive-looking answer for a comprehensive account; computational output for objective authority; synthetic testimony for witnessed testimony; and repeated machine-generated language for independent corroboration. The danger is intensified when these confusions become institutionalized. [56]

Epistemic injustice adds another dimension. AI systems may reproduce existing credibility hierarchies, marginalize certain forms of testimony, encode dominant conceptual frameworks and contribute to hermeneutical exclusion. Fricker's account of testimonial and hermeneutical injustice remains indispensable here, but recent work demonstrates that algorithmic and generative systems require further conceptual development. [ 57 ] AI can participate in epistemic injustice not because a machine possesses human prejudice in precisely the same way, but because socio-technical systems can encode, amplify and normalize patterns of epistemic exclusion. [ 58 ] The issue of epistemic agency is equally decisive. Human beings may become dependent upon AI without becoming better knowers. Dependence is not itself a defect; human knowledge is profoundly dependent upon testimony, institutions, experts and technological artefacts. The problem arises when dependence becomes deference and when deference becomes surrender of judgment. The emerging literature on artificial epistemic authority and automation bias demonstrates why this distinction matters. [59]

Accordingly, the epistemic propriety of AI must be judged not merely by output accuracy but by the quality of the epistemic relationships that AI establishes or transforms. An epistemically responsible system should be appropriately reliable, transparent enough for its purpose, honest about uncertainty, clear about provenance, open to correction, sensitive to epistemic diversity, resistant to manipulative persuasion, and subordinate to justified rather than merely presumed authority. The ultimate philosophical question is therefore not whether artificial intelligence can answer questions. It plainly can answer many questions, sometimes with remarkable competence. The more important question is whether its participation in inquiry leaves human beings capable of understanding why an answer should be accepted, when it should be rejected, what evidence supports it, whose perspective it represents, what assumptions structure it, and what remains unknown. The proper relationship between humanity and artificial intelligence is consequently neither one of domination nor one of surrender. AI should function as an epistemic instrument and collaborator within a human practice of responsible inquiry, while the final responsibility for belief, judgment and action remains with human agents and the institutions to which they belong.

## BIBLIOGRAPHY


- Ahlstrom-Vij, Kristoffer. *Constructing the Epistemic Landscape: Aesthetics and Epistemic Norms*. Oxford: Oxford University Press, 2013.
- Alvarado, Ramón. "AI as an Epistemic Technology." *Science and Engineering Ethics* 29, no. 5 (2023): Article 32. https://doi.org/10.1007/s11948-023-00451-3.
- Arora, Caterina, *et al.,* "Experts or Authorities? The Strange Case of the Presumed Epistemic Superiority of Artificial Intelligence Systems." *Minds and Machines* (2024).
- Ballantyne, Nathan, Joseph B. Celniker, and David Dunning. "Do Your Own Research." *Social Epistemology* 38 (2024): 302–317.
- Bender, Emily M. "Resisting Dehumanization in the Age of AI." *Journal of Social Computing* 4, no. 1 (2023): 3–13.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021): 610–623. https://doi.org/10.1145/3442188.3445922.
- Clark, Andy, and David J. Chalmers. "The Extended Mind." *Analysis* 58, no. 1 (1998): 7–19. https://doi.org/10.1111/1467-8284.00096.


[55]Alvarado, "AI as an Epistemic Technology," *Science and Engineering Ethics* 29, no. 5 (2023), article 32, https://doi.org/10.1007/s11948-023-00451-3

[56]Hauswald, "AI and the Philosophy of Expertise and Epistemic Authority." Hauswald argues that AI will increasingly form part of the division of epistemic labour while generating new problems for epistemological theories of expertise and authority.

[57] Floridi and Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences"; Bender *et al.,* "On the Dangers of Stochastic Parrots."

[58] Fricker, *Epistemic Injustice*, 1–4; El Kassar, "Epistemische Ungerechtigkeiten in und durch Algorithmen"; Mollema, "A Taxonomy of Epistemic Injustice."

[59]Mollema, "A Taxonomy of Epistemic Injustice"; see also McInerney, "The Algorithmic Construction of Epistemic Injustice."

- Coady, C. A. J. *Testimony: A Philosophical Study*. Oxford: Clarendon Press, 1992.
- Coeckelbergh, Mark. *AI Ethics*. Cambridge, MA: MIT Press, 2020.
- Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.
- Dotson, Kristie. "Conceptualizing Epistemic Oppression." *Social Epistemology* 28, no. 2 (2014): 115–138.
- El Kassar, Nadja. "Epistemische Ungerechtigkeiten in und durch Algorithmen – ein Überblick [Epistemic Injustices in Algorithms – An Overview]." *Zeitschrift für Praktische Philosophie* 9, no. 1 (2022): 279–304. https://doi.org/10.22613/zfpp/9.1.11.
- Eubanks, Virginia. *Automating Inequality*. New York: St. Martin's Press, 2018.
- Fallis, Don. "The Epistemic Threat of Deepfakes." *Philosophy & Technology* 34 (2021): 623–643.
- Floridi, Luciano. *The Philosophy of Information*. Oxford: Oxford University Press, 2011.
- Floridi, Luciano, and Massimo Chiriatti. "GPT-3: Its Nature, Scope, Limits, and Consequences." *Minds and Machines* 30 (2020): 681–694.
- Floridi, Luciano, and J. W. Sanders. "On the Morality of Artificial Agents." *Minds and Machines* 14 (2004): 349–379.
- Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press, 2007.
- Fricker, Miranda. "Epistemic Injustice and the Ethics of Knowing." In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus Jr. New York: Routledge, 2017.
- Gettier, Edmund L. "Is Justified True Belief Knowledge?" *Analysis* 23, no. 6 (1963): 121–123.
- Goldman, Alvin I. *Experts: Which Ones Should You Trust?* Oxford: Oxford University Press, 2001.
- Goldman, Alvin I. *Knowledge in a Social World*. Oxford: Oxford University Press, 1999.
- Hauswald, Rico. "AI and the Philosophy of Expertise and Epistemic Authority." In *A Companion to Applied Philosophy of AI*, edited by Martin Hähnel and Regina Müller. Hoboken, NJ: Wiley, 2025. https://doi.org/10.1002/9781394238651.ch5.
- Hauswald, Rico. "Artificial Epistemic Authorities." *Social Epistemology* 39, no. 6 (2025): 716–725. https://doi.org/10.1080/02691728.2025.2449602.
- Jovchevski, Perica, Stefan Buijsman, and Mark Neerincx. "What Is Wrong With Automation Bias?" *Philosophy & Technology* 39 (2026): Article 84. https://doi.org/10.1007/s13347-026-01090-9.
- Lackey, Jennifer. *Learning from Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press, 2008.
- McInerney, Tomas. "The Algorithmic Construction of Epistemic Injustice." 2025.
- Medina, José. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford: Oxford University Press, 2013.
- Mollema, Warmhold Jan Thomas. "A Taxonomy of Epistemic Injustice in the Context of AI and the Case for Generative Hermeneutical Erasure." *AI and Ethics* 5 (2025). https://doi.org/10.1007/s43681-025-00801-w.
- Nguyen, C. Thi. "Echo Chambers and Epistemic Bubbles." *Episteme* 17, no. 2 (2020): 141–161. https://doi.org/10.1017/epi.2018.32.
- Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press, 2018.
- O'Neil, Cathy. *Weapons of Math Destruction*. New York: Crown, 2016.
- Pasquale, Frank. *The Black Box Society: The Secret al., gorithms That Control Money and Information*. Cambridge, MA: Harvard University Press, 2015.
- Rini, Regina. "Deepfakes and the Epistemic Backstop." *Philosophers' Imprint* 20, no. 24 (2020): 1–16.
- Vallor, Shannon. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press, 2016.