Jurnal Dekonstruksi
Online ISSN: 2797-233X
Volume 12, Nomor 3, Juli (2026)

Epistemmic Responsibility in the Age of Generative Artificial Intelligence: Trust, Authority, and the Ethics of Knowing

Gabriel Abdi Susanto
STF Driyarkara
abdisusanto@yahoo.com

Abstract

The rapid integration of generative artificial intelligence into knowledge-sensitive domains has prompted extensive ethical debate, largely focused on issues of accuracy, bias, and governance. This article argues that such approaches overlook a more fundamental normative challenge: the transformation of epistemic responsibility and authority in contemporary practices of knowing. Drawing on epistemic ethics and virtue epistemology, the article contends that generative AI systems lack epistemic agency and therefore cannot bear responsibility for knowledge claims. Through normative conceptual analysis, the article identifies a structural "responsibility gap" that emerges when epistemic reliance is conflated with trust. Furthermore, this article provides a theoretical in-depth analysis of human-AI assemblages and proposes a practical implementation framework for higher education and journalism. The framework emphasizes accountable agency, institutional responsibility, and the cultivation of epistemic virtues as necessary conditions for integrating generative AI without undermining democratic epistemic practices.

Keywords*: Epistemic Responsibility; Generative Artificial Intelligence; Epistemic Trust; Virtue Epistemology; Public Reason.*

1. Introduction

The rapid diffusion of generative artificial intelligence into domains of knowledge production marks a significant transformation in contemporary epistemic life. Systems capable of producing fluent text, plausible explanations, and context-sensitive responses are increasingly embedded in practices of research, education, journalism, administration, and public communication. Much of the ethical debate surrounding these technologies has focused on questions of accuracy, bias, transparency, and control (Mittelstadt et al., 2016; Floridi, 2023). While these concerns are undeniably important, they risk obscuring a more fundamental issue: the transformation of epistemic responsibility and authority in environments where knowledge-like outputs are generated by systems that lack epistemic agency.

Epistemology has traditionally understood knowledge not merely as a relation between beliefs and facts, but as an activity situated

within normative practices of inquiry, justification, and testimony (Dancy, 1985). To know is to participate in social practices governed by expectations of responsibility, answerability, and responsiveness to reasons. Epistemic authority, on this view, is not reducible to reliability or performance alone; it is grounded in the capacity of agents to stand behind their claims, to revise them in light of critique, and to be held accountable for epistemic error. These normative assumptions underpin epistemic trust, testimonial exchange, and the possibility of collective inquiry.

Generative artificial intelligence unsettles these assumptions. Although such systems can generate outputs that resemble the products of human cognition, they do so without understanding, judgment, or commitment to truth. Their apparent epistemic competence is derivative, grounded in statistical correlations across vast datasets rather than in the exercise of epistemic agency. Yet in practice, AI-generated outputs are often treated as epistemically authoritative. They are consulted, cited, and relied upon in ways that approximate trust, even as the systems themselves remain incapable of bearing responsibility for the epistemic consequences of their use.

This development raises a distinctively epistemic-ethical problem. When authority is accorded to outputs that are not anchored in accountable agency, responsibility for knowledge claims becomes diffuse and opaque. Decisions informed by AI-generated content may be difficult to contest, errors may be hard to attribute, and epistemic harms may lack clear sites of accountability. The result is what can be described as a responsibility gap: a structural condition in which epistemic outcomes are produced and circulated without clearly identifiable epistemic agents who can be held answerable for them.

The ethical significance of this gap extends beyond individual cases of misinformation or error. Epistemic trust, as a normative relation between agents, presupposes commitments to truthfulness, responsiveness to reasons, and accountability (Simpson, 2012). When reliance on system performance is conflated with trust, the moral infrastructure that sustains testimonial practices and shared inquiry is weakened. Moreover, because epistemic practices are embedded in relations of social power, the normalization of machine-mediated authority risks amplifying existing forms of epistemic injustice, particularly when AI-generated content is perceived as neutral or objective.

These concerns are especially pressing in the context of democratic public discourse (Habermas, 1996; Rawls, 1993). Deliberative democratic theory emphasizes that legitimate collective decision-making depends on practices of public reason in which claims can be justified, contested, and revised. Such practices presuppose interlocutors who can give reasons and respond to critique. Generative AI, however, cannot participate in this dialogical accountability. When AI-generated content enters public deliberation without clear attribution or normative framing, it may distort the conditions under which democratic legitimacy is formed, not through coercion or censorship, but through the subtle erosion of epistemic responsibility.

Existing scholarship on AI ethics has primarily focused on issues such as algorithmic bias, transparency, fairness, explainability, and governance (Mittelstadt et al., 2016; Coeckelbergh, 2020; Floridi, 2023). Other studies have examined the implications of artificial intelligence for democratic legitimacy, public discourse, and institutional accountability (Binns, 2018; Taddeo & Floridi, 2018). Within epistemology, discussions concerning epistemic injustice, testimonial authority,

and public reason have also provided important resources for evaluating technologically mediated knowledge practices (Fricker, 2007; Habermas, 1996). However, comparatively limited attention has been given to the specifically epistemic-ethical problem concerning the erosion of epistemic responsibility when generative AI systems are treated as epistemically authoritative despite lacking epistemic agency.

This article argues that these challenges cannot be adequately addressed through technical solutions or regulatory compliance alone. While transparency, explainability, and bias mitigation are necessary, they do not engage the normative foundations of knowledge practices. What is required is an epistemic-ethical analysis that foregrounds responsibility, agency, and trust as central evaluative criteria. By drawing on epistemic ethics and virtue epistemology, and by bringing these approaches into dialogue with critical theory and deliberative democratic theory, this article develops a normative framework for assessing the epistemic implications of generative artificial intelligence.

Methodologically, the article proceeds through normative conceptual analysis and critical reconstruction of key epistemic concepts, including epistemic responsibility, epistemic trust, testimony, and public reason. Rather than asking whether generative AI produces accurate information, the analysis asks how its integration reshapes the ethical conditions under which knowledge is recognized as legitimate. In doing so, the article seeks to clarify what is at stake epistemically in the rise of system-generated authority and to articulate the conditions under which generative AI might be integrated without eroding the moral architecture of knowing.

The structure of the article is as follows. The next section situates epistemic responsibility within epistemic ethics and virtue epistemology, establishing the normative background against which generative AI is evaluated. The discussion section then analyzes the emergence of a responsibility gap and its implications for epistemic trust, testimony, epistemic injustice, and democratic public reason. The final section synthesizes these insights and outlines an epistemic-ethical orientation toward generative AI that emphasizes accountable agency, institutional responsibility, and the preservation of epistemic trust as a condition of democratic life.

This article makes a distinctively epistemic-ethical contribution to contemporary debates on generative artificial intelligence by reframing the problem of responsibility beyond prevailing concerns with accuracy, bias, or governance. While much of the existing literature in AI ethics addresses accountability in terms of technical reliability, regulatory compliance, or institutional oversight, this article argues that such approaches overlook a more fundamental normative disruption: the erosion of epistemic responsibility and authority in practices of knowing. By drawing a systematic distinction between epistemic trust and epistemic reliance, and by situating generative AI within traditions of epistemic ethics and virtue epistemology, the article conceptualizes the responsibility gap not as a legal or technical deficit, but as a structural epistemic condition arising from the attribution of authority to systems that lack epistemic agency. This reframing clarifies why transparency or explainability alone cannot resolve the ethical stakes of generative AI and highlights the preservation of accountable agency as a necessary condition for sustaining public reason and democratic epistemic practices.

2. Methodology

This article employs a qualitative, normative, and conceptual methodology grounded in philosophical analysis. Rather than relying

on empirical data or experimental design, the study proceeds through critical examination of concepts, normative frameworks, and theoretical arguments relevant to contemporary epistemic practices. This methodological choice reflects the nature of the research question, which concerns not the empirical performance of generative artificial intelligence, but its ethical and epistemic implications for responsibility, trust, and democratic knowledge practices.

The primary method used is normative conceptual analysis. This involves clarifying and critically examining key concepts such as epistemic responsibility, epistemic trust, epistemic reliance, agency, testimony, and public reason. These concepts are not treated as fixed or purely descriptive, but as normatively loaded terms whose meanings are shaped by moral expectations and social practices. By analyzing how these concepts function within established traditions of epistemology and ethics, the article evaluates the extent to which generative AI can meaningfully be situated within existing epistemic norms.

In conducting this analysis, the article adopts an interpretive and reconstructive approach to established philosophical theories. Epistemic ethics and virtue epistemology provide the primary normative framework, particularly through their emphasis on responsible agency, intellectual virtues, and accountability in knowledge practices (Vallor, 2016). These approaches are brought into dialogue with critical theory and deliberative democratic theory, especially in relation to the concepts of epistemic injustice and public reason (Habermas, 1996). Rather than applying these theories deductively, the article reconstructs their core normative commitments and assesses their relevance in the context of technologically mediated epistemic environments.

A central methodological strategy of the article is comparative normative evaluation. Generative AI is analyzed by contrast with paradigmatic cases of human epistemic agency, testimonial practices, and institutional knowledge production. This comparison is not intended to anthropomorphize AI systems, but to illuminate the normative asymmetries between human agents and artificial systems. By highlighting these asymmetries, the analysis identifies the conditions under which epistemic authority is traditionally justified and examines how these conditions are disrupted when AI-generated outputs are treated as authoritative.

The article also employs a critical diagnostic method, characteristic of critical theory, to examine how generative AI interacts with existing social and institutional structures. This involves identifying structural tendencies—such as the diffusion of responsibility, the privileging of efficiency over accountability, and the normalization of machine-mediated credibility—that shape contemporary epistemic practices. The aim of this diagnostic approach is not to assign blame, but to reveal how epistemic norms may be subtly reconfigured through technological integration, often without explicit normative reflection.

Importantly, the methodology is non-instrumental and non-prescriptive in a narrow policy sense. While the article draws normative conclusions, it does not propose technical design solutions or regulatory mechanisms in detail. Instead, it seeks to articulate the ethical conditions under which such measures could be meaningfully evaluated. In this respect, the methodological orientation is foundational rather than applied: it aims to clarify what is at stake epistemically before questions of implementation or governance are addressed.

Finally, reflexivity constitutes an implicit methodological commitment throughout the

analysis. The article treats epistemic responsibility not only as an object of inquiry but as a guiding norm for philosophical practice itself. Claims are advanced with explicit acknowledgment of their normative standpoint, and conceptual distinctions are justified through reasoned argument rather than assumed consensus. This reflexive orientation aligns with the article's broader thesis that epistemic ethics remains indispensable in evaluating emerging forms of knowledge production, particularly in contexts where responsibility and authority risk becoming diffuse.

Philosophical rigor is ensured through systematic conceptual clarification, engagement with established epistemic traditions, and internal coherence of normative argumentation.

3. Discussion

3.1 Epistemic Responsibility, Trust, and the Moral Architecture of Knowledge in the Age of Generative Artificial Intelligence

Epistemic ethics concerns the normative dimensions of belief, inquiry, and knowledge (Dancy, 1985; Fricker, 2007). Rather than treating knowledge as a purely descriptive achievement—defined solely in terms of accuracy, reliability, or successful representation—epistemic ethics emphasizes the moral significance of epistemic practices themselves. To know is not merely to hold a true belief; it is to participate in practices of inquiry, justification, and communication that are governed by norms of responsibility, accountability, and respect for others as knowers. From this perspective, epistemic evaluation extends beyond the truth or falsity of beliefs to include the conditions under which those beliefs are formed, sustained, revised, and communicated.

Within this framework, epistemic responsibility occupies a central role. Epistemically responsible agents are not simply those who happen to arrive at correct conclusions, but those who are responsive to evidence, attentive to counterarguments, open to revision, and willing to stand behind their epistemic commitments. Responsibility here is irreducibly normative: it presupposes an agent who can recognize reasons as reasons, who can be addressed through critique, and who can be held accountable for epistemic failure. Epistemic practices are thus inseparable from agency, and epistemic norms are inseparable from moral expectations about how agents ought to conduct themselves in relation to truth and to one another.

Virtue epistemology articulates this insight by framing epistemic responsibility in terms of intellectual virtues (Vallor, 2016). On this view, knowledge is not merely a product but an achievement of character. Intellectual virtues such as intellectual humility, conscientiousness, honesty, and courage structure the way agents engage with evidence and with others in epistemic communities. These virtues presuppose capacities for reflection, self-correction, and moral accountability. To know responsibly, therefore, is not simply to generate true beliefs, but to inhabit epistemic practices that respect norms of justification, communicative integrity, and mutual recognition. Knowledge, understood in this way, is a fundamentally agential and social achievement.

Generative artificial intelligence poses a profound challenge to this normative picture. Although such systems are capable of producing outputs that resemble the products of human cognition—coherent arguments, plausible explanations, and context-sensitive responses—they do not possess epistemic agency in the relevant sense. They do not assess reasons as reasons, commit themselves to truth, or recognize their own fallibility. Their apparent epistemic competence is derivative, grounded in statistical pattern recognition across vast *corpora* of human-

generated data rather than in the exercise of judgment or understanding. The appearance of knowing, in this case, is not accompanied by the normative capacities that make knowing an ethically significant activity.

This disjunction gives rise to what may be described as a responsibility gap. When generative AI systems are deployed in epistemically sensitive domains—education, journalism, public administration, healthcare, or democratic deliberation—their outputs often inform decisions, shape interpretations, and influence public discourse. Yet responsibility for these epistemic outcomes cannot be straightforwardly attributed to the systems themselves. Instead, responsibility is dispersed across developers, data curators, platform designers, institutional adopters, and end-users. This dispersion frequently obscures accountability rather than distributing it transparently. The result is not merely practical confusion about who is to blame for error, but a structural weakening of the norms that govern responsible epistemic practices.

The emergence of this responsibility gap has direct consequences for epistemic trust. Within epistemic ethics, trust is not reducible to confidence in the reliability or efficiency of an information source. Epistemic trust is a normative relation between agents. To trust someone epistemically is to regard them as committed to truth, responsive to reasons, and answerable for error. Trust involves expectations of sincerity, competence, and accountability, and it enables practices of testimony, cooperation, and collective inquiry. These practices form the moral infrastructure of shared knowledge.

Generative AI complicates this infrastructure by producing outputs that appear trustworthy while lacking the normative commitments that ground epistemic trust. Users may rely on AI-generated content because it is fluent, coherent, and contextually appropriate, even when they are aware that the system does not "understand" in a human sense. This reliance represents a shift from trust in persons to reliance on systems. While reliance may be pragmatically useful, it lacks the reciprocal accountability that characterizes trust. The ethical danger lies in conflating the two—treating reliance as though it were trust and allowing functional performance to substitute for normative responsibility.

The distinction between epistemic reliance and epistemic trust thus becomes crucial. As Simpson (2012) argues, reliance concerns functional dependability, whereas trust presupposes accountability and answerability. In this light, reliance refers to a pragmatic dependence on the functional reliability of a system; one relies on a calculator, a navigation system, or a spell checker without attributing to it moral or epistemic agency.

Trust, by contrast, presupposes a normative relationship grounded in accountability and responsiveness to reasons. To trust is to assume that the trusted party can be called to account for error or misjudgement. When this distinction collapses, the ethical core of epistemic practices is hollowed out. What remains is efficiency without answerability, output without ownership, and credibility without responsibility.

Generative AI actively encourages such a collapse. Its outputs are often treated as epistemically authoritative because they are statistically reliable or contextually appropriate. Yet this treatment obscures the fact that AI systems cannot participate in the normative practices that sustain epistemic trust. They cannot recognize criticism as criticism, revise commitments in light of reasons, or acknowledge responsibility for epistemic harm. What appears as epistemic trust is therefore a technologically mediated reliance that lacks moral reciprocity. Over time, this shift subtly reshapes expectations

about what knowledge is and how it should function, replacing accountability with performance as the primary epistemic value.

This transformation has significant implications for the structure of testimony. Traditionally, testimonial knowledge involves a speaker who presents a claim and assumes responsibility for its truth. Even in cases of institutional or expert testimony, epistemic authority ultimately rests on chains of human accountability. Someone can be questioned, challenged, corrected, or sanctioned. When generative AI functions as a source of testimonial content, these chains become opaque. The question of who speaks is displaced by the question of how the system performs. This displacement weakens the normative conditions under which testimony can function as a reliable source of knowledge, not because the content is necessarily false, but because responsibility for its truth is no longer clearly situated.

From the perspective of epistemic injustice, the implications of this shift are particularly troubling. As Miranda Fricker has argued, epistemic practices are embedded in social relations of power, and failures of epistemic justice occur when individuals or groups are unfairly discredited or marginalized as knowers (Fricker, 2007). Generative AI systems, trained on historically sedimented data, risk reproducing and amplifying these injustices while presenting themselves as neutral or objective. When AI-generated content is accorded default credibility, existing asymmetries of epistemic authority may be intensified. Marginalized voices risk being further displaced, not through overt silencing, but through the quiet normalization of machine-mediated credibility.

This form of epistemic marginalization is especially insidious because it operates under the guise of neutrality. The apparent objectivity of AI outputs can mask the ways in which historically embedded biases are reproduced at scale. Moreover, because responsibility for epistemic harm is dispersed across technical and institutional actors, it becomes difficult for affected individuals or groups to contest or rectify injustice. Epistemic injustice, in this context, is not merely a matter of individual prejudice, but a structural feature of technologically mediated epistemic environments.

At this point, the argument moves from individual epistemic responsibility to its implications for public and democratic epistemic practices. The widespread epistemic reliance on generative AI does not merely affect individual knowers, but reconfigures the conditions under which public justification and collective reasoning take place.

These concerns become particularly acute in the context of public discourse and democratic deliberation. Democratic societies depend on a shared epistemic space in which reasons can be exchanged, contested, and revised. Public reason, as articulated in deliberative democratic theory, presupposes not only access to information, but participation in communicative practices governed by norms of sincerity, justification, and mutual recognition. These norms are inherently dialogical: they assume interlocutors who can be challenged and who can respond.

Generative AI, however, cannot participate in such dialogical accountability. Although it can simulate argumentative structures and respond to prompts, it does not stand behind its claims in the manner required by public reason. It cannot offer reasons as reasons, nor can it engage in the reciprocal practices of justification and critique that underpin democratic legitimacy. When AI-generated content enters public deliberation without clear attribution or contextualization, it risks distorting the communicative conditions under which democratic legitimacy is formed. What appears as an expansion of

informational resources may in fact be a contraction of epistemic responsibility.

From a Habermasian perspective, this distortion can be understood as a colonization of communicative rationality by system-level processes (Habermas, 1996). Generative AI operates according to instrumental rationality, optimizing outputs based on probabilistic correlations rather than reasons that can be publicly justified. When such outputs are integrated into communicative spaces without normative safeguards, they may subtly reshape expectations about what counts as a reason, who counts as a speaker, and how disagreement is resolved. The danger is not the overt suppression of deliberation, but its gradual reconfiguration in ways that privilege efficiency, fluency, and plausibility over accountability and justification.

At the same time, it would be misleading to frame generative AI solely as an external threat to democratic epistemic life. These systems are developed, deployed, and legitimated within existing social and institutional arrangements. The displacement of epistemic responsibility is therefore not a technological inevitability, but a reflection of broader tendencies to outsource judgment, accelerate decision-making, and defer accountability. In this sense, generative AI functions as a magnifying lens, revealing latent weaknesses in contemporary epistemic cultures—weaknesses that predate AI but are intensified by it.

An epistemic-ethical response must therefore move beyond calls for technical fixes or individual vigilance. While transparency, explainability, and bias mitigation are important, they do not address the normative core of the problem. What is required is a rearticulation of epistemic responsibility that acknowledges the collective and institutional dimensions of knowledge production in technologically mediated environments. Responsibility must be understood not only as an individual virtue, but as a social obligation embedded in practices, institutions, and norms.

Institutions that deploy generative AI in knowledge-sensitive domains bear a particular burden of epistemic responsibility. This includes responsibility for clarifying the role of AI in epistemic processes, for establishing mechanisms of contestation and correction, and for ensuring that human judgment remains central where normative evaluation is required. Such measures are not merely procedural safeguards; they express a commitment to preserving the ethical conditions under which knowledge can function as a public good.

At the level of individual practice, epistemic responsibility entails cultivating critical awareness of the limits of generative AI. Users must resist the temptation to treat AI outputs as epistemically self-sufficient. Instead, AI-generated content should be approached as a provisional resource, subject to evaluation, contextualization, and revision by human agents. This orientation aligns with virtue epistemology's emphasis on intellectual humility and attentiveness to one's epistemic environment. The ethical task is not to reject AI assistance, but to integrate it in ways that reinforce, rather than undermine, responsible epistemic agency.

Virtue epistemology offers important resources for understanding what is at stake in this integration. Intellectual virtues are cultivated within epistemic environments that reward certain forms of engagement over others. If epistemic environments prioritize speed, fluency, and output volume over reflection, justification, and dialogue, intellectual virtues may be systematically discouraged. Conversely, when AI is integrated in ways that foreground human judgment and critical engagement, it may support more reflective epistemic practices. The ethical question is therefore not

whether AI can generate knowledge-like outputs, but whether its integration sustains the virtues required for responsible knowing.

Bernard Williams' analysis of truthfulness further illuminates this point. Williams distinguishes between sincerity and accuracy as the two fundamental virtues that sustain practices of truth. Generative AI complicates both. While such systems can be optimized for accuracy in a statistical sense, they lack sincerity, understood as a commitment to truth grounded in agency. When epistemic practices prioritize accuracy detached from sincerity, the moral basis of truthfulness is weakened. Truth risks becoming a technical achievement rather than a shared ethical commitment.

This shift has long-term consequences for trust in institutions. Democratic institutions rely on epistemic legitimacy, which depends not only on producing correct outcomes, but on being perceived as answerable to those affected by their decisions. When institutional knowledge practices rely heavily on AI systems whose operations are not normatively transparent, legitimacy may erode even in the absence of overt error. Citizens may come to experience knowledge claims as imposed rather than justified, fueling skepticism, disengagement, and epistemic cynicism.

An epistemic-ethical framework thus calls for a recalibration of responsibility across human–AI assemblages. Rather than attempting to assign responsibility to AI systems themselves, such a framework emphasizes the obligation of human agents and institutions to remain epistemically answerable. This includes making explicit the role of AI in knowledge production, preserving spaces for contestation, and resisting the temptation to treat algorithmic outputs as epistemically self-authenticating.

By situating generative AI within a broader moral economy of knowledge, epistemic ethics resists the reduction of ethical concerns to matters of efficiency or compliance. It insists that the value of knowledge lies not only in its utility, but in the ethical practices through which it is produced and shared. In the absence of such practices, the promise of generative AI risks being overshadowed by a quiet but profound erosion of epistemic responsibility.

a. Theoretical In-depth Analysis: Recalibrating Responsibility within Human-AI Assemblages

This article does not advocate the rejection of generative AI from epistemic practices. Rather, it argues that generative AI should be integrated within a normative framework that preserves human epistemic agency and accountability. The central position defended here is that AI may function as an epistemic aid, but never as an autonomous bearer of epistemic authority. Human agents and institutions must therefore remain the ultimate locus of epistemic responsibility.

At this stage, the discussion deepens the normative implications of the preceding analysis by clarifying the conditions under which epistemic responsibility and authority can be meaningfully preserved in AI-mediated practices.

The integration of generative artificial intelligence into knowledge practices demands a profound redefinition of our conventional understanding of epistemic responsibility. The fundamental challenge that arises is not merely the potential for misinformation, but the creation of a structural responsibility gap. Because generative AI systems operate through statistical correlations without possessing consciousness, will, or a commitment to truth, they lack epistemic agency. Consequently, AI cannot bear the moral burden for the knowledge claims it produces, even though these claims often appear fluent and authoritative.

In this context, it is crucial to distinguish between epistemic reliance and epistemic

trust. Reliance is a form of pragmatic relationship regarding the functional reliability of a tool—similar to how we rely on a calculator. In contrast, trust is a normative relationship between agents that assumes accountability and responsiveness to reasons. The greatest ethical danger arises when society begins to blur this boundary, treating system efficiency as a substitute for moral responsibility.

Furthermore, the use of AI without responsible oversight risks exacerbating epistemic injustice. By basing its output on historical data that may be biased, AI can normalize the marginalization of certain voices under the guise of 'machine objectivity'. In a democratic public sphere, this threatens public reason because AI cannot participate in the dialogical accountability necessary to form collective legitimacy.

This article does not advocate the rejection of generative AI from epistemic practices. Rather, it argues that generative AI should be integrated within a normative framework that preserves human epistemic agency and accountability. The central position defended here is that AI may function as an epistemic aid, but never as an autonomous bearer of epistemic authority. Human agents and institutions must therefore remain the ultimate locus of epistemic responsibility.

As a mitigation strategy, this article proposes a reorientation that emphasizes:

- Accountable Agency: Epistemic responsibility must remain centered on the human subjects and institutions that deploy AI systems, ensuring that every claim remains contestable and correctable.
- Cultivation of Epistemic Virtues: Users must develop intellectual virtues such as humility and attentiveness to avoid treating AI as a self-authenticating authority.
- Institutional Integrity: Institutions must ensure that AI is used as a provisional resource, rather than a final authority that replaces human judgment in normatively sensitive domains."

b. Practical Implementation Framework: Higher Education and Journalism

The preceding theoretical analysis provides the normative grounding for considering how epistemic responsibility can be institutionally and practically sustained in contexts increasingly shaped by generative AI.

To operationalize the proposed epistemic-ethical framework, this article outlines practical implementations for two critical knowledge-sensitive sectors: higher education and journalism. These measures aim to bridge the "responsibility gap" by shifting from mere functional reliance to authentic epistemic trust.

Higher education and journalism are selected because both sectors function as foundational epistemic institutions within democratic societies. Universities are responsible for the cultivation of intellectual virtues and the production of legitimate knowledge, while journalism mediates public reason and testimonial circulation in the public sphere. These domains are therefore especially vulnerable to the erosion of epistemic responsibility caused by uncritical reliance on generative AI.

3.3.1. Higher Education Institutions In the educational context, AI integration must support, rather than replace, the formation of epistemic virtues:

- Human-in-the-Loop Assessment Policies: Institutions should mandate that AI-based evaluations remain provisional, requiring final human verification to ensure "accountable agency."

- Virtue Epistemology Curriculum: Education must go beyond technical AI literacy to cultivate "intellectual humility" and "attentiveness," training students to treat AI as a resource that requires contextualization rather than a self-validating authority.
- Mandatory Attribution: Explicit disclosure of AI's role in scholarly production is essential to maintain "communicative integrity" and "public reason" within the academic environment.

3.3.2. Journalism and Media Organizations In journalism, the challenge is to prevent the "normalization of machine-mediated credibility" from obscuring professional accountability:

- Verification Protocols and Accountability Chains: Human editors must be established as the primary "epistemic agents" responsible for factual claims, even when data is synthesized by AI.
- Bias Audits for Epistemic Justice: Media institutions must conduct regular audits of AI systems to ensure algorithms do not reinforce "epistemic injustice" or marginalize minority voices under the guise of "machine objectivity."
- Epistemic Authority Labeling: Clear labeling of AI-generated content is required to distinguish between performative outputs and information backed by a human journalist's "commitment to truth."

4. Conclusion

Generative artificial intelligence marks a decisive turning point in contemporary epistemic life. Its ability to produce fluent, coherent, and context-sensitive outputs challenges long-standing assumptions about how knowledge is generated, who may legitimately claim epistemic authority, and where responsibility for epistemic outcomes ought to reside. As this article has argued, the ethical significance of this transformation is not exhausted by concerns about accuracy, misinformation, or technical bias. At its core, the challenge posed by generative AI is epistemic and normative: it concerns the displacement and diffusion of epistemic responsibility that accompanies the emergence of system-generated authority.

By situating generative AI within the framework of epistemic ethics and virtue epistemology, and by bringing these perspectives into dialogue with critical theory and deliberative democratic theory, this article has shown that responsible knowing presupposes accountable agency. Epistemic trust, testimonial authority, and public reason depend on agents who can recognize reasons as reasons, respond to critique, and be held answerable for error. Generative AI systems, however sophisticated their outputs may appear, lack these normative capacities. They can assist inquiry, augment informational access, and support certain epistemic tasks, but they cannot themselves bear epistemic responsibility.

The ethical risk arises when AI-generated outputs are treated as epistemically authoritative without adequate contextualization or attribution of responsibility. In such cases, reliance on system performance is tacitly transformed into a form of trust that lacks moral reciprocity. This confusion between epistemic reliance and epistemic trust weakens the normative foundations of knowledge practices, obscures chains of accountability, and threatens the conditions under which testimony, justification, and democratic deliberation can function responsibly. The resulting responsibility gap is not merely a technical or organizational problem, but a structural feature of emerging epistemic environments.

Addressing this challenge requires more than technical safeguards or appeals to individual vigilance. While transparency, explainability, and bias mitigation remain important, they do not reach the normative core of the issue. What is required is a reorientation of AI governance toward the ethical conditions of knowledge production. Such a reorientation foregrounds responsibility rather than efficiency, accountability rather than mere reliability, and justification rather than plausibility. It calls for institutional arrangements that make the role of generative AI in epistemic processes explicit, preserve spaces for contestation and correction, and ensure that human judgment remains central wherever normative evaluation is at stake.

At the same time, an epistemic-ethical response emphasizes the cultivation of intellectual virtues within technologically mediated epistemic environments. Virtues such as intellectual humility, conscientiousness, and attentiveness are not optional personal traits, but dispositions essential to sustaining responsible epistemic practices. When epistemic environments reward speed, fluency, and output volume at the expense of reflection and justification, these virtues are systematically undermined. Conversely, when generative AI is integrated as a provisional resource rather than an epistemic authority, it may support more reflective and inclusive forms of inquiry.

More fundamentally, this analysis reaffirms that truth in democratic epistemic cultures is not merely a technical achievement, but a normative one. Truth functions as a regulative ideal sustained through social practices of justification, critique, and mutual recognition. When responsibility for knowledge claims becomes diffuse and opaque, the possibility of holding anyone answerable for epistemic failure diminishes, and the social trust upon which democratic institutions depend is placed at risk.

In the age of generative artificial intelligence, the task of epistemic ethics is therefore not to resist technological innovation, nor to romanticize pre-digital epistemic ideals. Rather, it is to ensure that the moral architecture of knowing—responsibility, accountability, and trust—is not eroded in the process of technological integration. Only by preserving these normative commitments can societies incorporate generative AI in ways that reinforce, rather than undermine, the ethical and democratic conditions under which shared knowledge remains possible.

References

Binns, R. (2018). Algorithmic Accountability and Public Reason. *Philosophy & Technology*, 31(4), 543–556.

Bridges, I. (2020). *Epistemic Injustice and the Ethics of Knowing*. Oxford: Oxford University Press.

Coeckelbergh, M. (2020). *AI Ethics*. Cambridge, MA: MIT Press.

Dancy, J. (1985). *Introduction to Contemporary Epistemology*. Oxford: Blackwell.

Deleuze, G., & Guattari, F. (1987). *A Thousand Plateaus: Capitalism and Schizophrenia*. Minneapolis: University of Minnesota Press.

Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge, MA: MIT Press.

Habermas, J. (2006). Religion in the public sphere. *European Journal of Philosophy*, 14(1), 1–25.

Jasanoff, S. (2016). *The ethics of invention: Technology and the human future*. New York: W. W. Norton.

Kitcher, P. (2011). *Science in a Democratic Society*. Amherst, NY: Prometheus Books.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21.
Morozov, E. (2013). *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.
O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.
Rawls, J. (1993). *Political Liberalism*. New York: Columbia University Press.
Ricoeur, P. (1992). *Oneself as Another*. Chicago: University of Chicago Press.
Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking.
Sandel, M. J. (2020). *The Tyranny of Merit: What's Become of the Common Good?* New York: Farrar, Straus and Giroux.
Simpson, T. W. (2012). Between Trust and Reliance. *Archiv für Rechts- und Sozialphilosophie*, 98(4), 495–508.
Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, 361(6404), 751–752.
Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press.
Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.