DOI: https://doi.org/10.14393/DLv19a2025-66
e019066

Article

Agency and Epistemic Integrity in Distant Writing

Agência e Integridade Epistêmica na Escrita Distante

Tiegue Vieira RODRIGUES[*]

ABSTRACT: In this article, we examine the epistemic and ethical dimensions of "distant writing," a form of composition assisted by Artificial Intelligence (AI), in the context of philosophical and academic production. Faced with the challenge that Large Language Models (LLMs) pose to traditional notions of authorship, the text develops a positive vision for their responsible use. We argue that, rather than diminishing human agency, distant writing reconfigures the role of the author to that of an epistemic curator, whose intellectual responsibility manifests itself through a new form of cognitive work. Based on concepts from virtue epistemology, we rigorously distinguish the philosopher's intentional epistemic agency from the mere epistemic instrumentality of the machine. While the human agent operates in the normative space of reasons, responding to epistemic commitments such as truth and justification, LLMs function as instruments that facilitate cognitive processes without possessing intentional mental states. In this context, we frame curation as a demanding cognitive act that involves three dimensions of knowledge: procedural knowledge (know-how) manifested in the sophisticated design of prompts; evaluative knowledge necessary for the critical assessment of outputs; and propositional knowledge (know-that) that underpins coherent argumentative integration. The article also analyzes the epistemic risks inherent in this practice, particularly the generation of epistemic simulacra, texts that exhibit the superficial form of philosophical discourse without genuine conceptual substance. We propose that the cultivation of specific intellectual virtues, such as epistemic vigilance, intellectual humility, and a commitment to deep understanding, is indispensable for mitigating these risks. Finally, we argue that authorial integrity can no longer be governed exclusively by traditional notions of plagiarism as misappropriation, but requires a new imperative of transparency, for which we develop a gradual taxonomy of ethical disclosure that recognizes different levels of technological assistance in textual production.
KEYWORDS: Distant Writing. Epistemic Agency. Epistemic Curation. Artificial Intelligence. Authorship.

RESUMO: Neste artigo, examinamos as dimensões epistêmicas e éticas da "escrita distante", uma forma de composição assistida por Inteligência Artificial (IA), no contexto da produção filosófica e acadêmica. Perante o desafio que os Modelos de Linguagem Ampla (LLMs) colocam às noções tradicionais de autoria, o texto desenvolve uma visão positiva para o seu uso responsável. Argumentamos que, em vez de diminuir a agência humana, a escrita distante reconfigura o papel do autor para o de um agente epistêmico curador, cuja responsabilidade

[*] PhD in Philosophy from the Pontifical Catholic University (PUCRS), Adjunct Professor of Philosophy in the Department of Philosophy at the Federal University of Santa Maria (UFSM), Santa Maria, RS, Brazil. tiegue.rodrigues@gmail.com

intelectual se manifesta através de uma nova forma de trabalho cognitivo. Com base em conceitos da epistemologia da virtude, distinguimos rigorosamente a agência epistêmica intencional do filósofo da mera instrumentalidade epistêmica da máquina. Enquanto o agente humano opera no espaço normativo das razões, respondendo a compromissos epistêmicos como verdade e justificação, os LLMs funcionam como instrumentos que facilitam processos cognitivos sem possuir estados mentais intencionais. Neste contexto, enquadramos a curadoria como um ato cognitivo exigente que envolve três dimensões de conhecimento: o conhecimento processual (saber-como) manifestado no design sofisticado de prompts; o conhecimento avaliativo necessário para a avaliação crítica dos outputs; e o conhecimento proposicional (saber-que) que fundamenta a integração argumentativa coerente. O artigo analisa também os riscos epistêmicos inerentes a esta prática, particularmente a geração de simulacros epistêmicos, textos que exibem a forma superficial do discurso filosófico sem a substância conceitual genuína. Propomos que o cultivo de virtudes intelectuais específicas, como a vigilância epistêmica, a humildade intelectual e o compromisso com a compreensão profunda, é indispensável para a mitigação destes riscos. Por fim, defendemos que a integridade autoral já não pode ser governada exclusivamente por noções tradicionais de plágio como apropriação indevida, mas exige um novo imperativo de transparência, para o qual desenvolvemos uma taxonomia gradual de divulgação ética que reconhece diferentes níveis de assistência tecnológica na produção textual.
PALAVRAS-CHAVE: Escrita Distante. Agência Epistêmica. Curadoria Epistêmica. Inteligência Artificial. Autoria.

Received: 11 October 2025
Accepted: 14 November 2025

1 Introduction

The recent and rapid proliferation of Large Language Models (LLMs)[1] heralds a significant reconfiguration of the epistemic foundations of authorship, challenging established assumptions about creativity, intentionality, and intellectual labor. This transformation, which affects all areas of textual production, takes on a particularly provocative character in the domain of philosophical writing, a practice historically associated with rigorous thinking, conceptual clarity, and individual reflective effort.

In this scenario of transformation, the concept of *"distant writing*," as proposed by Luciano Floridi (2025a), offers a robust theoretical framework for understanding

[1] Hereinafter, we will use the acronym *LLM* (or *LLMs*, in the plural), derived from *the* original *Large Language Models*, to abbreviate the terms *"Large Language* Model" and *"Large Language Models*," respectively.

this new mode of textual production. Unlike *"distant reading,"* proposed by Franco Moretti (2013), which uses computational tools to retrospectively analyze large text corpora[2] , distant writing is fundamentally prospective and generative, leveraging LLMs to create new works. In it, the human author transitions from a direct composer to a *designer* or curator, guiding an AI system through the creation of *prompts* and iterative refinement. The author thus moves from writing in text to writing through the machine, transforming the nature of the creative act.

The emergence of this practice raises urgent epistemological questions, especially in the philosophical context: if an author uses an LLM to generate parts of an article, who is responsible for the epistemic status of the final product? In light of these questions, and recognizing that the use of this technology seems to be here to stay, this text, while problematizing distant writing, seeks to construct a positive vision for its appropriate application in the context of philosophical and academic writing[3]. Rather than adopting a simply critical or prohibitive stance, we will defend the thesis that, in distant philosophical writing, the author does not relinquish their epistemic agency, but rather *reconfigures* it, assuming the crucial role of an epistemic curator.[4] Epistemic agency, as conceived from an epistemic perspective, refers to a subject's capacity to be a knower that is active and responsible for their beliefs and assertions (Zagzebski, 1996). We argue that the use of LLMs in philosophy does not

[2] The term *"distant reading,"* proposed by Moretti (2013), does not refer to inattentive reading, but rather to a methodological approach in the digital humanities that uses computational methods (such as topic modeling, network analysis, and data mining) to analyze vast literary archives. The goal is to identify large-scale patterns, trends, and structures that would not be visible through *close reading* of individual texts. For an in-depth exploration, see Moretti (2013), *Distant Reading*.

[3] In this article, we focus only on the epistemic and ethical dimensions of authorship. For a broader discussion of the aesthetic implications and the debate about whether AI can be genuinely "creative," see Boden's (2004) seminal work on creativity, as well as more recent analyses of AI-generated art, such as Zylinska (2020).

[4] The choice of the term "curator" rather than "editor" or "supervisor" is deliberate. While an editor typically refines a work already conceived by an author, and a supervisor guides the work of another agent, the curator (in the museological sense) selects, contextualizes, and organizes elements to create a new exhibition with its own meaning. The term "curator" better captures the author's high-level agency in distant writing, which involves not only correction but strategic selection and construction of a coherent whole from externally generated materials.

eliminate this type of agency, but shifts its exercise to a type of curation, a process that is not a passive act of selection, but a cognitively demanding epistemic activity. To develop this constructive thesis, the article will first unveil the epistemic architecture of distant philosophical writing, distinguishing the author's epistemic agency from the epistemic instrumentality of the LLM. It will then analyze the division of cognitive labor and the risks inherent in this practice, such as the generation of epistemic simulacra, texts that simulate the form of knowledge without its substance5. It will be argued that intellectual virtues such as vigilance, epistemic humility, and critical self-reflection are indispensable for mitigating these risks (Battaglia, 2022; Kidd; Battaglia, 2023). Finally, the article will address the implications for authorship, originality, and academic integrity in philosophy, proposing that transparency about AI mediation become a norm of intellectual responsibility. The ultimate goal is to provide a theoretical framework that preserves the centrality of epistemic agency, offering a path for the ethical and productive use of this new technology.

2 The Epistemic Architecture of Distant Philosophical Writing

In order to properly access the implications that arise from the use of distant writing in philosophy, it is necessary to unravel its epistemic architecture, that is, the underlying structure of knowledge, agency, and justification that governs it. This new mode of textual production does not represent a simple updating the philosopher's tools, as if replacing a typewriter with a computer. On the contrary, it fundamentally reconfigures the distribution of cognitive work, forcing us to clearly divide the contribution of the human agent and that of the artificial system. It can be noted that at the heart of this architecture lies a crucial philosophical distinction: the

[5] The concept of "epistemic simulacra" is inspired by Jean Baudrillard's theory (1994), which describes the progression of simulacra in four stages, culminating in "pure simulation," in which the image has no relation to an underlying reality, becoming its own reality. An "epistemic simulacrum" functions in a similar way: it is a text that perfectly imitates the form of knowledge (arguments, evidence, citations) without being based on the substance of knowledge (understanding, justification, truth), becoming a simulation of academic discourse.

difference between epistemic agency, which remains firmly with the human author, and epistemic instrumentality, which characterizes the function of the LLM. Ignoring this distinction leads to a misunderstanding of authorship and responsibility, while its elucidation reveals the demanding nature of the philosopher's new role as an epistemic curator.

2.1 Epistemic Agency *versus* Epistemic Instrumentality

When we consider the analytical tradition of epistemology, especially with regard to virtue epistemology[6], we identify that the term "epistemic agency" is defined as the capacity of an agent not only to hold beliefs, but to form them, evaluate them, and take responsibility for them in an active and rational manner (Greco, 2010; Zagzebski, 1996). In this sense, an epistemic agent is one who is capable of deliberating, responding to reasons, and justifying their statements. This is someone who positions themselves in the normative space of knowledge, being worthy of credit for their true beliefs and responsible for their errors. This conception of agency necessarily implies intentionality, consciousness, and a capacity for reflective evaluation, attributes that are normally taken to be constitutive of philosophical practice. The philosopher not only asserts a thesis but commits to its defense, anticipates objections, and situates it within a tradition of thought.

LLMs, on the other hand, operate entirely outside this normative space. As argued by Floridi (2025b), AI systems can be seen as a form of agency without intelligence (non-epistemic agency), capable of performing complex tasks autonomously, but devoid of understanding, consciousness, or intentional mental

[6] Virtue epistemology is a broad field that can be roughly divided into two main strands. The first, for example, virtue reliabilism (associated with theorists such as Ernest Sosa and John Greco), focuses on reliable cognitive faculties (such as vision or memory) as "virtues." The second, virtue responsibility (associated with theorists such as Linda Zagzebski), focuses on intellectual character traits (such as open-mindedness, curiosity, and intellectual courage) as the central virtues. In this text, we mainly take the second approach, emphasizing the character and dispositions of the agent. For an overview, see Baehr (2011).

states. Their functioning is not guided by inferential reasoning, but by mathematical optimization and prediction of the next word based on statistical correlations identified in vast training datasets (Buckner, 2019). Consequently, an LLM does not know or believe anything in the propositional sense. When it generates text, it is not expressing a justified belief, but rather composing a linguistically plausible sequence; analogous to a "parrot" that can mimic human speech without any understanding of its meaning (Bender *et al.*, 2021).

In this sense, the LLM is, at best, an epistemic instrument. We can consider an epistemic instrument to be something that constitutes a tool capable of facilitating or expanding our cognitive abilities, such as a telescope, a microscope, or a calculator.[7] These tools clearly play a causal role in the acquisition of knowledge, but no one would be willing to attribute epistemic agency to the telescope for the discovery of a new moon, or to the calculator for the solution of a given equation. The responsibility and credit belong to the astronomer and mathematician who use them. LLMs are undoubtedly instruments of unprecedented sophistication, whose generative capacity far exceeds that of previous tools. However, from an epistemic point of view, they remain in the category of instruments. Their complexity does not give them agency; it merely expands the scope of tasks that can be delegated by the human agent.

An analogy can be drawn with Plato's criticism of writing in *the Phaedrus* dialogue. Socrates argues that the written word is like a painting: it seems alive, but if we ask it a question, it remains majestically silent. Written text cannot defend itself, clarify its ambiguities, or adapt to different interlocutors; it is an inert instrument, devoid of the living soul of dialectical discourse. The LLM is, in many ways, the

[7] The notion of "epistemic instrument" discussed here correlates with Clark and Chalmers' (1998) extended mind thesis. In this perspective, an external object (such as a notebook or an LLM) can become a constitutive part of a cognitive system, functioning as an extension of the agent's mind. It is crucial to note, however, that even as part of an "extended mind," the object does not acquire agency. It remains an instrument, and the epistemic agency of the entire hybrid cognitive system remains exclusively with the human user.

ultimate manifestation of this Platonic text. It can simulate a response, generating more text in reaction to a new *prompt*, but, like Plato's text, it lacks a genuine understanding to "defend" itself based on reasons. Its response does not arise from a reflection on meaning, but from a new probabilistic calculation.

Just as writing, for Plato, was an instrument that threatened memory and true knowledge (dialectics), LLMs are instruments that, if misunderstood, threaten the correct attribution of agency and epistemic responsibility. They are not interlocutors in philosophical dialogue, but rather sophisticated automata that organize symbols, echoing Searle's famous *Chinese Room* argument (1980)[8], according to which the syntactic manipulation of symbols does not, in itself, constitute semantic understanding.

Consider a practical example: a philosopher who uses an LLM with the *prompt*: "Write a Plato-style dialogue in which Socrates discusses the ethics of artificial intelligence with a young programmer named Technos". The model can generate an impressive text, imitating the maieutic structure, Socrates' interrogative tone, and the deference of his interlocutor. It can even articulate arguments that seem coherent, such as the concern that AI, by focusing on efficiency, may neglect virtue. However, the LLM does not understand what ethics, virtue, the Socratic method, or AI itself are. The text is a stylistic and thematic collage based on patterns. The coherence is apparent, the depth is simulated. The philosophical value of the text does not lie in the raw output of the machine, but rather in what the philosopher (agent) does next: their evaluation, their criticism, their decision to discard, refine, or

[8] The *Chinese Room Argument* is a thought experiment proposed by John Searle (1980) to challenge the strong AI thesis, according to which a properly programmed computer can have a genuine mind and understanding. Searle imagines himself inside a room, manipulating Chinese symbols according to a rule book in English, in such a way that, to an outside observer, it appears that he understands Chinese. However, Searle argues that he does not understand a single word of Chinese. His conclusion is that the purely syntactic manipulation of symbols is not sufficient for semantics, that is, for understanding. For an in-depth discussion of the argument and its numerous criticisms, see the *Stanford Encyclopedia of Philosophy* entry on the subject (Cole, 2020).

integrate it into their own argument. The machine provides the syntactic marble; the philosopher, as an epistemic agent, is the sculptor who imposes (or not) a conceptual form with meaning and validity.

2.2 The Epistemic Work of Curation

If the LLM is an instrument, then the philosopher's work in distant writing cannot be reduced to the simple operation of a tool. Instead, the author's epistemic agency manifests itself through demanding curatorial work. This term, borrowed from the art world, is redefined here to describe a multifaceted intellectual process that goes far beyond mere selection. A museum curator does not merely choose works; he interprets, contextualizes, and organizes them to construct a coherent narrative that generates knowledge and provokes an informed response in the viewer (Elgin, 2017). Similarly, the writer-curator (philosopher) not only selects AI-generated text excerpts, but actively orchestrates the entire dialogical process with the machine to construct a robust and original philosophical argument. This curatorial work is epistemically generative and can be broken down into at least three types of interdependent knowledge that the human author must possess and apply.

The first is procedural knowledge, or "know-how" (Ryle, 1949). This refers to the practical competence to interact effectively with the LLM. In philosophy, this translates into the art of constructing *prompts* that are conceptually precise and philosophically fertile. A vague *prompt* such as "talk about free will" will produce a generic and useless summary. An expert *prompt*, such as "Develop a compatibilist argument that responds to Derek Pereboom's *Manipulation Objection*, using an analogy with autonomous agent programming in AI", is a form of epistemic *design*. It requires deep knowledge of the debate to guide the machine to a latent space of textual possibilities that is relevant.

The second is what we might call *evaluative knowledge*, the ability to judge the quality, coherence, and validity of the generated output. This is where philosophical

expertise is most critical. An LLM can produce a paragraph on Kant's categorical imperative that is grammatically perfect and uses the correct terminology. However, only the agent-curator can evaluate: is the generated formulation of the categorical imperative a faithful representation or a misleading simplification? Does the example provided by the AI to illustrate the concept (e.g., about a false promise) capture the essence of the universalization test or does it subtly fail? Does the text adequately distinguish between perfect and imperfect duties? These are judgments that require a keen philosophical sensibility, an ability to detect nuances, ambiguities, and conceptual errors that the machine, focused on superficial plausibility, is unable to recognize (Grasso, 2023).

The third, and most fundamental, is semantic or propositional knowledge, the *"knowing that"*. This is the body of philosophical knowledge that the author possesses on the subject in question. In traditional writing, this knowledge is the direct source of content. In distant writing, it serves as the final verification criterion and the foundation for curatorial work. The human author remains responsible for the veracity of statements, the soundness of inferences, and the integrity of sources (given that LLMs are notoriously known for "hallucinating" or fabricating) (Strzelecki, 2023). Without this prior knowledge, the author becomes a mere transcriber of a black box, running the risk of propagating falsehoods or absurdities under a veneer of eloquence, a criticism that echoes the concept of "*bullshit*", proposed by Frankfurt (2005), in which statements are produced without any consideration for their truth.

A useful analogy to capture the dynamics of this process is that of a composer working with a brilliant improvisational pianist who is thematically amnesiac. The composer (the philosopher) establishes the theme, harmonic structure, and emotional goal of the piece. He may ask the pianist (the LLM) to improvise a variation on a specific motif. The pianist, with his prodigious technique, can produce a cascade of complex and beautiful melodies and harmonies. However, he has no memory of the

initial theme or an overall view of the composition. It is the composer who must listen carefully, select the fragments that serve his vision, reject those that deviate, and weave the chosen elements into a cohesive and meaningful work. The creative act resides both in the initial instruction and, perhaps more importantly, in the composer's evaluative and integrative judgment. The creative and epistemic agency is his. The LLM, like the amnesiac pianist, is a powerful generator of variations, but it is the writer-curator who gives it direction and purpose, exercising what might be called epistemic metacontrol. This iterative process of instructing, generating, evaluating, and refining reflects a cycle of epistemic calibration, in which the human agent continually adjusts their interaction with the instrument to ensure that the results align with their standards of truth and coherence (Pritchard, 2016). In short, the epistemic architecture of distant philosophical writing does not marginalize the human author; on the contrary, it requires them to operate at a higher level of supervision, exercising their agency not through the direct production of each word, but through the orchestration and critical curation of the entire process of meaning creation.

3 The Division of Cognitive Labor and Epistemic Risks

The characterization of distant writing as an interaction between a human epistemic agent and an artificial epistemic instrument lays the foundation for a deeper analysis of its operational dynamics. This interaction constitutes a new form of cognitive division of labor, a process in which the tasks of idea generation, linguistic articulation, and argumentative refinement are distributed across the boundaries of the human brain and the computer system. Although this human-machine collaboration can unlock remarkable efficiency and creativity, it is inherently fraught with vulnerabilities. The fluidity and simulation capabilities of LLM create fertile ground for subtle and sometimes dangerous epistemic risks.

Navigating this new ecosystem therefore requires not only technical competence, but also a keen awareness of its pitfalls and the mobilization of a robust set of intellectual virtues to safeguard the integrity of philosophical work.

3.1 The Distributed Nature of Philosophical Writing with AI

The interaction between a philosopher and an LLM, in the practice of distant writing, inaugurates a new and complex form of distributed cognition. The concept of the "extended mind", originally introduced by Andy Clark and David Chalmers (1998), challenges the traditional view that cognition is a process that occurs solely and exclusively within the limits of the human brain and body. To better understand this notion of the "extended mind", consider an example. Imagine that Otto, a man with Alzheimer's, uses a notebook to record important information about his daily life. During his visit to the museum, Otto makes a point of writing down in his notebook the name of his guide and all the works he has seen, so that he does not risk forgetting them. Clark and Chalmers argue that the notebook, in this case, is not just an external tool, but part of Otto's cognitive system; the process of consulting the notebook performs the same function as biological memory, according to the *"Principle of Parity"*[9]. The relevant cognitive system is not just "Otto", but the coupled and functionally integrated "Otto-notebook" unit.

Distant writing operates according to a similar principle, but with an order of magnitude greater complexity and interactivity. The LLM is not a passive repository like Otto's notebook; it is a dynamic and generative partner in the thinking process. It functions as an *epistemic scaffold*. "Epistemic scaffold" is a term defined as a structure external to the cognitive agent that allows the agent to perform cognitive

[9] The *"Principle of Parity,"* formulated by Clark and Chalmers (1998), is the central criterion for determining whether an external process counts as part of a cognitive system. According to this principle, if a part of the world functions as a process that, if performed in the mind, we would not hesitate to recognize as part of the cognitive process, then that part of the world is (we assert) part of the cognitive process. It is this principle that allows us to argue that Otto's notebook is not just a tool, but an extension of his memory.

tasks that would otherwise be too demanding or even impractical (Clark; Chalmers, 1998). Let us imagine a philosopher working on the relationship between Husserl's phenomenology and Merleau-Ponty's. Traditionally, this work would require hours of reading and synthesis to map the points of convergence and divergence. By using an LLM strategically, the philosopher can delegate the initial task of summarizing each author's positions on a specific concept, such as intentionality or the *world of life. The* LLM can quickly generate structured summaries that serve as a basis, or scaffolding, for analysis. This does not replace philosophical thinking; on the contrary, it frees up cognitive resources. By externalizing the more laborious work of initial collection and organization, the philosopher can focus their mental energy on higher-order tasks: critical analysis, the discovery of a new synthesis, the formulation of an original objection. The LLM becomes part of a cognitive *feedback* loop, in which the author solicits and evaluates the scaffolding generated, refines their own understanding, and formulates new questions in an iterative process of knowledge construction.

This dynamic establishes a clear, albeit unprecedented, division of cognitive labor. Tasks are distributed according to the nature of each component's capabilities within the hybrid system. The LLM (the instrument) is delegated functions that depend on largescale data processing, pattern recognition, and syntactic generation, such as: the initial drafting of text sections, the exploration of different stylistic formulations to increase clarity, the summarization of relevant academic literature, or the articulation of standardized objections and counterexamples. However, functions that require genuine epistemic agency remain the exclusive domain of the human philosopher. Such tasks include: defining the purpose and thesis of the argument (the *telos* of the investigation), formulating hypotheses and creative *prompts*, evaluating the logical soundness and validity of the philosophical relevance of the generated content, judging conceptual nuance and accuracy, verifying the

veracity of sources, and, most crucially, assuming full intellectual responsibility for the final product.

The division of labor finds an illuminating parallel in the way "Big Science" operates[10]. Consider a particle physicist at CERN (the European Organization for Nuclear Research) analyzing data from a collision in *the Large Hadron Collider* (LHC). No single scientist understands every detail of the millions of components of the detector or *the software* that processes the *petabytes* of data. Knowledge is necessarily distributed across a vast collaboration of experts and complex computational tools. The role of the principal scientist is not to perform the calculations, but rather to design the experiment, interpret the results provided by the computers in light of theory, and ultimately construct an argument that justifies a discovery.

Similarly, the philosopher in distant writing may not understand the internal mechanisms of the LLM, but their expertise lies in knowing how to "operate" the tool in order to produce meaningful results and, more importantly, in knowing how to interpret, validate, and contextualize these results within a rigorous philosophical framework. The epistemic authority of the article on the *Higgs boson*[11] does not belong to computers, but to the scientists who signed it. Similarly, the authorship of a philosophical article produced through distant writing belongs unequivocally to the philosopher who conceived it, guided it, and takes responsibility for it.

[10] The analysis of science as a form of distributed knowledge is a central topic in the sociology and philosophy of science. Authors such as Peter Galison, in works such as *Image and Logic* (1997), have explored how knowledge in fields such as particle physics is produced through vast collaborations between theorists, experimentalists, and complex technical apparatus, in which no single individual possesses all the knowledge of the system. This perspective offers a useful model for thinking about *human-AI* collaboration in writing.

[11] The *Higgs boson* is the elementary particle associated with the Higgs field, whose function is to explain the origin of the mass of fundamental particles in the Standard Model of physics. Its existence was proposed in the 1960s and experimentally confirmed in 2012 by the ATLAS and CMS detectors at the Large Hadron Collider.

3.2 Cognitive Vulnerabilities and "Epistemic Simulacra"

The deep integration between human thought and machine generation, while powerful, opens the door to a set of cognitive vulnerabilities that are particularly acute in philosophical writing. The most insidious risk of LLMs lies not in their obvious errors, but in their deceptive fluency. Modern models are exceptionally proficient at mimicking the conventions of academic writing, producing texts that are grammatically flawless and stylistically polished. This quality alone can mask gaps in content and logic, presenting a veneer of authority that disarms critical vigilance.

This fluency creates a dangerous epistemic illusion, a *"halo effect,"* in which the polished form of the text confers a false aura of rigor on its content. In philosophy, where the nuance of language and rhetorical structure are often inseparable from the substance of the argument itself, this capacity for imitation constitutes a trap of the first order. A well-structured paragraph with precise technical terminology can easily bypass the initial critical defenses of a reader or, more dangerously, a hasty author who is overly confident in the tool.

This characteristic gives rise to what can be appropriately designated as epistemic simulacra: results that perfectly simulate the form of knowledge without possessing its substance. The concept, which echoes Jean Baudrillard's (1994) notion of "simulacrum" as a copy without an original, describes a text that has all the external markers of legitimate philosophical discourse (precise concepts, argumentative structure, logical connectors), but which is empty at its epistemic core[12]. It is, in essence, an epistemic zombie: it looks and behaves like a living philosophical argument, but it

[12] Although we have used the term "simulacrum" in a general way, Baudrillard's (1994) theory is more granular, describing four stages of representation: (1) the reflection of a basic reality; (2) the mask and perversion of a basic reality; (3) the mask of the absence of a basic reality; and (4) the pure simulacrum, which has no relation to any reality. The epistemic simulacra generated by LLMs can vary across this spectrum, from paraphrasing (stage 2) to the complete fabrication of facts and arguments (stage 4).

lacks the mind, that is, the understanding, intentionality, and justification that would give it intellectual
life.

Consider, for example, a philosopher who asks an LLM for "an analysis of the relationship between *Dasein* and temporality in Heidegger's Being and Time." The model can generate a convincing text, correctly using terms such *as "being-towards-death," "authenticity," and "care."* However, the LLM has no phenomenological understanding of how these concepts interconnect in Heidegger's complex ontology. The text is a syntactic montage based on patterns extracted from thousands of texts on the subject, a perfect imitation of a philosophical explanation, but devoid of the thought that should animate it. For an author who is not a deep expert on Heidegger, this simulacrum may be indistinguishable from a genuine analysis, leading to its uncritical incorporation into the work.

A powerful philosophical analogy can be drawn with the Sophists of Ancient Greece. Figures such as Gorgias and Protagoras were masters of rhetoric, capable of construct eloquent speeches on any topic with the aim of persuading an audience, regardless of the truth. Their focus was on opinion (*doxa*), not knowledge (*episteme*). LLMs function as the supreme Sophists of the digital age. Their architecture is optimized not for truthfulness, but for textual plausibility, the modern correlate of persuasion. When asked to defend a thesis, the LLM assembles the most likely sequence of words that, in its training data, corresponds to that defense. It is, in essence, a Gorgian machine, capable of generating a discourse with equal apparent conviction, but without any commitment to the truth. The philosopher who uses this tool without relentless vigilance risks becoming a sophist's apprentice.

Beyond these conceptual risks, there are practical and immediate dangers, such as the well-documented phenomenon of AI *"hallucinations*." These occur when a model generates factually incorrect information but presents it with complete confidence, such as attributing a quote to the wrong philosopher or inventing non-

existent academic articles to support a particular argument (Strzelecki, 2023). In philosophical writing, the integrity of sources and the correct genealogy of an idea are not mere formalities; they are central to intellectual honesty. A fabricated reference is not simply an error, it is a corruption of the academic record. The responsibility for meticulously verifying every statement and source generated by AI rests uncompromisingly with the human author.

Finally, LLMs can perpetuate and amplify systemic biases present in their vast training data (Friedman; Nissenbaum, 1996). A model trained predominantly in the Western philosophical canon may, when asked to compare ethical theories, marginalize or caricature the perspectives of African or Asian philosophy. This is not done out of malice, but rather reflects the statistical asymmetry of the texts with which it learned. The distributed cognitive system thus becomes a zone of vulnerability, where the delegation of cognitive work carries the risk of a corresponding, and dangerous, delegation of epistemic and ethical vigilance.

3.3 Intellectual Virtues as an Antidote

If remote writing, practiced without due care, is an area of epistemic vulnerability, the solution to its dangers does not lie in technophobic rejection or purely procedural checklists. We believe that the most robust and lasting response lies in the realm of ethics and epistemology, specifically in the character of the agent themselves. Virtue epistemology, with its focus on the dispositions, character traits, and intellectual excellences of the knower, offers an ideal normative framework to guide the philosopher in this new landscape (Baehr, 2011; Zagzebski, 1996). To navigate responsibly in collaboration with LLMs, the author must actively cultivate and exercise a set of intellectual virtues that serve as the necessary antidote to the cognitive vulnerabilities inherent in this process.

The first and most fundamental of these provisions is epistemic vigilance. This virtue transcends mere fact-checking; it is an active and habitual posture of

methodical skepticism. The vigilant philosopher approaches each fragment of text generated by an LLM not as a statement to be accepted, but as a hypothesis to be rigorously tested. In an analogy with the Cartesian method, the author must initially remain skeptical of the machine's results. Every argument, every definition, and every example, however plausible it may seem, must be submitted to the court of reason and the author's prior knowledge. Vigilance requires incessant questioning: is the chain of inferences logically sound or merely rhetorically persuasive? Is the analogy offered by AI truly enlightening or just a misleading simplification? Is the source cited real and has its content been accurately represented? This virtue transforms the author from a passive consumer of text to a tireless interrogator, ensuring the epistemic hygiene of their work.

Closely linked to vigilance is epistemic humility, a virtue with two sides. On the one hand, it implies a clear and constant awareness of the limitations of LLM. The epistemically humble philosopher must never forget that he is interacting with a nonsentient tool, a statistical *"black box"* prone to errors, biases, and the fabrication of convincing absurdities. This attitude prevents blind trust and the tendency to delegate critical intellectual tasks.

On the other hand, humility manifests itself in the recognition of the limitations of the author's own knowledge. No philosopher is an expert in all fields. Humility leads the author to ask themselves: "Do I have the depth of knowledge necessary to critically evaluate this analysis of modal logic that the LLM has just generated?" This awareness fosters intellectual honesty and encourages the search for external verification from other human experts, acting as an essential safeguard against the arrogance that the apparent omniscience of AI could induce (Nyholm, 2023).

The third indispensable virtue is *critical self-reflection*, which functions as a kind of *"meta-virtue"* that oversees the author's own cognitive process. Its focus is not on the machine's output, but on the author's interaction with it. The self-reflective

philosopher questions their own motivations and susceptibility to cognitive biases.[13] They must be aware of *confirmation bias: the* tendency to uncritically accept AI results that reinforce their own theses, while subjecting dissonant results to much more severe scrutiny. This virtue implies asking: "Am I using LLM to genuinely explore a problem or to find shortcuts that avoid the hard work of original thinking? My satisfaction with this generated paragraph is due to its quality intrinsic or to the fact that he conveniently articulated something I had difficulty expressing?" This type of introspection is the antidote to the destitution of human agency that some critics fear (Clowes, 2023), as it reaffirms the author's conscious and deliberate control over his own intellectual process.

These three core provisions are complemented by other equally important virtues. Intellectual courage is necessary to discard a piece of AI-generated text that is eloquent and well-written but philosophically inaccurate or superficial, resisting the temptation to keep it because of the effort already invested. Intellectual thoroughness manifests itself in the patience to meticulously check every detail, resisting the pressure for productivity that these tools can create.

Together, this mosaic of virtues constitutes the core of epistemic responsibility in the age of AI. These are not optional qualities, but rather necessary conditions for the practice of writing in a way that is intellectually honest and philosophically valuable. The final product of a human-machine collaboration reflects not only the capabilities of the technology, but more decisively, the intellectual character and integrity of the human agent who conducted it.

[13] The study of *cognitive biases* is a prolific area of research in psychology and behavioral economics. Confirmation bias, the tendency to seek, interpret, and remember information in a way that confirms pre-existing beliefs, is one of the best documented. For an accessible introduction to the main biases that affect human reasoning, see Daniel Kahneman (2011), *Thinking, Fast and Slow*.

4. Implications for Authorship, Originality, and Philosophical Integrity

The reconfiguration of writing as an act of curation and *design*, mediated by powerful generative tools, transcends questions of process and enters the normative heart of academic practice. The emergence of remote writing forces a critical reassessment of the fundamental concepts that underpin the integrity of philosophy and other disciplines: authorship, originality, and plagiarism. This new reality of intellectual work imposed by remote writing requires us to rethink and reinforce the epistemic, ethical, and normative structures that have traditionally been based on the assumption of a single, intentional, and unique human author.

Our interest is not to advocate abandoning these principles, but to adapt them to a new context. This necessarily implies a shift in focus, moving away from the notion of plagiarism as "theft" and toward an ethic of transparency, responsibility, and epistemic trust.

4.1 Rethinking Plagiarism and Attribution

In addition to transforming our modes of intellectual production, the introduction of distant writing into philosophical practice also destabilizes the ethical foundations that guide this activity, requiring a profound revision of the notions of authorship, plagiarism, and attribution.

According to tradition, the notion of plagiarism is fundamentally focused on the idea of misappropriation of a product (text, idea, etc.) that belongs to another author. In general terms, plagiarism is understood as a form of intellectual theft: an act that takes away someone's recognition, authorship, and the result of their effort, presenting them as if they belonged to the plagiarist (Scanlon, 2003). This model is based on the premise that there is an author who holds moral and epistemic rights over their work. However, this normative structure collapses when the source of the text is a Large Language Model (LLM), whose nature does not fit into the traditional categories of human authorship.

As has been established, an LLM is not an agent in the philosophically relevant sense. It is a sophisticated instrument, a statistical machine devoid of intentionality, consciousness, rights, or the ability to "own" the content it generates. In this context, the accusation of "plagiarizing from AI" becomes a categorical error[14]. You cannot steal intellectual property from an entity that cannot, in the first place, possess it. Trying to apply the traditional model of plagiarism to this new reality is like trying to apply traffic laws to a school of fish: the categories simply do not correspond. This observation, however, does not absolve us of the ethical problem. On the contrary, it forces us to locate it more precisely. If the error does not lie in the violation of the rights of a non-existent machine author, then what is the nature of the transgression? The answer lies in a change of framework: the fundamental ethical error of the undeclared use of LLMs is not appropriation, but deception and misrepresentation. The damage is not inflicted on the machine, but on the community of readers and the very integrity of philosophical discourse.

Academic practice is based on an implicit trust between the author and the reader. The reader trusts that the work presented under an author's name is an honest reflection of their cognitive effort, expertise, and intellectual voice. Failure to disclose the substantial use of an LLM fundamentally violates this trust. The reader is led to believe that a complex argument or eloquent prose originated in the author's mind, when in fact it may have been generated by a computer system. This omission constitutes a serious epistemic misrepresentation, as it prevents the community from correctly assessing the origin, merit, and authority of the work (Roig, 2006).

14 This philosophical position is increasingly being reflected in the legal domain. The U.S. Copyright Office, for example, issued guidance in 2023 stating that a work generated entirely by AI without any human creative input cannot be protected by copyright, as U.S. copyright law only protects the fruits of intellectual labor that are based on the creative powers of the human mind. The absence of human authorship at the source of AI-generated text reinforces the idea that plagiarism, as a violation of copyright, is an inapplicable category. See U.S. Copyright Office (2023).

Here we can draw an illuminating analogy: the practice of academic *ghostwriting*[15]. Consider certain "sensitive" areas of knowledge[16]. For example, let's take the field of medicine. In this sense, the publication of a scientific article written by a renowned doctor, but secretly written by a company communication contracted by a pharmaceutical company would constitute a serious ethical violation (Ross *et al.*, 2008). The central problem is not exactly the delegation of writing to another agent, but rather the fact that this delegation is kept secret. Concealment creates a false attribution of authority, misleads reviewers and readers, and undermines the credibility of the scientific process. The undeclared use of an LLM in philosophy is a form of technological *ghostwriting*. Similarly, by presenting the generated text as their own, philosophers lend their credibility and name to a product whose cognitive origin is, at best, hybrid and, at worst, entirely delegated to the machine. In this way, the reader is deprived of the knowledge that they are engaging with the thoughts of a human being or with the statistical result of a machine.

In order to deepen our understanding of the term "authorship" in this new context, we can resort to an analogy: the work *Fountain* (1917), produced by Marcel Duchamp[17]. Duchamp is not considered the manufacturer of the urinal in any relevant sense; all he did was select this object (which is normally mass-produced), remove it from its original functional context, give it a new context, and, in doing so, sign it and present it as art. His act of authorship was a gesture of selection,

[15] G*hostwriting* is the practice of hiring someone to write a text that will be officially credited to another person. In the academic context, it is considered a serious violation of integrity, as the named author (who signs the work) receives credit and authority for an intellectual effort they did not perform. This practice deceives the community of readers and reviewers, misrepresenting the true origin of the work.

[16] The term "sensitive" refers to those areas of knowledge that deal directly with human well-being and life, and their impact is costly.

[17] The discussion about authorship in conceptual art is vast. Duchamp's intervention with his *readymades* challenged the idea that art resides in the artist's manual skill, shifting the focus to the idea, the choice, and the institutional gesture. This shift is central to understanding how authorship in distant writing may reside more in conceptual *design* and curation than in linguistic production. For an in-depth philosophical analysis, see Goldie and Schellekens (2007).

recontextualization, and conceptual framing, not of manufacture. The originality of the work lies in this act of *design and* not in the creation of the physical object.

The philosopher's practice in distant writing is, in many ways, analogous. Their authorial act may lie not in the generation of prose, but in the formulation of a brilliant *prompt*, the careful selection of the AI's output, and its ingenious integration into a larger argument. However, the analogy only holds under one crucial condition: the transparency. Duchamp never pretended to have sculpted the urinal. His honesty about the nature of the "found object" was an integral part of his artistic intervention. The philosopher who conceals the use of an LLM is committing the equivalent of Duchamp claiming to have molded the porcelain with his own hands. The error lies not in the nontraditional method of creation, but in the dishonesty about the use of that method.

This leads us to rethink the very notion of *"originality"* in philosophical writing. The era of distant writing may accelerate the shift away from a romantic ideal of *creatio ex nihilo* (creation from nothing) toward a more sophisticated appreciation of originality as an act of synthesis, curation, and connection. An author's "voice" may manifest itself less in their idiosyncratic prose style and more in their higher-order choices: in the originality of the questions, they pose to the machine (the *prompts)*, in the depth of their critical judgment in evaluating the results, and in the overall architecture of the argument they construct from the generated materials. The author's unique contribution thus becomes their ability to orchestrate productive collaboration with technology and take full responsibility for the final product.

In short, the inadequacy of traditional models of plagiarism requires the adoption of a new regulatory framework for authorship in the age of AI. This framework must abandon the logic of *"theft"* by a non-agent and focus on the ethical imperative of transparency. The use of a tool as powerful as an LLM is a methodologically relevant fact that affects the conditions of knowledge production. As

such, its declaration should not be seen as an admission of guilt, but as a practice of intellectual rigor and honesty, similar to the way scientists declare their methods, the *software* they use, and their sources of funding (Nature, 2023). Only through a culture of radical transparency can we preserve trust, which is the basis of all philosophical and academic endeavor.

4.2 A Taxonomy for Ethical Disclosure

Having established transparency as the central ethical imperative for distant writing, the pragmatic question that follows is: what should this transparency look like in practice? A generic, binary statement such as "AI was used in the preparation of this manuscript" is clearly insufficient. Such a statement is too vague to be informative, obscuring the crucial difference between using AI to correct a comma and using it to generate a central argument. For disclosure to be ethically meaningful, it must be granular and proportional to the nature and depth of the machine's involvement in the intellectual work. What is proposed here is a taxonomy of AI involvement, a conceptual framework that distinguishes different levels of collaboration and can guide philosophers in their disclosure responsibility[18]. This taxonomy works in a similar way to film credits: instead of a single credit for *"authorship,"* it recognizes that creation is a process with different roles (director, editor, screenwriter), each deserving a specific type of recognition.

At the most basic level of taxonomy, we find minor lexical and stylistic assistance. This encompasses the use of LLMs for tasks that are functionally equivalent to those of established writing tools, albeit more powerful. It includes grammar and spelling correction, synonym suggestion, or paraphrasing individual

[18] Many academic institutions and publishers are beginning to develop their own policies on the use of AI, although these vary significantly. While some prohibit the use of AI to generate text, others allow it on condition of disclosure. The taxonomy proposed here aims to offer a more robust conceptual framework than current, often binary approaches, allowing for a more informed and consistent disclosure policy. For an example of an influential editorial policy, see Nature (2023).

sentences to improve clarity or fluency without altering their semantic substance. For example, a philosopher who has written a dense and complex sentence about Kant's "transcendental unity of apperception" could ask the LLM: "Rewrite this sentence in a simpler way, while maintaining the original meaning." In this case, the LLM acts as a text editor or an advanced thesaurus. T h e structure of the argument, ideas, a n d authorial intent remain entirely human. The machine's contribution is superficial and does not affect the intellectual core of the work. Consequently, the need for disclosure at this level is minimal or nonexistent, just as one does not usually credit a word processor for its grammatical suggestions.

One step up, we find stylistic enhancement. At this level, the author uses the LLM not just for minor corrections, but to significantly shape the "voice," tone, or rhetorical effect of entire passages, while the underlying argument remains the author's. Imagine a philosopher who has written an academic analysis of Aristotle's virtue ethics and wants to adapt it for a wider audience. He could instruct the LLM: "Rewrite this paragraph on *phronesis* in a more engaging and narrative style, using a contemporary analogy." Here, the central idea about practical wisdom is the author's, but its stylistic execution and rhetorical packaging are the product of a hybrid collaboration. The voice of the text is no longer purely that of the author. As this may affect the reader's perception of the author's style and communication skills, discreet disclosure becomes a practice of good intellectual honesty. A footnote or a mention in the acknowledgments section, such as "AI tools were used to assist in the stylistic refinement of some sections," would be appropriate and sufficient.

The third level, which marks the entry into a territory of deep collaboration, is argumentative generation. Here, the LLM ceases to be a mere stylist and becomes an active participant in the construction of philosophical content. The author can use the tool as a *brainstorming* partner or a hypothesis generator. For example, a philosopher working on a critique of Cartesian dualism might request: "Generate an original argument against the divisibility of the body as proof of its distinction from

the mind, using concepts from modern physics." LLM could produce an outline of an argument that the author had not considered on their own. The philosopher could then adopt this line of reasoning, refine it, reinforce it, and integrate it into their work. In this scenario, the genesis of part of the intellectual substance of the article is non-human. Presenting this argument as being solely the result of one's own deliberation would be a clear misrepresentation of the creative process. Consequently, this level requires explicit and prominent disclosure, preferably in a methodology section or introductory note, detailing the role of AI in the development of specific sections of the argument.

At the top of the taxonomy lies the deepest level of collaboration: thematic innovation. In this mode, LLM acts as a catalyst for discovering new connections, analogies, or even theses. A philosopher interested in comparative philosophy could use an exploratory and open *prompt*: "Identify and describe possible, but little explored, structural resonances between the concept of *'becoming'* in Deleuze and the notion of *'impermanence'* in Buddhist thought." By processing and synthesizing patterns from vast text *corpora*, the LLM can highlight an analogy or conceptual tension that the author had not glimpsed, thus opening up an entirely new and original line of investigation. In this case, AI played a role similar to that of a research collaborator or a computational muse. Its contribution to the originality of the central thesis is undeniable. Concealing this contribution would be the most serious form of epistemic misrepresentation. Disclosure must therefore be complete and explicit, openly acknowledging the role of the tool as a partner in the process of conceptual discovery, an act that preserves the author's honesty and enriches the community's understanding of the new ways in which knowledge can be generated.

The adoption of such a taxonomy is not merely an exercise in procedural rigor; it is the practical manifestation of an ethic of authorship adapted to the 21st century. By applying it, philosophers actively exercise the virtues related to intellectual honesty, taking responsibility both for the content of their work and for the

transparency of the process employed, ensuring that their relationship with the community of readers remains based on trust. In this way, the integrity of philosophical authorship is not diminished by technology, but, on the contrary, is reinforced through a deeper commitment to clarity and truth in all phases of knowledge production.

5 Conclusion

The advent of distant writing, as theorized by Luciano Floridi (2025a), represents more than a mere technological innovation in textual production; it imposes a fundamental reassessment of the practices and assumptions that underpin philosophical writing. In this article, we defend the thesis that, given the generative capacity of Large Language Models (LLMs), the role of the philosopher is neither diminished nor replaced, but rather reconfigured to that of an epistemic curator. Far from being a mere operator of a tool, the author in the age of AI is called upon to exercise their agency at a higher level of abstraction, orchestrating, evaluating, and taking responsibility for a process of meaning creation that is intrinsically hybrid.

To support this thesis, our analysis began by establishing a crucial distinction in the epistemic architecture of distant writing: the difference between epistemic agency of author and the epistemic instrumentality of LLM. We argue that, although the LLM is an instrument of unprecedented power, it remains devoid of understanding, intentionality, or responsibility, functioning as a sophisticated syntactic engine. The genuinely philosophical work of justifying beliefs, constructing valid arguments, and generating knowledge remains the exclusive domain of the philosopher-curator, who mobilizes his procedural, evaluative, and semantic knowledge to guide and validate the process.

However, this new form of distributed cognition is not without its dangers. We analyze how the rhetorical fluency of LLMs can create epistemic simulacra, texts that mimic the form of philosophical knowledge without its substance, introducing

vulnerabilities such as misinformation, fallacious arguments, and hidden biases. The response to these risks, we argue, is not purely technical, but normative, and lies in the cultivation of intellectual virtues. Vigilance, humility, and critical self-reflection were presented as the essential dispositions that enable philosophers to navigate this ecosystem safely and responsibly, ensuring the integrity of their work.

Finally, we explore the implications of this new reality for the concepts of plagiarism and authorship. We conclude that traditional models of plagiarism, based on the idea of theft, are inadequate. The ethical error of undeclared AI use lies in deception and a breach of trust with the reader community. To resolve this issue, we propose a taxonomy for ethical disclosure, a framework that allows for granular and honest transparency about the level of AI involvement, from simple stylistic assistance to thematic innovation. The application of this taxonomy is not merely a formality, but an exercise in responsibility and epistemic awareness.

Ultimately, distant writing does not signal the end of philosophical authorship, but rather its evolution. It challenges us to be more rigorous, not only in the content of our arguments, but also in our reflection on the methods we use to produce them. The responsibility for the coherence, truthfulness, and originality of a philosophical work remains, indivisibly, on the shoulders of the human author. The age of AI in philosophy is not an invitation to abdicate intellectual effort, but a call for a deeper form of mastery: that of the thinker who not only masters their ideas, but also critically and virtuously masters the powerful tools that help shape them.

Translated by the author.

References

ARISTOTLE. **Nicomachean ethics**. Translated by W. D. Ross. c. 350 BCE.

ATLAS Collaboration; CMS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson at the LHC. **Physics Letters B**, v. 716, n. 1, p. 1–29, 2012.

BAEHR, J. The epistemic virtues and the nature of knowledge. *In*: PRITCHARD, D. (ed.). ***Epistemology****:* The key thinkers. London: Routledge, 2011. p. 148-164.

BATTAGLIA, F. Intellectual virtues, extended cognition, and the epistemology of AI. **Philosophies**, v. 7, n. 4, p. 84, 2022. DOI

BAUDRILLARD, J. **Simulacra and simulation**. Ann Arbor: University of Michigan Press, 1994. DOI https://doi.org/10.3998/mpub.9904

BENDER, E. M. *et al*. On the dangers of stochastic parrots: Can language models be too big? *In*: **Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.** 2021. p. 610–623. DOI https://doi.org/10.1145/3442188.3445922

BODEN, M. A. **The creative mind**: Myths and mechanisms. 2. ed. London: Routledge, 2004. DOI https://doi.org/10.4324/9780203508527

BUCKNER, C. Deep learning: A philosophical introduction. **Philosophy Compass**, v. 14, n. 10, e12625, 2019. DOI https://doi.org/10.1111/phc3.12625

CLARK, A.; CHALMERS, D. The extended mind. **Analysis**, v. 58, n. 1, p. 7–19, 1998. DOI https://doi.org/10.1093/analys/58.1.7

CLOWES, R. W. AI tools, epistemic dependence and the demotion of human agency. **AI and Ethics**, p. 1–13, 2023. DOI

COLE, D. The Chinese Room Argument. *In*: ZALTA, E. N. (ed.). **The Stanford Encyclopedia of Philosophy** (Winter 2020 Edition). 2020. Disponível em: https://plato.stanford.edu/archives/win2020/entries/chinese-room/. Acesso em: 11 out. 2025.

EATON, S. E.; GÁLIK, M. A view on generative artificial intelligence and the future of academic integrity. **Journal of Education, Culture and Society**, v. 15, n. 1, p. 11–26, 2024. DOI

ELGIN, C. Z. **True enough**. Cambridge: MIT Press, 2017. DOI https://doi.org/10.7551/mitpress/9780262036535.001.0001

FLORIDI, L. Distant writing: Literary production in the age of artificial intelligence. **SSRN**, 2025a. DOI https://doi.org/10.2139/ssrn.5232088

FLORIDI, L. AI as agency without intelligence: On artificial intelligence as a new form of artificial agency and the multiple realisability of agency thesis. **Philosophy & Technology**, v. 38, n. 30, 2025b. DOI https://doi.org/10.1007/s13347-025-00858-9

FRANKFURT, H. G. **On bullshit**. Princeton: Princeton University Press, 2005.

FRIEDMAN, B.; NISSENBAUM, H. Bias in computer systems. **ACM Transactions on Information Systems,** v. 14, n. 3, p. 330-347, 1996. DOI https://doi.org/10.1145/230538.230561

GALISON, P. **Image and logic:** A material culture of microphysics. Chicago: University of Chicago Press, 1997. DOI https://doi.org/10.29173/istl1396

GOLDIE, P.; SCHELLEKENS, E. (ed.**). Philosophy and conceptual art**. Oxford: Oxford University Press, 2007. DOI https://doi.org/10.1093/oso/9780199285556.001.0001

GOLDMAN, A. I. **Knowledge in a social world**. Oxford: Oxford University Press, 1999. DOI https://doi.org/10.1093/0198238207.001.0001

GRASSO, F. On the epistemic risks of large language models. **Minds and Machines**, v. 33, p. 685–690, 2023. DOI

GRECO, J. **Achieving knowledge**: A virtue-theoretic account of epistemic normativity. Cambridge: Cambridge University Press, 2010. DOI https://doi.org/10.1017/CBO9780511844645

JAKESCH, M.; LEDERER, M.; MEYER, M. Epistemic risks of AI-generated texts. **Digital Society**, v. 2, n. 1, p. 10, 2023. DOI

KAHNEMAN, D. **Thinking, fast and slow.** New York: Farrar, Straus and Giroux, 2011.

KIDD, I. J.; BATTAGLIA, F. AI and intellectual virtue: An introduction. **Philosophies**, v. 8, n. 6, p. 110, 2023. DOI

MORETTI, F. Conjectures on world literature. **New Left Review,** n. 1, p. 54–68, 2000. DOI https://doi.org/10.64590/hxj

MORETTI, F. **Distant reading.** London: Verso Books, 2013.

NATURE. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. **Nature**, v. 613, p. 612, 2023. DOI https://doi.org/10.1038/d41586-023-00191-1

NYHOLM, S. The ethics of AI ethics: An introduction to the special issue. **Ethics and Information Technology**, v. 25, n. 1, p. 1, 2023. DOI

PRITCHARD, D. Epistemic risk. **Journal of Philosophy**, v. 113, n. 11, p. 550-571, 2016. DOI https://doi.org/10.5840/jphil20161131137

ROIG, M. **Avoiding plagiarism, self-plagiarism, and other questionable writing practices***:* A guide to ethical writing. U.S. Department of Health & Human Services, Office of Research Integrity, 2006. Disponível em: https://ori.hhs.gov/plagiarism-0. Acesso em: 11 out. 2025.

ROSS, J. S. et al. Guest authorship and ghostwriting in publications related to rofecoxib: A case study of industry documents from rofecoxib litigation. **JAMA**, v. 299, n. 15, p. 1800–1812, 2008. DOI https://doi.org/10.1001/jama.299.15.1800

RYLE, G. **The concept of mind**. Chicago: University of Chicago Press, 1949.

SCANLON, P. M. Student online plagiarism: How do we respond? **College Teaching**, v. 51, n. 4, p. 161–165, 2003. DOI https://doi.org/10.1080/87567550309596432

SEARLE, J. R. Minds, brains, and programs. **Behavioral and Brain Sciences**, v. 3, n. 3, p. 417–457, 1980. DOI https://doi.org/10.1017/S0140525X00005756

STRZELECKI, A. To use or not to use ChatGPT in research? Researchers' acceptance of AI-based conversational agents. **Journal of Information Science**, v. 0, n. 0, 2023. DOI

U.S. COPYRIGHT OFFICE. Copyright registration guidance: Works containing material generated by artificial intelligence. **Federal Register**, v. 88, n. 59, p. 16190-16194, 2023. Disponível em: https://www.federalregister.gov/documents/2023/03/16/2023-05321. Acesso em: 11 out. 2025.

ZAGZEBSKI, L. **Virtues of the mind***:* An inquiry into the nature of virtue and the ethical foundations of knowledge. Cambridge: Cambridge University Press, 1996. DOI https://doi.org/10.1017/CBO9781139174763

ZYLINSKA, J. **AI art***:* Machine visions and warped dreams. London: Open Humanities Press, 2020.