ISSN: 0798-1171 / e-ISSN: 2477-9598

Dep. Legal ppi 201502ZU4649

Esta publicación científica en formato digital
es continuidad de la revista impresa
Depósito legal pp 197402ZU34 / ISSN 0798-1171

REVISTA DE FILOSOFÍA

Universidad del Zulia
Facultad de Humanidades y Educación
Centro de Estudios Filosóficos
"Adolfo García Díaz"
Maracaibo - Venezuela

Nº 115
2026 - 1
Enero - Marzo

Revista de Filosofía
Vol. 43, Nº115, 2026-1, (Ene-Mar) pp. 95-109
Universidad del Zulia. Maracaibo-Venezuela
ISSN: 0798-1171 / e-ISSN: 2477-9598

Agrippa's Trilemma Revisited: Opacity, Circularity, and Structural Dogmatism in High-Dimensional Algorithmic Models

El trilema de Agripa revisitado: opacidad, circularidad y dogmatismo estructural en modelos algorítmicos de alta dimensión

Fabio Morandín-Ahuerma
ORCID: https://orcid.org/0000-0001-6082-2207
Benemérita Universidad Autónoma de Puebla
Puebla - México
fabio.morandin@correo.buap.mx

DOI: https://doi.org/10.5281/zenodo.19835158

Abstract

This article reviews Agrippa's trilemma in the context of artificial intelligence systems, specifically in high-dimensional algorithmic models: the classic epistemological challenge is that any justification of knowledge inevitably leads to infinite regression, circular reasoning, or arbitrary or dogmatic foundations, and therefore knowledge is not possible. This resurfaces strongly in machine learning models characterized by opacity, recursive optimization, and performance-based validation. Large language models and recommendation systems are paradigmatic cases; thus, the study shows how algorithmic inference often operates without semantic grounding or explicit logical structure, or with access to reasons intelligible to humans, thereby generating statistically robust but epistemically opaque results. It is argued that the trilemma appears algorithmically in the combination of three interrelated phenomena, namely: 1) opacity caused by the architecture of machine learning models and decision pathways themselves, 2) circularity, when training, validation, and performance measures feed back into each other without external epistemic reference, and 3) structural dogmatism, which involves taking for granted that high-performance results are correspondently true, even though the inside of the "black box" cannot be visualized. On this basis, the article proposes a structural-pragmatic epistemology, suggesting that justification is interpreted not negatively in terms of access to internal reasons, but positively in terms of satisfying the minimum requirements of coherence, traceability, and human accountability. The paper argues that justification in AI, both in relation to human users (as end recipients) and among AI agents, must be situated—accountable and corrigible—within a sociotechnical system that can ensure epistemic legitimacy without, therefore, presupposing total transparency or ideal rational subjects. Finally, it is argued that epistemic accountability in AI requires technical robustness on the one hand, and on the other, constant oversight based on philosophical and normative reflection on its outcomes and consequences.

Keyword: Agrippa's Trilemma, Epistemic opacity, Algorithmic justification, Structural-pragmatic epistemology, Epistemic accountability in AI.

Recibido 26-10-2025 – Aceptado 21-02-2026

Resumen

Este artículo analiza el trilema de Agrippa en el contexto de los sistemas de inteligencia artificial, específicamente en modelos algorítmicos de alta dimensión: el desafío epistemológico clásico reside en que cualquier justificación del conocimiento conduce inevitablemente a una regresión infinita, un razonamiento circular o fundamentos arbitrarios o dogmáticos, y por lo tanto, el conocimiento no es posible. Esto resurge con fuerza en los modelos de aprendizaje automático caracterizados por la opacidad, la optimización recursiva y la validación basada en el rendimiento. Los modelos de lenguaje extensos y los sistemas de recomendación son casos paradigmáticos; por lo tanto, el estudio muestra cómo la inferencia algorítmica a menudo opera sin fundamento semántico ni estructura lógica explícita, o con acceso a razones inteligibles para los humanos, generando así resultados estadísticamente robustos pero epistémicamente opacos. Se argumenta que el trilema aparece algorítmicamente en la combinación de tres fenómenos interrelacionados, a saber: 1) la opacidad causada por la arquitectura de los modelos de aprendizaje automático y las propias vías de decisión; 2) la circularidad, cuando las medidas de entrenamiento, validación y rendimiento se retroalimentan sin una referencia epistémica externa; y 3) el dogmatismo estructural, que implica dar por sentado que los resultados de alto rendimiento son correlativamente verdaderos, aunque no se pueda visualizar el interior de la "caja negra". Sobre esta base, el artículo propone una epistemología estructural-pragmática, sugiriendo que la justificación se interpreta no negativamente en términos de acceso a razones internas, sino positivamente en términos de satisfacer los requisitos mínimos de coherencia, trazabilidad y responsabilidad humana. El artículo argumenta que la justificación en IA, tanto en relación con los usuarios humanos (como destinatarios finales) como entre los agentes de IA, debe situarse —de forma responsable y corregible— dentro de un sistema sociotécnico que pueda garantizar la legitimidad epistémica sin, por lo tanto, presuponer una transparencia total ni sujetos racionales ideales. Por último, se sostiene que la rendición de cuentas epistémica en IA requiere solidez técnica por un lado y, por el otro, una supervisión constante basada en la reflexión filosófica y normativa sobre sus resultados y consecuencias.

Palabras clave: Trilema de Agripa, Opacidad epistémica, Justificación algorítmica, Epistemología pragmática estructural, Responsabilidad epistémica en IA.

1 Introduction

Recent AI systems, particularly complex machine learning models, have altered the ways we draw conclusions, make decisions, and create knowledge. Rather than merely enhancing human thinking, these systems, especially in their more complex forms, provide outputs that do not fit into normal forms of logical reasoning, cause-and-effect reasoning, or meaning-seeking understanding. This algorithmic turn presents a problem for epistemology (Heersmink et al., 2024): how can we provide an explanation for the justification of a belief, inference, or prediction when it is not generated by an epistemic subject, i.e., a human being, but rather by a black box of computational infrastructure (Durán et al., 2022; Mittelstadt et al., 2019; Ortmann, 2025).

In this situation, Agrippa's old trilemma —saying that all reasons for belief lead to endless questioning, circular reasoning, or blind acceptance—shows up in new ways in data-driven systems (Machuca, 2022; Nescolarde-Selva et al., 2025). Regression appears in the constant dependence on data to validate results (Ma & Yang, 2024); circularity manifests itself in algorithmic feedback that reinforces its patterns (Atkinson & Peijnenburg, 2017); and dogmatism emerges in the acceptance of highly performative outputs without access to explanatory mechanisms (Lipton, 2018). Faced with this scenario, many technical developments have opted for an implicit pragmatic epistemology: if it works, it's valid

(Fassio et al., 2024; MacKenzie, 2023; Sinclair, 2023). However, this position evades the philosophical demands for rigorous epistemic justification, especially in areas where algorithmic decisions affect human rights, opportunities, for example labor, or even lives, in the case of criminal law (Côté-Bouchard, 2024).

The epistemology of artificial intelligence, as argued in this article, is in need of reconceptualization. Instead of demanding total nonpragmatic transparency or complete evidence-something difficult or even impossible to provide in highly complex and algorithmically opaque systems-an attempt will be made to outline a structural-pragmatic epistemology capable of formulating minimal conditions for justification in low-traceability systems (Heersmink et al., 2016; Ortmann, 2025). Even within each of these ontological configurations, one reads categories that hinder the argument for the neutrality of algorithmic judgment and call for a better understanding of it as a distributed, situated, and governed epistemic practice (Russo et al., 2024) rather than an inherent property of the computational system (Conner, 2024).

The structure of this article is as follows: A brief introduction is provided in Section 1; Section 2 presents the theoretical framework that links Agrippa's trilemma with epistemic tensions in AI models (Machuca, 2022; Nescolarde-Selva et al., 2025). Section 3 analyzes two paradigmatic cases—large-scale language models and recommender systems—that illustrate contemporary algorithmic forms of inference without explicit justification (Durán et al., 2022; Heersmink et al., 2024). In section 4, a structural–pragmatic epistemic proposal is presented: a set of operational normative criteria for assessing the legitimacy of the output of algorithms, alongside debates on explainability and accountability (Wachter et al., 2017; Novelli et al., 2024). Lastly, section 5 explores the philosophical and practical repercussions of this proposal for the governance of automated knowledge, particularly its relevance in sensitive domains, including medicine, law, and education.

2 Conceptual framework and background

2.1 Agrippa's Trilemma and Justification in Technical Contexts

Agrippa's trilemma, often referred to as the Münchhausen problem (Schurz, 2021), poses a serious difficulty for any theory of epistemic justification. If one were to claim knowledge, one should present one of three possibilities: (a) an infinite regress of justifications, (b) a dogmatic point devoid of any other justification, or (c) justificatory circularity (Machuca, 2022; Albert, 1968; Popkin, 2003). In the traditional field of logical inference or rational argumentation, the trilemma has been examined in terms of foundationalism, coherentism or infinitism, without a definitive consensus (Nescolarde-Selva et al., 2025).

In algorithmic contexts, the trilemma takes on a new form. Machine learning systems—particularly those trained on large volumes of unstructured data—do not offer a traditional justificatory architecture (Durán et al., 2022). Their outputs do not derive from first principles or are articulated in explicit justificatory chains, as is the case with human deductive reasoning (Lipton, 2018). Instead, they are created from patterns found in data

using complex computer methods, which are often hard to understand even for the people who built them (Liu, 2025). This situation demands rethinking the very meaning of "justification" when dealing with non-symbolic agents or statistically optimized black boxes, which operate with practical success but without traditional epistemic validation criteria (Goodman & Flaxman, 2017).

2.2 Epistemic Justification in Machine Learning

In data-driven AI models, the notion of justification has been functionally replaced by that of predictive performance. From an externalist perspective, particularly the reliabilist variant, it has been proposed that a belief or computational output is justified if it is the product of a reliable process (Atkinson & Peijnenburg, 2017). However, with artificial intelligence, the focus shifts to the accuracy, robustness, and generalizability of models, reflecting a performance-oriented evaluation (O'Connor & Weatherall, 2018; Domingos, 2015). This shift demands more sophisticated understanding that goes beyond conventional reliabilist standards and recognizes the specific challenges posed by artificial intelligence systems (Floridi, 2014).

Although reliabilism, viewed from the perspective of artificial intelligence, has some shortcomings—for example, a) algorithmic opacity prevents the measurement of internal evaluations of inferences (Heersmink et al., 2024); b) bias due to the data used generates a loss of confidence in the conclusions (Durán et al., 2022; Noble, 2018); c) the circularity of model evaluation invalidates the epistemic independence of the outcome, since the success criteria are drawn from the same environments that will train the model (Crawford, 2021)— it is possible to reinterpret these characteristics of artificial intelligence as episteme.

This leads to a central normative problem: when is it epistemically responsible to take an algorithmic inference to be knowledge? One of the main insights of the article is that the answer cannot rely on purely technical performance, but rather on wider justificatory frameworks better, i.e., contrasted notions of traceability, human agency/ control, and epistemic sensitivity, but let's take this step by step (Ma & Valton, 2024; Diakopoulos, 2016).

3 A Structural and Transcendental Change in Epistemology

According to this structural realism, which evokes a Kantian line of thought, algorithms involve the organization of the very conditions of possibility for computational knowledge; they do not merely generate results (Vivas-Reyes, 2024; MacKenzie, 2023; Floridi, 2014). Therefore, machine learning models predict patterns in data while simultaneously identifying the representational structures through which conclusions acquire relevance, thus enacting a form of algorithmic reasoning —algorithmic rationality— that operates independently of human intentionality. This idea is also supported by recent philosophical reflections suggesting that epistemic practices are being implemented by data-driven systems, in the absence of semantic access or reflection, and that they replace human

judgment with technical rationality as a kind of autonomous process (Bender and Koller, 2020; Vallverdú, 2024). As Searle (1980) made clear long ago, syntax is not semantics.

This point is raised, for example, in Bernard Stiegler's critique of technical rationality two and a half decades ago (1994, 2004), who warned that algorithmic systems could replace epistemic judgment with automation procedures that would bypass normative reflection. This concern is also found in Floridi (2024), who highlights how automated information systems reorganize the conditions of knowledge, which other authors express in a similar vein, for example, when referring to autonomous information systems (Latour, 2004) as constitutive conditions of knowledge production.

When it comes to changes in the very way epistemology is done, it is no longer argued that it will eventually be replaced by technology in general, but rather by a separate (automated and specific) algorithmic counterpart, which could imply a different type of normative control and another type of philosophical critique of AI (Ma and Valton, 2024; Mittelstadt et al., 2016; Diakopoulos, 2016).

3.1 High-Dimensional Language Models: Coherence without Justification

Large-scale language models, such as GPT, Gemini, or Claude, are paradigmatic examples of non-transparent algorithmic inference. Their operation is based on the adjustment of millions (or billions) of parameters based on statistical correlations over massive textual corpora (Floridi, 2023; MacKenzie, 2023; Vallverdú, 2024). From a structural standpoint, these systems stabilize distributional regularities between language sequences, following the logic of predictive statistics rather than interpretative semantics, hence negating strong sense meanings (Bender & Koller, 2020; Huang & Wu, 2024; Ma & Valton, 2024; Searle, 1980).

Their outputs, which seem to be “intelligible” —that is, coherent responses, reasonable translations, or instructive summaries— should not be used for credit towards actual epistemic justification (Appriou et al., 2024; Zanboori et al., 2024). Not a visible connection between premises and conclusion, but rather an internal coherence inside an opaque inferential architecture, is generated (Durán et al., 2022). In this regard, the subject of epistemic justification cannot be satisfactorily addressed in conventional terms as neither the system nor the user can explain why a response is legitimate outside of its contextual or functional relevance (Ortmann, 2025; Bratman, 1992).

This phenomenon might be seen as an algorithmic expression of weak epistemic coherentism, in which beliefs (or outputs) are confirmed by their successful insertion into a local inferential network, yet lack anchoring in a global normative framework (Lehrer, 1990). From a Kantian perspective, this is a technical reason devoid of reflective judgment whereby inference is absorbed by its practical efficacy instead of being susceptible to critical assessment.

3.2 Recommendation Systems: Circularity, Performativity, and Epistemic Loops

Another area where the concept of epistemic justification is questioned is that of recommendation systems found in services such as Netflix, YouTube, and Amazon. These systems use users' current and historical activities, audience segmentation, and other forms of collaborative filtering to recommend content that optimizes dwell time or click-through rates (Schuster and Lazar, 2025; Mittelstadt et al., 2019; Wachter et al., 2017).

However, their epistemic functioning incurs a structural circularity: a) recommendations determine future user behavior, b) validated behavior feeds back into the model, and c) the model confirms its "effectiveness" based on the feedback it has induced (Côté-Bouchard, 2024).

This performative epistemic loop helps to avoid separating statistical self-justification from genuine reasoning. Building on Agrippa's trilemma, this scenario shows a type of naturalized algorithmic circularity in which the source of justification is integrated into the system itself without the possibility of external validation (Machuca, 2022; Nescolarde-Selva et al., 2025).

While this may sound exaggerated in the case of platforms with recommendation algorithms, it is not so much so in cases where algorithmic decisions can permanently affect people's lives, for example, in the case of judicial systems using AI or in human resources departments that decide who to hire or who to dismiss, the same thing happens in the banking and financial system (Vuković et al., 2025).

The pragmatic epistemology of success (if it works, therefore, it is valid) conflicts with normative frameworks that demand transparency, fairness, and the possibility of rational dissent (Ma & Valton, 2024; 2024; Novelli et al., 2024). This highlights the need to develop forms of epistemic governance that integrate explanation, human intervention, and institutional control (Daoust & Côté-Bouchard, 2023).

3.3 Explainability, Robustness, and Meaningful Human Control

In the face of these tensions, proposals have emerged that focus on algorithmic explainability as an epistemic criterion. However, not all explainability is equivalent to justification. Models that offer attention visualizations, saliency maps, or counterfactual paths may be useful from a technical or legal perspective, but they do not resolve the core of the epistemic problem: In what sense is what the model infers as output justified?

Some proposals — e.g., contrastive explanation (Lipton, 2018) or local causal attribution (Wachter et al., 2017) — attempt to fill this gap, enabling the user to compare inferential alternatives as well as a causal dependence of each input in determining the output. But these approaches make an assumption about a meaningful human control architecture, in which the human agent is able to intervene and understand the system's inference, be able to interpret, and ultimately reject it. In the absence of such an architecture,

then, justification becomes an internal feature of the system: technically effective, perhaps, but philosophically wanting (Heersmink et al., 2016).

4 Proposal: A Structural-Pragmatic Epistemology Applicable to AI

4.1 Minimal Conditions for the Epistemic Justification of Opaque Models

To the extent that high-dimensional models are not structurally transparent, a theory of epistemic justification is proposed here that neither presupposes total transparency nor completely abandons normative standards (Appriou et al., 2024; Dong & Zhou, 2025; Huang & Wu, 2024). This plan combines aspects of structural realism (Ladyman & Ross, 2007), epistemic pragmatism (MacKenzie, 2023), and a theory of action that intersects technical artifacts with action, derived from distributed and situated cognitive epistemology (Longino, 2020. Therefore, three conditions are required for an algorithmic outcome to be epistemically justified:

a) Internal structural coherence; the system must be embedded in a structurally consistent network; that is, the inference must be consistent with the regularities inferred by the model and show stability in the face of small perturbations, as demonstrated by adversarial robustness tests (Ye, 2023; Yu et al., 2025; Zhou et al., 2025).

b) Supervised external validation; not in terms of absolute truth, but in terms of reliable performance under human-traceable criteria, the system must be subject to supervised external validation processes that allow its behavior to be audited under counterfactual or alternative contexts (Mittelstadt et al., 2019; Durán et al., 2022).

c) Epistemic sensitivity to human intervention; the model must be adaptable to human agents who can evaluate, modify, or suspend its inferences. The algorithmic episteme must remain open to the practical and ethical judgment of the human environment in which it operates (Wachter et al., 2017).

This idea focuses on viewing AI systems as part of a broader distributed epistemic practice, in which people and technical objects together form a specifically constituted epistemic community that generates meaningful judgments, rather than as epistemic agents in their own right (Longino, 2020; Vivas-Reyes, 2024). In this respect, justification was not a mere internal aspect of the system, which one could simply read off at face value, but a stituated relation between model, target, institutional envrionments, and evaluative agents in which normative arrangement, interpretation, and situtated evaluation frame factors intervened (Barad, 2007; Ravish & Sirola, 2023).

4.2 Algorithmic Knowledge as a Situated Epistemic Practice

This idea focuses on viewing AI systems as part of a broader distributed epistemic practice, in which people and technical objects together form a specifically constituted epistemic community that generates meaningful judgments, rather than as epistemic agents in their own right (Longino, 2020; Vivas-Reyes, 2024). In this respect, justification was not a mere

internal aspect of the system, which one could simply read off at face value, but a stituated relation between model, target, institutional envrionments, and evaluative agents in which normative arrangement, interpretation, and situtated evaluation frame factors intervened (Barad, 2007; Ravish & Sirola, 2023).

The idea of situated epistemic practice is based on recent work on distributed epistemology (Longino, 2020; Haraway, 1991; Balayla, 2024) and translated to the ontological realm, since models are not epistemologically neutral, but rather situate—and reproduce—the social, political, and epistemic conditions of their design, formation, and deployment (Côté-Bouchard, 2024; Rossi, 2025). Therefore, their validity cannot be assessed in isolation but must take into account the data collection methods, possible forms of intervention, and the accountability architecture in which they are embedded (Novelli et al., 2024).

4.3 Epistemic Governance and Algorithmic Accountability

This proposal means that it is possible to observe how algorithms make decisions without taking into account how they are managed and the ethical issues involved (Goodman and Flaxman, 2017). The technical functioning of the system cannot be separated from its value in terms of knowledge if we want to avoid a collapse of knowledge standards in systems that, although functioning correctly, may present biases, hide errors, or limit the possibility of reasonable disagreement. This implies that:

a) AI architectures should meet certain accountability requirements at both the functional and epistemic levels such as transparency, traceability, and critical auditability (Ma & Valton, 2024).

b) Explainability needs to open up to mechanisms of epistemic dissent and human review, with genuine normative bite, not just as an advisory council on wheels (Lipton, 2018).

c) Validation should be diverse and include different kinds of knowledge, and not be homogeneous and referential to itself. Hence, what is true and relevant is context dependent from a variety of fields of knowledge and judgment that extend beyond algorithmic "reason" (Floridi, 2024).

The argument here, therefore, argues for a move away from the machine—which treats the epistemic value of technical production as a given—and toward a sociotechnical systems epistemology, in which what counts as knowledge is the product of a distributed practice, historically situated and normatively susceptible to human scrutiny (Barad, 2007; Haraway, 1991; Longino, 2020; Minazzi, 2022; Novelli et al., 2024).

4.4 Possible Objections and Responses

A first objection that could be raised against this approach is that, by privileging criteria such as structural coherence or sensitivity to human intervention, the notion of truth would be replaced by a mere functional operability (Merwe, 2025). From a classical epistemological

point of view, this would be considered a surrender of the ideals of objectivity and rational justification to instrumental pragmatism (Sinclair, 2023). Wouldn't truth be eliminated as an epistemic measure if the rationality of a system could be justified in the sense that its inference pattern is coherent or useful? It is worth noting that the strategy proposed here does not deny the importance of truth, but rather seeks to reinstate it within a situated epistemic practice (Longino, 2020; Haraway, 1991; Zhang & Li, 2024). Truth cannot be seen as direct correspondence in algorithmic environments lacking direct access to intentional grounds or representations; rather, it can only be seen as the structural stabilization of reliable connections (Ladyman & Ross, 2007). In this paradigm, the concept of justification is transformed rather than disappeared; it is not a given, but a continuous activity of situated, traceable, and correctable judgment (MacKenzie, 2023).

A second pertinent criticism is that adding human intervention as an epistemic condition seems to be a retreat toward subjectivity or relativism. If an inference can be presented in a way that is susceptible to human inspection or correction, wouldn't this inject a subjective element into what is supposed to be an impartial process? This is an objection rooted in an overly narrow conception of objectivity (Zoglauer, 2023). The image of human intervention presented here is not that of an arbiter, but that of an episteme (Wachter et al., 2017): it is precisely what prevents systems from closing their own inferences back on themselves, thus producing self-referential loops (Ortmann, 2025). If this principle is transferred to contexts where algorithms impact social, political, or legal decisions, the ability to interrupt, redirect, or audit inference does not weaken its epistemic value but rather guarantees its inclusion in a broader rational community, where disagreements, alternatives, and corrections are possible (Floridi, 2024; Canale, 2021).

Finally, it could be objected that this epistemology, although philosophically sophisticated, offers few operational criteria for the technical design and evaluation of real-world systems. How can this approach concretely guide the creation of AI models if it does not specify applicable metrics, protocols, or standards? The answer is that the purpose of the article is not to dictate technical procedures, but rather to provide a normative philosophical framework to guide design, implementation, and evaluation practices (Mittelstadt et al., 2019; Ma & Valton, 2024). This framework can—and should—be translated into functional requirements, such as decision traceability, the possibility of counterfactual simulations, the creation of meaningful human review mechanisms, or the use of structural robustness tests (Zhou et al., 2025). Rather than replacing technical design, this epistemology proposes conditions of intelligibility and legitimacy that allow algorithmic developments to be responsibly integrated into human knowledge practices in favor of humanity itself.

5 Conclusion

Artificial intelligence systems based on high-dimensional machine learning represent a profound transformation in the way knowledge and the traditional epistemic approach are conceived and justified. Given the impossibility of applying traditional criteria of justification—such as transparent deduction, foundational evidence, or explicit causal

inference—it is proposed to standardize a shift toward functional, empirical, and performative forms of validation. However, this shift is not without epistemic and even normative risks. Agrippa's trilemma does not disappear with algorithmic efficiency. What is argued here is that the trilemma is reformulated in technical-computational environments as an unresolved conflict between opacity, circularity, and structural dogmatism. In response, a structural-pragmatic epistemology is suggested that allows for the articulation of minimum criteria for epistemic justification without relying on perfect transparency or an ideal epistemic subject. Technical pragmatism and situated practice theory are based on structural realism. This method allows us to evaluate algorithmic conclusions as products of a broader sociotechnical architecture, which includes formal, institutional, and ethical-normative characteristics.

In this sense, epistemic justification becomes a matter of distributed governance and accountability, rather than a mere internal attribute of the model. This suggestion has several ramifications: on the one hand, it presents a compelling philosophical basis for addressing explainability, reliability, and accountability in AI systems; on the other, it provides a conceptual framework for building institutional epistemic practices that accompany the application of algorithmic technologies in sensitive settings, such as medicine, education, law, and public administration. Future research could explore the concrete implications of these perspectives on normative imperatives and how cooperation among philosophers, engineers, and policymakers could generate a more robust, open, and equitable epistemic field for algorithmic decision-making in society.

The algorithmic shift in modern artificial intelligence has transformed human tools of inference and reinterpreted the basic conditions under which knowledge claims are proposed, accepted, and applied. Machine learning systems reflect ways of thinking that evade common sense ideas about justification and responsibility and do not promote human rationality. Any discourse prior to, one might say, the AI boom we are currently experiencing is, so to speak, obsolete.

As argued throughout this essay, an epistemic problem arises that is not merely technical but philosophical: it demands a reconsideration of how belief, inference, and prediction can be justified in the absence of an epistemic subject. Agrippa's classic trilemma receives updated expression in the form of computational infrastructures that produce regression without understanding, circularities without context, and dogmatisms hidden behind the rhetoric of performance. In response, a structural-pragmatic epistemology has been proposed that redefines justification as a practice situated, governed, and correctable within sociotechnical systems. For AI to meaningfully participate in knowledge production, its epistemic outputs must be open to critique, review, and oversight. Only by anchoring algorithmic inference in practices of collective justification can we ensure that knowledge transformation in this domain remains compatible with the ideals of reason, responsibility, and situated algorithmic legitimacy, yet, at the same time, epistemically justified.

References

Albert, H. (1968). Traktat über kritische Vernunft. J.C.B. Mohr (Paul Siebeck).

Appriou, T., Rullière, D., & Gaudrie, D. (2024). High-dimensional Bayesian optimization with a combination of Kriging models. Structural and Multidisciplinary Optimization, 67, 196. https://doi.org/10.1007/s00158-024-03906-8

Atkinson, D., & Peijnenburg, J. (2017). Epistemic Justification. In Fading Foundations (Synthese Library, Vol. 383). Springer, Cham. https://doi.org/10.1007/978-3-319-58295-5_2

Atkinson, D., & Peijnenburg, J. (2017). Loops and Networks. In Fading Foundations (Synthese Library, Vol. 383). Springer, Cham. https://doi.org/10.1007/978-3-319-58295-5_8

Atkinson, D., & Peijnenburg, J. (2017). The Probabilistic Regress. In Fading Foundations (Synthese Library, Vol. 383). Springer, Cham. https://doi.org/10.1007/978-3-319-58295-5_3

Balayla, J. (2024). Applications in Bayesian Epistemology and Artificial Intelligence (AI). In Theorems on the Prevalence Threshold and the Geometry of Screening Curves. Springer, Cham. https://doi.org/10.1007/978-3-031-71452-8_11

Barad, K. (2007). Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning. Duke University Press.

Bender, E. M., & Koller, A. (2020, July). Climbing towards NLU: On meaning, form, and understanding in the age of data. In Proceedings of the 58th annual meeting of the association for computational linguistics (pp. 5185-5198). https://aclanthology.org/2020.acl-main.463.pdf

Bratman, M. E. (1992). Practical reason and acceptance in a context. Ethics, 102(3), 441–456. https://www.jstor.org/stable/2254116

Canale, D. (2021). The Opacity of Law: On the Hidden Impact of Experts' Opinion on Legal Decision-making. Law and Philosophy, 40, 509–543. https://doi.org/10.1007/s10982-021-09408-8

Carr, N. (2010). The Shallows: What the Internet Is Doing to Our Brains. W. W. Norton & Company.

Conner, W. (2024). Radical epistemology, theory choice, and the priority of the epistemic. Synthese, 203, 33. https://doi.org/10.1007/s11229-023-04448-0

Côté-Bouchard, C. (2024b). Should we Trust Our Feeds? Social Media, Misinformation, and the Epistemology of Testimony. Topoi, 43, 1469–1486. https://doi.org/10.1007/s11245-024-10116-w

Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs. Yale University Press.

Diakopoulos, N. (2016). Algorithmic accountability: Journalistic investigation of black boxes. Digital Journalism, 4(7), 802-811. https://doi.org/10.7916/D8ZK5TW2

Durán, J. M., Sand, M., & Jongsma, K. (2022). The ethics and epistemology of explanatory AI in medicine and healthcare. Ethics and Information Technology, 24 , 42. https://doi.org/10.1007/s10676-022-09666-7

Daoust, M. K., & Côté-Bouchard, C. (2023). Epistemic Consequentialism, Veritism, and Scoring Rules. Erkenntnis, 88, 1741–1765. https://doi.org/10.1007/s10670-021-00426-5

Domingos, P. (2015). The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books.

Dong, Y., & Zhou, X. (2025). Ah-knockoff: false discovery rate control in high-dimensional additive hazards models. Journal of the Korean Statistical Society. Advance online publication. https://doi.org/10.1007/s42952-025-00317-3

Durán, J. M., Sand, M., & Jongsma, K. (2022). The ethics and epistemology of explanatory AI in medicine and healthcare. Ethics and Information Technology, 24, 42. https://doi.org/10.1007/s10676-022-09666-7

Fassio, D., Tang, W. H., & Ye, R. (2024). Introduction to Current Themes in Epistemology: Asian Epistemology Network. Asian Journal of Philosophy, 3, 87. https://doi.org/10.1007/s44204-024-00218-y

Floridi, L. (2014). The Fourth Revolution: How the Infosphere is Reshaping Human Reality. Oxford University Press.

Floridi, L. (2023). AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models. Philosophy & Technology, 36(15), 1-7. https://doi.org/10.1007/s13347-023-00621-y

Floridi, L. (2024). Introduction to the Special Issues: The Ethics of Artificial Intelligence: Exacerbated Problems, Renewed Problems, Unprecedented Problems. American Philosophical Quarterly, 61(4), 301-307. https://doi.org/10.5406/21521123.61.4.01

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". AI magazine, 38(3), 50-57. https://doi.org/10.1609/aimag.v38i3.2741

Haraway, D. J. (1991). Simians, cyborgs, and women: The reinvention of nature. Routledge.

Hayles, N. K. (2017). Unthought: The Power of the Cognitive Nonconscious. University of Chicago Press.

Heersmink, R., de Rooij, B., Clavel Vázquez, M. J., et al. (2024). A phenomenology and epistemology of large language models: transparency, trust, and trustworthiness. Ethics and Information Technology, 26, 41. https://doi.org/10.1007/s10676-024-09777-3

Huang, J., & Wu, Y. (2024). High-dimensional robust inference for censored linear models. Science China Mathematics, 67, 891–918. https://doi.org/10.1007/s11425-022-2070-2

Ladyman, J., & Ross, D. (2007). Every Thing Must Go: Metaphysics Naturalized. Oxford University Press.

Lehrer, K. (1990). Theory of Knowledge. Routledge.

Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36–43. https://doi.org/10.1145/3233231

Liu, B. (2025). Demystifying the black box: AI-enhanced logistic regression for lead scoring. Applied Intelligence, 55, 574. https://doi.org/10.1007/s10489-025-06430-4

Longino, H. E. (2020). Interaction: a case for ontological pluralism. Interdisciplinary Science Reviews, 45(3), 432-445. https://doi.org/10.1080/03080188.2020.1794385

Ma, J., & Yang, S. (2024). High-dimensional stochastic control models for newsvendor problems and deep learning resolution.3 Annals of Operations Research, 339, 789–811. https://doi.org/10.1007/s10479-024-05872-2

Ma, W., & Valton, V. (2024). Toward an Ethics of AI Belief. Philosophy & Technology, 37, 76. https://doi.org/10.1007/s13347-024-00762-8

Machuca, D. E. (2022). The Agrippan Modes and the Challenge of Disagreement. In Pyrrhonism Past and Present (Synthese Library, Vol. 450). Springer, Cham. https://doi.org/10.1007/978-3-030-91210-9_4

MacKenzie, A. (2023). Postdigital Epistemology. In P. Jandrić (Ed.), Encyclopedia of Postdigital Science and Education. Springer, Cham. https://doi.org/10.1007/978-3-031-35469-4_9-1

MacKenzie, D. (2023). Trading at the speed of light: Thoughts on automation, explanation, and epistemology in financial AI. Theory, Culture & Society, 40(1), 125–144. https://doi.org/10.1177/02632764221113422

Mao, J., Gao, Z., Jing, B. Y., et al. (2024). On the statistical analysis of high-dimensional factor models. Statistical Papers, 65, 4991–5019. https://doi.org/10.1007/s00362-024-01557-x

Merwe, R. (2025). Perspectives and meta-perspectives: context versus hierarchy in the epistemology of complex systems. European Journal for Philosophy of Science, 15, 14. https://doi.org/10.1007/s13194-025-00641-9

Minazzi, F. (2022). For a Historical-Critical Epistemology. From the Criticism of Epistemology to Critical Epistemology. In Historical Epistemology and European Philosophy of Science (Studies in Applied Philosophy, Epistemology and Rational Ethics, Vol. 62). Springer, Cham. https://doi.org/10.1007/978-3-030-96332-3_3

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. In Proceedings of the conference on fairness, accountability, and transparency (pp. 279-288). https://doi.org/10.48550/arXiv.1811.01439

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), https://doi.org/10.1177/2053951716679

Nescolarde-Selva, J. A., Usó-Doménech, J. L., Segura-Abad, L., et al. (2025). Beliefs, Epistemic Regress and Doxastic Justification. Foundations of Science, 30, 109–147. https://doi.org/10.1007/s10699-023-09927-8

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. New York University Press.

Novelli, C., Taddeo, M., & Floridi, L. (2024). Accountability in artificial intelligence: what it is and how it works. AI and Society, 39, 1871–1882. https://doi.org/10.1007/s00146-023-01635-y

O'Connor, C., & Weatherall, J. O. (2018). The Misinformation Age: How False Beliefs Spread in the Age of AI. Yale University Press.

Ortmann, J. (2025). Of opaque oracles: epistemic dependence on AI in science poses no novel problems for social epistemology. Synthese, 205, 80. https://doi.org/10.1007/s11229-025-04930-x

Popkin, R. H. (2003). The history of scepticism from Savonarola to Bayle (Revised Edition). Oxford University Press.

Ravish, S., & Sirola, V. S. (2023). Can Social Reflective Equilibrium Delineate Cornell Realist Epistemology? Philosophia, 51, 2015–2033. https://doi.org/10.1007/s11406-023-00654-9

Rossi, E. (2025). What Can Epistemic Normativity Tell us About Politics? Ideology, Power, and the Epistemology of Radical Realism. Topoi, 44, 77–88. https://doi.org/10.1007/s11245-024-10142-8

Russo, F., Schliesser, E., & Wagemans, J. (2024). Connecting ethics and epistemology of AI. AI and Society, 39, 1585–1603. https://doi.org/10.1007/s00146-022-01617-6

Schurz, G. (2021). Der Begriff des Wissens. In: Erkenntnistheorie. J.B. Metzler, Stuttgart. https://doi.org/10.1007/978-3-476-04755-7_2

Schuster, N. & Lazar, S. (2025). Attention, moral skill, and algorithmic recommendation. Philosophical Studies, 182, 159–184. https://doi.org/10.1007/s11098-023-02083-6

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–457. https://doi.org/10.1017/S0140525X00005756

Shi, H., Yang, W., Sun, B., & Guo, X. (2025). Tests for high-dimensional partially linear regression models. Statistical Papers, 66, 59. https://doi.org/10.1007/s00362-025-01679-w

Sinclair, R. (2023). Précis of Quine, Conceptual Pragmatism, and the Analytic-Synthetic Distinction. Asian Journal of Philosophy, 2, 63. https://doi.org/10.1007/s44204-023-00115-w

Stiegler, B. (1994). La technique et le temps. Tome 1: La faute d'Épiméthée. Galilée.

Stiegler, B. (2004). De la misère symbolique. Tome 1: L'époque hyperindustrielle. Galilée.

Vallverdú, J. (2024). Generative AI and Causality. In Causality for Artificial Intelligence. Springer, Singapore. https://doi.org/10.1007/978-981-97-3187-9_6

Vivas-Reyes, R. (2024). Clashing perspectives: Kantian epistemology and quantum chemistry theory. Foundations of Chemistry, 26, 291–300. https://doi.org/10.1007/s10698-024-09508-y

Vuković, D. B., Dekpo-Adza, S., & Matović, S. (2025). AI integration in financial services: a systematic review of trends and regulatory challenges. Humanities and Social Sciences Communications, 12(1), 562. https://doi.org/10.1057/s41599-025-04850-8

Wachter, S., Mittelstadt, B. D., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76–99.1 https://doi.org/10.1093/idpl/ipx005

Ye, F. (2023). Epistemology and Methodology. In Studies in No-Self Physicalism. Springer, Singapore. https://doi.org/10.1007/978-981-19-8143-2_6

Yu, K., Guo, X., & Luo, S. (2025). Group inference for high-dimensional mediation models. Statistical Computation, 35, 61. https://doi.org/10.1007/s11222-025-10591-0

Zanboori, A., Zanboori, E., Mousavi, M., et al. (2024). Bayesian stein-type shrinkage estimators in high-dimensional linear regression models. São Paulo Journal of Mathematical Sciences, 18, 1889–1914. https://doi.org/10.1007/s40863-024-00473-0

Zhang, X., & Li, Z. (2024). Linear hypothesis testing in ultra high dimensional generalized linear mixed models. Journal of the Korean Statistical Society, 53, 791–814. https://doi.org/10.1007/s42952-024-00268-1

Zhou, Y., Zhang, X., & Kwong, S. (2025). Introduction of High Dimensional Machine Learning. In Computational Intelligence for High-Dimensional Machine Learning (SpringerBriefs in Computer Science). Springer, Singapore. https://doi.org/10.1007/978-981-96-2687-8_1

Zoglauer, T. (2023). Post-Truth Epistemology. In Constructed Truths. Springer Vieweg, Wiesbaden. https://doi.org/10.1007/978-3-658-39942-9_2

UNIVERSIDAD
DEL ZULIA

REVISTA DE FILOSOFÍA

Nº 115 - 2026 - 1 ENERO - MARZO

Esta revista fue editada en formato digital y publicada en MARZO de 2025

por el ***Fondo Editorial Serbiluz, Universidad del Zulia. Maracaibo-Venezuela***

www.luz.edu.ve **www.serbi.luz.edu.ve**
www.produccioncientificaluz.org