Science, Education and Innovations in the Context of Modern Problems
Issue 1, Vol. 9, 2026

RESEARCH ARTICLE

The Problem of Epistemic Opacity in Machine Learning Models

Yamina Abid
Doctor
University of Amar Thlidji
Algeria
Email : a.abid@lagh-univ.dz , Orcid: 0009-0003-3243-2114

Ameur Brahimi
Doctor
University of Amar Thlidji
Algeria
Email: ma.brahimi@lagh-univ.dz ; Orcid:0001-0008-3243-2113

Issue web link https://imcra-az.org/archive/389-science-education-and-innovations-in-the-context-of-modern-problems-issue-1-vol-9-2026.html

Keywords Epistemic Opacity, Machine Learning, Scientific Explanation, Philosophy of Science, Black Box, Epistemic Understanding.

Abstract

This research aims to investigate the radical epistemological challenges posed by the concept of "Opacity" in machine learning models within the context of contemporary philosophy of science. The core problem lies in the widening gap between the high predictive efficiency of these models and their failure to provide transparent, human-intelligible "scientific explanations." The study employs a descriptive-analytical method to deconstruct the algorithmic "black box" structure and question the ability of these models to generate objective knowledge that transcends mere statistical correlations. The key findings indicate that epistemic opacity is not merely a temporary technical hurdle but a structural feature of data-intensive science, leading to the erosion of the classical "deductive-nomological" model of explanation. The research concludes that the concept of "scientific understanding" must be redefined to align with the era of algorithmic mediation, suggesting a shift from representational realism to a structural realism that acknowledges the limits of human cognition in the face of computational complexity.

Citation

Yamina A; Ameur B. (2026). The Problem of Epistemic Opacity in Machine Learning Models. *Science, Education and Innovations in the Context of Modern Problems*, 9(1), 810-816. https://doi.org/10.56334/sei/9.1.73

Received: 24.05.2025 | Accepted: 25.11.2025 | Published: 06.01.2026 (available online)

Introduction

Classical philosophy of science is grounded in an epistemological presupposition according to which *scientificity* is conditional upon **inferential transparency**. Since the formulation of the deductive–nomological model, philosophical consciousness has largely converged on the view that scientific explanation attains legitimacy only insofar as the relation between premises and conclusions is governed by logical rules of derivation that are clear and cognitively accessible to human understanding. Within this framework, the function of scientific theory is to provide a transparent ontological representation of the mechanisms governing phenomena, thereby ensuring that the transition from empirical data to predictive outputs remains open to epistemic scrutiny and critical evaluation.

However, the beginning of the twenty-first century has witnessed a radical transformation in the structure of scientific practice, driven by the rise of machine learning models and data-intensive artificial intelligence. This shift has generated an epistemological rupture with the traditional explanatory paradigm, as scientific inquiry has moved away from models

grounded in prior theoretical hypotheses toward approaches centered on the extraction of probabilistic patterns from massive data streams. This transformation has given rise to a problematic condition commonly referred to as **"epistemological opacity."** Such opacity does not denote a lack of accuracy or predictive adequacy ; rather, it signifies the absence of epistemic traceability and interpretive accessibility.

Indeed, deep neural networks—with their non-linear transformations and millions of latent variables—produce a structural gap between instrumental success and theoretical understanding. The knowing subject is no longer capable of reconstructing the logical justifications underlying scientific outputs, nor of penetrating the internal architecture of the "black box," which operates beyond the traditional bounds of human cognitive categories. This situation confronts us with a crisis of legitimacy that affects the very notion of scientific rationality, now oscillating between remarkable predictive efficiency and a structurally embedded explanatory deficit. Consequently, this development necessitates a fundamental re-examination of the nature of scientific inference—particularly inference to the best explanation—in an epistemic landscape increasingly dominated by complex statistical correlations rather than transparent causal laws. At stake, therefore, is a profound philosophical question concerning whether such opaque models can still be regarded as sources of scientific knowledge in the strict sense, or whether they instead signal a shift toward a new form of **structural realism** that relinquishes the aspiration to comprehensive understanding in favor of functional control over complexity.

Literature Review

In "*Extending Ourselves*: *Computational Science, Empiricism, and Scientific Method* (Oxford University Press, 2004, Oxford, UK), **Paul Humphreys** argues that science has entered an "Epistemically Opaque Regime," where machines surpass human cognitive limits, decoupling operational success from human understanding. Similarly, **Zachary C. Lipton**, in "*The Mythos of Model Interpretability*" (ACM Queue, vol. 16, no. 3, 2018, pp. 31-57, New York, USA), deconstructs the transparency myth, viewing XAI explanations as "post-hoc narratives" that satisfy human intuition rather than reflecting actual algorithmic logic. This is further supported by **Rozenblit** and **Keil** in "*The misunderstood limits of folk science: an illusion of explanatory depth*" (Cognitive Science, vol. 26, no. 5, 2002, pp. 521-562, NJ, USA), who demonstrate that epistemic opacity is exacerbated by a cognitive fallacy where the subject mistakenly believes they understand complex systems.

Research Contribution

This research transcends the technical and psychological descriptions of opacity by offering a Transcendental Critique. It treats opacity as an "Ontological Fact" redefining the essence of science and critiques "Digital Anthropomorphism" as an "Ontological Twilight of Reason." Ultimately, it proposes an "Ethics of Opacity" for the post-human sovereignty era, moving beyond the mere attempt to open the "black box.

The central research problem of this study lies in examining the **epistemological reliability** of opaque scientific models, through addressing the following question: *Can a scientific model legitimately qualify as knowledge in the absence of explanatory transparency?* And how does the acceptance of epistemological opacity as a constitutive feature reshape the logical and ontological structure of scientific inquiry in the twenty-first century?

Sub-questions:

- What is the logical and semantic distinction between opacity as a technical property resulting from computational complexity and opacity as an epistemological limit confronting the knowing subject?
- To what extent does reliance on opaque models contribute to the marginalization of "theory" in favor of instrumental modeling within contemporary natural and cognitive sciences?
- How can a form of **structural realism** be articulated that remains coherent with the presence of non-interpretable algorithmic mediations at the core of scientific practice?

Research Hypothesis: This study proceeds from the hypothesis that epistemological opacity is not a temporary technological obstacle, but rather a structural transformation in the criteria of scientific validity—from **explanatory understanding** to **functional efficiency**. Such a shift necessitates the adoption of a **perspectival epistemology** that acknowledges the limits of human rationality and treats algorithmic outputs as objective relational structures that do not require intrinsic transparency.

Significance of the Study: The significance of this research lies in its engagement with the most recent international debates in the philosophy of science concerning the limits of **data-driven science**. By advancing a rigorous epistemological critique that transcends purely technical approaches, the study contributes to a redefinition of the concept of **scientific understanding** in the age of high-performance computation, and offers an original theoretical framework for addressing the "black box" problem as a fundamentally philosophical and logical issue.

In order to comprehensively address the problematics of our study and to delve more deeply into their theoretical and epistemological dimensions, we have adopted a research plan structured around two main sections. The first section examines **the ontology of opacity and the collapse of the paradigm of explanatory transparency**, focusing on the deconstruction of the concept of epistemological opacity, the formulation of a rigorous epistemological characterization of algorithmic opacity, the elucidation of the disintegration of classical explanatory models, and the analysis of the dialectic between understanding and knowledge within opaque models that have profoundly reshaped the conditions of

contemporary scientific rationality.The second section is devoted to interrogating **the illusion of explanation and the psychology of false understanding**, through a critical examination of the paradigm of explainable artificial intelligence as a new epistemological mask, an analysis of the psychology of the illusion of explanatory depth, and an exposure of the forms of epistemological responsibility collapse in the age of algorithmic opacity.The study concludes with a synthetic conclusion in which the most significant theoretical findings are highlighted, accompanied by a set of epistemological and ethical recommendations, thereby opening a renewed critical horizon for rethinking knowledge and rationality in the era of artificial intelligence.

1. The Ontology of Opacity and the Collapse of the Paradigm of Explanatory Transparency

The radical transformation brought about by artificial intelligence in the structure of contemporary science represents a genuine *rupture* in the epistemological fabric that has long governed scientific rationality. What we are facing is not merely a technical difficulty, but a profound shift in the *regime of truth*: knowledge increasingly appears as an autonomous entity no longer inhabited by human *understanding*, but rather by *computational efficiency*. This section constitutes an archaeological attempt to dismantle the structure of *opacity*, not as a mere lack of information, but as a *regime of truth* that undermines the historical sovereignty of the knowing subject.

1.1 The Concept of Epistemological Opacity

Epistemological opacity (*Epistemological Opacity*) manifests itself as a state of *cognitive concealment* that affects the relationship between the rational subject and the object of knowledge. At its linguistic core, as indicated by Ibn Manẓūr, the concept is associated with two central meanings: delay and darkness. He explains that *"al-ʿatma"* refers to "the first third of the night after the disappearance of twilight; it is also said to be the moment when the darkness of the night intensifies," and adds that *"ʿatama ʿan al-shayʾ"* means "to refrain from something after having intended it, or to delay it" (Ibn Manẓūr, 1994, p. 385).

This semantic field resonates with what André Lalande presents in his philosophical vocabulary, where *obscurité* appears as the opposite of clarity and distinction, designating representations that lack rational penetrability and logical transparency (Lalande, 2010, p. 697).

From this etymological foundation, the concept moved from its *physical* domain to an *epistemological* one, where Gaston Bachelard reconceptualized opacity as an *epistemological obstacle*. In this sense, opacity is the accumulation of common-sense beliefs and unexamined assumptions that render reason itself "opaque" and incapable of producing the necessary *epistemological rupture* required for scientific clarity (Bachelard, 1986, p. 14) .

However, the concept acquired a more technical and contemporary meaning with Paul Humphreys, who offered a rigorous definition according to which: *"a process is epistemically opaque for a cognitive agent X at time t if X does not know, at time t, all of the epistemically relevant elements of the process"* (Humphreys, 2009, p. 618). Humphreys thus distinguishes between *weak opacity*, resulting from contingent or temporary limitations, and *essential opacity*, which stems from the inherent impossibility of the human mind grasping the structural complexity of certain systems (Pedro, 2020, p. 4).

Within the same analytical framework, Wendy Parker emphasizes that such opacity becomes unavoidable when humans are unable to inspect and justify every computational step performed by a computer due to its sheer scale (Parker, *Values and Evidence in Simulation*, 2014) (Parker, 2014, p. 5). This leads directly to the problem of the *black box*, wherein internal operations remain epistemically concealed despite the transparency of inputs and outputs.

This synthesis of the linguistic heritage emphasizing delay and concealment, together with contemporary epistemological accounts acknowledging non-transparency in the face of technological complexity, renders epistemological opacity a central concept for understanding the limits of contemporary rationality when confronted with systems that exceed the capacity of human traceability.

1.2 The Epistemological Characterization of Algorithmic Opacity

The problem of opacity (*Opacity*) constitutes a condition of *structural non-transparency* that strikes at the very core of the relationship between the subject and the object of knowledge. It is not a contingent ambiguity, but an epistemological threshold separating the computational logic of machines from the human cognitive apparatus.

In this context, Paul Humphreys grounds what he calls *essential epistemological opacity*, arguing that computational science has ushered us into a condition in which the human mind can no longer encompass the underlying logical processes. According to Humphreys, *"a process is epistemically opaque if some of the ontological or logical elements relevant to the process are not computationally accessible to any human agent"* (Humphreys, 2004, p. 147).

This analysis reveals that knowledge has become *external* to consciousness. Jenna Burrell reinforces this view by arguing that opacity is not the result of deliberate concealment, but rather of a *structural mismatch* between high-dimensional mathematics and human cognition. As she states: *"Machine learning opacity is the result of a mismatch between mathematical optimization in high-dimensional spaces and human modes of reasoning, which tend toward simplification"* (Burrell, 2016 , p. 5).

This condition marks a sharp rupture between computation and intuition. Opacity is not merely a *technical shortcoming*, but a crisis at the very heart of human *understanding*, which is unable to keep pace with digital flux. It is a moment of

epistemological alienation in which reason loses its ability to follow the internal logic of machine-generated results, transforming science from a *construction of concepts* into a mere *reception of outputs*. Daniel Andler captures this rupture succinctly when he writes: *"Opacity is not a lack of data, but a break in the thread of meaning; it is the condition in which understanding disappears behind a wall of mute digital operations"* (Andler, 2016 , p. 210).
Accordingly, opacity becomes the ontological gap separating *computational intelligence* from the *meaning* constructed by the knowing subject. This situation may be understood as a failure of the language of thought itself: we possess the technical keys but lack the conceptual language required to interrogate the machine. Computational science has thus produced a semantic gap between what machines can calculate and what humans can understand.

1.3 The Collapse of Classical Explanatory Models

Explanatory transparency long constituted the organizing framework of the relationship between science and reality since Galileo. Today, however, this paradigm is collapsing under the dominance of correlation-based models that replace the question *"why?"* with *"what?"*.
This collapse becomes evident when confronting Carl Hempel's model, which required scientific explanation to take the form of a logical argument deducing the explanandum from universal laws. Hempel maintained that scientific explanation must be a logical argument deriving the phenomenon from general natural laws; without this connection, science degenerates into mere surface description (Hempel, 1965, p. 337). Yet algorithmic reality shatters this logical bridge, revealing the fracture of the classical explanatory model that has prevailed since the scientific revolution.
The breaking of this *logical bridge* entails the dismantling of deductive reasoning that connects universals (laws) to particulars (phenomena) through logical necessity. In this context, the algorithm ceases to function as an instrument of demonstration and becomes an instrument of *prediction*. It no longer concerns itself with essences or final causes, but confines itself to detecting recurrent patterns in *Big Data.*
This shift from causality to correlation empties the concept of *natural law* of its ontological weight, reducing it to a surplus or even a dispensable entity in the world of digital simulation. Here, epistemological opacity reaches its apex: algorithms achieve remarkable statistical success while leaving reason in a state of *explanatory emptiness*. Mario Alai explicitly acknowledges this when he writes: *"The deductive bridge has collapsed; algorithms do not infer results but predict them statistically, rendering the concept of natural law superfluous"* (Alai, 2021, p. 8). We thus find ourselves before an eminently *instrumentalist* scientific model, in which theory dissolves behind the brute efficiency of computation.
From the same perspective, Henk de Regt emphasizes that the absence of transparency precludes the formation of any *explanatory argument*. In his award-winning book *Understanding Scientific Understanding*, he states: *"In opaque models, we possess outputs but lose logical justifications; this explanatory vacuum transforms science into an instrumental practice devoid of a transparent theoretical core"* (De Regt, 2017, p. 92) .
This epistemic imbalance culminates in what is often referred to as the *end of theory*, where the deluge of data renders the scientific method obsolete: correlation alone is deemed sufficient, especially when data abundance allows numbers to "speak for themselves." Sabina Leonelli forcefully criticizes this correlational tendency, arguing that it turns science into a form of *blind empiricism*. She shows that an exclusive focus on patterns leads to tracking phenomena without understanding their causal mechanisms, there by threatening to reduce science to the mere efficient management of data (Leonelli, 2016).

1.4 The Dialectic of Understanding and Knowledge in Opaque Models

The epistemic crisis manifests most clearly in the widening gap between the possession of information and the apprehension of meaning. Can algorithmic outputs legitimately be called *knowledge* if they lack human understanding?
This question implicitly demands a reassertion of *intelligibility* as a necessary condition for scientific understanding, distinguishing it from mere predictive adequacy, which may remain procedurally opaque. Understanding, in this sense, is not simply the logical derivation of phenomena from laws, but a cognitive capacity enabling the knowing subject to grasp the internal relations within a representational model and deploy them inferentially. When De Regt describes opacity as the enemy of understanding, he refers to the functional rupture that occurs when scientists can no longer conceptualize the internal mechanisms of a model, thereby transforming science into a *black box* that generates accurate predictions without providing epistemic insight. Consequently, scientific understanding requires the existence of *cognitively accessible structures* that function as a conceptual bridge between theoretical abstraction and mental representation. Understanding is thus context-sensitive and profoundly pragmatic, refusing to settle for mere statistical correlation and demanding instead the ontological transparency of explanatory mechanisms. *"Opacity," De Regt concludes, "is the enemy of scientific understanding because it severs the link between the representational model and cognitively accessible structures"* (De Regt, 2017, p. 101) .
This rupture situates us within what Frank Pasquale terms the *"black box society,"* where authority shifts from methodological transparency to pragmatic success in outputs. Pasquale explains that the black box represents a transformation from the ideal of scientific transparency to the brute effectiveness of results, whereby epistemic authority is exercised without the need for rational justification. (Pasquale, 2015, p. 44)

The absence of semantic clarity turns science into a form of *technological magic*, in which interpretive wisdom is replaced by blind algorithmic control. Precise prediction without understanding severs the ontological link between model and reality, reducing knowledge to a purely instrumental yet epistemically opaque tool. We thus acquire the capacity to tame phenomena without grasping their essence, entrenching a divide between technical efficiency and philosophical insight into the mechanisms of nature. This alienation transforms the world into an object of control rather than understanding, allowing predictive power to masquerade as a false substitute for critical vision and conceptual clarity. As Michela Massimi observes, *"Opaque models provide a mathematically successful and objective perspective, yet one that remains epistemically inaccessible to the human mind; this confronts us with the fact that machines do not see the world as we do"* (Massimi, 2022, p. 310).

This shift ultimately signals the end of understanding as a condition of knowledge, insofar as understanding presupposes cognitive grasp, and opacity blocks such grasp. The knowing subject is thus reduced from a producer of truth to a passive consumer of machine-generated outputs that it does not comprehend (Elgin, 2017, p. 44).

Epistemological opacity marks the ontological displacement of the scientific object from the realm of 'intelligibility' to the realm of 'computational efficiency.' The collapse of the transparent model signifies a radical rupture: science is no longer a mirror reflecting causal essences, but a functional mediation that manages complexity without dissolving it. In this new regime, we possess 'knowledge-without-understanding,' where the predictive success of the black box stands as a silent monument to the limits of human cognitive sovereignty.

2. The Deception of Interpretation and the Psychology of Illusory Understanding

In this section, we move from describing *opacity* as an ontological structure to analyzing the *deception of interpretation* as an epistemological and psychological phenomenon. The contemporary problem does not lie merely in the opacity of models, but in the fact that attempts to "humanize" them through so-called *Explainable Artificial Intelligence* (XAI) have produced a form of *artificial transparency*—one that grants the human agent a false sense of understanding while the underlying mechanisms remain fundamentally inaccessible.

2.1. The Explainable AI (XAI) Paradigm as an Epistemological Mask

Technical attempts to render algorithms interpretable amount, in most cases, to reductive translations that do not necessarily reflect the truth embedded within the black box. What emerges here is a deep tension between *truthfulness* and *interpretability*.

In this context, Zachary Lipton offers a radical critique of the concept of interpretation in machine learning, arguing that what we commonly call an explanation is, in fact, a post hoc narrative that fails to express the system's internal logic. Lipton emphasizes that the pursuit of interpretability often results in the construction of simplified models that deceive users into believing that the system operates according to human-like reasoning, even though such explanations may be profoundly unfaithful to the actual computational mechanisms that produced the decision (Lipton, 2018, pp. 31-57.).

We are thus confronted with a genuine *epistemological deception*: interpretation here does not function as a disclosure of truth but rather as a *mask* that conceals opacity. Algorithmic explanations may be mobilized to justify biases instead of exposing them. As noted in the literature, "explanations in opaque systems can function as instruments of false legitimacy; people tend to trust outputs when they are accompanied by a seemingly logical justification, even if that justification bears no relation to how the neural network actually processed the data" (Bender, 2021, p. 12).

2.2. The Psychology of the Illusion of Explanatory Depth

Why do we accept the superficial explanations provided by machines? The answer lies in a specifically human *cognitive limitation*: a tendency to favor *narratives* over complex *computational processes*. From this perspective, Leonid Rozenblit and Frank Keil analyze the phenomenon known as the *illusion of explanatory depth*—a condition in which individuals believe they understand a complex system while, in reality, they grasp only its surface features.

Rozenblit and Keil argue that humans possess an inflated confidence in their understanding of complex systems; when an algorithm provides a visual explanation or a heat map, it reinforces this illusion, causing us to suspend deeper inquiry under the false impression that we have grasped the underlying cause (Rozenblit, 2002, p. 522).

This illusion is precisely what designers of opaque systems exploit to legitimize machine decisions. As meaning-seeking beings—even within randomness—the human mind is predisposed to constructing coherent stories from fragmented data. Consequently, any algorithmic output accompanied by a *reason* will appear convincing, because we tend to prefer *illusory certainty* over complex and opaque truth (Kahneman, 2011, p. 199).

In a similar vein, Catherine Elgin argues in *True Enough* that *understanding* can become detached from *truth* in favor of *epistemic adequacy*. She maintains that we often accept models that are "true enough" because they facilitate practical engagement with reality; however, in the context of artificial intelligence, this acceptance turns into an epistemological trap that obscures the fact that we have lost logical control over our objects of inquiry (Elgin C. Z., 2017, p. 15).

2.3. The Collapse of Epistemological Responsibility in the Age of Opacity

When we abandon genuine understanding in favor of illusory comprehension, we engage in what may be called *epistemological delegation*, a process that leads to the erosion of both ethical and scientific responsibility. Black boxes are not merely technical artifacts; they are *socio-technical assemblages* that absolve us from the labor of thinking. Once a

system is deemed successful, the black box is sealed, and accountability ceases. As Bruno Latour warns, "the danger of algorithmic science lies in having built machines that relieve us of the obligation to justify, rendering us 'silent witnesses' to truths we can no longer defend" (Latour, 1987, p. 131).

This silence resonates with what Michel Foucault elsewhere describes as the transformation of the *will to knowledge* into a *will to power*: the algorithm does not explain in order to reveal, but in order to impose a reality. Yuval Noah Harari likewise analyzes this erosion of human sovereignty, arguing that we are witnessing a transition from human authority grounded in understanding and intuition to algorithmic authority grounded in data. The explanations we receive, he contends, function as an *epistemological opiate* designed to soothe our anxiety over the loss of control (Harari, 2018, p. 45). Thinking has thus been progressively delegated to intelligent systems that act on our behalf in choosing, ranking, interpreting, and even shaping opinions themselves. The dominance of artificial intelligence over decision-making processes manifests not only in its technical power, but in its acceptance as a comfortable epistemic substitute—one that spares us the effort of reflection and persuades us to settle for ready-made, rapid, and conceptually shallow outputs. In this way, the logical sovereignty of the knowing subject gradually erodes: we lose the capacity to delimit our objects, construct our problematics, and interrogate our assumptions, in favor of a computational rationality governed by probability and optimization rather than truth and meaning.

It is for this reason that Sabina Leonelli warns of the *death of the knowing subject*. She argues that "if science continues to rely on opaque models wrapped in spurious explanations, we will end up producing knowledge without knowers—that is, an accumulation of successful results without human agents capable of assuming epistemic responsibility for them" (Leonelli, 2016, p. 189).

We are transitioning from Epistemic Agency to Algorithmic Subjugation. The Black Box does not explain; it imposes. What we term 'understanding' is merely a Post-human Alienation, where the human subject is reduced to a silent witness of a computational truth it can no longer verify. This is the Indigence of Reason in the age of optimization.

The real danger, then, does not lie in the intelligence of the machine, but in the inertia of the human mind that has consented to transform itself from an epistemic agent into a cognitive consumer, from a thinking subject into a mere interface through which algorithmic suggestions pass without critical resistance or rational—and ethical—interrogation.

Conclusion

In concluding this study, we find ourselves confronted not merely with a technical transformation in the tools of scientific research, but with an **"ontological eclipse"** that has redefined the distance between reason and reality. Algorithmic opacity, whose dimensions we have deconstructed in the preceding sections, is not a mere procedural obstacle ; rather, it constitutes a zero point that announces the end of the sovereignty of the **"knowing subject"** in its Kantian and Hempelian sense, and the advent of an era of **knowledge detached from understanding**.

Overview of the General Findings of the Study

On the basis of the analysis presented, the core results of this research can be articulated in the following analytical points :

- **The collapse of the transparency paradigm :**

The study has demonstrated that *explanatory transparency* is no longer a necessary condition for the production of scientific truth. Contemporary science, as Paul Humphreys argues, has shifted toward an **inherently opaque epistemological regime**, in which models achieve remarkable predictive success without passing through the "filter of human understanding." This means that opacity has become a structural feature of knowledge production rather than a temporary cognitive deficiency.

- **The decline of the classical explanatory model :**

We have concluded that Carl Hempel's *deductive-nomological model* has collapsed in the face of an emerging *correlational turn*. Science no longer primarily seeks universal laws, but rather statistical patterns within massive datasets. This transformation threatens to reduce science to a form of **blind empiricism**, devoid of a transparent theoretical core, thereby producing a historical rupture between *power* (prediction) and *wisdom* (understanding).

- **The fallacy of post-hoc explanations:**

The analysis in the second chapter revealed that attempts to "humanize" algorithms through *explainable artificial intelligence (XAI)* are, in most cases, efforts to construct **artificial transparency**. Our findings support Zachary Lipton's thesis (Lipton, 2018) that such explanations function as *simplified narratives* that mislead human agents by creating an illusion of explanatory depth, while the underlying computational logic remains fundamentally resistant to genuine understanding.

- **The fragmentation of the knowing subject:**

The study concludes that we are witnessing an **epistemological fragmentation**: the human is no longer the "legislator" of nature, but has become a *consumer* of outputs generated by a machine that exceeds human cognitive limits. As Michela Massimi notes (Massimi, 2022, p. 310), we now inhabit a form of **perspectival realism**, in which the machine constructs a reality of its own that does not necessarily intersect with classical human rationality.

Recommendations

In light of these unsettling yet intellectually provocative findings, the study recommends the adoption of new intellectual and methodological strategies for engaging with the **"age of opacity"**:

- **Redefining "scientific understanding":**

The study calls for a decoupling of *knowledge* from *immediate human understanding*. Philosophy of science must develop new conceptualizations of understanding appropriate to *machine mediation*, such that a model need not be "mentally surveyable" to be scientifically legitimate. Instead, alternative criteria of **functional transparency** should be explored.

- **Establishing a new epistemological humility:**

Rather than adhering to a form of scientific hubris that claims comprehensive mastery, scientists and researchers must adopt a stance of **epistemological humility**. This entails acknowledging the limits of human rationality in the face of algorithmic complexity and engaging with black-box outputs through sustained critical caution, without succumbing to the *illusion of explanatory depth*.

- **Ethical governance of opacity**

Opacity must not remain a *black hole* from which decision-makers escape accountability. We recommend the development of **epistemological regulations** that compel technical institutions to provide *responsibility-sufficient explanations*, rather than merely technical ones. When transparency fails to be technical, it must become **political**.

- **Integrating the humanities into data laboratories:**

The design of algorithms should not be left to engineers alone. The study recommends integrating philosophers, linguists, and epistemologists into the stages of model construction, to ensure that *human values* and *semantic clarity* remain embedded within design criteria, rather than being superficial adjustments added after the fact.

We stand today on the threshold of a **second Enlightenment**—one that does not place the human at the center, but situates humanity within a **complex dialogue** with a digital "other" unlike itself. Algorithmic opacity is not the end of science; it is the beginning of a new science—one that demands the courage to relinquish illusions of total control and to seek truth within the gray zones between what we know and what machines are capable of computing. The greatest challenge is not to *open the black box*, but to learn how to live, create, and ground ethics in a world in which **opacity has become the new light**. Understanding, in this sense, is the capacity to inhabit a world whose complete map we no longer possess: it is the adventure of reason in confrontation with the digital infinite.

Ethical Considerations

This study is theoretical and philosophical in nature and is based exclusively on the analysis of existing scientific literature, conceptual frameworks, and publicly available academic sources related to machine learning and epistemology. No empirical data involving human participants, personal data, experiments, or proprietary algorithms were used. Consequently, ethical approval was not required. The research adheres to principles of academic integrity, transparency, and responsible scholarship, ensuring accurate citation, faithful interpretation of sources, and avoidance of misrepresentation.

Author Contributions

Yamina Abid contributed to the conceptualization of the study, development of the theoretical framework, philosophical analysis, and drafting of the manuscript.
Ameur Brahimi contributed to the methodological design, critical review of epistemological arguments, refinement of the analytical sections, and revision of the manuscript.
Both authors jointly reviewed, edited, and approved the final version of the article.

Acknowledgements

The authors would like to express their sincere appreciation to the Laboratory of Philosophy and Social Development at the University of Amar Thlidji, Laghouat, for providing an intellectually supportive research environment. They also thank colleagues for their constructive discussions and scholarly feedback that contributed to the refinement of this study.

Funding Statement

This research did not receive any specific grant from public, commercial, or non-profit funding agencies.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Bibliography

Alai, M. (2021). *Models and Theories in Science.* Urbino, Italy: University of Urbino (Department of Pure and Applied Sciences).
Andler, D. (2016). *La Silhouette de l'humain: Enquête sur la ligne de front de l'IA.* Paris, Éditions Gallimard.
Bachelard, G. (1986). *La Formation de l'esprit scientifique: contribution à une psychanalyse de la connaissance objective* (éd. 13e édition). Paris: Librairie Philosophique J. Vrin.
Bender, E. M. (2021, March). "On the Dangers of Stochastic Parrots". *FAccT '21,*. Récupéré sur https://dl.acm.org/doi/10.1145/3442188.3445922

Burrell, J. B. (2016 , January-June). Understanding opacity in machine learning algorithms. *Big Data & Society, Vol. 3*(Issue 1).
De Regt, H. W. (2017). *Understanding Scientific Understanding.* New York, USA: Oxford University Press.
Elgin, C. Z. (2017). *True Enough.* Cambridge: Massachusetts, MIT Press. Récupéré sur https://mitpress.mit.edu/9780262036535/true-enough/
Elgin, C. Z. (2017). *True Enough.* Cambridge, MA, USA: MIT Press.
Harari, Y. N. (2018). *21 Lessons for the 21st Century.* New York, USA: Spiegel & Grau. Récupéré sur https://www.ynharari.com/book/21-lessons-for-the-21st-century/
Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* New York, USA: The Free Press (Macmillan Publishing Co.
Humphreys, P. (2004). *xtending Ourselves: Computational Science, Empiricism, and Scientific Method.* New York: Oxford University Press.
Humphreys, P. (2009). The Philosophical Novelty of Computer Simulation Methods,. *Synthese, Vol. 169, No. 3*(No. 3), p. 618.
Ibn Manẓūr, d. a.-D. (1994). *Lisān al-ʿArab (The Tongue of the Arabs).* Beirut: Ṣādir publications.
Kahneman, D. (2011). *Thinking, Fast and Slow.* New York, USA: Farrar, Straus and Giroux.
Lalande, A. (2010). *Vocabulaire technique et critique de la philosophie* (éd. 18e éd .coll. « Quadrige »). Paris: Presses Universitaires de France (PUF).
Latour, B. (1987). *Science in Action.* Cambridge (MA), USA: Harvard University Press. Récupéré sur https://www.hup.harvard.edu/books/9780674792913
Leonelli, S. (2016). *Data-Centric Biology: A Philosophical Study.* Chicago, IL, USA: University of Chicago Press.
Leonelli, S. (2016). *Data-Centric Biology: A Philosophical Study.* Chicago, USA: University of Chicago Press.
Lipton, Z. C. (2018, June 15). "The Mythos of Model Interpretability". *ACM Queue, vol. 16*(no. 3). Récupéré sur https://dl.acm.org/doi/10.1145/3236386.3241340
Massimi, M. (2022). *Perspectival Realism.* Oxford, UK:: Oxford University Press.
Parker, W. (2014). Values and Evidence in Simulation. *The Oxford Handbook of Philosophy ofScience*, p. 5.
Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information.* Cambridge, MA, USA: Harvard University Press.
Pedro, I. S. (2020, p. 4). *Degrees of Epistemic Opacity*. Récupéré sur PhilSci-Archive: https://philsci-archive.pitt.edu/17124/
Rozenblit, L. &. (2002, Sept). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science, vol. 26, no. 5,*. Récupéré sur https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog2605_1