IDS-RNI0335 | NAN-RDG000288

Nanopublication — L'opacité de l'IA n'a personne pour en divulguer le codex

par Arnaud Quercy

La limite que cette source trace pour le cadre tient à ceci : l'invariant institué exige-t-il un sujet intentionnel et interrogeable ? Tout l'appareil de divulgation du cadre — explicabilité (P19a), ricochet (P18a), induction du proto-codex (P14a) — est construit pour des pratiques où quelqu'un, même maladroitement, détient la règle. La littérature en épistémologie de l'IA recensée ici soulève, sans la trancher, l'hypothèse de cas où la règle n'a jamais été détenue par personne : induite à partir de données, lisible seulement (si tant est qu'elle le soit) par une interprétabilité mécaniste encore immature. Le cadre n'a pas de vocabulaire pour un invariant institué par personne ; la source nomme précisément la condition (ouvrir la boîte noire) qui trancherait si un tel cas est réel ou seulement apparent.

CE QUE DIT LA SOURCE [1]

L'article recense les travaux philosophiques sur la question de savoir si les humains sont épistémiquement justifiés à former des croyances, à revendiquer un savoir ou à atteindre une compréhension à partir des sorties de l'IA. Il traite l'IA à la fois comme objet d'enquête — qu'est-ce que l'opacité, l'interprétabilité, l'explicabilité, d'où viennent-elles, comment les réduire — et comme outil ou quasi-agent fondant la croyance humaine. Il passe en revue les théories de la justification (fiabilisme, connaissance instrumentale, justification inductive/statistique des résultats d'apprentissage automatique) et se demande si des modèles boîte noire peuvent néanmoins produire une compréhension scientifique, concluant que la plupart de ces questions restent ouvertes et disputées.

OÙ LA LENTILLE MORD

L'axe donné/institué de P3a et la condition d'explicabilité de P19a touchent directement le traitement de l'opacité en section 2. La définition de Humphreys — l'opacité comme ignorance par un agent des « éléments épistémiquement pertinents » d'un processus — se lit presque comme une version formelle de la distinction donné/institué : ce qu'un capteur ou un utilisateur peut relever en surface (le donné) contre la règle qui élit des relations dans cette surface et qui exigerait une divulgation pour être lue (l'institué). L'inquiétude de l'article lui-même quant à savoir si les modèles d'apprentissage automatique « contiennent de telles hypothèses », qui resteront « inconnues tant que la boîte noire n'est pas ouverte », est structurellement la même inquiétude que nomme P19a : un invariant qui existe mais reste muet tant qu'il n'est pas articulé dans un dialogue.

OÙ ELLE CASSE

La prise se dérobe précisément là où la source elle-même signale l'indécision : la condition d'explicabilité de P19a et le ricochet de P18a présupposent tous deux un sujet qui détient la règle et peut, en dialogue, répondre « pourquoi ces choix ». L'article rapporte que pour beaucoup de réseaux de neurones, un tel sujet pourrait ne pas exister même en principe — « les humains n'implémentent pas d'hypothèses substantielles sur leurs cibles dans les RNA », si bien que « la question de savoir si de telles hypothèses sont correctement implémentées ne se pose pas ». Ce n'est pas un artiste qui tairait ou ne saurait pas articuler un invariant (le cas de P19a) ; c'est l'absence possible de quiconque, humain ou non, ayant jamais détenu la règle. Cela se dérobe SI l'on lit « invariant institué » comme exigeant un bâtisseur de codex intentionnel et interrogeable en dialogue (le modèle du ricochet, P18a) ; cela tient — ne se brise pas — SI l'on lit l'invariant institué de P3a simplement comme « la règle qui élit des relations dans la surface », indépendante d'un agent, auquel cas les poids d'un réseau entraîné sont simplement un invariant institué que personne n'a encore déchiffré, non différent en nature du proto-codex de P14a, juste bien plus difficile à induire. La source n'arbitre pas entre ces deux lectures ; elle les laisse comme une question ouverte.

GRILLE PROPOSITIONNELLE [2]

Lire cette grille. « Non évaluée » ne signifie **pas** que la source n’a rien à dire sur cette proposition. Cela signifie que **la lecture ne l’a pas instruite**. Ce n’est pas un résultat,

et cela ne se compte dans aucun dénominateur. Un silence, lui, est un constat : la lecture a touché la proposition sans conclure, ou a relevé elle-même son absence.

Évaluées

proposition		statut d'audit	verdict relationnel
P3a	Double Structure of the Invariant	**audité**	met sous pression
P14a	The Proto-Codex	non audité	laisse ouvert
P18a	The Ricochet Effect	non audité	laisse ouvert
P19a	Explicability as Condition	non audité	met sous pression

Ce qui a été instruit. Ces propositions ont été jugées et portent un énoncé courant : c'est **lui** que la lecture a testé, pas l'énoncé fondateur.

- **P3a** — jugée en Note 05 · énoncé courant : say GIVEN invariant and INSTITUTED invariant, never 'physical' and 'intentional' — the axis is MODE OF EXISTENCE, read as regime of access: the given is what a sensor measures on the surface WITHOUT the codex; the instituted is the rule that elects relations in that surface and that only the disclosed codex deciphers

Non évaluées (33) — P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P14b, P15, P16, P16a, P17, P18, P19, P20, P21, P22, P23, P24, P25, P26, P27, P28, P29, P30, P31

Verdict d'ensemble : Bounded / Open — boundary case — taxonomie de la passe d'août 2026, antérieure au passage au verdict relationnel. Conservé comme métadonnée : l'écart entre les deux formes est une donnée, pas un résidu. Le jugement de cette lecture est dans la grille ci-dessus.

Lecture du 2026-10-17 · modèle claude-sonnet-5 · source 138,094 car. · 1 tentative(s)

RÉFÉRENCES

[1] Epistemology of Artificial Intelligence. Stanford Encyclopedia of Philosophy, 2026-10-17. https://plato.stanford.edu/entries/epistemology-ai/

[2] Quercy, A. (2026). The 31 Propositions of Ideamorphism. https://ideamorphism.org/fr/publications/2026/03/les-31-propositions-de-lideamorphisme.html

PROFIL ÉPISTÉMIQUE

Type de revendication	boundary reading
Voix	third person analysis
Statut épistémique	bounded grip
Méthodologie	Boundary-finding lens applied to one source, model claude-sonnet-5, 1 attempt(s), traceability check passed.

SOMME DE CONTRÔLE (SHA-256)

01a71fb159193b3ab75ae62ec4b1693187d3f88aa12e459384bc98fb046a1fb5

Sous licence Creative Commons Attribution 4.0 International (CC BY 4.0)

Auteur Arnaud Quercy
Asset code IDS-RNI0335
Identifiant NAN-RDG000288
Version 1
Publié le 2026-10-17

ISSN : demande en cours — Centre ISSN américain, Library of Congress

© 2026 Études idéamorphiques

Publié par Art Quam Anima Publishing New York,
un nom commercial d'AQA Publishing LLC

c/o Northwest Registered Agent, 418 Broadway Ste N
Albany, NY 12207, USA
+1 917-764-5470

NY DOS ID 7624376 · Assumed Name ID 7658709

publishing.artquamanima.com

Dernière mise à jour : 2026-10-17

URI persistante : https://ideamorphism.org/fr/nanopubs/2026/10/IDS-RNI0335-ai-opacity-has-no-codex-bearer-to-disclose-it.pdf