

THE HERMENEUTICS IN THE MACHINE: RETHINKING ETHICS AND POLITICS IN KNOWLEDGE INFRASTRUCTURES

ALBERTO ROMELE*

Abstract: *What does it mean to interpret in the age of algorithmic knowledge? This keynote addresses the ethical and political stakes of knowledge infrastructures shaped by artificial intelligence, with a focus on classification systems — from traditional documentation practices to complex AI-driven data architectures. Against the prevailing myth of automation and neutrality, I propose a digital hermeneutics as both a method and a critique. On one hand, digital hermeneutics operates as a repair mechanism: it helps make sense of misunderstandings, reveals hidden assumptions, and fosters reflexivity within algorithmic systems. On the other hand, it functions as a critical tool for unmasking the implicit worldviews encoded in datasets, taxonomies, and recommendation algorithms. Classification, I argue, is always a political and interpretive act — one that shapes not only what is knowable, but also what becomes visible, actionable, or excluded. Artificial intelligence should be understood as a machine of classification and world-making: a producer of data imaginaries. These imaginaries, often presented as smooth, seamless, and apolitical, participate in organizing not only information, but also our collective hopes, fears, and expectations. In this context, the ethical and political challenge is not merely to improve transparency or fairness in AI systems, but to rethink the very frameworks through which knowledge is structured and legitimized. By reframing ethics and politics as hermeneutics, and knowledge organization as a site of symbolic and material struggle, this talk calls for the development of interpretive infrastructures — ones that make room for agonism, contestation, and epistemic pluralism within the machine.*

Keywords: Artificial Intelligence; Classification systems; Digital hermeneutics.

Resumo: *O que significa interpretar na era do conhecimento algorítmico? Esta comunicação aborda os desafios éticos e políticos das infraestruturas de conhecimento moldadas pela inteligência artificial, com foco nos sistemas de classificação — desde práticas tradicionais de documentação até arquiteturas complexas de dados impulsionadas por IA. Contra o mito predominante da automação e neutralidade, proponho uma hermenêutica digital como método e crítica. Por um lado, a hermenêutica digital funciona como um mecanismo de reparação: ajuda a compreender mal-entendidos, revela suposições ocultas e promove a reflexividade dentro dos sistemas algorítmicos. Por outro lado, funciona como uma ferramenta crítica para desmascarar as visões de mundo implícitas codificadas em conjuntos de dados, taxonomias e algoritmos de recomendação. A classificação, argumento eu, é sempre um ato político e interpretativo — que molda não apenas o que é cognoscível, mas também o que se torna visível, actionável ou excluído. A inteligência artificial deve ser entendida como uma máquina de classificação e criação de mundos: uma produtora de imaginários de dados. Esses imaginários, frequentemente apresentados como suaves, contínuos e apolíticos, participam da organização não apenas da informação, mas também das nossas esperanças, medos e expectativas coletivas. Neste contexto, o desafio ético e político não é apenas melhorar a transparência ou a equidade nos sistemas de IA, mas repensar as próprias estruturas através das quais o conhecimento é estruturado e legitimado. Ao reformular a ética e a política como hermenêutica, e a organização do conhecimento*

* IRMÉCCEN, Sorbonne Nouvelle. Email: alberto.romele@sorbonne-nouvelle.fr. ORCID: <https://orcid.org/0000-0001-8648-6619>.

como um local de luta simbólica e material, esta comunicação apela ao desenvolvimento de infraestruturas interpretativas — que abram espaço para o agonismo, a contestação e o pluralismo epistémico dentro da máquina.

Palavras-chave: Inteligência Artificial; Sistemas de classificação; Hermenêutica digital.

1. FROM CLASSIC HERMENEUTICS TO MATERIAL HERMENEUTICS

To interpret is to bridge a gap — between signs and meanings, data and sense, part and whole. Since its origins, hermeneutics has been concerned with the conditions under which understanding becomes possible. From the exegesis of sacred texts to the interpretation of legal or literary works, hermeneutics developed as a discipline of interpretation: a reflection on how we come to make sense of something that speaks to us, but not in our own voice. Yet in the present moment, this question takes on a new urgency. As knowledge increasingly circulates through algorithmic infrastructures — search engines, recommendation systems, large language models — the act of interpretation is no longer a solely human prerogative. Machines now «read», «classify», and «respond». They produce meaning-effects, and in doing so they participate in shaping what counts as knowledge, what becomes visible, and what remains silent or invisible. The age of algorithmic knowledge, therefore, calls for a renewed reflection on the very meaning of interpretation.

Classically, hermeneutics has been understood in two main senses: a technical or methodological sense and an ontological one. The technical (or methodological) sense of hermeneutics refers to the set of rules, procedures, and interpretive principles that allow for a correct understanding of texts. It is, in a sense, a technology of interpretation: a craft that seeks to establish reliable methods for reconstructing meaning. From Schleiermacher to Dilthey, hermeneutics was conceived as an epistemological enterprise — a way of securing understanding within the human sciences, analogous to explanation (*Erklären*) in the natural sciences. Its goal was to codify the conditions of interpretive validity, that is, how to reconstruct an author's intention, how to move between the parts and the whole of a text, how to distinguish misunderstanding from comprehension.

The ontological turn, inaugurated by Heidegger and radicalized by Gadamer, displaced this perspective. Hermeneutics was no longer about *how* we interpret, but *what it means to be* an interpreting being. Understanding is not one activity among others — it is the very mode of our existence in the world. For Heidegger, *Dasein* is always already interpretive: we do not first exist and then interpret, but exist *as* interpretation. Gadamer extended this view by showing that understanding is always

historically situated, conditioned by language, tradition, and prejudice (*Vorurteil*). There is no «neutral» standpoint outside interpretation. Every act of understanding is embedded in a horizon of meanings that both enables and limits it.

The distinction between a technical and an ontological hermeneutics has often been perceived as a stark alternative — as if one had to choose between *method* and *truth*. Paul Ricoeur famously captured this tension in his reading of Gadamer's *Truth and Method*: for him, the conjunction «and» in the title should rather be read as a disjunction — *Truth OR Method*. Ricoeur saw in Gadamer's project an explicit refusal of the methodological ideal inherited from the human sciences, a turn toward the ontological dimension of understanding that resists codification.

Yet Ricoeur also sought to move beyond this opposition. His proposal can be summed up in a simple yet profound formula: we must explain more in order to understand better. In other words, the path to understanding passes through explanation. Ricoeur's «long route» (*la voie longue*) is not a rejection of method but its reorientation. It consists in taking seriously the methodological rigor of the human and social sciences — for him, especially the formal precision of structural linguistics — while acknowledging that such rigor ultimately opens onto something irreducible: an ontological remainder that cannot be fully captured by any methodological apparatus. It is in this remainder, in this slight but decisive gap between explanation and understanding, that truth emerges. Interpretation thus oscillates between two poles: the analytic and the existential, the procedural and the revelatory. And perhaps what makes hermeneutics so vital today is precisely this capacity to hold the tension between them.

In the contemporary landscape, this gap appears to be narrowing. With the rise of digital humanities and computational sociology, methodological formalization has reached unprecedented levels. Algorithms now perform what once seemed the most human of interpretive acts: detecting patterns, modeling behaviors, even identifying the «style» or «signature» of a text or an image. These methods demystify the notion of genius and challenge traditional ideas of uniqueness in creation and conduct. They show that meaning, style, and behavior are distributed phenomena, emerging from networks, data, and repetition.

And yet, even within these large-scale approaches, something resists full formalization. Franco Moretti's celebrated essay *The Slaughterhouse of Literature* (2000) offers a revealing example. By comparing dozens of forgotten nineteenth-century detective novels, Moretti adopts a quantitative and morphological perspective to explain why certain narrative forms survived while others disappeared. In this large-scale comparison, Arthur Conan Doyle's fiction emerges not as the expression of a singular genius, but as a formal mutation within a literary system. What distinguishes it is the introduction of a new narrative device — the trace — which reorganizes the relation between text and reader, transforming reading itself into an act of inference and interpretation. Paradoxically,

it is through such systematic, even quasi-algorithmic analysis that the interpretive dimension of literature reappears: method leads us back to meaning.

Classical hermeneutics placed texts at the center of interpretation. To interpret meant to engage with a work of language — a text, a discourse, a symbol — in order to disclose the world it opens up. The text was both medium and object of understanding. Yet in the late twentieth century, a new strand of thought began to shift this focus from texts to technologies. Don Ihde's *Technology and the Lifeworld* (1990) is exemplary of this transformation. Ihde proposed what he called a material or technological hermeneutics, which can be understood in both a narrow and a broad sense.

In its narrow sense, material hermeneutics refers to a particular class of technologies — those that mediate our access to the world through representations that themselves require interpretation. These are devices that produce a view of reality, but not reality itself: the book, the map, the microscope, the radar screen, or the cockpit of an airplane. Each offers a depiction of the world — or of a specific portion of it — that must be correctly read in order for the world to appear. In this sense, a novel and a navigation instrument share a family resemblance: both are interfaces of understanding. They do not replace the world but configure our way of being in it. In its broader sense, however, Ihde argues that all technologies are hermeneutic, for every technology mediates perception and action through what he famously calls magnification-reduction structures. To use a technology is always to gain and to lose: to amplify certain aspects of the world while obscuring or neglecting others. The telescope magnifies celestial detail but erases peripheral vision; the smartphone brings the distant near but compresses the spatiality of everyday life. Technologies thus interpret the world for us, shaping the very field of what can be seen, heard, or acted upon.

This broader hermeneutics displaces the human interpreter from the center of the scene. Interpretation is no longer a purely human affair that occurs after the fact of perception. It is a distributed process, involving humans, artifacts, and symbolic systems in complex networks of mediation. Between the subject who interprets and the thing interpreted, there always stand not only cultural frameworks — worldviews, beliefs, prejudices — but also material mediations: tools, instruments, interfaces. And these mediations are never neutral. They are saturated with symbolic and social meanings, often sedimented into their very design. Langdon Winner's (1980) well-known example of the low overpasses designed by Robert Moses on Long Island makes this point clear. According to Winner, these bridges were built deliberately low to prevent buses — and thus, at the time, lower-income and racialized communities — from accessing the parkways leading to the beaches. Whether or not the historical detail is fully accurate, the example illustrates the deeper insight: artifacts themselves can embody political intentions and exclusions.

2. FROM MATERIAL HERMENEUTICS TO DIGITAL HERMENEUTICS

Digital technologies, too, can be understood as hermeneutic — and indeed, in both the narrow and the broad senses outlined above. In the narrow sense, digital technologies produce representations of the world that must themselves be interpreted. They mediate our access to reality through data, models, and interfaces that do not simply depict but actively construct what we perceive as «the world». In this sense, there is perhaps nothing more hermeneutic than data itself. As has often been noted, data are never simply «given» (*data*), but «taken» (*capta*). They are the result of prior acts of selection, formatting, and categorization — acts that decide what counts as information and what is left aside. Every data set is thus an interpretation of the world before it becomes an input to further interpretation. The same applies to the methods by which data are processed, correlated, and visualized. A statistical model, a clustering algorithm, or a network graph performs what Paul Ricoeur would have called a *mise en intrigue*: it configures a story out of discrete elements, establishing relations of causality, relevance, and meaning. A data visualization, in turn, is not a transparent window but a condensed and selective rendering — an interpretive synthesis that both reveals and conceals. To «read» data, then, is to engage in a hermeneutic act: to move between part and whole, detail and pattern, signal and context. And to design data infrastructures is to define, in advance, the possible horizons of interpretation.

In the broader sense, digital technologies are hermeneutic because they organize how the world can be experienced, acted upon, and understood. The digital environment is not only a space of representation but a machine of classification. Every act of computation — from indexing a document to recommending a video or generating an image — involves a process of sorting, labeling, and correlating. Algorithms do not simply reflect our categories; they instantiate them, making distinctions operational. In this sense, classification systems are the backbone of the digital lifeworld. They determine what appears as relevant, similar, anomalous, or deviant. They constitute, in short, the conditions of intelligibility within the digital milieu. The digital does not merely process preexisting knowledge; it continuously reorganizes the space of what can be known. In doing so, it performs what Jacques Rancière (2019) calls a *partage du sensible* — a distribution of the sensible that determines what is visible or invisible, sayable or unsayable, audible or unheard. Every classificatory system, whether bureaucratic or algorithmic, participates in such a distribution. It decides what counts as relevant information, which patterns are recognized as meaningful, and which signals are dismissed as noise. These decisions, though often presented as technical or neutral, are in fact deeply political: they shape the contours of the common world, defining who or what can appear within it.

Before turning to the ethical and political implications of algorithmic knowledge infrastructures, I would like to pause briefly and clarify what I mean by digital hermeneutics. The term has accompanied my work for several years — first in *Digital Hermeneutics* (2019) and later in *Digital Habitus* (2023) — and over time it has acquired multiple, layered meanings. We can distinguish at least five of them: a deconstructive, an epistemological, an ontological, an existential, and a cultural sense.

In its first sense, digital hermeneutics is a deconstructive project. Here, the digital serves as a conceptual tool to challenge the classical presuppositions of hermeneutics — its preference for language and textuality, its linear and monomedial logic, its anthropocentric conception of understanding. The digital exposes the material, multimodal, and non-linear dimensions of meaning-making. It shows that interpretation is not confined to the text or to the dialogical relation between reader and author, but unfolds across heterogeneous media, platforms, and interfaces. In this deconstructive sense, the digital becomes an opportunity to rethink hermeneutics from the ground up: to move from a hermeneutics of texts to a hermeneutics of technical mediations.

Second, digital hermeneutics can be understood in an epistemological sense. In *Digital Hermeneutics*, I suggested replacing the notion of data with that of trace. A trace is the presence of an absence: it points to something that once was, but is no longer directly accessible. Drawing on Carlo Ginzburg's «evidential paradigm», I proposed an epistemology that is fragile and conjectural rather than absolute — an epistemology of clues rather than certainties. Similarly, I have spoken of a hermeneutic concept of information, positioned equidistant between Claude Shannon's mathematical definition and Luciano Floridi's semantic and truth-based one. Information, from a hermeneutic perspective, is neither a pure signal nor a proposition that must be true; it is a relation of meaning that always presupposes interpretation. Knowledge in the digital age, therefore, is not the elimination of uncertainty but its transformation into a dynamic process of sense-making.

Third, there is an ontological sense of digital hermeneutics. Here, the focus shifts to the interpretive agency of machines themselves. Digital technologies are not only mediators of human interpretation but, in a certain sense, interpreting entities — or at least proto-interpreting ones. Of course, we must distinguish between different levels of interpretation: from the syntactic operations of pattern recognition to the semantic and reflexive dimensions proper to human understanding. But it would be reductive to reserve the term «interpretation» exclusively for the latter. Algorithms interpret in their own way: they select, correlate, and assign meaning-like relations. They construct equivalences and establish contexts. In this sense, interpretation becomes a distributed phenomenon across human and non-human agents. The digital environment reveals that understanding has never been purely human — it has always been co-constituted by technical mediations.

Fourth, digital hermeneutics also has an existential dimension, explored more extensively in *Digital Habitus*. Algorithms can be understood as machines of *habitus*: they act as structures of proto-classification that, by exposing users to repetitive patterns of similarity, shape their dispositions, preferences, and behaviors. In this context, I have distinguished between *idem* and *ipse* identity, borrowing Ricoeur's terminology. Algorithmic systems tend to flatten subjectivity toward *idem* — toward sameness, repetition, predictability — rather than supporting *ipse*, the narrative and reflexive dimension of the self that allows for deviation and transformation. The digital *habitus* thus risks reducing the style of a life to a statistical profile, translating singularity into a cluster.

Finally, there is a cultural sense of digital hermeneutics. In this broader register, *habitus* can also mean mental habit, a collective disposition of thought and expectation. The widespread diffusion of AI today is not only the result of its technical capabilities, but also of the belief in those capabilities — a kind of social wager on the promises of intelligence, creativity, and automation. Digital technologies are always embedded in sociotechnical imaginaries: shared visions, hopes, and fears about what technology can and should do. And in a recursive movement, these imaginaries are themselves reshaped by the technologies they inspire. To study the hermeneutics of the digital is therefore also to study the circulation of meanings, values, and affects that orient our relation to machines — and through them, our relation to the world.

3. FROM DIGITAL HERMENEUTICS TO DIGITAL AGONISM

Let me now return to the idea of *habitus* machines and classification systems — the core of our digital condition. There is a long intellectual history of critique directed at classification, long before artificial intelligence made it an everyday phenomenon. Geoffrey Bowker and Susan Leigh Star's *Sorting Things Out: Classification and Its Consequences* (1999) remains a foundational text in this regard. Their central insight is that every classificatory system, whether administrative, medical, or digital, embodies moral and political decisions about what counts, who counts, and under what conditions. Classifications, they write, «both reflect and constitute social life»: they not only describe the world but actively produce its divisions, hierarchies, and exclusions. This insight has taken on renewed relevance in the age of AI. The large-scale data infrastructures that underpin machine learning systems have inherited, and often intensified, these dynamics. Kate Crawford's *Atlas of AI* (2021) offers a powerful example through her analysis of ImageNet, one of the most influential datasets in computer vision. ImageNet, built from millions of images scraped from the internet, was organized according to the lexical categories of WordNet — a linguistic ontology designed for computational efficiency, not for ethical or social nuance. The result was a massive machine-readable map of the world in which human

beings, faces, and bodies were sorted into categories that reflected existing social biases: racial, gendered, and cultural stereotypes embedded in the very architecture of visual knowledge. The problem is not simply that these systems contain biases that could, in principle, be corrected. It is that the act of classification itself — the very process of sorting and labeling — is inherently interpretive and political. To classify is to draw boundaries: to decide what is similar and what is different, what is normal and what is deviant, what is relevant and what can be ignored.

This is precisely where a digital hermeneutics can intervene. On the one hand, it functions as a repair mechanism. It helps detect and make sense of misinterpretations; it reveals hidden assumptions; it fosters reflexivity within algorithmic systems and design teams. Hermeneutics, in this sense, can be embedded into the processes of technological development — as a dialogical and reflective practice. It can take the form of workshops, participatory design sessions, or iterative feedback loops that reintroduce interpretation where automation threatens to suppress it. The goal is not to eliminate bias — which is impossible — but to cultivate awareness of the magnifications and reductions inherent in every system. On the other hand, digital hermeneutics must never become an *ancilla technologiae* — a servant to technology, a rhetorical ornament used to legitimize existing infrastructures. It must retain its critical edge, functioning as a practice of unmasking. From outside the systems it analyzes, it exposes the epistemic, ethical, and political assumptions that sustain them. In this sense, digital hermeneutics stands both *within* and *against* the machine: it is internal enough to understand its operations, but external enough to question its ends. Digital hermeneutics must be both restorative and suspicious. It repairs meaning where it has broken down, but it also interprets interpretation itself — asking who interprets, on whose behalf, and to what effect.

At the heart of what I call *digital agonism* lies a simple but decisive question: how can we reintroduce conflict into systems that appear smooth, neutral, and frictionless? Chantal Mouffe's theory of agonism (2013) provides a crucial framework for addressing this question. For Mouffe, democracy is not the elimination of conflict but its domestication — the transformation of antagonism, where enemies seek each other's destruction, into agonism, where adversaries recognize one another as legitimate opponents within a shared symbolic space. Politics, in this sense, is not the pursuit of consensus but the institution of conflictual consensus: a space where disagreement is visible, voiced, and productive. A similar insight can be found in Paul Ricoeur's notion of a «conflict of interpretations» (2004). For Ricoeur, interpretation is never a solitary act leading to one correct meaning. It is a dialogical and contested process in which different readings confront one another, each disclosing partial aspects of truth. Truth, here, does not reside in the triumph of one interpretation over another but in the space of tension that their coexistence creates.

Bringing these two perspectives together allows us to imagine the digital domain — and especially AI — as a new site for this conflictual consensus. If, as I have argued, artificial intelligence is a machine of classification and world-making, then its ethical and political stakes concern precisely the possibility of opening this machine to plural interpretations. The challenge is not merely to make AI systems transparent or explainable — both ideals that often reproduce the logic of closure — but to make them contestable.

Today, however, the opposite tendency prevails. The machine presents itself as friendly, seamless, and smooth. ChatGPT, for instance, offers an interface of perfect cordiality: it responds politely, fluently, and without apparent friction. Its very friendliness discourages suspicion. Everything seems to flow effortlessly — and precisely for this reason, we stop asking where its knowledge comes from, how its classifications are built, or what remains invisible behind its polished surface. Behind this apparent fluidity lie the latent spaces that make generative AI possible: immense architectures of classification, correlation, and reduction. These spaces are trained on massive datasets filled with historical biases, cultural stereotypes, and inherited imaginaries. The result is paradoxical: systems profoundly shaped by partial, situated, and exclusionary worldviews project an aura of certainty and neutrality.

We see this clearly when we examine the visual output of generative models. Despite their novelty, the images they produce often replicate the same stereotypes that populated pre-AI stock imagery — the same gendered, racialized, and technophilic imaginaries. In my own work on the visual culture of AI (Romele 2022), I have shown how stock images consistently represent artificial intelligence through a limited set of visual tropes: blue tones, glowing brains, humanoid robots, the handshake between human and machine. Generative AI systems, rather than subverting these conventions, frequently reinforce them — translating inherited clichés into new, hyperreal forms. This continuity is precisely what French artists Gwenola Wagon and Stéphane Degoutin have explored in their project *Everything Is Real* (everythingisreal.net), where they deliberately make the machine «slip» (*dérapier*) — forcing it to produce uncanny, unstable, or contradictory images that disrupt the illusion of transparency. Such artistic détournements are exemplary of what I mean by digital agonism: practices that introduce friction, hesitation, and plurality into an infrastructure that strives for smoothness.

In this sense, digital hermeneutics must cultivate forms of suspicion — not in the paranoid sense of unveiling hidden intentions, but in the critical sense of keeping the interpretive space open. To «open the machine» means not only to render it explainable to experts but also to enlarge the community of interpreters: to include designers and users, engineers and artists, scholars and citizens. Only by multiplying

the voices engaged in interpretation can we transform algorithmic infrastructures into truly interpretive infrastructures — spaces where meaning is not dictated by the machine, but negotiated through collective agonism.

REFERENCES

BOWKER, G. C., and S. L. STAR, 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge; London: MIT Press.

CRAWFORD, K., 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven; London: Yale University Press.

IHDE, D., 1990. *Technology and the Lifeworld: From Garden to Earth*. Bloomington; Indianapolis: Indiana University Press.

MORETTI, F., 2000. The slaughterhouse of literature. *Modern Language Quarterly*. **61**(1), 207-227.

MOUFFE, C., 2013. *Agonistics: Thinking the World Politically*. London; New York: Verso.

RANCIÈRE, J., 2019. *The Distribution of the Sensible: Politics and Aesthetics*. London; New York: Bloomsbury.

ROMELE, A., 2023. *Digital Habitus: A Critique of the Imaginaries of Artificial Intelligence*. New York; London: Routledge.

ROMELE, A., 2022. Images of artificial intelligence: A blind spot in AI ethics. *Philosophy & Technology*. **35**(4), 1-22.

ROMELE, A., 2019. *Digital Hermeneutics: Philosophical Investigations in New Media and Technologies*. New York; London: Routledge.

WINNER, L., 1980. Do artifacts have politics? *Daedalus*. **109**(1), 121-136.