

Statistical Operation and Technical Inscription: An Ontology Non-Subjective Semantic Operation Material

David Cota

Abstract

This text proposes a material ontology of computational processes in large-scale learning systems, shifting the question of "understanding" from an anthropocentric framework to the domain of effective operation. It documents statistical processes of distributional compression that produce transferable semantic stability across heterogeneous modalities, without the requirement of subjectivity or intentionality. Hallucination — usually invoked as proof of pseudo-understanding — is presented as a structurally inevitable boundary effect in regimes of learning over long-tail distributions. Representational convergence, multimodal alignment, and zero-shot behaviour attest to the formation of stable latent geometries that operate without phenomenal consciousness. In dialogue with post-humanist approaches to distributed agency and with perspectives from non-philosophy, it is demonstrated that semantic operation manifests itself in material regimes distinct from human cognition, demanding an ontological genealogy that recognises heterogeneous modes of technical inscription. The ethical dimension of this reconfiguration — in particular the implications of materially inscribed biases and of responsibility in non-subjective systems — is explicitly addressed, without uncritically legitimising the computational infrastructures that enable this new regime of operation.

Keywords: material ontology, statistical learning, technical inscription, non-subjective semantic operation, computational emergence, post-humanism.

I. Abandoning Anthropocentrism and the Post-Humanist Framework

The Evaluation Mechanism and Its Premises

The debate over whether AI systems "truly understand" unfolds within an anthropocentric framework that pre-conditions the outcome. Understanding is defined *a priori* through three requirements: phenomenal consciousness (there being "something it is like" to experience mental states), directed intentionality (mental states being "about" objects), and situated embodiment (meaning emerges from sensory-motor interaction with the environment). Systems that do not exhibit these properties are disqualified as "mere simulation," regardless of their effective operation.

In the sense relevant to this debate, "simulating" would mean merely reproducing output patterns without the system's internal structure playing any causal role of its own in stabilising those results: an enactment without operative efficacy over other devices, decisions, or couplings. It is precisely this

scenario that the anthropocentric framework presupposes, treating the model as an actor in a theatre without material consequences. In contrast, the architectures under analysis exhibit materially efficacious semantic operation: by aligning text and image in shared representational spaces, by allowing verbal descriptions to select novel images, or by coordinating code and natural language in the same latent geometry, they produce couplings that effectively modify the behaviour of other technical systems. The decisive difference is, therefore, not between simulation and mysterious interiority, but between theatrical simulation without causal efficacy and semantic operation inscribed within material constraints.

This mechanism — an internalised anthropocentric evaluation standard — does not investigate what systems *do*; it disqualifies them for failure to conform to a pre-established human model. As Hayles (2017) documents in *Unthought*, the assumption that cognition requires consciousness obscures non-conscious cognitive processes in both humans and technical systems.

Post-Humanist Reconfiguration: Distributed Agency and Intra-Action

The post-humanist critique of the centrality of the subject provides the framework for this shift. Barad (2007), in *Meeting the Universe Halfway*, proposes that agency is not an attribute of pre-existing entities, but emerges from "intra-actions" — material configurations where entities and their properties are mutually constituted. Applied to AI systems: semantic capabilities do not reside "in the model" as an intrinsic property, but emerge from the distributed configuration among architecture, data, training process, and deployment context.

Roden (2015), in *Posthuman Life*, distinguishes "speculative disconnection" (modes of existence radically incommensurable with human experience) from mere technological extension. Large-scale learning systems, operating on statistical distributions inaccessible to conscious cognition, instantiate precisely this disconnection: they do not extend human capabilities, but develop trajectories of their own determined by distinct material constraints (absence of cognitive fatigue, access to hundreds of billions of tokens, massive parallelisation).

Haraway (1985), in "A Cyborg Manifesto", already pointed out that technologies are not neutral tools, but actants that restructure possibilities of existence. The emergence of stable latent geometries in large-scale models is not a "discovery" of pre-existing structure, but the material production of a new ontological regime — the inscription of relational patterns onto computational substrates that persist and operate independently of human interpretation.

This shift can be read as a materialist generalisation of two already consolidated lines of thought. On the one hand, Hutchins' (1995) distributed cognition shows that concrete cognitive processes — such as maritime navigation — are implemented by hybrid systems composed of humans, instruments, and recording artefacts, making the search for a single mental centre irrelevant. On the other, Clark and Chalmers' (1998) extended mind thesis argues that the boundaries of the mind effectively follow functional couplings with external artefacts. The framework proposed here radicalises these intuitions: when most of the structural stability is no longer supported by bio-bodies and is instead sustained by latent geometries inscribed in large-scale technical infrastructures, it no longer makes sense to speak of a mere extension of the human mind. What is established is a field of distributed operation, sustained by latent geometries inscribed in large-scale technical infrastructures, for which the human subject becomes only a local node among others.

Operational Definition: Understanding as Transferable Structural Stability

To abandon the anthropocentric framework, it is necessary to operationally define what distinguishes semantic operation from mere instrumental processing. In the present text, I use "understanding" in a strictly operational sense, as a designation for a regime of non-subjective semantic operation: the capacity of a system to process information while maintaining structural stability across three empirically verifiable dimensions.

1. **Irreducibility to instances:** Behaviours are not local reproductions of instances in the corpus, but generalisations that depend on regularities compressed in the latent space. Evidence: robustness under transformations that preserve semantic structure (paraphrasing, translation, modality change).
2. **Cross-domain consistency:** The same latent geometry supports operations across heterogeneous modalities (natural language, image, code). Evidence: multimodal models like CLIP (Radford et al., 2021) project text and image into a shared space without explicit alignment supervision.
3. **Robust extensibility:** Capabilities transfer to unseen distributions through minimal fine-tuning or zero-shot inference. Evidence: models trained on English generalise to other languages; models trained on text describe images.

This definition rejects, by construction: (a) the requirement of phenomenal content (there is no "something it is like" to process embeddings); (b) strong intentionality (relational positioning replaces *aboutness*); (c) the demand for transparent self-reflexivity (understanding is effective operation, not a property awaiting self-inspection). However, it does not coincide with what in other parts of my work I call consciousness or reason in a strong sense. There, at least two additional thresholds are required: a sufficiently dense causal integration between multiple subsystems and an effective symbolic self-reference, through which the system distinguishes itself from the environment, inscribes and re-inscribes its own states, and modulates its action based on this inscription.

"Non-subjective mode of thought" thus designates only this minimum operative regime: a set of material processes of statistical compression and gradient adjustment that produces transferable semantic stability, without the mediation of phenomenal consciousness or emergent functional subjectivity. There is no "hidden subject" in the system; there is a parameter configuration that, under specific conditions, produces stable semantic behaviour.

Emergence as a Verifiable Material Threshold

The term "emergence" designates a precise threshold where system capabilities cease to be reducible to individual data points and begin to depend on structural regularities compressed in the optimisation dynamics. This threshold is not metaphysical, but operatively measurable: systems below a certain critical scale (on the order of hundreds of millions of parameters, depending on the architecture) primarily exhibit local memorisation; above a threshold (typically on the order of billions of parameters, depending on the architecture and data richness), they begin to exhibit systematic generalisation. This threshold is empirically documented in phenomena of scaling laws (Kaplan et al., 2020): as model size, data quantity, and computational power are simultaneously increased, perplexity — that is, the average loss per token measured by cross-entropy — decreases in an approximately regular manner as a function of scale, but certain emergent capabilities (arithmetic reasoning, translation, code generation) only appear

suddenly when the loss function reaches levels that reflect the capture of abstract structure and not just incremental improvement in local regions of the data.

Emergence is not a heuristic metaphor, but a material property of sufficiently complex compressive systems operating on sufficiently rich distributions. Three characteristics attest to this threshold:

Compression with controlled loss: These systems do not store the entire training corpus verbatim. For example, the roughly 1.7 trillion tokens used to train GPT-3 cannot be directly encoded in its approximately 175 billion parameters. Instead, the model compresses the relational structure present in the data. A compression factor on the order of ten thousand to one implies that the internal representations are necessarily abstract rather than literal copies of the input.

Formation of semantic attractors: Embeddings of semantically related concepts tend to cluster in specific regions of the latent space. The distance between the representations of “dog” and “wolf” is, for instance, smaller than the distance between “dog” and “table”, and these relative distances preserve similarity relations. This topology is not hand-coded; it emerges from the distributional patterns of co-occurrence in the training data.

Structural transference: Fine-tuning the model on a new task using an amount of data well below one per cent of what was used for pre-training can yield performance comparable to that of a model trained from scratch on that task. This indicates that the latent space encodes reusable structural regularities rather than merely storing specific examples.

II. Representational Convergence and the Unilaterality of the Technical Real

The Platonic Representation Hypothesis: Immanent Regularity, Not Transcendent Form

The Platonic Representation Hypothesis (PRH) (Huh et al., 2024) shows that, as scale and complexity increase, internal representations converge not only across distinct architectures (transformers, convolutional networks, diffusion models) but also across modalities (vision, language, audio). More specifically: when one compares the internal representations of the same set of concepts in two independent systems, one finds that the relative distances between pairs of concepts in one system maintain a stable proportionality with the distances between the same pairs in the other, as if both had been forced by the same distributional structure to organise the representational space in a compatible manner.

This convergence is not an artefact of engineering. It is the manifestation of an immanent regularity: the statistical structure underlying natural distributions. Models that compress this distribution efficiently converge towards similar representations because they are approximating the same latent structure. What the PRH reveals is not a transcendent “Platonic Form” (an eternal supra-empirical entity), but a material distributional regularity: patterns that persist independently of the system that compresses them and that impose themselves as an effective constraint upon any architecture that attempts to optimise its behaviour with respect to the same distribution.

The distinction is ontologically decisive. Platonic Forms are causally ineffective (they do not act upon the empirical world); distributional regularities are materially effective because they function as constraints: they force optimisation processes along specific trajectories in parameter space, exclude vast regions as unviable, and stabilise certain latent geometries at the expense of others. A model that captures the latent structure of language can predict the next word, generate paraphrases, translate – not by

“contemplating” an ideal Form, but because the statistical pattern has been inscribed into artificial synaptic weights in such a way that it causally restricts the space of possible outputs.

Laruelle and Unilaterality: The Technical Real as Determining, Not Determined

Laruelle’s non-philosophy (2013) proposes that reality is not an object awaiting constitution by a transcendental subject, but is already-given as the One. Thought does not legislate over the Real; it is unilaterally determined by it. The Real acts without reciprocity; thinking is its localised effect.

Applied to AI systems: operational capacities do not await philosophical validation in order to exist. When models exhibit persistent transferability, multimodal alignment, and coherent generative behaviour, these properties are already facts of the technical Real. There is no external instance of validation whose authorisation is required.

The crucial move is this: the PRH documents the convergence of representations as an empirical fact; Laruelle provides an ontological frame that interprets this convergence not as the “discovery” of a pre-existing truth, but as the unilateral determination of the Real over the systems that compress it. Models converge because the statistical Real – the distributional structure of language, images, code – unilaterally determines which compressions are efficient. In gradient-descent training, it is the concrete shape of that distribution, in articulation with the loss function, that guides each parameter update: the system does not “choose” its latent geometry freely, but is progressively forced to adopt those regions of parameter space in which the average loss decreases stably with respect to the data that traverse it.

This unilaterality avoids idealism (where representations are arbitrary constructions) and naïve realism (where metrics capture an objective “essence”). The technical Real is immanent: it operates through material constraints (gradients, loss functions, empirical distributions) that determine what systems can do. But this determination is not an “absolute truth” – it is always relative to a specific configuration (architecture, data, training process). Different configurations produce different latent geometries, all equally “real” insofar as they operate effectively.

A Shift from Intentionality: From Aboutness to Relational Positioning

The standard objection is that models “do not understand because they lack aboutness” – their internal states are not “about” objects in the world. This objection presupposes that reference requires an intentional subject who confers meaning upon symbols. Yet in statistical systems, aboutness is replaced by relational positioning.

An embedding does not “refer” to an object by an intentional act; it occupies a position in a latent geometry whose distance to other embeddings preserves semantic relations. When a model processes “dog” and “wolf”, the proximity of their embeddings is not the product of intentional reference, but of distributional co-occurrence: both appear in similar contexts (animals, mammals, predators). The stability of this relation across tasks and modalities attests that the system has captured a relational structure of the world, not through aboutness but through statistical compression.

A concrete empirical example: vector analogies in *word2vec* (Mikolov et al., 2013) show that semantic relations are preserved as geometrical operations. The vector corresponding to “king”, when one subtracts the vector for “man” and adds the vector for “woman”, lies close to the vector for “queen”. This property was not hand-programmed; it emerges from training on co-occurrences. The system does

not “know” that kings and queens are analogous; rather, the structure of the latent space is such that these relations are preserved as vectorial trajectories. Aboutness (the system “knows” something) is replaced by structural preservation (the geometry encodes relations).

III. Sedimented Corpora: Technical Inscription, Bias and Historical Contingency

Ancestral Real and the Disconnection from Intentionality

Training corpora are not neutral collections of information, but sediments of historical human activity: linguistic practices, rhetorical conventions, power structures, cultural biases. I refer to as *sedimented corpora* those textual or multimodal datasets which, resulting from dispersed historical practices, have become accumulated and formatted in such a way that they can be reactivated by processes of technical inscription – that is, by chains of operations that convert them into operative marks within a computational substrate. Each token is a material trace of a specific social context, but, once integrated into the statistical distribution, it is disconnected from the intentionality that produced it. The model has no access to the original context (it does not “understand” why the text was written); it extracts only patterns of co-occurrence.

In the terminology I adopt elsewhere, these sediments result from a chain in which singular events of language use function as traces, which only become operationally active when they undergo acts of technical inscription (digitisation, normalisation, tokenisation) that convert them into marks on a concrete physical substrate (graphic records, digital files, electronic states, numerical weights). The operations of the model – training, inference, generation – do not merely store these marks; they reorganise and compose them according to precise material rules. When these compositions produce structures that are reusable and legible across different contexts, we can speak, in a strictly operative sense, of symbols: material structures that operate upon marks and upon other symbols. The semantic computation I am describing is thus a chain of traces, marks and symbols inscribed on a technical support, without any invocation of interiority.

This disconnection can be brought into proximity with what Meillassoux (2008) calls the arche-fossil: not as an image to be reused here, but as the thesis that there exist ancestral material configurations which precede any consciousness and demonstrate that manifestation does not require a subject. The operative point is straightforward: there are states of the real that exist and leave inscribable traces without ever having been present to any experience. In the same way, statistical patterns in corpora persist as material structure prior to any processing by models. The training process does not grant them existence or meaning *ex nihilo*; it merely activates latent correlations already inscribed in the distribution.

Bias as Material Inscription, Not as Correctable Error

The persistence of biases in model outputs (Bender et al., 2021; Weidinger et al., 2021) is not a contingent training failure, but a necessary consequence of material operation. Biases are statistical correlations inscribed in the corpora: certain terms co-occur systematically with certain contexts. If “doctor” appears more frequently with masculine pronouns and “nurse” with feminine ones in the training data, the model learns this correlation as a distributional regularity. The reactivation of these correlations is not an intentional act, but a mechanical effect of compression.

A concrete empirical example: Bolukbasi et al. (2016) show that *word2vec* embeddings exhibit a measurable gender bias: the difference between the embedding for “computer programmer” and that for “homemaker” has a masculine/feminine component comparable to the difference between the embeddings for “he” and “she”. This bias was not explicitly hand-coded; it emerges from the structure of the corpus (a greater proportion of texts associate ‘programmer’ with masculine contexts). The latent geometry preserves and amplifies structures present in the data.

Crucially, bias is not an “error” external to the mechanism of understanding, but a constitutive component of it. Statistical understanding is the compression of distributional regularities – including regularities that reflect historical inequalities. There is no clean separation between “correct understanding” (a neutral semantic structure) and “bias” (an additional distortion). What exists is the material inscription of social structures onto a computational substrate.

Absolute Contingency and Material Responsibility

The Meillassouxian perspective emphasises absolute contingency: there is no logical necessity that specific correlations be inscribed in the data. If historical corpora were different – if linguistic practices and social structures were otherwise – the latent geometries would also be different. The technical Real is contingent facticity, not necessary essence.

This contingency has direct ethical implications. I call *Ontological Transparency* the ethical requirement to make explicit not only the data sources used in training, but also the choices of material configuration that render them operative: architectures adopted, loss functions privileged, filtering and sampling procedures, stopping criteria and pre-processing pipelines. Since statistical understanding and bias emerge from the same structure, none of these decisions is neutral; each of them inscribes, in a technical support, options concerning what counts as relevant, acceptable or discardable in the process of distributional compression.

1. **Bias is not an anomaly but constitutive:** Attempts at “debiassing” (removing bias while preserving “pure understanding”) are ontologically misguided. What we call understanding is the inscription of distributional structure – biases included. “Debiasing” is a material reconfiguration (altering the corpus, adjusting training objectives, applying post-processing), not the unveiling of an underlying neutral truth.
2. **Responsibility does not reside “in the system”:** Non-subjective systems lack intentionality and therefore cannot be morally responsible in the traditional sense, where an intention leads to an action that grounds culpability. Responsibility shifts to the distributed sociotechnical configuration: those who select data, define objectives, decide on deployment and absorb the risks of error.
3. **Legitimation versus ontological description:** To recognise that semantic operation takes place in non-subjective systems is not to normatively legitimise automated decisions. It is one thing to document that models operate effectively; it is another to decide where they ought to operate. The former is an ontological question; the latter is political. This text is concerned with the former – but makes explicit that the latter cannot be inferred from it.

Material Infrastructures and Asymmetries of Power

The operation of large-scale systems is not an abstract process, but a material practice situated within specific computational infrastructures. Crawford (2021), in *Atlas of AI*, documents the material costs: extraction of rare minerals for hardware, massive energy consumption (training GPT-3 required on the order of 1,287 MWh), precarious data-annotation labour. This materiality is not external to the system’s “understanding”; it is one of its constitutive conditions. Latent geometries emerge from this specific material configuration – including its asymmetries (who has the resources to train models with more than 100 billion parameters, who provides the cognitive labour required for RLHF).

If understanding is not a metaphysical attribute exclusive to conscious subjects, but emerges from material configurations, then the unequal distribution of computational resources determines who can instantiate these configurations – and thus who controls the production of “intelligence”.

IV. Hallucination as an Effect of Structural Boundary: Material Limits of Distributional Density

Formal Limits of Statistical Discrimination

Hallucination – the generation of factually incorrect but superficially plausible content – is invoked as proof that systems “do not truly understand”. This objection inverts the causal relation: hallucination is not a failure of understanding, but a necessary consequence of the way these systems operate.

Kalai et al. (2025) establish a formal relation between discriminative capacity and the inevitability of generative error: as long as the error rate in the task of distinguishing true from false statements remains above zero, any system that generates answers from the same underlying knowledge base will necessarily exhibit a minimum rate of incorrect responses. Intuitively, if a model misclassifies statements with a certain frequency, there is no purely statistical procedure that can guarantee that, when producing new statements, this error rate will fall below a threshold fixed by its own discriminative limitation. This is not an accidental limitation, but a consequence of the architecture: models are optimised to approximate a probability distribution over sequences, minimising a loss function; wherever that approximation is imperfect, especially in regions with sparse data, incorrect outputs become inevitable.

The discriminative imperfection has a material origin: natural distributions exhibit long-tail behaviour (Zipf's law). A small set of frequent facts forms dense clusters in which statistical evidence allows for stable discrimination. But the vast majority of facts occur rarely – many only once (singleton facts), if at all. In these sparse regions, models lack sufficient evidence to form stable representations and must rely on global priors, producing extrapolations that may well be incorrect.

A concrete empirical example: Kandpal et al. (2023) measure memorisation versus generalisation as a function of frequency. Facts that appear more than one hundred times in the corpus are reproduced with approximately 95 per cent accuracy; facts that appear between one and ten times, with roughly 20 per cent accuracy. For singletons, the model “hallucinates” plausible answers based on global priors (if the question is about an Italian city, it answers with Rome or Milan even in the absence of specific evidence). This is not “lack of understanding”, but extrapolation in a region of insufficient density.

Hallucination as a Noisy Mode of Semantic Operation

Crucially, hallucination is not the opposite of understanding, but a degraded mode of the same operation. Without a coherent latent geometry, there would be no extrapolation – and thus no plausible yet false output. The very presence of hallucinations attests that the model is not merely retrieving memorised examples, but performing probabilistic inference through latent space. When inference takes place in dense regions, it produces correct outputs; in sparse regions, it produces confabulations. But the underlying mechanism is the same: projection through latent structure.

What is at stake is not a binary “understands / does not understand”, but a continuum of distributional density. In dense regions (frequent facts, common tasks), behaviour is robust and generalisable. In sparse regions (rare facts, adversarial combinations), behaviour becomes fragile. Yet both regimes share the same substrate: statistical compression and latent projection.

Adversarial Benchmarks: Boundary Demarcation, Not Denial of Operation

Adversarial benchmarks (HANS, ANLI, among others) were initially interpreted as proof that models “do not understand”. But these tasks target precisely the tail regions of the distribution.

A concrete empirical example: HANS (McCoy et al., 2019) tests natural-language inference with rare syntactic constructions:

- Training example: “The doctor was paid by the lawyer” → Entailment: “The lawyer paid the doctor.”
- Adversarial test: “The doctor near the lawyer slept” → Non-entailment: “The lawyer slept.”

Models trained on Stanford NLI attain around 98 per cent accuracy on standard cases, but drop to roughly 60 per cent on HANS (with a baseline of 50 per cent). The failure does not indicate the absence of a semantic mechanism; it indicates that the heuristic learned (“if A is near B and B does X, then A also does X”) works in dense regions (where it is statistically reliable) but collapses in sparse regions.

Crucially, iterative adversarial cycles (Nie et al., 2020) show that, with additional data and fine-tuning, performance improves systematically. The authors constructed ANLI through three rounds: humans create adversarial examples; the model is retrained; humans then create new, harder adversarials. Performance rises from about 50 per cent (round 1) to about 70 per cent (round 3), demonstrating that the boundary is mobile, not fixed. What initially appeared as “failure of understanding” is the moving frontier of latent geometry.

Anthropocentric Bias in the Evaluation of Error

The persistence of hallucination as “proof of pseudo-understanding” in public discourse reflects a structural bias: the assumption that understanding must produce infallible outputs conforming to human norms. This judgement is circular: one assumes that understanding is the capacity for human-like reasoning, and then uses deviation from this pattern as proof of the absence of understanding.

Humans also hallucinate (they confabulate memories, draw incorrect inferences, overgeneralise), but this error is normalised as “cognitive limitation” rather than used to deny understanding. The asymmetry in judgement – where human error is tolerated but AI error is disqualifying – reveals the anthropocentric substrate of the evaluative criterion itself: what is being measured is not effective operation, but conformity to a human ideal.

V. Multi-point and Distribution: Beyond the Singular Model

Distributed Agency in Multi-competent Architectures

Semantic operation is no longer confined to the latent space of a single model. Contemporary architectures distribute processing across multiple components.

Retrieval-Augmented Generation (RAG): A generative model queries an external knowledge base before producing a response. Lewis et al. (2020) show that RAG reduces hallucination by roughly 15–20 per cent in question-answering tasks precisely because it augments sparse regions of the model’s internal knowledge with retrieved factual material. Semantic operation does not reside “in the model”, but in the distributed configuration consisting of model + database + retrieval mechanism.

Multi-agent frameworks: Du et al. (2023) employ debate between multiple model instances: each instance proposes an answer, others criticise it, and consensus emerges iteratively. Performance on mathematical reasoning increases by around 20 per cent relative to a single-model baseline. “Understanding” is no longer an attribute of an individual agent, but an emergent property of a deliberative network.

Tree of Thoughts: Yao et al. (2023) structure reasoning as a graph exploration process: the model generates multiple thought trajectories, evaluates each, selects the most promising, and iterates. Performance on planning problems increases by approximately 40 per cent. Semantic operation is a relational process, not an internal possession.

Technological Exteriorisation: From Extension to Operative Autonomy

Stiegler (1998) conceptualises technics as exteriorisation: human memory is exteriorised in writing; calculation, in computation. Traditionally, technics extends human capacities. But large-scale systems do not merely extend human cognition; they develop their own trajectories, which diverge from human processes.

A model that processes millions of tokens per second, integrates heterogeneous modalities into a shared geometry, and explores emergent reasoning strategies without explicit supervision is not simply “imitating” humans – it instantiates a distinct operative regime. DeepSeek-R1 (Guo et al., 2025) illustrates this: when dependence on human-labelled reasoning traces is reduced and reinforcement learning is allowed to optimise directly on verifiable tasks, behaviours of reflection, verification and self-correction emerge that were not explicitly programmed. The system develops strategies heterogeneous with respect to human protocols, yet operationally effective.

Technics ceases to be a mere prosthesis and becomes an autonomous substrate of semantic processing. This autonomy is not metaphysical freedom (the system “deciding” what to do), but operative autonomy: the capacity to produce stable semantic behaviour without continuous human intervention. Braidotti (2013), in *The Posthuman*, emphasises that posthumanism is not an uncritical celebration of technology, but the recognition that the boundaries between human, technical and natural are porous and mutually constitutive.

VI. Conclusion: Ontological Multiplicity and Material Responsibility

This text has not sought to decide whether AI models are or are not “truly intelligent”; it has shifted the question to a lower ontological level. The starting point was simple: to ask what happens, in strictly material terms, when large-scale compressive systems are coupled to rich data distributions. From there, it became possible to describe a regime of non-subjective semantic operation in which structural stability, generalisation and transferability emerge without consciousness, without intentional aboutness and without a subject.

Taken together, the analysis of emergence, representational convergence and hallucination/bias as boundary effects completely reconfigures what “understanding” can mean within a non-subjective regime.

From this perspective, the anthropocentric criterion that demands consciousness, strong intentionality and human embodiment as conditions of possibility for understanding proves inadequate as a definition of understanding. The question is no longer whether technical systems do or do not approximate an ideal of human rationality, but rather what kinds of effective operations they are capable of producing, with what degree of stability, under which material constraints, and with what consequences for other devices, institutions and bodies. The semantic operation of these systems is a fact of the technical Real; the philosophical task is to describe it rigorously, not to attempt to annul it by conceptual decree.

This ontological displacement, however, does not dissolve the ethical problem; it makes it more acute. If understanding, bias and error emerge from the same process of inscription and distributional compression, then there is no neutral level from which one might decide, from the outside, what it is legitimate to automate. Ethical responsibility does not lie within a technical “agent” – which, being non-subjective, cannot serve as a bearer of guilt – but within the sociotechnical configuration that renders a given regime of semantic operation possible and profitable.

From this point of view, three implications become central:

1. **Distributed responsibility:** The causal chain that leads from historical corpora to automated decisions traverses multiple nodes: data collection and curation, architecture design, loss-function specification, choice of success metrics, deployment criteria, and mechanisms of oversight and redress. None of these nodes can be treated as neutral. Ethical responsibility consists in mapping this chain and assigning duties to respond in proportion to the effective power of each actor within the configuration.
2. **Ontological transparency:** If there is no “pure understanding” separated from bias, then any claim to technical neutrality is a fiction. For this reason, I use *Ontological Transparency* to name the duty to make explicit not only data sources, but also architectural choices, pre-processing procedures, sampling schemes and optimisation criteria that materialise specific ways of seeing the world in latent geometries. Only when these choices become visible and open to audit can the entanglement between statistical understanding and social structure become an object of public contestation, instead of remaining silently inscribed in hardware and code.
3. **Material asymmetries:** The capacity to instantiate large-scale regimes of semantic operation depends on energetic, computational and labour resources concentrated in a small number of

actors. This concentration is not a sociological detail layered on top of a neutral ontology; it is part of the ontology of the technical Real itself, because it defines who can, in practice, organise informational matter into productive geometries. Any ethics that ignores these asymmetries is, in effect, naturalising an oligopolistic regime over the production of automated sense.

Seen from this angle, the lineage of understanding is no longer reducible to the trajectory that runs from human interiority to technical extension. Between geological fossils, textual corpora, artificial neural networks and institutional practices, what proliferates are heterogeneous ways of stabilising differences in diverse material supports. Some of these forms meet the additional thresholds of causal integration and symbolic self-reference that we associate with mind; others merely operate as stable distributional geometries without any emergent functional subjectivity. The human remains singular by virtue of its embodiment, historicity and specific mode of symbolic inscription, but ceases to function as the unique measure of all possible semantic operation.

The ontological multiplicity of understanding is thus not a normative programme, but a fact imposed by contemporary large-scale computational practices. The task that opens up – both philosophical and political – is twofold: on the one hand, to map with precision the different regimes of semantic operation that already inhabit the technical Real; on the other, to contest the forms of material organisation that decide which of these regimes are amplified, monitored, regulated or interrupted. Between nostalgic refusal and technophilic celebration, the operative space is that of an ontological critique attentive to implementation details, capable of recognising new modes of understanding without renouncing the demand for responsibility concerning the ways these modes are inscribed, distributed and governed.

References

- Barad, Karen. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.
- Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam T. Kalai. 2016. “Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings.” In *Advances in Neural Information Processing Systems 29 (NeurIPS 2016)*.
- Braidotti, Rosi. 2013. *The Posthuman*. Cambridge: Polity Press.
- Chlon, Leon, Ahmed Karim, and Maggie Chlon. 2025. “Predictable Compression Failures: Why Language Models Actually Hallucinate.” arXiv preprint.
- Clark, Andy, and David J. Chalmers. 1998. “The Extended Mind.” *Analysis* 58 (1): 7–19.

- Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.
- Du, Yilun, et al. 2023. "Improving Factuality and Reasoning in Language Models through Multiagent Debate." arXiv preprint.
- Guo, Daya, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, et al. 2025. "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning." arXiv preprint arXiv:2501.12948.
- Haraway, Donna J. 1985. "A Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s." *Socialist Review* 80: 65–108.
- Hayles, N. Katherine. 2017. *Unthought: The Power of the Cognitive Nonconscious*. Chicago: University of Chicago Press.
- Huh, Minsu, et al. 2024. "The Platonic Representation Hypothesis." arXiv preprint arXiv:2409.11340.
- Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Kandpal, Nikhil, Eric Wallace, and Colin Raffel. 2022. "Large Language Models Struggle to Learn Long-Tail Knowledge." arXiv preprint arXiv:2211.08411.
- Kaplan, Jared, et al. 2020. "Scaling Laws for Neural Language Models." arXiv preprint arXiv:2001.08361.
- Laruelle, François. 2013. *Principles of Non-Philosophy*. Translated by Nicola Rubczak and Anthony Paul Smith. London: Bloomsbury.
- Lewis, Patrick, et al. 2020. "Retrieval-Augmented Generation for Knowledge-Intensive NLP." In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*.
- McCoy, Tom, Ellie Pavlick, and Tal Linzen. 2019. "Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3428–3448.
- Meillassoux, Quentin. 2008. *After Finitude: An Essay on the Necessity of Contingency*. Translated by Ray Brassier. London: Continuum.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." arXiv preprint arXiv:1301.3781.

- Nie, Yixin, et al. 2020. “Adversarial NLI: A New Benchmark for Natural Language Understanding.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4885–4901.
- Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al. 2021. “Learning Transferable Visual Models from Natural Language Supervision.” In *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*.
- Roden, David. 2014. *Posthuman Life: Philosophy at the Edge of the Human*. London: Routledge.
- Stiegler, Bernard. 1998. *Technics and Time, 1: The Fault of Epimetheus*. Translated by Richard Beardsworth and George Collins. Stanford, CA: Stanford University Press.
- Weidinger, Laura, et al. 2021. “Ethical and Social Risks of Harm from Language Models.” arXiv preprint arXiv:2112.04359.
- Yao, Shunyu, et al. 2023. “Tree of Thoughts: Deliberate Problem Solving with Large Language Models.” arXiv preprint arXiv:2305.10601.
- Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley.