Ledger and Junk: Opacity as the Condition of Generativity in Computational Systems

Flyxion

August 28, 2025

Abstract

In the epistemology of science, from Pliny the Elder's reflections on marble quarries to modern controversies surrounding "junk DNA" and "hallucinating" artificial intelligence (AI) systems, a persistent tension emerges between the impulse to classify, formalize, and render transparent and the recalcitrant presence of remainder, opacity, and apparent waste. This paper argues that such opaque zones—historically marginalized as trivial or erroneous—are not merely incidental but constitutive of generativity across biological, linguistic, and computational systems. Drawing on insights from philosophy of science, science and technology studies (STS), and computational linguistics, we examine Large Language Models (LLMs) as a paradigmatic case. The "ledger" of LLMs—predictable, high-probability outputs such as factual responses or grammatical corrections—represents a traceable and instrumentalized subset of their functionality. However, their true generative capacity lies in the "junk": low-probability traversals of the grammatical Directed Acyclic Graph, which produce structurally constrained yet semantically unpredictable outputs often labeled as hallucinations. Far from being errors, these traversals enable creative synthesis in domains such as art, fiction, and conceptual innovation. The positivist demand for total transparency—interpretable model weights and fully legible outputs—misconstrues the lo-

cus of creativity, echoing historical missteps in genomics that dismissed non-coding DNA. By integrating philosophical critiques of reductionism, STS perspectives on the sociotechnical construction of knowledge, and computational linguistic analyses of probabilistic grammars, this paper contends that opacity is not a flaw but a foundational condition of generativity. To prioritize transparency over opacity is to risk dismantling the very substrate of emergent phenomena, mistaking the ledger for the mountain itself.

1   Introduction

In his *Naturalis Historia*, Pliny the Elder critiqued the extraction of marble from quarries, lamenting that "we penetrate into the bowels of the earth, digging veins of gold and silver, and ores of copper and lead; we seek also for gems and certain little pebbles, and drive shafts deep into the ground" (the Elder, 1855). This act of sorting, bucketing, and discarding—what we term the SBD (Shock, Boredom, Demand) cycle—captures a recurring epistemic pattern in scientific practice. Novel phenomena provoke shock, disrupting established frameworks. As familiarity sets in, they are categorized, becoming mundane (boredom). Institutional demands for utility and legibility then prioritize certain elements, relegating others to the status of remainder or "junk" (Latour, 1987). From ancient natural history to modern genomics and artificial intelligence (AI), this cycle produces opaque zones—material or phenomena dismissed as trivial or wasteful—that consistently prove to be sites of regulation, creativity, and emergence (Douglas, 1966; Haraway, 1988; **?**).

This paper proposes the ledger/junk dichotomy as an analytic lens to interrogate this pattern. The "ledger" represents ordered, transparent, and instrumental outputs—elements that can be classified, validated, and operationalized. The "junk," by contrast, encompasses the opaque, low-probability, or seemingly erroneous outputs that resist categorization. Far from peripheral, junk is the substrate of generativity, enabling novelty and emergence (Rheinberger, 1997).

This framework draws on philosophy of science (Kuhn, 1962; Feyerabend, 1975; Hacking, 1983), science and technology studies (STS) (Latour, 1987; Jasanoff, 2004; Haraway, 1988), computational linguistics (Chomsky, 1957; Manning and Schütze, 1999), critical data studies (?Bowker and Star, 1999), and postcolonial theories of opacity (Glissant, 1997; Anzaldúa, 1987). The argument unfolds through two case studies: the revaluation of "junk DNA" in genomics and the role of "hallucinations" in Large Language Models (LLMs). These cases demonstrate that opacity is a structural condition of generativity, not a flaw to be eradicated. This dynamic is mirrored in scholarly practice: bibliographies, often deferred as secondary "junk," become foundational ledgers when surfaced, underscoring the epistemic reversal at the heart of this study (Rheinberger, 1997; Shapin and Schaffer, 1985). The trajectory moves from Pliny's ancient critique to contemporary computational systems, with implications for epistemology, AI governance, and scientific practice.

2 The SBD Cycle as Epistemic Pattern

2.1 Defining the SBD Cycle

The SBD cycle—Shock, Boredom, Demand—describes a fundamental mechanism in scientific knowledge production. Novel phenomena initially provoke shock, challenging established epistemic frameworks by introducing anomalies that resist existing categories. As scientists become accustomed to these phenomena, they are bucketed into familiar classifications, entering a phase of boredom where their novelty fades. Finally, institutional and practical demands—driven by funding, publication pressures, or policy needs—prioritize legible, instrumental elements, discarding the rest as remainder (Latour, 1987; Scott, 1998). This cycle is not merely technical but sociopolitical, shaped by the need for legibility in bureaucratic and scientific systems (Jasanoff, 2004). For instance, the discovery of vast non-coding DNA regions shocked molecular biologists, only to

3

be categorized as "junk" to preserve the ledger of protein-coding genes (Ohno, 1972). This deferral of complexity reflects the broader logic of the SBD cycle, where opacity is sidelined to maintain epistemic order (Bowker and Star, 1999). The cycle's persistence across domains suggests a structural feature of knowledge production, one that privileges transparency while marginalizing the generative potential of the unledgerable (Star, 1990).

2.2 Historical Illustrations

Historical cases illuminate the SBD cycle's persistence across scientific domains. In eighteenth-century natural history, teratological specimens or "monsters" provoked shock by defying taxonomic order. These anomalies were soon classified as aberrations, excluded from systematic ledgers of species to maintain the coherence of naturalist frameworks (Daston and Park, 1998). Similarly, the phlogiston theory, once a cornerstone of chemical explanation, shocked scientists with its interpretive power before being relegated to junk as oxygen-based models gained dominance. Yet, phlogiston's legacy shaped modern chemistry, demonstrating the generative potential of discarded frameworks (Shapin and Schaffer, 1985). The luminiferous ether, posited as the medium for light propagation in the nineteenth century, followed a similar trajectory: initially shocking, then normalized, and finally discarded as relativity theory emerged, yet its conceptual role influenced modern physics (Whittaker, 1951). Thomas Kuhn argues that anomalies, initially shocking, are critical to paradigm shifts, as they expose the limits of existing models (Kuhn, 1962). Paul Feyerabend's pluralism further suggests that excluded perspectives harbor epistemic value, resisting the demand for a singular ledger (Feyerabend, 1975). Donna Haraway's concept of situated knowledges underscores that what is deemed junk is a matter of perspective, shaped by power and institutional context (Haraway, 1988). These cases reveal how the SBD cycle structures scientific inquiry, marginalizing opacity to prioritize legibility (Bowker and Star, 1999).

2.3   Implications for Science

The SBD cycle reveals knowledge production as a sorting mechanism, dividing phenomena into transparent ledgers and opaque remainders. This division is not neutral but institutional, driven by demands for legibility in funding, policy, and governance (Scott, 1998). For example, funding agencies often favor projects with clear, measurable outcomes, marginalizing exploratory research into opaque phenomena (Bowker and Star, 1999). This prioritization of the ledger risks overlooking the generative potential of junk, as seen in genomics and AI. The cycle's persistence underscores the need for an analytic lens that revalues opacity, recognizing it as a condition of scientific creativity (Rheinberger, 1997; Barad, 2007). By reframing junk as a generative substrate, we challenge the positivist fantasy of complete transparency and open new avenues for discovery

3   Ledger and Junk as an Analytic Lens

3.1   Ledger as Metaphor

The ledger metaphor originates in account books, colonial taxonomies, and bureaucratic records, where transparency and order are paramount (Poovey, 1998). In modern science, ledgers manifest as datasets, performance benchmarks, and explainability frameworks, designed to render systems legible and operational (**?**). These structures align with positivist ideals of complete knowledge, prioritizing predictability and utility (Hacking, 1983). In AI, for instance, benchmarks like accuracy metrics serve as ledgers, quantifying performance while marginalizing unmeasurable outputs (Bowker and Star, 1999). Similarly, in genomics, the gene-centric paradigm treated coding sequences as a ledger, cataloging functions while deferring non-coding regions as junk (Keller, 2000). The ledger metaphor thus enforces a logic of transparency, shaping what counts as valuable knowledge (Scott, 1998). This logic is evident in scholarly practice, where bibliogra-

phies, though essential, are often deferred as secondary until explicitly demanded (Rheinberger, 1997).

3.2   Junk as Epistemic Category

Junk encompasses the opaque and unledgerable: anomalies, waste, and infrastructural remainders. Mary Douglas's analysis of pollution highlights how junk is defined by its transgression of categorical boundaries, existing outside the ordered ledger (Douglas, 1966). Bruno Latour's concept of black-boxing shows how scientific practice excludes the messy or opaque to maintain epistemic order (Latour, 1987). Édouard Glissant's "right to opacity" frames junk as a site of resistance and creativity, irreducible to transparent systems (Glissant, 1997). Similarly, Gloria Anzaldúa's concept of borderlands highlights the generative potential of marginal spaces, where hybridity and ambiguity foster new forms of knowledge (Anzaldúa, 1987). A cultural parallel appears in Psalm 118:22: "The stone which the builders rejected has become the chief cornerstone" [1]. These perspectives reveal junk not as mere error but as a generative substrate, challenging the hegemony of the ledger (Barad, 2007).

3.3   Remainder and Reversal

The ledger/junk dichotomy mirrors scholarly practice: bibliographies, often deferred as secondary "junk," are structurally essential, surfacing as ledgers when demanded (Rheinberger, 1997). Hans-Jörg Rheinberger's concept of epistemic things emphasizes how remainders persist as material substrates for future inquiry, enabling paradigm shifts (Rheinberger, 1997). Steven Shapin's historical epistemology further suggests that discarded knowledge forms the foundation for new discoveries (Shapin and Schaffer, 1985). Susan Leigh Star's concept

---

[1]This motif of reversal underscores the generative potential of discarded elements, resonating across cultural and epistemic domains (Douglas, 1966; Latour, 1987; Glissant, 1997; Anzaldúa, 1987).

of boundary objects highlights how remainders mediate between communities, fostering innovation at the margins (Star, 1990). This framework challenges reductionist accounts, proposing that opacity is a condition of generativity, not a flaw

4   The Genomics Case: From Junk to Regulatory Substrate

The history of genomics provides one of the clearest instantiations of the ledger/junk dynamic. When molecular biology consolidated around the "central dogma" in the 1950s—that DNA encodes RNA, which in turn encodes protein—the genome was interpreted as a kind of linear ledger. Genes were conceptualized as discrete entries, each linked to a specific protein function. This framework shaped both experimental practice and epistemic expectation: the genome was a book of life, a catalog of elements that could, in principle, be exhaustively decoded.

Yet even at the earliest stages of large-scale genome sequencing, a striking anomaly emerged: only a tiny fraction of DNA—around 1–2 percent in humans—was directly involved in protein coding. The vast majority of sequences did not fit the ledger. In 1972, Susumu Ohno gave this remainder a name: "junk DNA" (Ohno, 1972). By designating the unledgerable portion of the genome as waste, biologists effectively deferred it into epistemic opacity. The ledger was preserved by pushing the anomaly into the category of junk.

This naming was not incidental. As Evelyn Fox Keller later argued, the twentieth-century gene was not merely a technical unit but a cultural construct, shaped by the promise of legibility and control (Keller, 2000). To admit that most of the genome had no obvious function threatened the very economy of legibility that underwrote molecular biology's success. Hence "junk" served as a placeholder category: it preserved the authority of the ledger while cordoning off remainder.

For decades, this designation structured both research priorities and funding.

Projects sought to identify coding sequences, map their functions, and annotate the genome's "useful" parts, while non-coding stretches were left unexplored. As Hans-Jörg Rheinberger has shown in his history of "epistemic things," scientific practice often relies on such material remainders—objects or phenomena that remain opaque within current conceptual frames but may later become central to new paradigms (Rheinberger, 1997). Junk DNA was precisely such an epistemic thing: a remainder initially excluded from the ledger, but persistent as a substrate for future inquiry.

By the late 20th century, however, evidence began to accumulate that non-coding DNA was far from inert. Comparative genomics revealed that many "junk" sequences were highly conserved across species, suggesting selective pressure to maintain them. Studies in developmental biology identified regulatory elements—enhancers, silencers, insulators—embedded in non-coding regions, controlling when and how coding genes were expressed. The "junk" was in fact the scaffolding of genomic regulation.

This epistemic reversal reached public prominence with the ENCODE (Encyclopedia of DNA Elements) project in the 2000s. In 2012, ENCODE researchers announced that as much as 80% of the human genome showed biochemical "function" by some measure (ENCODE Project Consortium, 2012; Pennisi, 2012). The headline "Eulogy for Junk DNA" captured the rhetorical stakes: what had been discarded as waste was now proclaimed essential. Though critics disputed ENCODE's expansive definition of "function," the project nonetheless symbolized a profound shift. Junk was no longer mere surplus; it was the cornerstone of genomic complexity.

Seen through the ledger/junk framework, this history illustrates how epistemic categories are not neutral but politically charged. The ledger—the coding genome—was valorized as transparent, legible, and useful. The junk—the non-coding majority—was marginalized as noise, even though it constituted the bulk of the material. Only when technical, institutional, and theoretical conditions

aligned did the junk become revalued as the very substrate of generativity. The stone rejected by the builders had indeed become the cornerstone.

This trajectory also complicates our understanding of reductionism in biology. By privileging discrete, legible coding units, the central dogma instantiated what Bruno Latour would call a "black-boxed" scientific object (Latour, 1987). Genes were enrolled as stable actants in the network of molecular biology, while the unledgered sequences were black-boxed as irrelevant. The revaluation of junk DNA re-opened the box, revealing that genomic function emerges not from isolated coding units but from distributed, context-dependent regulation. The ledger metaphor, while powerful, had obscured the generative role of opacity.

Moreover, this case resonates with feminist and STS critiques of legibility. Donna Haraway's insistence on "situated knowledges" (Haraway, 1988) reminds us that the very categories of ledger and junk are historically contingent, embedded in social and institutional practices. Non-coding DNA was "junk" not because it lacked function, but because its functions did not fit the dominant epistemic framework. The ledger/junk split thus reveals how power, perspective, and institutional authority shape what is seen as useful knowledge.

Finally, the genomics case demonstrates the productive role of remainder in scientific discovery. Just as Kuhn described anomalies as the seeds of paradigm shifts (Kuhn, 1962), and Feyerabend argued for the epistemic value of methodological pluralism (Feyerabend, 1975), junk DNA exemplifies how opacity sustains scientific creativity. The category of junk deferred complexity rather than erasing it, leaving space for future inquiry to reframe it as central.

In sum, the history of genomics shows how the ledger/junk dynamic is not a trivial matter of terminology but a deep epistemic logic. By relegating the vast majority of the genome to junk, molecular biology preserved the legibility of the ledger. Yet this very remainder became the site of novelty, enabling a richer understanding of gene regulation, development, and evolution. The lesson is clear: opacity is not a flaw to be eliminated but a condition of generativity. To

mistake the ledger for the entirety is to misrecognize the system, just as to quarry marble is to miss the mountain.

5 The LLMs Case: Hallucination as the Generative Substrate

If the history of genomics dramatizes the ledger/junk inversion in biology, then Large Language Models (LLMs) offer its clearest computational analogue. These models, which dominate the current landscape of artificial intelligence, are structured in a way that makes the distinction between "ledger" and "junk" not merely metaphorical but mechanical. At their core, LLMs are probabilistic machines for traversing the grammatical Directed Acyclic Graph (DAG) of natural language. Their high-probability outputs—factual responses, code completions, grammatical corrections—appear to exemplify the ledger side of the dynamic: reliable, transparent, and instrumental. Yet it is their so-called hallucinations, the low-probability continuations that defy easy categorization, that constitute the generative substrate. What is conventionally dismissed as error turns out to be the very condition of creativity.

5.1 The Ledger: High-Probability Continuations

Transformers, the architecture underlying modern LLMs, operate by assigning probabilistic weights to possible next tokens given a sequence of inputs (Vaswani et al., 2017). This mechanism effectively maps the space of language as a network of weighted dependencies. The strongest, most frequently reinforced connections— Paris → is → the capital → of → France—represent the ledger of the system: sequences that are not only probable but legible, easily traced to patterns in the training data.

From a user's perspective, this ledger manifests as predictability and utility.

LLMs can correct grammar, summarize articles, translate between languages, or write executable code. Each of these outputs is a ledger-like function: it can be checked against external references, verified for correctness, and instrumentalized for practical tasks. In software development, for example, code completions provided by LLMs are useful precisely because they align with established libraries and idioms. They reproduce known pathways through the DAG of programming languages, yielding outputs that can be ledgered as correct or incorrect (Chen et al., 2021).

The ledger functions of LLMs thus appear to fulfill the positivist dream of transparency. They suggest that AI can be a deterministic, legible machine, producing outputs that map cleanly onto existing knowledge systems. It is no surprise that early popular discourse emphasized these capabilities: LLMs as assistants, copilots, or productivity tools, valuable because they expand the ledger of what can be quickly produced and verified (Bommasani et al., 2021).

5.2 The Junk: Hallucination as Structured Exploration

Yet alongside these high-probability outputs lies a vast space of low-probability continuations, the so-called "hallucinations." When an LLM asserts that the capital of France is Rome, or invents a bibliographic citation, or produces a surreal analogy, it is traversing pathways that deviate from the ledger. To many observers, these outputs exemplify the unreliability of the system: they are opaque, unverifiable, misleading. They are junk (Bender et al., 2021).

But the crucial point is that these hallucinations are not random. They are structured explorations within the grammatical DAG. The model's probabilistic sampling process ensures that even low-probability tokens are drawn in ways constrained by grammar and context. Thus, hallucinations are almost always syntactically well-formed and rhetorically persuasive. They preserve the skeleton of legibility even as they drift into semantic novelty (Shanahan, 2023).

This is why hallucinations have become the source of both fascination and

concern. On the one hand, they undermine trust in LLMs as ledger-like machines for factual recall. On the other hand, they enable precisely those creative functions that distinguish generative AI: the production of fiction, poetry, visual art prompts, musical scores, and novel conceptual frameworks. Hallucination is the substrate of generativity, the condition under which LLMs can do more than reproduce the ledger (Hofstadter, 1995).

5.3   The Historical Misrecognition

Just as non-coding DNA was dismissed as "junk" because it did not fit the gene-centric ledger, so too have hallucinations been marginalized as bugs because they do not fit the paradigm of information retrieval. Researchers, journalists, and policymakers often frame hallucination as a problem to be solved: something to be eliminated through better alignment, tighter grounding, or more exhaustive training data (Bender et al., 2021; Marcus and Davis, 2022).

This framing reflects the positivist bias toward transparency. To the extent that LLMs are evaluated as factual systems—as search engines, encyclopedias, or oracles—hallucinations are errors. They compromise the ledger. Yet this perspective risks repeating the genomic mistake: mistaking the ledger for the system's entirety, and dismissing the remainder as irrelevant (Keller, 2000).

In practice, hallucinations drive the most compelling use cases for generative AI. Users turn to LLMs to draft stories, invent metaphors, design speculative scenarios, or imagine counterfactuals—tasks where factual correctness is secondary to creative exploration. Just as non-coding DNA regulates and orchestrates expression, hallucinations regulate and expand the expressive range of language models (Barad, 2007).

5.4   Mechanisms of Generativity

Understanding hallucination as generative requires looking closely at how LLMs operate. The transformer's attention mechanism calculates weighted dependencies among tokens, constructing a high-dimensional representation of context. This representation is not human-readable; it is an embedding space in which semantic relationships are encoded as geometric proximities (Vaswani et al., 2017). Within this space, low-probability continuations emerge not as noise but as explorations of less-traveled paths.

Consider a prompt like "Explain democracy as if it were a cooking recipe." The model is not retrieving a known ledger entry but recombining distant regions of its embedding space: political theory and culinary instruction. The hallucinated analogy is structurally well-formed, even if semantically surprising. Similarly, when asked to invent fictional references or synthesize competing frameworks, the model traverses paths that are improbable but grammatically constrained (Shanahan, 2023). These traversals mirror the recombinatory processes of imagination described in cognitive science (Hofstadter, 1995).

What appears as junk is in fact a form of structured search through the latent DAG of grammar and meaning. By drifting from high-probability continuations, the model reveals capacities for metaphor, analogy, and speculation—precisely the hallmarks of creativity (Bowker and Star, 1999).

5.5   The Politics of Transparency

The ledger/junk inversion in LLMs is not only technical but political. Demands for transparency and explainability dominate contemporary AI discourse, particularly in ethics and governance (Burrell, 2016). Policymakers call for interpretable model weights, audit trails, and guarantees against hallucination. These demands echo the positivist fantasy of a complete ledger: a system where every output can be traced, validated, and rendered legible (Scott, 1998).

Yet as STS scholars remind us, transparency is never neutral (Jasanoff, 2004; Haraway, 1988). To prioritize transparency is to privilege certain kinds of knowledge and exclude others. In the case of LLMs, it risks marginalizing the generative capacity of hallucination in favor of ledger-like functions. Just as "junk DNA" was devalued because it did not produce proteins, hallucinations are devalued because they do not produce facts. Both cases reveal how epistemic categories reflect institutional and cultural biases (Glissant, 1997; Anzaldúa, 1987).

To demand that LLMs eliminate hallucination is, in effect, to quarry away the mountain in search of marble. It is to misrecognize the system's true substrate of creativity.

5.6   Creativity as Structural Opacity

Reframing hallucination as generative also challenges the conventional opposition between creativity and error. In the ledger/junk framework, creativity is not a separate module added onto factual recall. It is the emergent property of opacity. Just as non-coding DNA orchestrates expression, hallucinations orchestrate the expressive capacities of language models (Rheinberger, 1997).

This perspective aligns with long-standing philosophical critiques of reductionism. Anomalies, remainders, and opacities are not obstacles to knowledge but conditions of its expansion (Kuhn, 1962; Feyerabend, 1975). In LLMs, the structured opacity of hallucination is precisely what enables novelty. To treat hallucination as junk is to miss its role as cornerstone (Star, 1990).

5.7   Illustrative Cases

A few concrete examples clarify this point:

Fiction and Storytelling. Asked to write a fairy tale about climate change, an LLM may invent anthropomorphic rivers or talking glaciers. These are hallucinations relative to factual climate science, but they generate powerful narrative

resources for public engagement (Bender et al., 2021).

Art and Design. Prompted with surreal or contradictory descriptions, LLMs (and their multimodal counterparts) generate imaginative outputs that inspire visual artists. Here, the very drift from ledgered reality enables aesthetic novelty (Manovich, 2020).

Philosophical Analogy. In conversation, LLMs often produce unexpected metaphors—likening algorithms to gardens, or governance to choreography. These analogies are not retrieved from any ledger of knowledge but hallucinated through low-probability traversals. They open new conceptual spaces for human thought (Hofstadter, 1995).

Speculative Science. LLMs sometimes generate hypotheses or frameworks that, while factually unfounded, can serve as starting points for exploration. Like non-coding DNA mutations, most of these are dead ends, but some spark genuine insight (Rheinberger, 1997).

In each case, the value lies not in correctness but in novelty. Hallucination functions as the substrate of creativity.

5.8   The Cornerstone Reversal

The ledger/junk inversion thus repeats itself: what was initially dismissed as error becomes recognized as essential. Just as "junk DNA" was reframed as regulatory architecture, hallucination in LLMs must be reframed as generative architecture. The epistemic reversal is captured in the same cultural motif: the stone the builders rejected has become the cornerstone (Glissant, 1997).

Recognizing this inversion requires a shift in how we evaluate AI. Rather than treating transparency as the supreme virtue, we must acknowledge opacity as a condition of creativity. Rather than quarrying away hallucination, we must learn to work with it, steering its generativity without reducing it to ledger-like functions.

5.9 Conclusion to Section

The LLM case demonstrates that opacity and remainder are not incidental but structural. Hallucination is not a bug to be eradicated but a feature to be cultivated. To treat LLMs as if their purpose were solely to expand the ledger of factual knowledge is to repeat the genomic error of mistaking the ledger for the mountain. The challenge is to design, evaluate, and govern these systems in ways that preserve their generative substrate while mitigating risks. This means respecting opacity not as a failure of legibility but as the condition of emergence.

6 Contribution and Implications

6.1 Epistemology

The ledger/junk dynamic reveals a recurring structure in knowledge systems: the division of phenomena into transparent ledgers and opaque remainders. From natural history to genomics and computational systems, opacity emerges as a condition of generativity, not a flaw (Kuhn, 1962; Rheinberger, 1997). This challenges positivist assumptions that complete legibility is possible or desirable, suggesting instead that remainders are the substrate of novelty (Barad, 2007; Shapin and Schaffer, 1985). By reframing junk as generative, we propose a new epistemology that values opacity as a driver of scientific creativity

6.2 Governance of AI

The demand for transparency in AI governance risks prioritizing ledger-like functions at the expense of generative opacity (Burrell, 2016). While mitigating harms like misinformation is critical, eliminating hallucination entirely would dismantle the creative potential of LLMs. Governance must balance accountability with the preservation of emergent capacities, recognizing opacity as a feature, not a bug (Jasanoff, 2004; Glissant, 1997). This requires policies that foster responsi-

ble exploration of AI's opaque zones, drawing lessons from genomics' embrace of junk DNA (Pennisi, 2012).

6.3 Science Policy

The genomics case offers lessons for science policy: funding and exploring the "junk" is as crucial as cataloging the ledger. Just as non-coding DNA proved essential, the opaque zones of AI systems may harbor future breakthroughs (Keller, 2000; ENCODE Project Consortium, 2012). Policies that incentivize exploration of remainders can foster paradigm shifts, as seen in historical cases like phlogiston

6.4 Philosophy of Science

This framework challenges reductionist epistemologies, proposing a theory of opacity as foundational to scientific creativity. By aligning with pluralist and feminist critiques, it values partial perspectives and situated knowledges (Feyerabend, 1975; Haraway, 1988). The ledger/junk lens offers a new way to understand the role of anomalies in knowledge production, with implications for how we conceptualize scientific progress

7 Conclusion

From Pliny's critique of quarrying to the revaluation of junk DNA and the generative hallucinations of LLMs, the ledger/junk dynamic reveals a persistent epistemic pattern. Scientific practice sorts phenomena into transparent ledgers and opaque remainders, yet it is the latter that often drive novelty. This dynamic is mirrored in scholarly practice: bibliographies, treated as secondary "junk," become foundational ledgers when surfaced (Rheinberger, 1997). To prioritize transparency over opacity is to quarry away the mountain in search of marble, mistaking the ledger for the system's entirety. Every ledger leaves a remainder;

in that remainder lies creativity. This paper calls for a revaluation of opacity across scientific domains, urging researchers, policymakers, and designers to embrace the generative potential of the unledgerable.

References

Anzaldúa, G. (1987). *Borderlands/La Frontera: The New Mestiza.* Aunt Lute Books, San Francisco.

Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning.* Duke University Press, Durham, NC.

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? pages 610–623.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Bowker, G. C. and Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1):1–12.

Chen, M., Tworek, J., Jun, H., Yuan, Q., de Oliveira Pinto, H. P., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., and Ray, A. (2021). Evaluating large language models trained on code. In *arXiv preprint arXiv:2107.03374*.

Chomsky, N. (1957). *Syntactic Structures.* Mouton, The Hague.

Daston, L. and Park, K. (1998). *Wonders and the Order of Nature, 1150–1750*. Zone Books, New York.

Douglas, M. (1966). *Purity and Danger: An Analysis of Concepts of Pollution and Taboo.* Routledge, London.

ENCODE Project Consortium (2012). An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57–74.

Feyerabend, P. (1975). *Against Method.* Verso, London.

Glissant, É. (1997). *Poetics of Relation.* University of Michigan Press, Ann Arbor, MI. Translated by Betsy Wing.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science.* Cambridge University Press, Cambridge.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3):575–599.

Hofstadter, D. R. (1995). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought.* Basic Books, New York, NY.

Jasanoff, S., editor (2004). *States of Knowledge: The Co-Production of Science and the Social Order.* Routledge, London.

Keller, E. F. (2000). *The Century of the Gene.* Harvard University Press, Cambridge, MA.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* University of Chicago Press, Chicago.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society.* Harvard University Press, Cambridge, MA.

Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing.* MIT Press, Cambridge, MA.

Manovich, L. (2020). *Cultural Analytics.* MIT Press, Cambridge, MA.

Marcus, G. and Davis, E. (2022). A very preliminary analysis of dall-e 2. *arXiv preprint arXiv:2204.13807*.

Ohno, S. (1972). So much "junk" dna in our genome. *Brookhaven Symposia in Biology*, 23:366–370.

Pennisi, E. (2012). Encode project writes eulogy for junk dna. *Science*, 337(6099):1159–1161.

Poovey, M. (1998). *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*. University of Chicago Press, Chicago.

Rheinberger, H.-J. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford University Press, Stanford, CA.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press, New Haven.

Shanahan, M. (2023). Talking about large language models. *arXiv preprint arXiv:2312.10628*.

Shapin, S. and Schaffer, S. (1985). *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton University Press, Princeton, NJ.

Star, S. L. (1990). Power, technology and the phenomenology of conventions: On being allergic to onions. *The Sociological Review*, 38(1):26–56.

the Elder, P. (1855). *The Natural History*. Taylor and Francis, London.

Vaswani, A., Shazeer, N., Parmar, N., Uszoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30.

Whittaker, E. (1951). *A History of the Theories of Aether and Electricity*. Thomas Nelson and Sons, London.