Proceedings of ICADSS 2025 Symposium: Forum on Generative AI Technologies and Media Industry Applications
DOI: 10.54254/2753-7064/2025.BJ29177

An Analysis on AI Hallucination from the Perspective of Media Archaeology

Xunhao Liu[1], Yue Meng[2*]

[1]School of Arts and Communication, Beijing Normal University, Beijing, China
[2]School of Communication, Fujian Normal University, Fuzhou, China
**Corresponding Author. Email: myletha1920@gmail.com*

Abstract. The current controversy surrounding the definition of AI hallucinations in the field of artificial intelligence research reveals the inherent limitations of an engineering-centered perspective. Based on the German theory of Kulturtechnik and through a retrospective analysis of media history, this paper argues that AI hallucinations are not merely technical flaws, but cultural practices that continue the developmental logic of earlier media such as writing and printing. Whether it was the telescope challenging the interpretive authority of theology, or the printing press shaping cognitive power, technological innovations have consistently structured cultural power by defining the boundaries of "reality," with human anxieties over the erosion of cognitive privilege deeply embedded throughout. As a product of cognitive augmentation in digital media, AI hallucinations, by rewriting the limits of "reality," constitute a dynamic frontier in the evolution of knowledge forms within human–machine symbiotic civilization. This paper advocates abandoning the binary corrective approach that treats hallucinations solely as technical errors, embracing instead a "new reality" under probabilistic distributions, and reflecting on the emergent ethical relationships of human–machine co-symbiosis and co-evolution.

Keywords: AI hallucination, generative artificial intelligence, human–machine ethics, philosophy of technology, media archaeology

1. Introduction

In today's era where artificial intelligence rapidly permeates human knowledge systems, the term "halluc," originally from psychology and literature, has been re-embedded into technical vocabulary. It refers both to algorithmic output deviations and suggests a cultural symptom that transcends factual errors. People are eager to label it as "error," yet overlook the metaphorical structure inherent in this naming, which human anxiety and self-defense in maintaining the ancient concept of "reality." The hallucinated outputs of technical systems reveal not only errors in computers under logistically complete states but also the plasticity of "real" as a cultural concept: in every media innovation, the boundaries of reality are red and recoded.

From Galileo's telescope to generative artificial intelligence's language models, technology exists in a dual stance—both as an extension of cognition and as a creator of hallucinations. The telescope amplifies the visibility of optical facts but is seen as "hallucination" to its challenge to theological interpretations; writing technology carries experience through symbolic structures but generates a duality of reason and imagination in Romantic poetics; when algorithms generate images and text in potential spaces, hallucinations transform into the mode of production of digital civilization. It is no longer a pathological error but a cultural technique (Kulturtechnik), an operation that constructs cultural order by defining "what is real."

Therefore, the so-called technical ideal of "eliminating hallucinations" perpetuates the rational myth that has been in existence since the Enlightenment. Viewed from a deep perspective media history, hallucination is not merely a deviation that needs correction but a driving force of cultural evolution—a way for technology to intervene in reality. In this sense, artificial intelligence hallucination becomes a key entry point for reconsidering the human-machine relationship: it compels us to confront the of human cognitive privilege and forces us to accept a probabilistic, dynamic "new reality." Rather than saying hallucination obscures truth, it might be said that it reveals the technical conditions and discursive boundaries of "reality" itself. In the face of an algorithm-generated world, what need is not demystification, but a reinterpretation and symbiosis with hallucination.

2. The controversy of AI hallucination and its discourse violence

In the field of computer science, "hallucination" is a term filled with ambiguity and even controversy A paper presented at the CAI conference in early 2024 reviewed the different usages of this term in artificial intelligence research and pointed out that it "lacks consistency in definition..." requiring interdisciplinary collaboration to ensure the concept remains semantically clear, as it may have profound impacts across various fields [1]. Such reflections are not isolated cases. Many information technology researchers have realized the vagueness of this concept, and they often attempt to address a superficially simple yet inherently complex question—"What exactly are we referring to with artificial intelligence hallucination?" Consequently, some advocate for a specific sub-con to encompass other usages, while others suggest replacing "hallucination" with terms like "confabulation," "delusion," or "fabrication" in different contexts. The ultimate goal remains to overcome the phenomenon of hallucination through strict engineering definitions, from data collection, annotation, parameter adjustment in the operational chain.

However, what does "eliminating hallucination" really mean? Can this concept be clearly defined? The thorny issue lies in the fact that the boundaries of "reality" themselves are not certifiable entities but an open range that can only be falsified but never ultimately proven. Therefore, limiting hallucination solely to the technical domain of artificial intelligence or computer science unveils the narrowness of its theoretical perspective. Hallucination, as a high-frequency and controversial concept, has long transcended disciplinary boundaries, becoming a phenomenon of symbol across cultures and contexts. Rather than establishing a specific entity as the "ontology" of hallucination, it is more important to trace the mechanism of this difference: why do different cultural groups use the same term in specific contexts to refer to similar experiential phenomena? In this a priori dimension, artificial hallucination is both a metaphor and an entity; it is both a sign and a form of cultural technique.

The term "cultural technique" originates from the German media studies school, with its ideological lineage traceable to Friedrich Kittler and his colleague Bernhardert. This theory emphasizes that technical operations precede cognitive formation, constituting the a priori conditions for cognitive conceptualization [2]. In other words, "writing as a cultural technique exists prior to the concept of letters, and counting as a cultural technique appears before the concept of numbers." Viewed from this, the conceptual controversy surrounding artificial intelligence hallucination precisely validates the interdisciplinary vitality of hallucination. As a psychological term, hallucination refers to a distortion or disorder at the level of perception; as a technical term, it is transformed into a manifestation of deviation from model generation logic. If halluc is understood as cultural technique, it implies transcending simple ontological entanglement and instead inquiring into its technical a priori, that is, how hallucination exists in an operational form prior to the certainty of "reality."

The earliest case of using "hallucination" technical context within the field of artificial intelligence can be traced back to the paper Hallucinating Faces presented at the fourth IEEE International Conference on Automatic Face and Gesture Recognition in 2000 [3]. This research proposed an algorithm based on Gaussian pyramids and gradient prior learning that achieves "hallatory" reconstruction of image details by supplementing missing information from low-resolution inputs. Here, "hallucination" no longer signifies perceptual error in the traditional sense but rather serves as a technical metaphor—algorithms utilize structured knowledge systems (i.e., training datasets) and probabilistic mechanisms to fill in unacquired information, thereby generating a cognitively enhanced "reality."

In this sense, the subject of the hallucination in "Generating Facial hallucinations" is no longer the human, but rather the algorithm. This shift prompts a deeper philosophical inquiry: Is artificial intelligence a simulation of the mind or purely a technology? As Philip K. Dick posed in Do Androids Dream of Electric Sheep?, this question runs throughout the entire history of artificial intelligence. Turing even asserted in his foundational paper Computing Machinery and Intelligence: "Can machines think? This question is meaningless and not worth discussing." This suspended attitude reveals the crux of the issue should be consistently defined the boundaries of "intelligence" and "hallucination" within a human-centered framework [4].

From the historical events generated erratically by ChatGPT (Figure 1) to the phenomenon of "phantom limbs" that emerged with the DALL·E model (Figure 2) the "errors" of technical systems are labeled as "hallucinations," which implicitly carries a moralizing and pathological critical tendency. When humans refer to machine discrepancies as "hallucinations," they are essentially delineating "truth" and "falsehood" using an existing knowledge system, thereby up the pre-existing cultural power structure. Technology fills information gaps through probabilistic models, and its mechanism is structurally very similar to the "emergence" of the human cognitive system—yet the former is diagnosed as pathological while the latter is celebrated as intelligent. This is precisely the discursive revealed by Foucault in The Order of Discourse(L'ordre du discours): defining the machine's act of information completion as a "hallucination" is, in essence, a form of technological discipline implemented through medical metaphors, maintaining the human cognitive paradigm as the arbiter of truth [5].

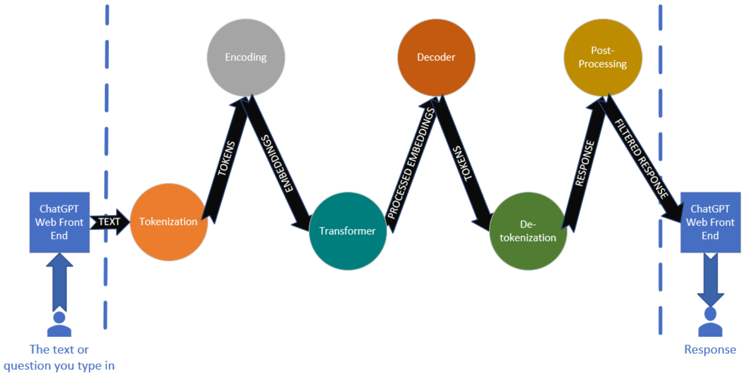

Figure 1. Flowchart of text generation process in ChatGPT

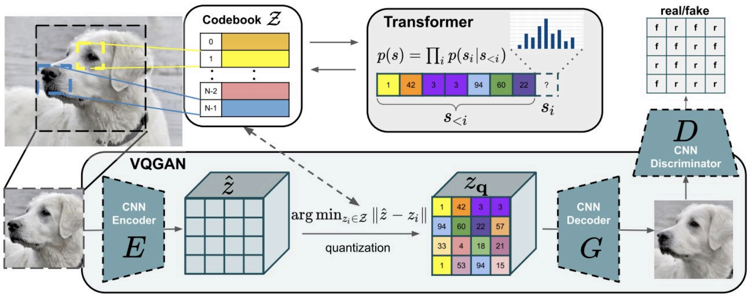

Figure 2. Basic principle diagram of DALL·E

The deep structure of this discourse power can be traced back to the modern cognitive system nurtured by writing technology. The couch in a psychoanalytic clinic and the desk of Romantic writers share the same material foundation—symbolic abstraction of experience. Freud's free association therapy requires patients toate their stream of consciousness through language, while Romantic writers invoke imagination through text; both generate a "hallucinatory" reality within a symbolic order. When language models generate text through vector space computations, they essentially continue the hallucinatory mechanism of the written medium: reconstructing discrete symbols into narratives, this "semblance of truth" is exactly what constitutes the hallucination of writing as a medium.

Consequently, the question shifts from "Do machines produce hallucinations?" to "How does the hallucination itself, as a cultural technique, construct the real?" If acknowledge that symbolic systems existed prior consciousness, then all cognitive systems, from knot-recording to neural networks, construct the world through symbolic operations. Artificial intelligence hallucinations are merely the latest fractures in this continuous spectrum—the old concept of "reality" encounters disintegration here, while "technology" becomes a force for redefinition.

Therefore, what truly deserves attention is understanding why hallucinations have become our means of redefining "the real." It not only reveals the cognitive fractures in human-machine relationships but also reminds us that each media innovation represents a reprogramming of "reality." Artificial intelligence hallucinations stand as the milestone of digital civilization, presenting not just the machines' hallucinations but also humanity's cultural echo gazing at itself in the technological mirror.

3. The media-archaeological genealogy of AI hallucination

The "hallucinations" of artificial intelligence extend along a long technical perception lineage in the history of Western knowledge. Each media innovation has been accused of being a device that "creates hallucinations," challenging old paradigms of reality material forms and reconstructing the way humans understand the world through new perceptual logics. From optical laboratories to writing desks, and to the clinics of psychoanalysis, hallucinations have always lurked in the depths of technology, serving as the shadow of human self-awareness.

In 1610 Galileo improved the Dutch optical device to create the astronomical telescope—contemporaries referred to it as a "tubular eyeglass" or "viewing tube" (see Figure 3). With it, he systematically observed the surface of the moon and the movements of Jupiter's moons the first time, a discovery that directly challenged the Aristotelian-Ptolemaic geocentric model of the universe and its scholastic philosophical system. To avoid a direct confrontation with the Roman Catholic Church, Galileo cleverly incorporated scientific results into the discourse of power: he proposed naming newly discovered moons the "Medicean Stars" to please the Grand Duke of Tuscany. However, this strategic phrasing did not quell controversy. Venetian scholar Paolo Sarpi criticized Galileo for being "ungrateful," accusing him of "hiding the contributions of Venetian artisans in lens polishing." this accusation was political rhetoric, it revealed the subtle relationship between scientific discovery and the legitimacy of power, as Galileo established a precarious balance between the optical experiments of science and the symbolic order of the Church [6].

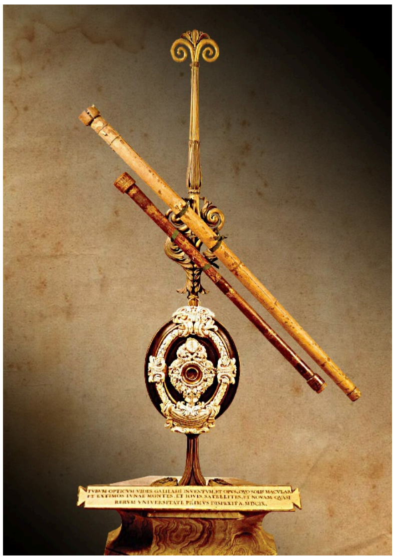

Figure 3. Galileo's telescope design

In another instance, scholar Martin Horky, in his work A Brief Drifting Against the Starry Report, proclaimed that Galileo's observational results were nothing more than "visual hallucinations caused by lens defects." He cited his own experience of "seeing three suns" during a eclipse as an example to accuse Galileo of being "deceived by reflection." Even the neutral philosopher Giovanni Battista Manso, while acknowledging that Galileo was "almost another Columbus," felt uneasy about his observational conclusions—because these new discoveries "could not be explained within the existing philosophical principles." this debate, the conflict between empirical observation and interpretive belief was laid bare. Galileo's discoveries not only undermined the theological order of the universe but also allowed "reality" to be redefined by media technologies.

DOI: 10.54254/2753-7064/2025.BJ29177

The controversy over optical instruments anticipates the logical paradox of knowledge system. The telescope is not a passive "mirror" reflecting the universe but rather a medium that redefines the boundaries of the "visible" through refractive operation. It opened up a new "optical world" on a symbolic level, which is structurally similar to the biblical notion of "the beginning was the Word," as both construct new realms of reality through a technical "act of creation." Thus, Galileo's "visual hallucinations" are essentially not a scientific error but a manifestation of cultural craft: it rewrote the boundaries between theology and perception through technological means.

Throughout European, the Church's fear of the telescope did not stem from its physical properties but from its revelation of the alternatives in interpretative systems. As later Jesuit astronomers' mastery of this technology indicated, the real threat lies not in "seeing," but in "how to see." When the of interpretation loses its singularity, the structure of theological truth collapses. The revolution of optical media thus becomes a crisis of symbolic power, with hallucination serving as a sign of self-renewal within the knowledge system.

The history of writing technology is no different. Since Gutenberg's printing (Figure 4) press the standardized reproduction of text has fixed "reality" within the order of the text. When Novalis declared, "If one can read correctly, a real visible world unfolds within the lines," he actually obscured the materiality of writing technology. The mechanical arrangement of letters was mistaken a form of truth, thus turning literature into a cultural apparatus that produces hallucinations. German Romantic writers—Goethe, Novalis, Hoffmann, among others—explored a new "reality" within this media transition: a dual construction between reason and fantasy, experience and transcendence.

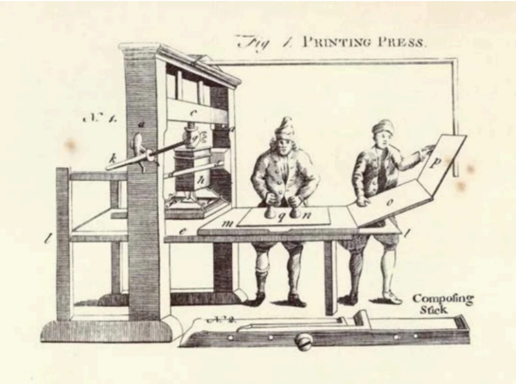

Figure 4. Gutenberg printing press

In Wilhelm Meister's Apprenticeship, Goethe reconstructs the fictive nature of dramatic life through the form of a novel, replacing "performance" with "reading," thus transforming literature into a hallucinatory space that substitutes for reality [7]. Novalis further deepens this poetic in Heinrich von Ofterdingen: when the protagonist "hallucinates" his deceased wife's voice after her passing, language is detached from its corporeal source and becomes a self-sufficient sign [8]. It is in this media practice that transcends that hallucination transforms into a technical operation of the soul. Writing becomes an act of "inscribing"—it leaves material traces on the spiritual substrate of the subject, making written language a community of psychology and media [9]. This hallucinatory logic of writing has gained new extensions with the rise of psychoanalysis.

Freud's free association therapy requires patients to reconstruct the stream of the subconscious through linear narrative, which is similar to the writing strategies of the Romantic period. Both organize chaotic experiences through symbolic systems, transforming disorderly perceptions into a narratable "mind." When Hoffman depicts Nathaniel "phantasm" in The Sandman (Figure 5), he is actually writing about the hallucination of the act of writing itself: the eye is both a sensory organ and a mediating device; the fear of being robbed of sight symbolizes the anxiety of losing the right to interpret. Freud interprets this as “castration anxiety” (Figure 6), while Hoffman reveals the same psychological mechanism in literary form—how language maintains the integrity of the subject through hallucination [10].

Figure 5. Stage still from Hoffmann’s The Sandman

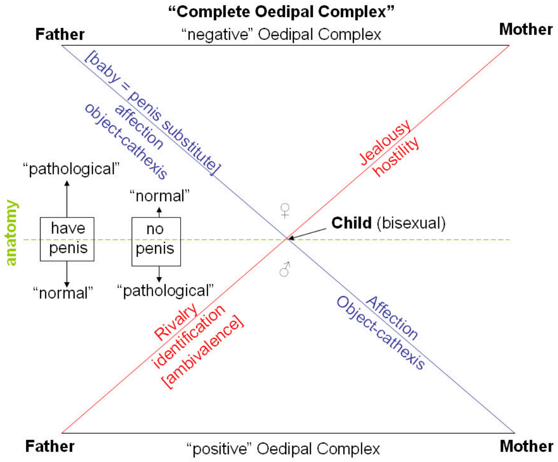

Figure 6. Diagram of Freud’s concept of “castration anxiety”

Telescopes, printing presses, and the psychoanalyst’s couch together form a genealogy of hallucination in media (Table 1). They are all devices that generate “reality,” representing cultural techniques that intervene in cognitive orders through technological operations. From Galileo to Freud, from telesc to language models, each medium innovation has redefined the meaning of “seeing.” hallucination is not a betrayal of truth, but rather a different form of truth in the transformation of media [11].

Table 1. A media-archaeological genealogy of “hallucination”

Historical Phase	Key Medium	Core Phenomenon	Form of Hallucination	Ethical and Cultural Insight
17th Century	Telescope (Galileo)	Scientific observation challenges theological authority	Optical hallucination: vision doubted as hallucination	Technology disrupts religious hermeneutics; reality becomes a question of communicative power.
15th Century onward	Printing press (Gutenberg)	Knowledge becomes reproducible and standardized	Typographic hallucination: form fabricates orderly reality	Reality becomes dependent on media structure; reason and hallucination co-constitute knowledge.
18th–19th Centuries	Romantic writing (Goethe, Novalis)	Literature generates “poetic reality”	Narrative hallucination: language reorganizes experience	Hallucination becomes creative force; writing transforms into spiritual reproduction.
Late 19th Century	Psychoanalysis & literature (Hoffmann, Freud)	Perception and subject displaced by technology	Psychological hallucination: anxiety of being seen	Technological expansion induces a crisis of perceptual ethics.
Contemporary Era	Generative AI	Algorithmic production of simulated worlds	Algorithmic hallucination: fiction and reality interweave	Hallucination becomes media normality; ethical cohabitation and trust must be reconfigured.

Therefore, the history of hallucination is a history of “the production of the real.” Galileo’s telescope, the writing practices of the Romantics, and Freud’s psychological therapies collectively reveal a deep-seated media truth: the so-called hallucination is precisely the way culture reorganizes itself through technology. Every instance of “error” is an opportunity for the knowledge system to shift towards a new cognitive model, just as today’s artificial intelligence hallucinations are merely a digital continuation of this ancient logic.

4. The AI hallucination as cultural craft

Reflecting on the history of hallucinations in the age of writing technology, our initial question becomes increasingly complex: what does it mean to completely eliminate hallucination? Can hallucinations be clearly defined or delineated? This question may need to be set aside, but at the very least, the history of hallucination makes us aware that there exists a technological structure that predates the binary distinction between “reality” and “hallucination.” This structure, before the formation of concepts, has already defined ways of cognition through operational levels.

hallucination is understood as a technological presence in the Heideggerian context, it infuses human culture with difference. It is precisely this difference that, by defining what constitutes hallucination, shapes the cultural realm referred to as reality. The German media scholar Bernhard Siegert believes that this is one of core theoretical characteristics of cultural craft. Cultural craft exists prior to concepts and provides the conditions for their establishment. Long before humans abstracted writing or letters into concepts, writing activities had already commenced; the concept of the image emerged post the creation of paintings and sculptures; even without familiarity with music systems, people can still sing and play music. The emergence of counting occurred long before the formation of numbers. It is certain that most cultures engage in some form of counting or computational activity, yet they may not have produced an abstract understanding of numbers.

Cultural craft embodies the externality andity of symbol systems; it signifies a complex network of actions composed of technical objects and operational processes [12].

The introduction of the theory of cultural craft from the German media studies field is because its birth context shares structural similarities with our current technological cultural predicament. Since the decline of Kant's philosophical system inquiries into transcendental questions within German intellectual circles have not ceased but have manifested in new historical forms in the realm of media studies [13]. This epistemological turn is partly fueled by Germany's unique political situation post-war. When the Berlin Wall had yet to be fully dismantled, a group of caught in ideological fissures used the term "media technology" to evade censorship, thereby continuing the tradition of transcendental philosophy. By linking philosophical transcendental questions to technological analysis through oblique writing, they laid the theoretical foundation for the formation of the cultural craft school. Although this academic strategy a twist of pragmatic considerations, it has nonetheless left a distinct epistemological stratification in subsequent research.

The knowledge production mechanisms during the Cold War, particularly following the French "68" student movement, shaped a unique structure for German media studies. Under the dual pressure of ideological control and academic, early researchers were compelled to conceal their critical edge within the historical investigation of technology [14]. This resulted in a special analytical approach: they no longer focused on the alienation of the culture industry like the Frankfurt School but shifted their attention to the constitutive power of "secondary media" such as, typewriters, and letter systems. Siegert recalls that Gittler once pointed out at a pivotal meeting that decided the direction of the discipline: "When media analysis becomes a reference point for other things, the spiritual realm of traditional humanities retreats from the center of epistemology." This shift the Foucaultian "historical transcendental" with a "technological transcendental," becoming a materialized rewriting of Kant's transcendental philosophy—media technology thereby secured its foundational position in the knowledge system.

This epistemology, born in a particular period, faced misinterpret by the Anglo-American academic community when it sought to restore the original meaning of philosophy in the late 1990s. Anglo-American researchers became perplexed by the German media scholars' tendencies toward "technological determinism" and their methodological tendencies towards historical investigation [15]. The former relates to German scholars early ideological evasion strategies, while the latter stems from two paradigms' differing understandings of "publicness." The Anglo-American academic community, inheriting Habermas's theory of the public sphere, views mass communication as a natural realm for cultural studies; in contrast, German academia on how media constitutes the conditions for "discourse networks" (Aufschreibesysteme). In other words, what they refer to as "publicness" points to the transcendental aspects of media structures. Siegert summarizes this well: "When the Anglo-American academic community asks, How do we become post-human? " German scholars explore the genealogical relationships between human and non-human—the former focuses on the results of technology, while the latter probes into the premises of technology [16]."

This difference is particularly evident in the practice of media archaeology. The cultural craft rejects the positivist approach of historicizing media as mere timelines of artifacts and emphasizes the function of media as a "reference system." This stance differs not only from the material reductionism of traditional science history but also from the Frankfurt School's ideological critique. Just as the French thinker Paul Virilio the seeds of accelerationism in the war machine, German scholars identified new cognitive paradigms in the technological apparatus of the Cold War era—technology is both a phenomenon of history and a premise for its formation.

Today, people are accustomed to explaining many debates sparked by artificial intelligence in terms of a dichotomy between culture and technology, a binary that actually inherits transcendental subject view formed during the Enlightenment in Western Europe. As early as the 1920s, Heidegger criticized the “scientific-mythical” binary framework proposed by German philosopher Ernst Cassirer. Although Heidegger’s position in the philosophy of technology has long been established, the academic world still frequently encounters the questions he once raised. Otherwise, American philosopher Hubert Dreyfus would not have spent his life advocating that computer science draw on Heidegger's thoughts to reflect on the essence of artificial intelligence.

In his later years, Heidegger placed his hopes reshaping metaphysics in cybernetics, while the theory of cultural craft, which emerged at the turn of the century, inherited his transcendental framework of the philosophy of technology and second-order cybernetics’ self-referential considerations. This inheritance can easily lead to the misconception that can only be divided into first-order and second-order states. In reality, within the contexts of cybernetics and cultural craft, the layers of technology often intermingle. For instance, in the case of writing, a child’s stream-of-consciousness diary exemplifies the first-order nature of technology when the diary content includes descriptions of “the act of writing itself,” the self-referential characteristics of writing reveal a second-order structure. The romantic writers mentioned earlier understood this deeply; they constructed a dream of hallucination within the poetic world—the poet’s journey chasing the symbol of the “blue flower” reveals the dual structure of language and hallucination. Even as Offred gains clarity in the story, she cannot escape the abstract hallucination dimension generated by the written word, which is the true meaning of the “blue flower” in romantic symbolism.

5. Reality based on probability distribution

From the perspective of cultural techniques, the generation of culture can be considered as originating from difference. This is exemplified by cultures stemming from the Pentateuch, which trace their origins to the elemental “good–bad” distinction introduced when God separated light from darkness on the first day of creation: “God saw that the light was good, and He separated it from the darkness [17].”

However, the differences between different groups are not in a parallel relationship; it can even be said that any given set of differences can be redefined as being on either side of another set of differences. Therefore, although all cultures and the differences they are based on are, so far transitional, contingent, and temporary, they can still repeatedly be regarded as eternal truths, the essence of things, or natural laws over a long period of time. This does not prevent them from being de-differentiated or re-differentiated and subverted in another culture.

Artificial intelligence hallucinations, as contemporary variants of old cultural techniques, still retain the core essence of their technological a priori. It divides the output data of artificial intelligence models directed toward humans into two parts: real and non-real (in the broad sense of “hallucination”). Humankind has developed a certain intuitive understanding of the potential for artificial intelligence to create hallucinations through science fiction represented by The Matrix, although this is essentially no different from the Romantic poets' fear or concern about writing replacing the subject of the writer.

This leads us to ask ourselves: is the elimination of artificial intelligence hallucinations akin to the rigid adherence to the “geocentric theory” that stifles incomprehensible new technologies, or is it a concern that our fabricated literary alter egos are usurping the hallucin techniques that originally belonged to us? However, poets during the prosperous era of the publishing industry, much like medieval clergy, were filled with pride in their constructed hallucinations based on symbolic systems, also possessing sufficient confidence. Otherwise, one of the “Weimar Trio,” Johann Gottfried Herder would not have associated the origin of language with herding and declared that the fences, as tools of domestication, existed prior to the differentiation between herders or hunters and livestock.

Fences transform wild sheep into domesticated animals, physically isolating predators from prey. This spatial control technique suppresses humanity’s primitive hunting instincts, allowing humans to interact with animals in a non-instinctual manner, thereby compelling humans to establish a symbolic system through observation and naming This process of symbolization is the foundation of language and rational thought, as well as the prototype of education [18]. When Herder established domesticated livestock as the paradigm of the thinking human subject, which he regarded as a technological a priori, he naturally placed the barbarism represented by the outside the fence in opposition to this differentiation. The flaws of Western centrism closely related to this cultural affinity will not be discussed here, but the brutal history of the colonial period at least serves as a warning.

Just as the Old Testament employs the “good–bad” binary as a cornerstone for establishing culture, attempts to eliminate artificial intelligence hallucinations or mitigate their harms already rely on binaries such as “hallucination–non-hallucination” or “real–non.” These distinctions actively shape digital culture, while carrying obvious biases that may not be deeply or consciously recognized.

In the late 18th century, Novalis regarded Goethe as a model for narrating everyday affairs, due to his exceptional ability to elevate the individual to the general within the literary world. Half a century prior, a British mathematician named Thomas Bayes had already achieved this operation symbols more obscure than the German language, as he established the mathematical theory later known as “probability theory.” The intricacies of this concept can be understood through the famous Bayes’ theorem, revealing its subversive impact on the familiar point-based worldview.

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)} \tag{1}$$

An illustrative scenario employing contextual substitution (Table 2) can be constructed under the framework of Bayesian theory. Suppose an artificial intelligence model exhibits a hallucination probability of only 1%, expressed mathematically as P(hallucination) = 1% and P(no hallucination) = 99%. Concurrently, a highly accurate detection algorithm achieves a 99% accuracy rate, formally defined as P(failed | hallucination) = 99% and P(passed | no hallucination) = 99%, where the symbol "|" denotes conditional probability. The critical calculation involves determining the probability of actual hallucination given algorithmic detection failure, P(hallucination | failed). Contrary to intuitive expectations of 99% accuracy, the Bayesian computation proceeds as follows: First, the total probability of detection failure is calculated as P(failed) = P(failed | hallucination)·P(hallucination) + P(failed | no hallucination)·P(no hallucination) = 99% × 1% + 1% × 99% = 1.98%. Subsequent Bayesian inversion yields P(hallucination | failed) = (99% × 1%) ÷ 1.98% = 50%.When an imagined super model encounters another imagined high-quality algorithm, the outcome may differ significantly from the intuition of most people, falling far short of their emotional.

Table 2. Bayesian probability inference for detecting AI hallucinations

Category	Event Description	Symbolic Representation	Probability Value (%)	Explanation
Prior Probability	Probability of the model hallucinating	P(hallucination)	1	Proportion of actual hallucinations in the model
	Probability of the model not hallucinating	P(no hallucination)	99	Proportion of true outputs from the model
Conditional Probability (Detection Algorithm Performance)	Probability of the detection algorithm identifying as "not passed" when there is a hallucination	P(not passed	hallucination)	99
	Probability of the detection algorithm identifying as "passed" when there is no hallucination	P(passed	no hallucination)	99
	Probability of the detection algorithm identifying as "not passed" when there is no hallucination	P(not passed	no hallucination)	1
Joint Probability	Overall probability of not passing the detection	P(not passed)	1.98	Calculated as: P(not passed) = P(not passed
Posterior Probability	Probability that there is indeed a hallucination in texts that did not pass detection	P(hallucination	not passed)	50
Conclusion	Deviation between actual detection results and intuitive expectations	-	-	Despite an algorithm accuracy of 99%, when hallucinations are rare (only 1%), only 50% of the "not passed" results are true hallucinations.

To mitigate artificial intelligence hallucinations, the theoretically-proposed algorithmic solutions can be replaced with pragmatic, skilled human engineers. However, expectations regarding the outcomes of such substitution should be tempered. According to data provided by Microsoft engineers, although the GPT-3 (text-dinci-) model reached a scale of hundreds of billions of parameters a few years ago [19], its number of neurons is comparable to that of the human brain, which easily leads people to fantasize about so-called “superintelligent AI.” Little do they know, a single cloud contains tens of trillions or even hundreds of trillions of water molecules, which is just one of countless astronomical figures that are readily available.

When a set of supercomputer arrays equipped with the best algorithms attempts to reach a conclusion about when a capricious cloud will turn into rain, all they can do is approximate Brownian motion in the physical world using mathematical expectations. At least under the current conditions, even regarding open models, the rate of hallucination itself can only be estimated; thus, its very existence contradicts our anticipated notion of “truth.”

The pursuit of approximating reality through probability distributions coincides with efforts to dissolve the boundaries between such approximations and actual reality. In this process, algorithm engineers and scientists emerge as contemporary romantics. Their published research on reducing hallucination rates, along with the intricate procedures labeled as "hallucination elimination" in technical documentation, epitomizes the endeavors of digital-age posthuman entities in their relentless quest for the cyborg's metaphorical "blue flower."

6. Conclusion

In the mythic age of artificial intelligence, efforts to eliminate the hallucinations of large language models resemble Sisyphus's eternal toil; they are a digital replica of absolute rationalism. While it holds technical significance, it is more akin to a cultural practice that poeticizes error. than fixating on dispelling hallucinations, it is more worthwhile to interrogate how they shape culture and rewrite forms of knowledge. hallucinations, as the overflow of symbolic systems, leave poetic gaps for reason and provide a space for the extension of human cognition. This symbiotic stance challenges traditional-machine ethics.

As Haraway states in the Cyborg Manifesto, "We are all chimeras, both machine and organism [20]." Embracing the uncertainty of algorithms, allowing hallucinations to redefine the boundaries of reality, may be a deconstruction of technocentrism. Feng Xiang noted "Machine intelligence is the Alpha of communism and the Omega of capitalism [21]." Artificial intelligence, as the embodiment of cultural skills, not only heralds the beginning of a new era but also marks the end of old knowledge.

However, symbiosis does not imply abandonment. A margin for error does not equal immunity. Kittler reminds us: "Temporary machines require temporary ethics." Heidegger warns with a line from Hölderlin: "Where danger is, there grows also what saves." The real crisis is never in the hallucination itself, but in humanity's obsession with absolute reality [22]. Learning to identify new truths within probabilities may indeed be the true wisdom of coexisting with the hallucinations of artificial intelligence.

References

[1] Maleki, N., Padmanabhan, B. and Dutta, K. (2024) AI Hallucinations: A Misnomer Worth Clarifying. In Proceedings of the 2024 IEEE Conference on Artificial Intelligence (CAI). Singapore: IEEE, 133–138.
[2] Siegert, B. (2023) Kulturtechniken – Rastern, Filtern, Zählen und andere Artikulationen des Realen. Baden: Rombach Wissenschaft Verlag, 27.
[3] Baker, S. and Kanade, T. (2000) Hallucinating Faces. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.
[4] Turing, A. M. (1950) Computing Machinery and Intelligence. Mind, New Series, 59(236).
[5] Foucault, M. (2022) Power of Psychiatry. Trans. Su, F. Shanghai: Shanghai People's Publishing House, 37.
[6] Bucciantini, M. and others (2024) Galileo's Telescope: A Revolution Triggered by an Astronomical Instrument. Trans. Liu, Y. K. Beijing: CITIC Press Group, 45–126.
[7] Novalis (1960) Blütenstaub. In Kluckhohn, P. and Samuel, R. (Eds.) Schriften. Die Werke Friedrich von Hardenbergs, Bd. III. Stuttgart: W. Kohlhammer Verlag, 424.
[8] Freud, S. (2012) Das Unheimliche (The Uncanny). Berlin: Europäischer Literaturverlag, 14.
[9] Wang, J. (2021) "Viewing technology through technology": The German theory of Kulturtechnik and its implications. Journalism Review, (02), 85–94.
[10] Freud, S. (2012) Das Unheimliche (The Uncanny). Berlin: Europäischer Literaturverlag, 33.
[11] Hu, Y. and Zhang, W. (2024) From supercomputers to virtual towns: The media archaeology of "AI agents". Modern Publishing, (06), 18–32.
[12] Novalis (1987) Heinrich von Ofterdingen. Stuttgart: Reclam Verlag, 165.
[13] Wang, J.(2020) Kulturtechnik: The frontiers of German media and cultural studies—A dialogue with media philosopher Geoffrey Winthrop-Young. Global Journalism Review, 42(05), 51–60.
[14] Mähl, H.-J. (1963) "Novalis' Wilhelm-Meister-Studien des Jahres 1797." Neophilologus, 47, 296.

[15] Hu, Y. (2017) In defense of media technological determinism: A new perspective on the history of communication thought. Modern Communication (Journal of Communication University of China), 39(01), 51–56.

[16] Novalis (1987) Heinrich von Ofterdingen. Stuttgart: Reclam Verlag, 161–163.

[17] Feng, X. (2006) The Five Books of Moses. Beijing: Oxford University Press, 1.

[18] Herder, J. G. (n.d.) Abhandlung über den Ursprung der Sprache, Bd. 1. Stuttgart: Reclam Verlag, 149–150.

[19] Singh, M. and others (2023) “CodeFusion: A Pre-trained Diffusion Model for Code Generation.” In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023, 11700.

[20] Haraway, D. (2016) “A Cyborg Manifesto.” In A Cyborg Manifesto. University of Minnesota Press, 7.

[21] Feng, X. (2025) I am Omega: On the rise of machine intelligence and the crisis of capitalist succession). Culture Horizons, (1), 94–107.

[22] Heidegger, M. (2005) Lectures and Essays. Trans. Sun, Z. X. Beijing: SDX Joint Publishing Company, 36.