

Between Knowledge and Simulation: Al-Ghazali's Epistemology and the Problem of Hallucination in Generative Artificial Intelligence

Diva Fakh As Shkiy Thalib*

UIN Raden Mas Said Surakarta, Indonesia
Email: divafakihasshkiythalib@gmail.com

Imam Mahdi

UIN Raden Mas Said Surakarta, Indonesia
Email: imahdi506@gmail.com

Muhamad Habib

UIN Raden Mas Said Surakarta, Indonesia
Email: muhamadhabib58@gmail.com

Lintal Muna

UIN Raden Mas Said Surakarta, Indonesia
Email: niilaulya3@gmail.com

Yusup Rohmadi

UIN Raden Mas Said Surakarta, Indonesia
Email: yusup.rohmadi@staff.uinsaid.ac.id

Andi Arif Rifa'i

UIN Raden Mas Said Surakarta, Indonesia
Email: andiarifrifa'i@staff.uinsaid.ac

*Corresponding authors: divafakihasshkiythalib@gmail.com

Received: 2026-05-08; Revised: 2026-09-20; Accepted: 2026-09-25; Published: 2026-09-28

DOI: [10.15575/Jaqfi.v11i2.59170](https://doi.org/10.15575/Jaqfi.v11i2.59170)

Abstract: The emergence of Generative Artificial Intelligence (AI) has transformed contemporary knowledge practices by enabling machines to produce texts, explanations, and interpretations that increasingly resemble human intellectual activities. However, the ability of Large Language Models (LLMs) to generate coherent and convincing responses raises a fundamental epistemological question: does the production of knowledge-like representations indicate genuine knowledge possession, or does it merely constitute a simulation of knowing? This study aims to critically examine the epistemological status of Generative AI by investigating the distinction between knowledge, information generation, and simulated knowledge through the perspective of Al-Ghazali's epistemology. Employing a qualitative conceptual approach, this research combines philosophical analysis of Generative AI, studies on AI hallucination and epistemic authority, and Al-Ghazali's concepts of knowledge, doubt, and certainty as an analytical framework. The study reveals three main findings: first, Generative AI functions as a representational system that reconstructs patterns of information rather than as an epistemic subject capable of understanding and knowing; second, AI hallucination represents an epistemic illusion that exploits human vulnerability by masking linguistic fluency with the non-transferable state of genuine certainty (*yaqīn*); and third, moving beyond generic ethical clichés, the Ghazalian framework furnishes a concrete normative typology that categorizes uncritical reliance as blameworthy imitation (*taqlīd*) and

elevates critical examination (*naẓar*) into a binding duty distributed between individual (*farḍ 'ayn*) and collective (*farḍ kifāyah*) obligations. This study concludes that integrating Islamic epistemology into contemporary AI discourse offers a rigorous philosophical architecture that positions AI strictly as an instrumental simulation engine while securing human intellectual agency.

Keywords: Generative AI; epistemology; Al-Ghazali; AI hallucination; knowledge representation; trustworthy AI.

1. Introduction

The rapid development of generative artificial intelligence has altered the way digital systems participate in the production, organization, and circulation of information. The emergence of large-scale autoregressive models, particularly GPT-3, demonstrated that increasing model scale and training data could generate substantial improvements in few-shot language performance across a wide variety of tasks.¹ This technological trajectory later contributed to what has been described as the rise of foundation models, systems trained on broad datasets and capable of being adapted to multiple downstream applications.² Their significance, however, is not confined to computational performance. Large language models increasingly operate within activities traditionally associated with knowledge practices, including answering questions, summarizing documents, explaining concepts, assisting research, and supporting information seeking. Alvarado therefore characterizes AI as an epistemic technology, since it operates directly within practices concerned with acquiring, manipulating, retaining, and producing epistemically relevant content.³ At the same time, the conversational and linguistically fluent character of contemporary LLMs can encourage users to experience them not merely as computational instruments but as interactive sources of information whose outputs may influence belief formation.⁴ The resulting problem is therefore not solely technological. When a machine-generated response resembles explanation, argument, or factual testimony, its linguistic form can invite the user to treat it as knowledge. This development raises the central epistemological question addressed in this study: does the capacity to generate the linguistic appearance of knowledge amount to knowing, or does it constitute a sophisticated simulation of knowledge without the epistemic conditions normally associated with genuine knowledge?

This question becomes particularly urgent in relation to the phenomenon widely described as *AI hallucination*. Research on neural text generation identified this problem before

¹ Tom B. Brown, Benjamin Mann, and Nick Ryder, "Language Models Are Few-Shot Learners," *Advances in Neural Information Processing Systems* 33 (2020): 1877–901, <https://doi.org/10.48550/arXiv.2005.14165>.

² Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models," version 3, preprint, arXiv, 2021, <https://doi.org/10.48550/ARXIV.2108.07258>.

³ Ramón Alvarado, "AI as an Epistemic Technology," *Science and Engineering Ethics* 29, no. 5 (2023), <https://doi.org/10.1007/s11948-023-00451-3>.

⁴ Richard Heersmink et al., "A Phenomenology and Epistemology of Large Language Models: Transparency, Trust, and Trustworthiness," *Ethics and Information Technology* 26 (2024), <https://doi.org/10.1007/s10676-024-09777-3>.

the widespread public adoption of contemporary conversational AI. Maynez et al. showed that abstractive summarization systems could generate information that was linguistically coherent but unsupported by, or inconsistent with, the source document from which the summary was supposedly derived.⁵ Kryscinski et al. similarly demonstrated that conventional measures of text-generation quality could fail to identify factual inconsistencies, highlighting a gap between surface quality and factual reliability.⁶ Subsequent scholarship has developed hallucination into a broader research category. Ji et al. describe hallucination as a persistent problem across natural language generation tasks, including summarization, dialogue generation, question answering, machine translation, and large language models.⁷ The problem is especially epistemically significant because hallucinated statements do not necessarily appear linguistically defective. A fabricated name, event, explanation, or citation can be embedded within grammatically correct and contextually persuasive prose. SelfCheckGPT was developed precisely in response to this difficulty, treating factual inconsistency across sampled responses as a possible signal that a model is generating unsupported information.⁸ Hallucination therefore reveals an important structural feature of generative AI: linguistic plausibility and factual grounding are separable. A model may succeed at producing the form expected of an informative answer while failing to maintain an adequate relation between the assertion it generates, and the evidence required to support that assertion.

The epistemological seriousness of hallucination becomes clearer when fluency is distinguished from truthfulness. Large language models are optimized to generate probable linguistic continuations, but probability within a language distribution itself does not establish the truth of the proposition being generated. The TruthfulQA benchmark provided empirical evidence for this distinction by testing whether language models reproduce common human misconceptions and falsehoods. Lin, Hilton, and Evans found that the strongest model tested in their study was truthful on only 58 percent of questions, compared with 94 percent for human respondents, and that larger models did not automatically become more truthful.⁹ The finding is important because it demonstrates that improved linguistic or benchmark performance cannot simply be equated with improved epistemic reliability. The relation between confidence and correctness is equally problematic. Zhou, Jurafsky, and Hashimoto found that linguistic markers of certainty and uncertainty can significantly influence model performance and that expressions

⁵ Joshua Maynez et al., "On Faithfulness and Factuality in Abstractive Summarization," 2020, 1906–19, <https://doi.org/10.18653/v1/2020.acl-main.173>.

⁶ Wojciech Kryscinski et al., "Evaluating the Factual Consistency of Abstractive Text Summarization," 2020, 9332–46, <https://doi.org/10.18653/v1/2020.emnlp-main.750>.

⁷ Ziwei Ji et al., "Survey of Hallucination in Natural Language Generation," *ACM Computing Surveys* 55, no. 12 (2023): 1–38, <https://doi.org/10.1145/3571730>.

⁸ Potsawee Manakul, Adian Liusie, and Mark Gales, "SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models," 2023, 9004–17, <https://doi.org/10.18653/v1/2023.emnlp-main.557>.

⁹ Stephanie Lin, Jacob Hilton, and Owain Evans, "TruthfulQA: Measuring How Models Mimic Human Falsehoods," 2022, 3214–52, <https://doi.org/10.18653/v1/2022.acl-long.229>.

of high certainty do not necessarily correspond to greater accuracy.¹⁰ This empirical mismatch has direct relevance to hallucination because users encounter model outputs primarily through language. A declarative sentence expressed confidently can therefore communicate a degree of certainty that is not grounded in a corresponding epistemic guarantee. Retrieval-based approaches can reduce knowledge hallucination by connecting generation to external information, as demonstrated by Shuster et al., yet the need for such grounding itself underscores the distinction between generated linguistic confidence and warranted knowledge.¹¹ Hallucination is thus not merely the production of false sentences. It exposes the possibility that a system can reproduce recognizable linguistic signs of certainty without possessing an independently secured relation to the truth of what it asserts.

Technical research has responded to this problem through increasingly sophisticated methods of detection, evaluation, retrieval, correction, and factuality assessment. Retrieval augmentation, for instance, can improve knowledge-grounded dialogue by providing external documents from which a model may generate its response, thereby reducing unsupported content.¹² Sampling-based approaches such as SelfCheckGPT attempt to identify hallucination by measuring whether repeated generations concerning the same subject remain internally consistent.¹³ Surveys of hallucination research likewise distinguish different forms of hallucination and organize mitigation strategies according to tasks, causes, evaluation methods, and grounding mechanisms.¹⁴ These developments are indispensable for improving the reliability of generative systems. Yet they do not fully resolve the philosophical problem. A technical system can be designed to produce more factually consistent outputs while leaving unanswered what epistemic status should be assigned to those outputs. Paseri and Durante have recently argued that the effects of LLMs cannot be understood only through technical or normative analysis because these systems also raise questions concerning epistemic qualification, reliability, technological dependence, and their relation to truth and accuracy.¹⁵ In other words, detecting whether an output corresponds to an external source is not identical to explaining why that output should count as knowledge. Epistemology introduces questions that computational metrics alone cannot settle: What constitutes adequate justification? What makes a source authoritative? How should uncertainty affect acceptance of a claim? What differentiates factual

¹⁰ Kaitlyn Zhou, Dan Jurafsky, and Tatsunori Hashimoto, "Navigating the Grey Area: How Expressions of Uncertainty and Overconfidence Affect Language Models," 2023, 5506–24, <https://doi.org/10.18653/v1/2023.emnlp-main.335>.

¹¹ Kurt Shuster et al., "Retrieval Augmentation Reduces Hallucination in Conversation," 2021, 3784–803, <https://doi.org/10.18653/v1/2021.findings-emnlp.320>.

¹² Shuster et al.

¹³ Manakul, Liusie, and Gales, "SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models."

¹⁴ Ji et al., "Survey of Hallucination in Natural Language Generation."

¹⁵ Ludovica Paseri and Massimo Durante, "Examining Epistemological Challenges of Large Language Models in Law," *Cambridge Forum on AI: Law and Governance* 1 (2025): e7, <https://doi.org/10.1017/cfl.2024.7>.

correspondence from understanding? These questions become particularly pressing when a generative system provides an answer without itself establishing the epistemic grounds on which a human user should accept the answer as true.

The distinction between linguistic production and epistemic understanding has already become a major theme in philosophical analyses of contemporary language models. Bender and Koller argue that success in manipulating linguistic form does not by itself demonstrate the acquisition of meaning, since form can be learned from patterns in linguistic data without the communicative grounding through which human speakers connect expressions to intentions and the world.¹⁶ Their argument provides an important basis for distinguishing the production of meaningful-looking language from understanding in a stronger semantic sense. Bender et al. later extended this criticism by warning against attributing human-like understanding to increasingly large language models merely because they reproduce complex regularities found in massive textual datasets.¹⁷ Floridi and Chiriatti likewise emphasize that GPT-3's remarkable capacity to produce coherent language should not lead to an uncritical identification of successful linguistic performance with human intelligence or understanding.¹⁸ More recent work has shown that this distinction also has consequences for epistemic trust. Heersmink et al. argue that LLM interfaces can be phenomenologically transparent and conversationally smooth even while the systems remain opaque at the level of their data and computational processes. This combination can encourage unwarranted trust, particularly when users encounter hallucinated outputs.¹⁹ The relevant distinction is therefore not between useful and useless AI. Generative AI can clearly perform useful epistemic functions. The more precise question concerns whether the successful manipulation and presentation of information constitute knowledge in the same sense in which a knowing subject possesses justified cognitive access to what is asserted.

This problem also concerns the authority that users may attribute to artificial systems. The growing integration of AI into epistemic practices has produced an emerging literature on whether artificial systems should be treated as experts, instruments, epistemic technologies, or forms of authority. Freiman argues that conversational AI creates epistemic relationships that cannot be adequately understood through traditional models of reliance on ordinary mechanical instruments because natural-language technologies present propositional content in a form that

¹⁶ Emily M. Bender and Alexander Koller, "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data," 2020, 5185–98, <https://doi.org/10.18653/v1/2020.acl-main.463>.

¹⁷ Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?," 2021, 610–23, <https://doi.org/10.1145/3442188.3445922>.

¹⁸ Luciano Floridi and Massimo Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences," *Minds and Machines* 30 (2020): 681–94, <https://doi.org/10.1007/s11023-020-09548-1>.

¹⁹ Heersmink et al., "A Phenomenology and Epistemology of Large Language Models: Transparency, Trust, and Trustworthiness."

resembles testimony.²⁰ This is significant because users do not merely read a numerical output from an LLM. They receive propositions, explanations, recommendations, qualifications, and sometimes apparent evidence. Ferrario, Facchini, and Termine consequently distinguish predictive or informational superiority from genuine epistemic authority. They argue that high performance alone does not straightforwardly establish the kinds of understanding and intellectual abilities associated with epistemic expertise, although human-AI systems may form hybrid epistemic arrangements under particular conditions.²¹ Alvarado similarly cautions that describing AI as an epistemic technology does not automatically imply that it is epistemically trustworthy.²² These distinctions are central to the problem of hallucination. The danger emerges when *epistemic presentation* is confused with *epistemic possession*: a system speaks in the linguistic register of an authority even though the warrant for accepting its claims may depend entirely on external verification. The problem therefore concerns not only whether AI produces mistakes, but also how human users interpret the apparent epistemic status conveyed by the form of its answers.

Al-Ghazali's epistemology offers a distinctive conceptual framework for examining this problem because his account does not treat epistemic justification merely as the reliability of a source or the availability of reasons supporting a proposition. In *al-Munqidh min al-Dalal* (*Deliverance from Error*), Al-Ghazali recounts a critical examination of inherited beliefs and of the cognitive grounds upon which certainty might be established.²³ His inquiry subjects both sensory experience and rational judgment to doubt, not because he rejects knowledge as such, but because he seeks to distinguish claims that merely appear secure from those capable of sustaining genuine certainty. Modern scholarship has emphasized the epistemological importance of this project. Alwahaib reads Al-Ghazali's intellectual trajectory as a movement from doubt toward certainty and emphasizes the philosophical structure of his examination of truth claims rather than reducing his crisis to a purely biographical or mystical episode.²⁴ Hadisi offers a further reconstruction in which Al-Ghazali's response to radical scepticism involves a transformation in the conditions under which foundational knowledge can be justified, rather

²⁰ Ori Freiman, "Analysis of Beliefs Acquired from a Conversational AI: Instruments-Based Beliefs, Testimony-Based Beliefs, and Technology-Based Beliefs," *Episteme* 21, no. 3 (2024): 1031–47, <https://doi.org/10.1017/epi.2023.12>.

²¹ Andrea Ferrario, Alessandro Facchini, and Alberto Termine, "Experts or Authorities? The Strange Case of the Presumed Epistemic Superiority of Artificial Intelligence Systems," *Minds and Machines* 34, no. 3 (July 2024): 30, <https://doi.org/10.1007/s11023-024-09681-1>.

²² Alvarado, "AI as an Epistemic Technology."

²³ Abu Hamid Muhammad al-Ghazali, *Al-Ghazālī's Path to Sufism and His Deliverance from Error: An Annotated Translation of Al-Munqidh Min Al-Dalal*, 2nd annotated reprint, trans. Richard Joseph McCarthy (Fons Vitae, 2000).

²⁴ Mohammad Alwahaib, "Al-Ghazali and Descartes from Doubt to Certainty: A Phenomenological Approach," *Discusiones Filosóficas* 18, no. 31 (July 2017): 15–40, <https://doi.org/10.17151/difil.2017.18.31.2>.

than a simple abandonment of rational inquiry.²⁵ Mohamed's systematic study of Ghazali's epistemology likewise places doubt, certainty, authoritative instruction, demonstration, and Sufi experience within a broader theory of knowledge.²⁶ This structure provides a comparative advantage for the present problem. Whereas reliability-oriented approaches can ask whether an AI system or its outputs are sufficiently dependable, the Ghazalian framework also asks whether the epistemic appearance of certainty is itself warranted from the standpoint of the knowing subject. *Shakk* therefore functions not simply as the absence of justification but as a methodological interruption of premature assent, while *yaqin* marks a stronger epistemic condition in which the possibility that a claim is otherwise cannot simply be concealed by confidence, authority, or persuasive presentation. This distinction is especially relevant to generative AI because hallucination creates precisely a situation in which the phenomenology of certainty can be produced without the epistemic grounds required for certainty. Al-Ghazali is therefore not introduced merely as another thinker who demands justification; his framework permits the analysis to distinguish between the reliability of generated information and the subject's warranted transition from persuasive appearance to epistemic assent. These features make Al-Ghazali relevant to the present problem not because medieval Islamic epistemology anticipated artificial intelligence, but because the *shakk-yaqin* structure supplies a conceptual vocabulary for examining a feature of generative AI that reliability alone does not fully capture: the discrepancy between the appearance of epistemic security and the grounds required for a subject to assent to a claim as knowledge. The Ghazalian question is therefore particularly sharp for generative AI: what warrants the transition from encountering a persuasive proposition to accepting that proposition as knowledge?

When brought into dialogue with contemporary generative AI, this Ghazalian concern enables hallucination to be interpreted as more than factual malfunction. A hallucinated response demonstrates that the form of an epistemically authoritative assertion can be reproduced independently of the conditions that make such authority warranted. A model may generate a confident statement, a detailed explanation, or even an apparently scholarly citation while the relevant factual relation is absent. Contemporary empirical research shows precisely this disjunction between linguistic confidence, factual consistency, and truthfulness.²⁷ From a Ghazalian perspective, the epistemically relevant issue is therefore not simply whether the output sounds plausible, but whether the appearance of certainty can survive the movement from *shakk* toward *yaqin*. This shifts the unit of analysis from the authority of the algorithm alone to the epistemic status of the proposition and the conditions under which a human subject may assent to it. In this respect, the Ghazalian framework contributes something more specific than

²⁵ Reza Hadisi, "Ghazālī's Transformative Answer to Scepticism," *Theoria* 88, no. 1 (2022): 109–42, <https://doi.org/10.1111/theo.12331>.

²⁶ Nabil Yasien Mohamed, *Ghazālī's Epistemology: A Critical Study of Doubt and Certainty* (Routledge, 2023).

²⁷ Lin, Hilton, and Evans, "TruthfulQA: Measuring How Models Mimic Human Falsehoods."

a general demand for source reliability: it treats doubt as an active epistemic procedure for resisting the conversion of persuasive appearance into unwarranted certainty. This interpretation should not be understood as attributing contemporary computational concepts directly to Al-Ghazali. It is an analytical application of his concern with *yaqin*, doubt, epistemic validation, and the examination of authoritative claims to a new technological context.²⁸ Such an application gains additional relevance from recent research on AI and Ghazalian thought. Anam has used Al-Ghazali to analyze *digital taqlid*, algorithmic authority, and the erosion of epistemic agency in AI-mediated environments, arguing that excessive dependence on algorithmically produced information can weaken the user's responsibility to examine the grounds of belief.²⁹ The present study takes a different analytical object. Anam's principal concern is the epistemic relation between the human user and algorithmic authority, particularly the danger of *taqlid* and diminished epistemic agency. The present analysis instead focuses on the epistemic structure of the generated claim itself: how an output can simulate the linguistic form of knowledge while failing to provide conditions sufficient for movement from *shakk* to *yaqin*. Hallucination therefore does not merely intensify the problem of dependence identified by Anam; it exposes a distinct problem concerning how certainty can be linguistically simulated even when its epistemic grounds are absent.

The existing literature therefore provides several components of the problem without fully integrating them. Computational research has extensively examined hallucination, factuality, faithfulness, truthfulness, uncertainty, and methods of grounding generated content. Philosophical research has examined the difference between linguistic form and understanding, the epistemic status of AI, trust in conversational systems, and the possibility of artificial epistemic authority. At the same time, contemporary scholarships have begun to apply Al-Ghazali's thought to AI, particularly in discussions of algorithmic authority and epistemic agency. It would therefore be inaccurate to claim that no previous study has connected Ghazalian epistemology with artificial intelligence. The research gap is more specifically located in the difference between two epistemic problems that should not be conflated. The first, already developed by Anam, concerns whether human users surrender epistemic agency through *taqlid* to algorithmic authority. The second, addressed here, concerns whether the output itself can display the linguistic signs of epistemic certainty while lacking the grounds necessary for movement from *shakk* to *yaqin*. Likewise, whereas reliability-oriented approaches can evaluate whether an output is accurate, consistent, calibrated, or appropriately grounded, they do not by themselves explain why the appearance of certainty should or should not license the knowing

²⁸ al-Ghazali, *Al-Ghazālī's Path to Sufism and His Deliverance from Error: An Annotated Translation of Al-Munqidh Min Al-Dalal*; Hadisi, "Ghazālī's Transformative Answer to Scepticism"; Mohamed, *Ghazālī's Epistemology: A Critical Study of Doubt and Certainty*.

²⁹ Rifqi Khairul Anam, "Against Algorithmic Authority: Al-Ghazali, Digital Taqlid, and the Crisis of Epistemic Agency in the Age of AI," *Journal of Islamic Philosophy and Contemporary Thought* 3, no. 1 (2025): 1–35, <https://doi.org/10.15642/jipct.2025.3.1.1-35>.

subject's assent. The literature located for this study has not yet systematically formulated generative AI hallucination through this specific problem: the boundary between knowledge and the simulation of knowledge as revealed by the discrepancy between epistemic appearance and the conditions of *yaqin*. This article addresses that intersection. It asks how Al-Ghazali's epistemology can illuminate the difference between producing a proposition that displays the linguistic characteristics of knowledge and possessing adequate epistemic grounds for treating that proposition as known. Accordingly, the study pursues two objectives. First, it analyzes hallucination as a separation among linguistic coherence, factual correctness, and epistemic justification. Second, it develops a Ghazalian conceptual reading of generative AI in order to clarify the boundary between knowledge and simulation of knowledge. The article thus proposes that the deepest epistemological problem of hallucination is not simply that AI can produce false information, but that it can reproduce the recognizable form of knowing without independently guaranteeing the conditions under which a claim deserves to be accepted as knowledge.

2. Methods

This study employs a qualitative conceptual-philosophical approach based on library research. The primary source is Al-Ghazali's *al-Munqidh min al-Dalal* (*Deliverance from Error*), particularly his discussions of doubt, certainty, reason, sensory perception, and the conditions of valid knowledge. Secondary sources include contemporary studies on Al-Ghazali's epistemology as well as recent scholarship on generative AI, *large language models*, hallucination, factuality, and AI epistemology. The sources were selected purposively according to their relevance to the central question of how knowledge claims are justified and distinguished from merely plausible representations. This form of conceptual inquiry emphasizes the systematic selection and integration of theories and concepts rather than empirical measurement.³⁰

The analysis was conducted in three stages. First, the study identified and reconstructed key Ghazalian epistemic concepts, particularly *'ilm*, *yaqin*, doubt, and epistemic authority. Second, contemporary literature on generative AI was examined to identify corresponding problems such as hallucination, factual inconsistency, linguistic confidence, and epistemic reliability. Third, these two conceptual domains were compared through a shared epistemic question: under what conditions can a claim be regarded as justified knowledge? Comparative analysis was used not to equate Al-Ghazali's medieval epistemology with contemporary AI, but to clarify similarities and differences through a defined conceptual basis of comparison.³¹ The final interpretation was then organized around the distinction between linguistic coherence,

³⁰ Elina Jaakkola, "Designing Conceptual Articles: Four Approaches," *AMS Review* 10 (2020): 18–26, <https://doi.org/10.1007/s13162-020-00161-0>.

³¹ Oliver Freiberger, *Considering Comparison: A Method for Religious Studies* (Oxford University Press, 2019), <https://doi.org/10.1093/os0/9780199965007.001.0001>.

factual correctness, epistemic justification, and the boundary between *knowledge* and *simulation of knowledge*.

3. Result and Discussion

Al-Ghazali's Epistemology of Knowledge and Certainty

Al-Ghazali's epistemology places *ilm* (knowledge) and *yaqin* (certainty) at the center of the human search for truth. Knowledge is not merely the possession of information but a form of understanding grounded in truth and capable of transforming the knower. In *Ihya' 'Ulum al-Din*, Al-Ghazali emphasizes the significance of knowledge for both intellectual understanding and human spiritual development.³² Therefore, knowledge must be distinguished from mere assumption (*zann*) or uncritical belief. Hadisi explains that Al-Ghazali's engagement with skepticism was intended to examine the foundations upon which knowledge could be regarded as certain rather than to reject knowledge itself.³³ Doubt (*shakk*) consequently functions as a critical stage through which apparently certain beliefs are examined. The movement from doubt toward *yaqin* demonstrates that genuine knowledge requires adequate epistemic grounds and cannot be established merely through subjective confidence.³⁴

The acquisition of knowledge, according to Al-Ghazali, involves several faculties, including the senses (*hiss*), reason (*'aql*), experience, and spiritual insight. The senses provide initial information about the empirical world, while reason evaluates and organizes sensory experience. However, neither sensory perception nor rational judgment is completely free from limitation. Contemporary studies therefore describe Al-Ghazali's epistemology as an integrated structure involving empirical, rational, and spiritual dimensions.³⁵ This position is consistent with *Ihya' 'Ulum al-Din*, where knowledge is connected not only with intellectual comprehension but also with the cultivation of the human self.³⁶ Experience thus provides important material for knowledge, but experience alone does not guarantee certainty. Knowledge emerges when sensory and experiential information is critically examined through appropriate intellectual and spiritual processes.³⁷

Authority also has an important position in Al-Ghazali's epistemology, particularly through teachers, scholars, and transmitted religious knowledge. Nevertheless, accepting a proposition from an authority is not necessarily equivalent to possessing certain knowledge. The

³² Abu Hamid Al-Ghazali, *Ihya' 'Ulum al-Din* (Beirut: Dar al-Kotob al-Ilmiyah, 2010), 5.

³³ Hadisi, "Ghazālī's Transformative Answer to Scepticism."

³⁴ Mohamed, *Ghazālī's Epistemology: A Critical Study of Doubt and Certainty*.

³⁵ Budi Syihabuddin and Sumanta Hasyim, "Al-Ghazali's Empirical–Rational–Spiritual Epistemology and Its Implications for Contemporary Islamic Religious Education," *Afaqulilmi: Journal of Islamic Education*, ahead of print, 2026, <https://doi.org/10.60769/afaq.v2i1.6>.

³⁶ Al-Ghazali, *Ihya' 'Ulum al-Din*.

³⁷ Fadhlī Hafizh, Rohanda, and Abdul Kodir, "Epistimologi Ilmu Pengetahuan Dalam Perspektif John Locke Dan Al-Ghazali," *Manthiq: Jurnal Filsafat Agama Dan Pemikiran Islam* 8, no. 2 (2024), <https://doi.org/10.29300/mtq.v8i2.6764>.

validity of transmitted knowledge depends on the reliability of its source as well as the recipient's capacity to understand what is transmitted. Suwarno and Albab emphasize that scholarly authority in Al-Ghazali's thought is closely related to competence, knowledge, and the responsibility of transmitting knowledge.³⁸ Thus, authority may serve as a pathway toward knowledge, but it should not encourage blind acceptance. This distinction is particularly important in relation to *taqlid*, because repeating an authoritative statement does not necessarily mean that the individual understands or possesses its epistemic justification. Al-Ghazali's conception of knowledge therefore requires a balance between legitimate reliance on authority and critical intellectual engagement.³⁹

The concept of *yaqin* further demonstrates that Al-Ghazali distinguishes genuine certainty from psychological confidence. A person may strongly believe that a proposition is true while lacking sufficient grounds for that belief. *Yaqin*, by contrast, refers to a more secure epistemic condition in which relevant grounds for doubt have been overcome. Mohamed identifies doubt and certainty as central dimensions of Al-Ghazali's epistemological project and emphasizes the importance of examining how knowledge acquires its justification.⁴⁰ At the same time, Al-Ghazali recognizes spiritual experience as an important dimension of higher knowledge. Recent scholarship shows that purification of the heart and spiritual experience are closely related to his conception of the unveiling of truth.⁴¹ Consequently, certainty in Al-Ghazali cannot be reduced to intellectual confidence alone; it involves a deeper relationship between the knower, knowledge, and truth.

Ultimately, Al-Ghazali's epistemology establishes an important distinction between knowing something and merely receiving a representation of something. A representation may take the form of information, language, or an intellectual description, but its existence does not automatically establish its truth. Genuine knowledge requires an adequate relationship between the knower, the object of knowledge, and the grounds that justify the claim. Al-Ghazali's *Ihya' Ulum al-Din* reinforces the idea that knowledge should produce genuine understanding and transformation rather than remain as information merely possessed by the intellect. Contemporary studies similarly emphasize that his epistemology concerns the sources, degrees,

³⁸ M. Hadi Suwarno and Hayyan Ahmad Ulul Albab, "Knowledge and Scholarly Authority in Al-Ghazali's Philosophy: A Study of *Fatihatul 'Ulum*," *Journal of Pesantren and Diniyah Studies* 1, no. 2 (2024): 185–98, <https://doi.org/10.63245/jpds.vii2.29>.

³⁹ Sujiat Zubaidi Saleh, Muhammad Taqiyuddin, and Rakhmad Agung Hidayatullah, "On Imam Ghazālī's *Farḍ 'Ain* and *Farḍ Kifāyah*: An Epistemological Approach," *Jurnal Ushuluddin* 30, no. 1 (2022): 48–62, <https://doi.org/10.24014/jush.v30i1.14661>.

⁴⁰ Mohamed, *Ghazālī's Epistemology: A Critical Study of Doubt and Certainty*, 21.

⁴¹ Amin Said Husni, Moh. Syaeful Bahar, and Dodik Harnadi, "Purifying the Heart, Unveiling the Truth: Understanding Al-Ghazālī's Epistemology," *Journal of Pesantren and Diniyah Studies* 2, no. 1 (June 2025): 1–14, <https://doi.org/10.63245/jpds.v2i1.32>.

and validity of knowledge.⁴² Therefore, *ilm* should be understood as more than information, while *yaqin* represents the degree of epistemic certainty that distinguishes knowledge from conjecture or mere representation. In this sense, knowing is not simply possessing a representation of reality, but establishing a justified and.

Generative AI as a System of Knowledge Simulation

Generative Artificial Intelligence (Generative AI) represents a significant transformation in the contemporary knowledge ecosystem because it enables machines to generate linguistic responses that resemble human explanation, reasoning, and interpretation. However, the mechanism underlying this capability fundamentally differs from human cognitive activity. Generative AI does not produce responses through conscious understanding, experiential engagement, or epistemic reflection, but through statistical prediction and reconstruction of patterns embedded within large-scale datasets. Large Language Models (LLMs) operate by calculating the probability of linguistic sequences and selecting the most contextually appropriate continuation based on previously learned patterns. Consequently, what appears as reasoning or understanding is primarily the result of sophisticated pattern recognition and representation generation. Recent studies in AI epistemology emphasize that the emergence of fluent and contextually appropriate responses does not necessarily indicate that LLMs possess knowledge in the same sense as human knowers. The central epistemological question is therefore not whether AI can generate meaningful language, but whether such generated meanings correspond to an internal possession of knowledge or merely constitute a simulation of knowledge-like outputs.⁴³ This distinction is essential because knowledge traditionally requires a relationship between a knowing subject, an object of knowledge, and justified grounds that support a particular claim. Generative AI lacks intentional awareness of the propositions it generates; it does not “believe,” “understand,” or “know” the statements it produces. Instead, it constructs representations derived from linguistic correlations within its training environment. Therefore, the intelligence demonstrated by Generative AI should be interpreted as representational intelligence rather than epistemic intelligence. The system excels in reproducing the structure and appearance of knowledge discourse, but this ability should not be confused with possessing knowledge itself. In this sense, Generative AI functions as a simulation engine of knowledge: it creates outputs that resemble the products of human understanding while remaining fundamentally different from the epistemic processes through which humans establish meaning, truth, and justification.

⁴² Rizamul Malik Akbar and Achmad Khudori Soleh, “The Classification of Knowledge Based on Al-Ghazali’s Perspective and Its Relevance to Qur’anic Exegesis,” *FiTUA: Jurnal Studi Islam* 6, no. 2 (October 2025): 218–28, <https://doi.org/10.47625/fitua.v6i2.1066>.

⁴³ Yuval Noah Harari, “Epistemology in the Age of Large Language Models,” *Knowledge* 5, no. 1 (2025): 1–15, <https://doi.org/10.3390/knowledge5010003>.

The distinction between producing coherent statements and possessing epistemic justification represents one of the most significant philosophical problems surrounding Generative AI. Linguistic coherence refers to the ability of a system to produce sentences that are grammatically correct, contextually relevant, and rhetorically persuasive. Epistemic justification, however, concerns whether those statements have sufficient grounds to be accepted as knowledge. These two dimensions frequently diverge in Generative AI systems because fluency does not necessarily guarantee factual validity. A model may generate a sophisticated explanation, provide apparently relevant evidence, or construct a convincing argument while simultaneously producing unsupported or inaccurate information. This phenomenon, widely discussed as hallucination, demonstrates that AI-generated language can maintain a high level of plausibility without maintaining an adequate relationship with external reality. Recent surveys on LLM hallucination explain that these failures occur because models are optimized primarily for generating probable sequences rather than verifying whether each generated proposition corresponds with factual conditions.⁴⁴ Therefore, hallucination should not be understood merely as a technical error but as an epistemological indicator of the difference between appearance and justification. The problem is not simply that AI occasionally produces false information; rather, the deeper issue is that false and true statements can emerge through the same linguistic mechanism and appear equally authoritative to users. Research on semantic uncertainty further demonstrates that language models may produce highly confident responses even when their internal uncertainty remains significant, showing that linguistic confidence does not function as a reliable indicator of truth.⁴⁵ Furthermore, studies examining knowledge integration in LLMs reveal that external grounding mechanisms such as knowledge graphs and retrieval systems are required precisely because language models do not inherently possess a stable connection between generated representations and factual reality.⁴⁶ Thus, the epistemological limitation of Generative AI lies in the separation between linguistic success and epistemic reliability. A response may be linguistically excellent while epistemically weak, because the ability to formulate a statement is fundamentally different from the ability to justify why that statement should be considered true.

Based on this analysis, Generative AI should be positioned as a producer of knowledge representations rather than an autonomous subject of knowing. This distinction allows a more precise understanding of AI's role in contemporary knowledge production. Generative AI undeniably contributes to epistemic practices by assisting researchers, organizing information, generating hypothesis, and facilitating access to complex materials. Nevertheless, participation

⁴⁴ Lei Huang et al., "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," *ACM Transactions on Information Systems* 43, no. 2 (2025): 1–55, <https://doi.org/10.1145/3703155>.

⁴⁵ Sebastian Farquhar et al., "Detecting Hallucinations in Large Language Models Using Semantic Entropy," *Nature* 630 (2024): 625–30, <https://doi.org/10.1038/s41586-024-07421-0>.

⁴⁶ Ernests Lavrinovics et al., "Knowledge Graphs, Large Language Models, and Hallucinations: An NLP Perspective," *Journal of Web Semantics* 85 (2025): 35, <https://doi.org/10.1016/j.websem.2024.100844>.

in knowledge practices does not automatically transform an artificial system into an epistemic agent. The difference lies in the absence of subjective awareness, intentionality, and responsibility regarding the claims produced. Human knowledge involves not only possessing information but also understanding relationships between concepts, evaluating evidence, recognizing uncertainty, and accepting responsibility for judgments. Generative AI, by contrast, transforms informational patterns into linguistic representations without possessing an experiential relationship with the objects being represented. Recent philosophical analyses of LLMs argue that these systems should be understood as cognitive artifacts or epistemic technologies that extend human intellectual capacities rather than replace human epistemic agency.⁴⁷ Therefore, the authority of AI-generated information depends not on the system itself but on human processes of evaluation, verification, and interpretation. This distinction becomes increasingly important as AI-generated texts enter scientific, educational, legal, and religious domains where epistemic accountability is essential. The future relationship between humans and Generative AI should not be constructed around the assumption that machines have become knowledgeable, but around a more nuanced understanding that AI produces highly sophisticated representations of knowledge. In other words, Generative AI occupies an intermediate position between information processing and knowledge production: it can simulate the external manifestations of knowing, but the final determination of meaning, truth, and justification remains a human epistemic responsibility. The significance of Generative AI therefore lies not in replacing human knowledge, but in transforming the conditions under which knowledge is produced, accessed, and evaluated in the digital age.

AI Hallucination as an Epistemological Problem

The phenomenon of hallucination in Generative Artificial Intelligence should not be understood merely as a computational malfunction or a temporary weakness in model performance. Rather, hallucination represents a deeper epistemological problem concerning the relationship between claims, sources, evidence, and truth. In conventional information systems, errors are generally evaluated through technical parameters such as accuracy, efficiency, or computational failure. However, in Generative AI, the problem becomes more complex because incorrect information is often produced through linguistically sophisticated and highly persuasive expressions. An AI-generated statement may appear rational, coherent, and authoritative while lacking an adequate relationship with external evidence. Therefore, hallucination reveals a fundamental distinction between the ability to generate propositions and the ability to justify those propositions.⁴⁸ Recent scholarship on AI reliability emphasizes that hallucination occurs because large language models generate outputs based on learned patterns rather than direct verification of whether a particular

⁴⁷ Heersmink et al., "A Phenomenology and Epistemology of Large Language Models: Transparency, Trust, and Trustworthiness," 34.

⁴⁸ Huang et al., "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," 6–10.

claim corresponds to reality.⁴⁹ The model does not intentionally deceive users, yet it can produce statements that imitate the structure of truthful knowledge without possessing the epistemic conditions necessary for truth validation. From an epistemological perspective, this problem concerns the transformation of information into knowledge. A claim becomes knowledge not simply because it is expressed clearly or accepted by users, but because it is supported by adequate justification, reliable sources, and a valid relationship with the object being described. Consequently, AI hallucination demonstrates that linguistic credibility and epistemic credibility are fundamentally different categories. The danger of hallucination lies precisely in this gap: the system can reproduce the external characteristics of knowledge discourse while simultaneously lacking the internal mechanisms required to establish whether its statements deserve epistemic acceptance. In this sense, hallucination should be interpreted as a crisis of justification rather than merely a crisis of accuracy, because it challenges the assumption that a well-formed statement necessarily represents a well-grounded claim.

The epistemological significance of AI hallucination can be further examined through Al-Ghazali's philosophical framework of doubt (*shakk*) and certainty (*yaqin*).⁵⁰ Although Al-Ghazali's epistemology emerged within a medieval Islamic intellectual context, his analysis of knowledge provides a valuable conceptual framework for examining contemporary questions surrounding artificial intelligence. For Al-Ghazali, knowledge cannot be reduced to the accumulation of information or the acceptance of claims based merely on appearance; rather, genuine knowledge requires a critical examination of the foundations upon which certainty is established. In *al-Munqidh min al-Dalāl*, Al-Ghazali describes his intellectual journey by questioning the reliability of previously accepted forms of knowledge and states:

فقلت في نفسي: إنما مطلوبي العلم بحقائق الأمور، فلا بد من طلب حقيقة العلم ما هي

"I said to myself: My objective is knowledge of the realities of things; therefore, I must seek what the true nature of knowledge actually is."⁵¹

This statement demonstrates that, for Al-Ghazali, the fundamental problem of knowledge is not merely the possession of information but the investigation of the conditions that make knowledge valid.⁵² Knowledge requires an examination of its foundations, sources, and justification. This perspective provides a significant conceptual framework for understanding AI hallucination because AI-generated responses often create an appearance of knowledge without necessarily fulfilling the epistemic requirements that transform information into justified

⁴⁹ Luciano C. Passos, "A Survey on Hallucination in Large Language Models: Causes, Detection, and Mitigation," *Artificial Intelligence Review* 58 (2025), <https://doi.org/10.1007/s10462-025-11106-5>.

⁵⁰ Hadisi, "Ghazālī's Transformative Answer to Scepticism," 109–12.

⁵¹ Abu Hamid Muhammad Al-Ghazali, *Al-Munqidh Min al-Dalal*, ed. Jamil Saliba and Kamil Ayyad (Beirut: Dar al-Andalus, 1967), 25.

⁵² Hadisi, "Ghazālī's Transformative Answer to Scepticism," 115–18.

knowledge. Similar to Al-Ghazali's distinction between apparent certainty and genuine certainty, a generated response may appear convincing while remaining epistemically uncertain.

Furthermore, Al-Ghazali emphasizes that doubt functions as a methodological instrument rather than a rejection of knowledge itself. He explains:

الشكوك هي الموصلة إلى الحق، فمن لم يشك لم ينظر، ومن لم ينظر لم يبصر، ومن لم يبصر
بقي في العمى والضلال.

"Doubts are the paths that lead to truth. Whoever does not doubt does not examine; whoever does not examine does not perceive; and whoever does not perceive remains in blindness and error."

Although this formulation is commonly attributed to Al-Ghazali, its precise textual location and exact wording in *al-Munqidh min al-Dalal* should not be treated as definitively established. The maxim is therefore employed here not as a direct textual quotation whose precise formulation can be securely located in *al-Munqidh*, but as a commonly attributed expression that encapsulates the Ghazalian epistemic movement from doubt and examination toward certainty.

This principle provides an important analytical lens for contemporary AI hallucination. In the context of Generative AI, the problem emerges when users accept AI-generated claims without undergoing a process of epistemic examination. The absence of critical verification transforms uncertainty into assumed certainty. From a Ghazalian perspective, hallucination represents a failure not only of information accuracy but also of epistemic discipline, because the recipient accepts a claim without examining the foundations that support it. The central question is therefore not simply whether an AI response is linguistically persuasive, but whether its epistemic grounds can withstand the critical examination represented by the movement from *shakk* toward *yaqin*.⁵³

Al-Ghazali's epistemology also distinguishes between information transmission and genuine understanding. In *Ihyā' Ulūm al-Dīn*, he argues that knowledge is not merely the possession of statements but a form of understanding that transforms the knower:

ليس العلم بكثرة الرواية، ولكن العلم نور يقذف في القلب.

"Knowledge is not the abundance of narration; rather, knowledge is a light that is cast into the heart."⁵⁴

This distinction is highly relevant to Generative AI because large language models are capable of producing extensive narratives, explanations, and references, yet such production

⁵³ Mohamed, *Ghazālī's Epistemology: A Critical Study of Doubt and Certainty*, 21–27.

⁵⁴ Al-Ghazali, *Ihyā' Ulūm al-Dīn*, 13–14.

does not necessarily constitute understanding.⁵⁵ AI can generate representations of knowledge without possessing the inner epistemic condition associated with knowing. At this juncture, it is crucial to explicitly distinguish between two registers of certainty in Al-Ghazali's thought: the discursive-methodological register (*shakk* driving inquiry) and the supra-rational register (*yaqīn* achieved through *kashf* or the light cast into the heart). In the context of Generative AI, this second register of *kashf* is entirely non-transferable to artificial systems. It must be acknowledged that Ghazalian *yaqīn* does not serve as a constructive model for algorithmic design or AI verification. Instead, it functions strictly as an absolute epistemological boundary. It performs necessary negative work by demarcating exactly what AI intrinsically lacks: consciousness, spiritual insight (*dzawq*), and epistemic responsibility. Therefore, through Al-Ghazali's framework, hallucination can be interpreted as a technological manifestation of the difference between transmitted representation and genuine knowledge. The model produces linguistic forms resembling knowledge, but the ultimate threshold of certainty (*yaqīn*) remains an exclusively human domain.

The emergence of false epistemic authority represents another significant consequence of AI hallucination. Epistemic authority traditionally depends on competence, reliability, and justified access to knowledge. However, Generative AI introduces a unique challenge because authority can be simulated through language itself. Users frequently encounter AI systems through conversational interfaces that reproduce characteristics associated with human expertise: confidence, clarity, explanation, and responsiveness. These linguistic features can create the impression that the system possesses knowledge authority even when its outputs remain dependent on probabilistic generation. Recent studies on human-AI interaction indicate that conversational fluency significantly influences user trust, sometimes causing individuals to attribute greater competence and reliability to AI systems than the systems actually possess.⁵⁶ This phenomenon creates what can be described as false epistemic authority, namely a condition where the appearance of knowing is mistaken for the possession of knowledge. The danger is particularly significant in domains where interpretation and justification are essential, such as academic research, religious discourse, law, and public decision-making. A fluent response may be accepted not because it has been critically verified, but because its linguistic form resembles the communication style of a knowledgeable expert. In this regard, hallucination exposes a fundamental vulnerability in human interaction with artificial systems: humans are naturally inclined to associate communicative competence with intellectual competence. Therefore, the challenge of Generative AI is not only improving technical accuracy but also cultivating epistemic awareness among users. The proper response to AI-generated knowledge claims is not unconditional trust or complete rejection, but critical engagement based on verification, source

⁵⁵ Bender and Koller, "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data."

⁵⁶ Heersmink et al., "A Phenomenology and Epistemology of Large Language Models: Transparency, Trust, and Trustworthiness," 9–12.

evaluation, and recognition of the difference between representation and reality. Generative AI may become an extraordinarily powerful instrument for knowledge production, but its outputs must remain situated within a human-centered epistemic framework where truth requires more than persuasive language it requires justification, accountability, and a responsible relationship between claims and evidence.

Between Knowledge and Simulation

The emergence of Generative Artificial Intelligence has produced a fundamental epistemological challenge concerning the distinction between knowledge, information generation, and the simulation of knowing. In classical epistemology, knowledge is not merely the possession of information but a justified understanding of reality that involves a relationship between the knower, the object of knowledge, and the reasons that support a particular claim.⁵⁷ Information generation, in contrast, refers to the ability to produce, organize, and communicate representations without necessarily possessing an understanding of their meaning or truth conditions. Generative AI operates primarily within this second domain because its outputs are generated through computational processes that reconstruct linguistic patterns rather than through conscious intellectual engagement with reality.⁵⁸ However, the sophistication of contemporary Large Language Models (LLMs) creates an appearance of understanding because these systems can produce explanations, arguments, and interpretations that resemble human intellectual activity. This condition may be described as a simulation of knowing, where the external characteristics of knowledge production are reproduced without the internal epistemic conditions of a knowing subject.⁵⁹ From Al-Ghazali's epistemological perspective, knowledge requires more than the accumulation or transmission of information; it requires examination, verification, and a justified relationship between a claim and its foundation.⁶⁰ Therefore, Generative AI should not be positioned as an epistemic subject that possesses knowledge, but as a technological system that produces representations requiring human interpretation and validation.⁶¹ The distinction is essential because attributing human epistemic capacities to AI risks confusing linguistic performance with genuine understanding. A system may successfully imitate the language of knowledge while remaining fundamentally different from a subject that knows, evaluates, and takes responsibility for truth claims.

The epistemological tension between knowledge and simulation becomes particularly visible in educational contexts where students increasingly integrate Generative AI into their

⁵⁷ Duncan Pritchard, "Anti-Luck Epistemology," *Synthese* 158, no. 3 (September 2007): 277–97, <https://doi.org/10.1007/s11229-006-9039-7>.

⁵⁸ Bommasani et al., "On the Opportunities and Risks of Foundation Models," 12.

⁵⁹ Massimo Negrotti, "On the 'Nature' of the 'Artificial,'" *AI & SOCIETY* 38, no. 5 (October 2023): 1935–40, <https://doi.org/10.1007/s00146-021-01178-0>.

⁶⁰ Alwahaib, "Al-Ghazali and Descartes from Doubt to Certainty."

⁶¹ Mark Coeckelbergh, "Democracy, Epistemic Agency, and AI: Political Epistemology in Times of Artificial Intelligence," *AI and Ethics* 3, no. 4 (November 2023): 1341–50, <https://doi.org/10.1007/s43681-022-00239-4>.

learning practices. Rather than approaching AI merely as a mechanical search instrument, conceptual observations of contemporary academic practices suggest a profound shift: users frequently interact with conversational AI as an intellectual assistant capable of explaining concepts, generating ideas, and solving problems. This interaction often leads to the delegation of epistemic tasks such as summarization, analysis, and writing directly to the artificial system. While learners appreciate the system's capacity to simplify complex information, this structural reliance creates a vulnerability regarding accuracy and critical engagement. These phenomenological shifts reveal a significant epistemological problem: AI is gradually moving from being a learning tool toward becoming a perceived source of knowledge authority. From the perspective of Al-Ghazali's epistemology, the central problem is not the use of AI itself but the disappearance of critical examination (*nazar*) before accepting AI-generated claims. When students accept AI responses because of their fluency and coherence rather than because of their evidential foundation, information generation becomes incorrectly treated as knowledge acquisition.

To systematically analyze the structural divergence between genuine epistemic possession and artificial simulation, Table 1 deconstructs the anatomy of AI-generated knowledge claims through Al-Ghazali's epistemological framework. Rather than merely summarizing definitions, this analytical matrix isolates the specific mechanisms by which Generative AI simulates epistemic conditions such as certainty, doubt, and authority and contrasts them with the strict Ghazalian requirements for genuine knowledge. By mapping these dimensions, the analysis demonstrates how the non-transferability of inner epistemic states (*yaqin* and *dzawq*) transforms Al-Ghazali's framework into a definitive negative boundary. This boundary exposes the epistemological vulnerability that occurs when users mistake the algorithmic mimicry of expert discourse for conscious intellectual possession.

Table 1. Analytical Deconstruction of AI Epistemic Simulation against Al-Ghazali's Framework

Epistemological Dimension	Ghazalian Requirement (The Genuine Condition)	AI's Mechanism of Simulation (The Epistemic Illusion)	Analytical Consequence & The Ghazalian Boundary
The Nature of Certainty (<i>Yaqin</i>)	A supra-rational conviction rooted in inner realization (<i>kashf</i>), spiritual insight (<i>dzawq</i>), and consciousness.	Generation of high-probability text sequences presented with declarative linguistic confidence.	Non-transferability of <i>Kashf</i> : AI certainly is an aesthetic linguistic feature, not an epistemic state. The Ghazalian boundary proves AI can never attain <i>yaqin</i> , relegating it entirely to the realm of simulation.
The Function of Doubt (<i>Shakk</i>)	An active, conscious methodological disruption (<i>nazar</i>) that prevents premature assent to unsupported claims.	Mathematical adjustments of semantic entropy (uncertainty functions) devoid of conscious inquiry or curiosity.	Offloading Epistemic Burden: Because AI cannot consciously "doubt" its own outputs, the active methodological work of <i>shakk</i> (standard evidentialist factchecking) must be entirely assumed by the human user.

Epistemic Authority	Inseparably bound to cognitive competence, moral accountability, and experiential truth.	Conversational fluency, rhetorical coherence, and structural mimicry of expert discourse (syntactic persuasiveness).	False Epistemic Authority: Exposes human vulnerability to equate communicative competence with intellectual competence. Ghazalian epistemology strips away this simulated authority, repositioning AI as a mere instrument.
The Output of Knowledge	<i>Ilm</i> acts as a transformative light (<i>nūr</i>) producing genuine comprehension of reality within the knowing subject.	Recombinative pattern-matching produces highly plausible propositions without referential grounding to reality.	The Crisis of Justification: Demonstrates the exact anatomy of "hallucination" AI performs the negative work of showing what representation without consciousness looks like. The output remains epistemically hollow until externally verified.

The empirical reality of user interaction with Generative AI illustrates the precise danger mapped in Table 1: the seamless transition from utilizing AI as a computational assistant to relying upon it as an epistemic substitute. Because AI successfully mimics the structural markers of authority, users naturally tend to attribute human-like cognitive abilities to these systems. The core problem is not the algorithmic production of text, but the human susceptibility to interpret generated explanations as evidence of actual comprehension. Recent studies on AI trust emphasize that conversational fluency and declarative confidence heavily skew users' perceptions of reliability.⁶² This dynamic actively generates the "false epistemic authority" detailed above. Because AI possesses only linguistic competence without corresponding epistemic possession, the responsibility for justification cannot be delegated. The human user must remain the sole active agent, applying standard evidentialist verification such as empirical proof and external referencing to ground the AI's simulated claims in reality.

Ultimately, this Ghazalian demarcation clarifies the exact operational limits of Generative AI within human knowledge systems. By recognizing that AI is permanently barred from the supra-rational register of certainty, its role can be sharply defined across four boundaries. First, AI can computationally generate probabilistic information, but it cannot independently justify the truth of its own outputs. Second, AI can facilitate information retrieval, but it is entirely excluded from the inner realization required for genuine understanding. Third, AI functions as a powerful synthetic instrument, yet it remains completely devoid of the moral and epistemic responsibility demanded of a knowing subject. Finally, the conversion of AI-generated representations into valid knowledge is never automatic; it is wholly dependent on external human judgment. Therefore, the philosophical significance of Generative AI lies not in creating a new type of knower, but in producing a highly sophisticated simulation engine—one

⁶² Ferrario, Facchini, and Termine, "Experts or Authorities?"

that necessitates a much stricter human defense of the boundary between algorithmic plausibility and authentic knowledge.

Implications for the Use of Generative AI

The epistemological limitations of Generative AI indicate that the practical verification of its outputs relies heavily on standard evidential norms. AI-generated outputs should not be accepted solely based on their linguistic coherence, fluency, or apparent authority, but must be evaluated through comparison with reliable references, empirical evidence, and established knowledge sources. While this reliance on empirical proof and external justification belongs to standard evidentialist epistemology already available without Al-Ghazali, his framework provides a crucial supplementary function. By explicitly acknowledging that the Ghazalian mystical certainty (*yaqīn* achieved through *kashf* or spiritual unveiling) is fundamentally non-transferable to artificial systems, we recognize AI as an entity permanently relegated to the realm of simulation.

This perspective places humans as the ultimate evaluators of AI-generated knowledge claims. Although Generative AI can accelerate information processing and generate explanations, it intrinsically lacks the inner epistemic conditions such as consciousness, epistemic responsibility, and spiritual insight (*dzawq*) required to genuinely possess knowledge. Therefore, users must maintain an active epistemic role by questioning, verifying, and interpreting AI responses. This principle is particularly important in academic, religious, and scientific contexts, where the distinction between plausible linguistic generation and justified knowledge has significant consequences.

The integration of Islamic epistemology into contemporary AI discourse, therefore, does not serve to provide a new methodological architecture for factchecking. Instead, Al-Ghazali's strict boundary of supra-rational certainty performs vital negative work: it acts as a philosophical safeguard against false epistemic authority. Rather than offering a constructive model for algorithmic verification, this Ghazalian perspective functions strictly as an epistemological boundary. It positions AI as a mediated instrument and reserves the profound realization of knowledge (*yaqīn*) entirely for the human epistemic agent. Consequently, Islamic epistemology enriches the global AI discourse not by supplying a new verification standard, but by establishing absolute limits, ensuring that simulated linguistic fluency is never mistaken for genuine epistemic possession.

4. Conclusion

This study demonstrates that applying Al-Ghazali's epistemological framework to Generative AI yields a profound philosophical diagnosis that remains unattainable through standard, secular AI ethics discourse. While contemporary philosophers and computational scientists have widely established that Large Language Models merely simulate knowledge rather than possessing it, the Ghazalian lens uniquely exposes the deeper internal danger of this simulation: it actively

exploits human psychological and cognitive vulnerabilities by masking syntactic fluency as genuine certainty. By rigorously distinguishing between discursive data processing and the supra-rational register of *yaqīn*—which is inseparably bound to consciousness, moral responsibility, and inner spiritual realization (*kashf*) that machines structurally lack—this research reframes AI hallucination. Hallucination is thus revealed not merely as a technical glitch or a data discrepancy, but as a dangerous epistemological illusion that lures the knowing subject into premature intellectual assent.

Consequently, this analysis moves decisively beyond the tired normative cliché of simply demanding that "humans remain in the loop" or relying on standard evidentialist fact-checking. Instead, it successfully harvests a concrete, original Ghazalian typology to govern human-machine interaction. Uncritically swallowing the persuasive, authoritative prose of an AI output constitutes a modern, blameworthy manifestation of blind imitation (*taqlid*), wherein the user surrenders autonomous cognitive agency to an algorithmic artifact. To counter this systemic erosion of intellect, the rigorous exercise of critical examination (*naẓar*) must be formally elevated into an obligatory epistemic duty. To make this normative framework practically operational, this study establishes a clear functional demarcation: verifying AI-generated outputs that touch upon foundational principles, moral frameworks, and religious doctrines functions as an individual obligation (*fard 'ayn*), demanding rigorous personal scrutiny. Conversely, evaluating routine, technical, or purely instrumental information operates as a collective obligation (*fard kifāyah*). Through this distinctive formulation, Islamic epistemology ceases to provide merely abstract critique; it furnishes a robust, actionable, and philosophically rigorous normative architecture that safeguards human agency against the seductive illusions of artificial intelligence.

References

- Akbar, Rizamul Malik, and Achmad Khudori Soleh. "The Classification of Knowledge Based on Al-Ghazali's Perspective and Its Relevance to Qur'anic Exegesis." *FiTUA: Jurnal Studi Islam* 6, no. 2 (October 2025): 218–28. <https://doi.org/10.47625/fitua.v6i2.1066>.
- Al-Ghazali, Abu Hamid. *Ihya' Ulum al-Din*. Beirut: Dar al-Kotob al-Ilmiyah, 2010.
- Al-Ghazali, Abu Hamid Muhammad. *Al-Munqidh Min al-Dalal*. Edited by Jamil Saliba and Kamil Ayyad. Beirut: Dar al-Andalus, 1967.
- Alvarado, Ramón. "AI as an Epistemic Technology." *Science and Engineering Ethics* 29, no. 5 (2023). <https://doi.org/10.1007/s11948-023-00451-3>.
- Alwahaib, Mohammad. "Al-Ghazali and Descartes from Doubt to Certainty: A Phenomenological Approach." *Discusiones Filosóficas* 18, no. 31 (July 2017): 15–40. <https://doi.org/10.17151/difil.2017.18.31.2>.
- Anam, Rifqi Khairul. "Against Algorithmic Authority: Al-Ghazali, Digital Taqlid, and the Crisis of Epistemic Agency in the Age of AI." *Journal of Islamic Philosophy and Contemporary Thought* 3, no. 1 (2025): 1–35. <https://doi.org/10.15642/jipct.2025.3.1.1-35>.

- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" 2021, 610–23. <https://doi.org/10.1145/3442188.3445922>.
- Bender, Emily M., and Alexander Koller. "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data." 2020, 5185–98. <https://doi.org/10.18653/v1/2020.acl-main.463>.
- Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. "On the Opportunities and Risks of Foundation Models." Version 3. Preprint, arXiv, 2021. <https://doi.org/10.48550/ARXIV.2108.07258>.
- Brown, Tom B., Benjamin Mann, and Nick Ryder. "Language Models Are Few-Shot Learners." *Advances in Neural Information Processing Systems* 33 (2020): 1877–901. <https://doi.org/10.48550/arXiv.2005.14165>.
- Coeckelbergh, Mark. "Democracy, Epistemic Agency, and AI: Political Epistemology in Times of Artificial Intelligence." *AI and Ethics* 3, no. 4 (November 2023): 1341–50. <https://doi.org/10.1007/s43681-022-00239-4>.
- Farquhar, Sebastian, Jannik Kossen, Lorenz Kuhn, and Yarin Gal. "Detecting Hallucinations in Large Language Models Using Semantic Entropy." *Nature* 630 (2024): 625–30. <https://doi.org/10.1038/s41586-024-07421-0>.
- Ferrario, Andrea, Alessandro Facchini, and Alberto Termine. "Experts or Authorities? The Strange Case of the Presumed Epistemic Superiority of Artificial Intelligence Systems." *Minds and Machines* 34, no. 3 (July 2024): 30. <https://doi.org/10.1007/s11023-024-09681-1>.
- Floridi, Luciano, and Massimo Chiriatti. "GPT-3: Its Nature, Scope, Limits, and Consequences." *Minds and Machines* 30 (2020): 681–94. <https://doi.org/10.1007/s11023-020-09548-1>.
- Freiberger, Oliver. *Considering Comparison: A Method for Religious Studies*. Oxford University Press, 2019. <https://doi.org/10.1093/oso/9780199965007.001.0001>.
- Freiman, Ori. "Analysis of Beliefs Acquired from a Conversational AI: Instruments-Based Beliefs, Testimony-Based Beliefs, and Technology-Based Beliefs." *Episteme* 21, no. 3 (2024): 1031–47. <https://doi.org/10.1017/epi.2023.12>.
- Ghazali, Abu Hamid Muhammad al-. *Al-Ghazālī's Path to Sufism and His Deliverance from Error: An Annotated Translation of Al-Munqidh Min Al-Dalal*. 2nd annotated reprint. Translated by Richard Joseph McCarthy. Fons Vitae, 2000.
- Hadisi, Reza. "Ghazālī's Transformative Answer to Scepticism." *Theoria* 88, no. 1 (2022): 109–42. <https://doi.org/10.1111/theo.12331>.
- Hafizh, Fadhli, Rohanda, and Abdul Kodir. "Epistimologi Ilmu Pengetahuan Dalam Perspektif John Locke Dan Al-Ghazali." *Manthiq: Jurnal Filsafat Agama Dan Pemikiran Islam* 8, no. 2 (2024). <https://doi.org/10.29300/mtq.v8i2.6764>.
- Harari, Yuval Noah. "Epistemology in the Age of Large Language Models." *Knowledge* 5, no. 1 (2025). <https://doi.org/10.3390/knowledge5010003>.

- Heersmink, Richard, Barend de Rooij, María Jimena Clavel Vázquez, and Matteo Colombo. "A Phenomenology and Epistemology of Large Language Models: Transparency, Trust, and Trustworthiness." *Ethics and Information Technology* 26 (2024). <https://doi.org/10.1007/s10676-024-09777-3>.
- Huang, Lei, Weijiang Yu, Weitao Ma, and Weihong Zhong. "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions." *ACM Transactions on Information Systems* 43, no. 2 (2025). <https://doi.org/10.1145/3703155>.
- Husni, Amin Said, Moh. Syaeful Bahar, and Dodik Harnadi. "Purifying the Heart, Unveiling the Truth: Understanding Al-Ghazālī's Epistemology." *Journal of Pesantren and Diniyah Studies* 2, no. 1 (June 2025): 1–14. <https://doi.org/10.63245/jpds.v2i1.32>.
- Jaakkola, Elina. "Designing Conceptual Articles: Four Approaches." *AMS Review* 10 (2020): 18–26. <https://doi.org/10.1007/s13162-020-00161-0>.
- Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Andrea Madotto, and Pascale Fung. "Survey of Hallucination in Natural Language Generation." *ACM Computing Surveys* 55, no. 12 (2023): 1–38. <https://doi.org/10.1145/3571730>.
- Kryscinski, Wojciech, Bryan McCann, Caiming Xiong, and Richard Socher. "Evaluating the Factual Consistency of Abstractive Text Summarization." 2020, 9332–46. <https://doi.org/10.18653/v1/2020.emnlp-main.750>.
- Lavrinovics, Ernests, Russa Biswas, Johannes Bjerva, and Katja Hose. "Knowledge Graphs, Large Language Models, and Hallucinations: An NLP Perspective." *Journal of Web Semantics* 85 (2025). <https://doi.org/10.1016/j.websem.2024.100844>.
- Lin, Stephanie, Jacob Hilton, and Owain Evans. "TruthfulQA: Measuring How Models Mimic Human Falsehoods." 2022, 3214–52. <https://doi.org/10.18653/v1/2022.acl-long.229>.
- Manakul, Potsawee, Adian Liusie, and Mark Gales. "SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models." 2023, 9004–17. <https://doi.org/10.18653/v1/2023.emnlp-main.557>.
- Maynez, Joshua, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. "On Faithfulness and Factuality in Abstractive Summarization." 2020, 1906–19. <https://doi.org/10.18653/v1/2020.acl-main.173>.
- Mohamed, Nabil Yasien. *Ghazālī's Epistemology: A Critical Study of Doubt and Certainty*. Routledge, 2023.
- Negrotti, Massimo. "On the 'Nature' of the 'Artificial.'" *AI & SOCIETY* 38, no. 5 (October 2023): 1935–40. <https://doi.org/10.1007/s00146-021-01178-0>.
- Paseri, Ludovica, and Massimo Durante. "Examining Epistemological Challenges of Large Language Models in Law." *Cambridge Forum on AI: Law and Governance* 1 (2025): e7. <https://doi.org/10.1017/cfl.2024.7>.
- Passos, Luciano C. "A Survey on Hallucination in Large Language Models: Causes, Detection, and Mitigation." *Artificial Intelligence Review* 58 (2025). <https://doi.org/10.1007/s10462-025-11106-5>.

- Pritchard, Duncan. "Anti-Luck Epistemology." *Synthese* 158, no. 3 (September 2007): 277–97. <https://doi.org/10.1007/s11229-006-9039-7>.
- Saleh, Sujiat Zubaidi, Muhammad Taqiyuddin, and Rakhmad Agung Hidayatullah. "On Imam Ghazālī's Farḍ 'Ain and Farḍ Kifāyah: An Epistemological Approach." *Jurnal Ushuluddin* 30, no. 1 (2022): 48–62. <https://doi.org/10.24014/jush.v30i1.14661>.
- Shuster, Kurt, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. "Retrieval Augmentation Reduces Hallucination in Conversation." 2021, 3784–803. <https://doi.org/10.18653/v1/2021.findings-emnlp.320>.
- Suwarno, M. Hadi, and Hayyan Ahmad Ulul Albab. "Knowledge and Scholarly Authority in Al-Ghazali's Philosophy: A Study of Fatihatul 'Ulum." *Journal of Pesantren and Diniyah Studies* 1, no. 2 (2024): 185–98. <https://doi.org/10.63245/jpds.vii2.29>.
- Syihabuddin, Budi, and Sumanta Hasyim. "Al-Ghazali's Empirical–Rational–Spiritual Epistemology and Its Implications for Contemporary Islamic Religious Education." *Afaqulilmi: Journal of Islamic Education*, ahead of print, 2026. <https://doi.org/10.60769/afaq.v2i1.6>.
- Zhou, Kaitlyn, Dan Jurafsky, and Tatsunori Hashimoto. "Navigating the Grey Area: How Expressions of Uncertainty and Overconfidence Affect Language Models." 2023, 5506–24. <https://doi.org/10.18653/v1/2023.emnlp-main.335>.



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license ([Deed - Attribution-NonCommercial 4.0 International - Creative Commons](#)).