P-ISSN : 2654-5721
E-ISSN : 2654-7546
DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2

TRUTH WITHOUT BELIEF: CAN LARGE LANGUAGE MODELS PRODUCE TRUTHFUL STATEMENTS WITHOUT KNOWING WHAT IS TRUE?

[1]Faiz Prama Dipta, [2]Rachel Ekarika Marbun,[3] Imel Damai Monica Girsang, [4]Swastika Nuraini Putri, [5]Adelia Paccecia Sitohang, [6]Johan Manurung

UNIVERSITAS NEGERI MEDAN

ABSTRACT

The emergence of Large Language Models (LLMs) has challenged conventional assumptions about the relationship between language, truth, belief, and knowledge. LLMs can generate statements that are factually correct and linguistically convincing, yet this capacity does not necessarily imply that the system believes, knows, or understands the truth of what it produces. This study examines how users interpret truthful statements generated by LLMs and how such interactions influence their attribution of belief, knowledge, assertion, and epistemic authority to artificial intelligence. The study adopts a qualitative interpretive phenomenological approach combined with reflexive thematic analysis. The proposed participants are users in higher-education contexts who have experience interacting with LLMs for academic or informational purposes. The analysis is organized around four provisional themes: (1) convincing-looking truth, (2) ruptures in trust, (3) being correct does not mean knowing, and (4) renegotiating epistemic authority. The conceptual analysis indicates that factual correctness and epistemic possession should be treated as distinct categories. An LLM may produce a true proposition without possessing belief or knowledge in the human epistemic sense. Consequently, the apparent truthfulness of an output should not automatically be interpreted as evidence of epistemic agency. The study contributes to the philosophy of language and AI ethics by proposing the concept of truth without belief and by highlighting the distinction between epistemic outsourcing and epistemic transfer. These findings also have implications for AI literacy in higher education, particularly in developing practices of verification, calibrated trust, and critical engagement with generative AI.

Keywords: Large Language Models; Truth; Belief; Knowledge; Epistemic Authority; Speech Acts; AI Literacy

DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2
P-ISSN : 2654-5721
E-ISSN : 2654-7546

1. INTRODUCTION

1.1 Background of the Study

The rapid development of Large Language Models (LLMs) has transformed the ways in which people search for information, formulate arguments, write academic texts, and interact with computational systems. Systems based on LLMs are increasingly capable of producing responses that are coherent, context-sensitive, fluent, and apparently informative. Their linguistic fluency, however, creates an important philosophical problem: when an LLM produces a statement that is true, in what sense, if any, does the system know that the statement is true?

This question becomes particularly significant because truth and knowledge are not necessarily identical. A proposition may be true even when the person who utters it does not believe it. Likewise, someone may accidentally produce a true statement without possessing knowledge of the relevant fact. The classical philosophical tradition has therefore distinguished truth from epistemic states such as belief and knowledge. In contemporary epistemology, knowledge has often been analyzed in relation to justified true belief, although this formulation remains philosophically contested (de Grefte, 2021).

The emergence of LLMs makes this distinction especially salient. An LLM can generate the sentence “Jakarta is the capital of Indonesia,” for example, and the proposition may be true. Yet the factual correctness of the statement does not by itself establish that the system believes the proposition or knows it to be true. The distinction becomes even more apparent when LLMs generate false but highly plausible statements. Research has demonstrated that LLMs may produce hallucinated information, including factual errors and fabricated bibliographic references (Farquhar et al., 2024; Li et al., 2024; Walters & Wilder, 2023).

The problem is therefore not merely whether LLM outputs are true or false. A deeper philosophical question concerns the epistemic status of the entity producing those outputs. Recent research has shown that language models can have difficulty reliably distinguishing belief, knowledge, and fact, particularly in tasks involving false beliefs and epistemic perspective-taking (Suzgun et al., 2025). Earlier research similarly investigated whether LLMs could model what humans know and believe, raising questions about whether successful performance should be interpreted as evidence of genuine epistemic states (Trott et al., 2023).

The issue is also connected to trust. Users may interpret fluent and conversational responses as more credible than equivalent information presented in a static format. Anderl et al. (2024), for example, found that conversational presentation can increase credibility judgments while reducing users’ ability to distinguish accurate from inaccurate information. Steyvers et al. (2025) further demonstrate a calibration problem in which people may overestimate what LLMs know and the reliability of their responses.

In educational contexts, this issue becomes particularly important. LLMs are increasingly used by students and educators for explanation, brainstorming, summarization, writing assistance, and information retrieval. Kasneci et al. (2023) emphasize both the opportunities and challenges of LLMs in education, including the need for critical thinking, verification, and appropriate AI competencies.

The philosophical problem can therefore be formulated as follows:

Can an LLM produce a truthful statement without believing or knowing that the statement is true?

This question requires a distinction between at least four concepts: truth, belief, knowledge, and assertion. Truth concerns the relation between a proposition and reality. Belief concerns an epistemic attitude toward a proposition. Knowledge involves a stronger epistemic relation to truth. Assertion, meanwhile, concerns the pragmatic act of presenting a proposition as true.

P-ISSN : 2654-5721
E-ISSN : 2654-7546
DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2

The problem becomes more complex when LLM-generated language is interpreted as an assertion. Gubelmann (2024) argues that current LLMs lack the intentional agency required for genuine speech acts in the ordinary human sense. Arora (2024), however, proposes the concept of proxy assertions to explain how machine-generated utterances may function within communicative practices even when the machine itself lacks the relevant human agency.

Consequently, this study does not simply ask whether LLMs know. Rather, it investigates how humans interpret and attribute epistemic status to LLM-generated language. The central concern is the gap between producing truth and possessing truth epistemically.

1.2 Problem Statement

The central problem of this study is the distinction between the production of a true proposition and the possession of epistemic knowledge. An LLM can generate a proposition that corresponds to reality, but this factual correctness does not necessarily establish that the LLM believes the proposition, knows the proposition, or intentionally asserts it.

The problem becomes particularly important because linguistic fluency can create an appearance of epistemic competence. An LLM may provide an answer that is coherent, detailed, grammatically correct, and confident. Users may consequently interpret the output as evidence that the system is knowledgeable. However, the production of a convincing linguistic response does not necessarily establish an epistemic relationship between the system and the proposition it produces.

The central problem can therefore be expressed as the difference between:

truth production and **epistemic possession**.

This distinction is especially important when LLMs are used as sources of information in educational and academic contexts.

1.3 Research Gap

Existing research has examined several aspects of LLM behavior and human interaction with AI systems, including hallucination, factuality, credibility, trust, epistemic reasoning, and the ability of language models to distinguish belief from knowledge and fact.

Research on hallucination has demonstrated that LLMs may produce factually incorrect or fabricated information while maintaining a linguistically convincing form (Farquhar et al., 2024; Li et al., 2024; Walters & Wilder, 2023). Research on credibility and trust has also shown that conversational presentation may influence how users evaluate information produced by LLMs (Anderl et al., 2024; Steyvers et al., 2025).

Other studies have investigated whether LLMs can model human knowledge and belief or distinguish between epistemic concepts (Trott et al., 2023; Suzgun et al., 2025). However, these lines of research do not by themselves resolve the deeper philosophical question of whether the ability to produce a true proposition should be interpreted as evidence that the producing system possesses knowledge of that proposition.

The present study addresses this gap by examining the distinction between **factual correctness** and **epistemic possession**. Rather than treating truth, belief, knowledge, and assertion as interchangeable, the study investigates how these concepts can be distinguished when applied to LLM-generated language and how users attribute epistemic status to such outputs.

1.4 Research Questions

This study addresses the following research questions:

1. How do users interpret truthful statements generated by LLMs?
2. How do users distinguish between an LLM producing a true statement and

an LLM knowing that the statement is true?
3. How does interaction with LLMs influence users' attribution of belief, knowledge, and assertion to AI systems?
4. How do users negotiate epistemic trust and authority when LLM outputs are simultaneously useful and potentially unreliable?

1.5 Research Objectives

The objectives of this study are:

1. To examine how users interpret truthful statements generated by LLMs.
2. To investigate how users distinguish factual correctness from epistemic knowledge.
3. To examine how interaction with LLMs influences the attribution of belief, knowledge, and assertion to AI systems.
4. To explore how users negotiate epistemic trust and authority when interacting with LLMs.

1.6 Significance of the Study

1.6.1 Theoretical Significance

This study contributes to contemporary philosophy of language and epistemology by examining whether truth-bearing linguistic output necessarily presupposes belief, knowledge, or intentional agency on the part of the producer. The study also proposes the conceptual distinction between truth production and epistemic possession.

1.6.2 Educational Significance

The study is relevant to higher education because LLMs are increasingly used for explanation, brainstorming, summarization, writing assistance, and information retrieval. Understanding the distinction between factual correctness and knowledge can support more critical and responsible use of LLMs.

1.6.3 Practical Significance

The study emphasizes the importance of calibrated trust. Users should neither automatically accept nor automatically reject LLM-generated information. Instead, the degree of verification should correspond to the importance and potential consequences of the claim.

2. THEORETICAL FRAMEWORK

2.1 Truth and the Truth of Propositions

At the most basic level, truth concerns whether a proposition accurately represents a state of affairs. A proposition can therefore be true independently of the psychological condition of the entity producing it.

This distinction is essential to understanding LLM-generated language. If an LLM produces a proposition that corresponds to reality, the proposition may be true even if there is no reason to attribute a human-like belief to the system.

Scharp (2021) emphasizes that the philosophical analysis of truth involves distinctions concerning aletheic properties and the conceptual structures through which truth is understood. For the present study, the important point is that truth should not automatically be equated with the psychological state of the producer.

Thus:

True statement ≠ statement produced by a believer.

This distinction provides the conceptual foundation for the phrase **"truth without belief."**

2.2 Belief

P-ISSN : 2654-5721
E-ISSN : 2654-7546
DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2

Belief is an epistemic attitude. When a person believes that p, the person stands in a particular psychological relation to proposition p.

Belief therefore concerns the relation between an agent and a proposition. The fact that a proposition is true does not automatically establish that an agent believes the proposition.

In the context of LLMs, this distinction is important because a system may generate a proposition without necessarily possessing a psychological belief state corresponding to that proposition.

2.3 Knowledge

Knowledge involves a stronger epistemic relation than belief. The classical formulation of knowledge as justified true belief remains influential, although contemporary epistemology has identified various problems with the formulation (de Grefte, 2021).

The concepts can be represented as follows:

Truth:
p is true.

Belief:
Person X believes that p.

Knowledge:
Person X knows that p.

These propositions should not be treated as interchangeable.

An LLM may produce:

“Water freezes at approximately 0°C under standard atmospheric pressure.”

The proposition may be true. But from this fact alone, we cannot infer:

The LLM believes that water freezes at approximately 0°C.

Nor can we infer:

The LLM knows that water freezes at approximately 0°C.

The output establishes, at most, that a linguistic system has generated a proposition that happens to correspond to a fact.

2.4 Truth, Belief, and Knowledge: Conceptual Distinction

The relationship among the three concepts can be summarized as follows:

Concept	Central Question
Truth	Is proposition p true?
Belief	Does agent X believe p?
Knowledge	Does agent X know p?

The central distinction is:

p is true ≠ X believes p ≠ X knows p

This distinction is fundamental to the argument of the present study. A true proposition does not necessarily require a believer, and a correct output does not necessarily establish knowledge on the part of the producer.

2.5 LLMs and Epistemic Attribution

Research increasingly demonstrates that LLM behavior complicates the attribution of epistemic states. Trott et al. (2023) examined whether LLMs could represent what humans know and believe. More recent research by Suzgun et al. (2025) found that language models cannot reliably distinguish belief from knowledge and fact across a range of epistemic tasks.

This evidence supports a cautious interpretation: behavioral success in epistemic tasks does not necessarily demonstrate possession of human-like epistemic states.

Moreover, LLMs may generate factual information inconsistently. Jiang et al. (2024) examined hallucinations concerning facts that models may appear to have previously represented. Zhao et al. (2024) similarly demonstrated that factual recall in language models involves multiple dimensions, including accuracy, consistency, and reliability.

Therefore, the epistemic status of an LLM output cannot be determined solely from its fluency.

2.6 Hallucination and Factuality

The phenomenon commonly called hallucination further complicates the relationship between language and truth. LLMs can produce statements that are linguistically well-formed but unsupported or false.

Farquhar et al. (2024) proposed semantic entropy as a method for detecting certain forms of hallucination or confabulation. Li et al. (2024) likewise investigated factuality hallucinations in LLMs.

Van Deemter (2024) argues that even the definition of hallucination is conceptually difficult. This is relevant to the present study because the philosophical issue is not simply that LLMs sometimes produce false information. More fundamentally, the problem concerns how humans interpret a linguistic output when its surface form suggests knowledge or confidence that may not exist.

Hallucination therefore demonstrates that:

linguistic fluency ≠ factual reliability, and: **linguistic confidence ≠ epistemic certainty.**

2.7 Assertion and Speech Acts

Human assertions are generally understood as communicative acts performed by agents. Assertions involve not merely producing words but participating in a normative communicative practice.

Gubelmann (2024) argues that current LLMs do not possess the intentional agency required for genuine speech acts in the conventional sense. The system can produce linguistic forms resembling assertions, but resemblance to assertion does not necessarily constitute an assertion performed by an intentional speaker.

Arora (2024) develops an alternative perspective through the concept of proxy assertions, whereby machine-generated statements may function as assertions within human communicative practices even if the machine is not itself a full epistemic agent.

This distinction allows the present study to separate:

linguistic output → apparent assertion → attributed epistemic authority

from:

intentional belief → knowledge → genuine human assertion.

2.8 Epistemic Authority and Trust

The philosophical problem is also social. Users do not interact with LLMs merely as language generators; they often treat them as sources of information.

Steyvers et al. (2025) show that people can overestimate what LLMs know. Anderl et al. (2024) demonstrate that conversational presentation can increase perceived credibility. Pataranutaporn et al. (2023) further show that beliefs about AI can influence perceptions of trustworthiness, empathy, and effectiveness.

This indicates that epistemic authority is not simply a property possessed by the system. It is also attributed by users.

P-ISSN : 2654-5721
E-ISSN : 2654-7546
DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2

The present study therefore conceptualizes epistemic authority as relational:

LLM output → user interpretation → attribution of competence → trust → epistemic authority

Epistemic authority can therefore develop through repeated experiences of usefulness, fluency, accuracy, and responsiveness, while it can also be weakened by errors and contradictions.

3. CONCEPTUAL FRAMEWORK

3.1 Conceptual Relationship

The conceptual framework of this study can be summarized as follows:

LLM produces a proposition

↓

The proposition may be true or false

↓

Truth does not automatically indicate belief

↓

Belief does not automatically indicate knowledge

↓

Knowledge is not identical to the ability to produce language

↓

LLM output can function like an assertion without necessarily having an intentional speaker

↓

Users attribute epistemic authority

↓

Trust is formed, calibrated, or ruptured

↓

Users renegotiate their relationship with knowledge

This framework places the LLM not simply as a system that either "knows" or "does not know," but as a linguistic system capable of producing propositions whose truth does not necessarily establish a human-like epistemic relation to those propositions.

3.2 Central Proposition

The central proposition of this study is:

Truth can be produced without belief, but epistemic authority cannot be assumed merely from the production of truth.

This proposition captures the main distinction developed throughout the study. A system can produce a proposition that is true without necessarily possessing belief, knowledge, or intentional assertion concerning that proposition.

3.3 Conceptual Model

The study proposes a conceptual distinction among four levels:

Level	Question	Status
Truth	Is the proposition correct?	Property of proposition
Belief	Does the agent believe the proposition?	Epistemic/psychological state
Knowledge	Does the agent know the proposition?	Epistemic status

UNIVERSITAS PERSADA INDONESIA YAI JAKARTA YAYASAN ADMINISTRASI INDONESIA

DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2
P-ISSN : 2654-5721
E-ISSN : 2654-7546

Level	Question	Status
Assertion	Is the proposition intentionally presented as true?	Speech act

An LLM may satisfy the first condition without necessarily satisfying the remaining three.

The central conceptual claim is therefore:

An LLM can produce a true proposition without possessing belief, knowledge, or intentional assertion concerning that proposition.

This proposition does not deny that future AI systems could possess richer forms of agency or cognition. Rather, it identifies a conceptual problem concerning how current LLM behavior should be interpreted.

4. METHODOLOGY

4.1 Research Design

This study employs a qualitative interpretive phenomenological approach combined with reflexive thematic analysis.

The phenomenological orientation is used to examine how users experience interactions with LLMs and how they interpret the meaning of truthful, uncertain, or false outputs. The approach is particularly relevant because the study is concerned not only with whether an output is objectively true but also with how users experience and interpret its truthfulness.

Reflexive thematic analysis is used to identify recurring patterns in participants' accounts concerning truth, belief, knowledge, trust, and epistemic authority.

4.2 Participants

The proposed participants are users who have experience interacting with LLMs in educational, academic, or informational contexts.

Purposive sampling is appropriate because participants need to have direct experience with LLM-based systems.

The proposed sample is approximately 15–20 participants.

The participant number and participant statements in the earlier analytical draft were simulated for conceptual development. They should be replaced with actual participant data before publication.

4.3 Data Collection

Data may be collected through:

1. Semi-structured interviews;
2. Elicitation scenarios involving true and false LLM responses;
3. Participants' reflections on their experiences using LLMs;
4. Observation of selected interaction episodes;
5. Relevant interaction artifacts, where ethically permissible.

The interviews may focus on questions such as:

- What makes an LLM response appear truthful?
- Does a correct answer mean that the LLM knows the answer?
- Can an LLM believe something?
- When does an LLM response become trustworthy?
- What happens to trust after an LLM gives an obviously false answer?
- Do users regard LLMs as sources of knowledge or merely tools for producing information?

4.4 Data Analysis

The analysis proceeds through several stages:

P-ISSN : 2654-5721
E-ISSN : 2654-7546
DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2

1. Familiarization with the data;
2. Identification of significant statements;
3. Initial coding;
4. Development of candidate themes;
5. Review and refinement of themes;
6. Interpretation of themes through philosophical concepts;
7. Development of a conceptual relationship among truth, belief, knowledge, assertion, and epistemic authority.

The interpretive process is guided by the following conceptual framework:

LLM output

↓

Interaction experience

↓

Interpretation of truthfulness

↓

Attribution of belief/knowledge

↓

Attribution of assertion

↓

Trust formation or rupture

↓

Reinterpretation of epistemic authority

4.5 Ethical Considerations

Because the proposed research involves human participants, the study should consider informed consent, anonymity, confidentiality, voluntary participation, responsible data handling, and appropriate protection of participant information.

Participant quotations and interaction artifacts should only be used where ethically permissible and should be anonymized where necessary.

5. PROVISIONAL ANALYTICAL THEMES

Note: The themes in this section represent a provisional analytical structure. They should be populated with genuine participant quotations and actual empirical data after the field research has been conducted. They should not be presented as completed empirical findings before the actual data are available.

Four major themes are proposed from the conceptual analysis.

5.1 Theme 1: Convincing-Looking Truth

LLM outputs may appear convincing because of their linguistic fluency, structure, completeness, and confidence. Users may initially associate linguistic quality with epistemic reliability.

A response that is:

- fluent,
- systematic,
- detailed,
- confident,
- grammatically correct,

may be interpreted as more knowledgeable than a shorter or less polished response.

This phenomenon corresponds with research showing that conversational presentation can increase perceived credibility (Anderl et al., 2024).

The philosophical significance is that the appearance of epistemic competence can emerge from linguistic performance without establishing genuine epistemic possession.

5.2 Theme 2: Ruptures in Trust

Trust may change when users encounter hallucinations, contradictions, fabricated citations, or obviously incorrect information.

The experience of a single false answer may cause users to reconsider previous assumptions about the reliability of the system.

This corresponds with research showing that LLMs may generate hallucinated or fabricated content, including bibliographic references (Farquhar et al., 2024; Walters & Wilder, 2023).

The resulting pattern can be described as:

confidence → contradiction → doubt → verification → calibrated trust

Trust therefore becomes conditional rather than absolute.

5.3 Theme 3: Being Correct Does Not Mean Knowing

This theme constitutes the central philosophical finding proposed by the study.

Users may recognize that an LLM can generate an answer that is correct while simultaneously questioning whether the system "knows" the answer.

The distinction can be represented as:

Correct output ≠ belief ≠ knowledge.

An LLM may produce a true proposition without possessing an epistemic relation to that proposition comparable to human knowledge.

This interpretation is strengthened by research showing that LLMs struggle to reliably distinguish belief, knowledge, and fact (Suzgun et al., 2025).

The result therefore supports a conceptual distinction between:

truth production

and

epistemic possession.

5.4 Theme 4: Renegotiating Epistemic Authority

Users do not necessarily abandon LLMs after encountering errors. Instead, they may change the way they use them.

LLMs may continue to be regarded as:

- brainstorming tools,
- explanatory assistants,
- drafting partners,
- language-processing tools,
- sources of possible answers,

while no longer being treated as unquestionable authorities.This process represents a renegotiation of epistemic authority.

Rather than asking:

"Can I trust the AI?"

users may increasingly ask:

"For what kinds of questions should I trust the AI, and to what degree?"

This represents a shift from unconditional trust to calibrated trust.

6. DISCUSSION

6.1 Truth Does Not Require Belief

The central argument of this study is that truth and belief should not be conceptually collapsed.

A statement can be true independently of whether its producer believes it. This becomes especially important for computational systems.

P-ISSN : 2654-5721
E-ISSN : 2654-7546
DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2

If an LLM generates a true proposition, the truth of that proposition is determined by the relevant relationship between proposition and reality, not by whether the machine possesses a corresponding psychological state.

Therefore:

An LLM can produce truth without necessarily possessing belief in that truth.

This does not mean that LLMs are truthful in the same sense as humans. Truthfulness often implies a relationship between what one believes and what one says. An LLM, however, may lack the relevant belief state.

The phrase **truth without belief** therefore describes a structural possibility rather than a claim that LLMs possess a new form of human-like truthfulness.

6.2 Being Correct Does Not Establish Knowledge

The second major implication concerns knowledge. Suppose an LLM produces proposition p, and p happens to be true.

From:

LLM produces p

and:

p is true

we cannot logically infer:

LLM knows p.

The inference requires additional premises concerning epistemic justification, cognitive access, or an appropriate knowledge relation.

This distinction is particularly important because users may interpret accurate answers as evidence of knowledge.

Suzgun et al. (2025) demonstrate why this assumption is problematic. Their findings show that contemporary language models have difficulty reliably distinguishing belief from knowledge and fact.

The present study therefore proposes that factual accuracy should be treated as a property of output, not automatically as evidence of epistemic status.

6.3 Hallucination as an Epistemic Warning

Hallucination makes the distinction more visible.

If a system produces an incorrect answer with high confidence, users may discover that linguistic confidence is not equivalent to epistemic certainty.

Farquhar et al. (2024) demonstrate that hallucination can be detected through semantic uncertainty, while Menczer et al. (2024) describe factuality challenges associated with LLM-generated content.

The philosophical lesson is that:

The linguistic form of certainty should not be confused with epistemic certainty.

An LLM may sound certain because of the statistical and generative characteristics of its output, while lacking the human capacity to possess or consciously experience certainty.

6.4 From Assertion to Proxy Assertion

The question of whether an LLM asserts something is more complicated.

In ordinary human communication, assertion is not merely the production of a sentence. It is an intentional communicative act.

Gubelmann (2024) therefore argues that current LLMs are not genuine speech-act agents in the ordinary sense because they lack the relevant intentional agency.

P-ISSN : 2654-5721
E-ISSN : 2654-7546

Arora (2024), however, suggests that machine-generated utterances can be understood as proxy assertions within broader human communicative arrangements.

This distinction is useful for interpreting LLM interaction.

The sentence produced by an LLM may function socially as an assertion even if the LLM itself is not an intentional speaker.

Thus:

machine output ≠ human assertion

but:

machine output → can function as a proxy assertion within human communication.

6.5 The Social Construction of Epistemic Authority

Epistemic authority is partly constructed through user interpretation.

A system may become authoritative when users repeatedly experience it as useful, fluent, accurate, and responsive. Yet authority may decline or collapse when users discover systematic errors.

Steyvers et al. (2025) show that users may overestimate LLM capabilities. Anderl et al. (2024) show that conversational presentation can increase credibility judgments.

Therefore, the authority of LLMs should be understood as relational and negotiated, rather than intrinsic.

The user may effectively perform an epistemic evaluation:

The answer sounds convincing.

then:

The answer appears correct.

then:

The system seems knowledgeable.

The danger lies in moving too quickly from the first two propositions to the third.

6.6 Epistemic Outsourcing Without Epistemic Transfer

The study proposes the concept of **epistemic outsourcing without epistemic transfer**.

When users ask an LLM a question, they may outsource part of the process of finding, organizing, or expressing information.

However, this does not mean that knowledge has been transferred from the LLM to the user or that the LLM itself possesses knowledge in the human sense.

This distinction can be expressed as:

Information production → epistemic outsourcing

but not necessarily:

Information production → epistemic possession.

The user may therefore outsource part of the cognitive labor while retaining responsibility for epistemic verification.

This principle is particularly relevant in higher education.

6.7 Implications for Higher Education

The findings have several implications for education.

First, AI literacy should not be limited to teaching students how to operate LLM tools.

Students should also understand the distinction between:

- information and knowledge;

P-ISSN : 2654-5721
E-ISSN : 2654-7546
DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2

- truth and belief;
- fluency and reliability;
- confidence and certainty;
- assertion and generated text;
- assistance and epistemic authority.

Second, educational institutions should emphasize verification practices.

Students should be encouraged to:

1. Verify factual claims;
2. Inspect original sources;
3. Check citations;
4. Compare alternative sources;
5. Identify uncertainty;
6. Distinguish generated content from established knowledge.

Kasneci et al. (2023) similarly emphasize the importance of critical thinking and appropriate human oversight in educational uses of LLMs.

Third, the goal should not be to teach students either to trust or distrust AI categorically.

The more appropriate goal is **calibrated epistemic trust**.

7. THEORETICAL CONTRIBUTION

7.1 Truth Without Belief

The study proposes the concept of **truth without belief** to describe a situation in which a linguistic system can generate a true proposition without necessarily possessing a corresponding human-like belief.

The concept does not claim that LLMs possess human-like truthfulness. Instead, it identifies a distinction between the truth of a proposition and the psychological or epistemic condition of the producer.

7.2 Truth Production and Epistemic Possession

The study distinguishes between:

truth production

and:

epistemic possession.

An LLM may generate a proposition that is factually correct. However, factual correctness is a property of the output and does not automatically establish that the system possesses knowledge of the fact.

The distinction can therefore be summarized as:

Correct output ≠ epistemic possession.

7.3 Epistemic Outsourcing Without Epistemic Transfer

The second major conceptual contribution is the distinction between epistemic outsourcing and epistemic transfer.

Users may delegate certain informational or linguistic tasks to LLMs. However, this does not necessarily mean that epistemic responsibility has been transferred to the machine.

The user may outsource part of the cognitive labor while remaining responsible for evaluating, verifying, and interpreting the information.

8. IMPLICATIONS

8.1 Theoretical Implications

The study contributes to philosophy of language by distinguishing:

- truth from belief;
- factuality from knowledge;
- linguistic production from assertion;
- machine output from intentional speech acts;
- information assistance from epistemic authority.

These distinctions help clarify how LLM-generated language should be interpreted without automatically treating linguistic

performance as evidence of human-like epistemic agency.

8.2 Educational Implications

Higher education should develop AI literacy that includes:

- source verification;
- epistemic calibration;
- critical reading;
- citation verification;
- recognition of hallucinations;
- awareness of the limits of machine-generated knowledge.

Students should understand why a fluent answer is not necessarily a reliable answer, why a correct statement does not necessarily demonstrate knowledge, and why apparent confidence does not constitute epistemic certainty.

8.3 Practical Implications

Users should treat LLMs as epistemically useful but epistemically non-infallible systems.

The appropriate response to an LLM answer is neither unconditional acceptance nor unconditional rejection, but critical verification proportional to the importance of the claim.

9. LIMITATIONS

This study has several limitations.

First, the philosophical distinction between truth, belief, knowledge, and assertion remains conceptually contested. Different epistemological and philosophy-of-language traditions may provide different interpretations.

Second, LLM capabilities evolve rapidly. Claims concerning agency, intentionality, and epistemic status must therefore be understood as historically situated claims concerning contemporary systems.

Third, user perceptions of LLMs may vary according to disciplinary background, AI literacy, previous experience, and task context.

Fourth, the provisional results presented in this draft should not be interpreted as completed empirical findings. The earlier participant matrix was developed as a simulated analytical scaffold. Actual publication requires replacing it with authentic interview data and verbatim quotations from participants.

10. CONCLUSION

This study examines the philosophical problem of whether Large Language Models can produce truthful statements without knowing what is true.

The analysis indicates that an LLM can generate a proposition that is true without this fact establishing that the system believes or knows the proposition.

The central distinction is therefore between **truth as a property of a proposition** and **knowledge as an epistemic relation**.

The analysis identifies four provisional themes: **convincing-looking truth, ruptures in trust, being correct does not mean knowing, and renegotiating epistemic authority**. These themes demonstrate that users' relationships with LLMs are shaped not only by factual accuracy but also by linguistic fluency, confidence, errors, verification practices, and changing perceptions of epistemic authority.

The philosophical contribution of the study is the concept of **truth without belief**. The concept highlights a condition in which linguistic systems can generate true propositions without necessarily possessing human-like beliefs about those propositions.

The study further introduces the idea of **epistemic outsourcing without epistemic transfer**. Users may delegate certain informational or linguistic tasks to LLMs without thereby transferring epistemic responsibility to the machine.

This distinction has practical implications for higher education. AI literacy should involve not merely technical competence but also philosophical and epistemic competence. Students need to understand why a fluent answer is not necessarily a reliable answer, why a correct statement does not necessarily demonstrate knowledge, and why apparent confidence does not constitute epistemic certainty.

Ultimately, the central question is not simply:

“Is the AI telling the truth?”

but:

“What does it mean for a system without belief to produce something true?”

This question places LLMs at an important intersection between philosophy of language, epistemology, artificial intelligence, and education.

11. FUTURE RESEARCH

Future research should investigate these questions using actual empirical data from students, lecturers, researchers, and professional users.

Comparative research could also examine whether different levels of AI literacy influence the attribution of belief, knowledge, and authority to LLMs.

Future studies may further examine how users distinguish between factual correctness and epistemic knowledge across different tasks and contexts. Such research could provide additional evidence concerning how epistemic authority is formed, challenged, and renegotiated during interactions with LLMs.

REFERENCES

Anderl, C., Klein, S. H., Sarigül, B., Schneider, F. M., Han, J., Fiedler, P. L., & Utz, S. (2024). Conversational presentation mode increases credibility judgements during information search with ChatGPT. *Scientific Reports, 14*, Article 17127. https://doi.org/10.1038/s41598-024-67829-6

Arora, A. (2024). Proxy assertions and agency: The case of machine-assertions. *Philosophy & Technology, 37*, Article 15. https://doi.org/10.1007/s13347-024-00703-5

de Grefte, J. (2021). Knowledge as justified true belief. *Erkenntnis, 88*, 531–549. https://doi.org/10.1007/s10670-020-00365-7

Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature, 630*, 625–630. https://doi.org/10.1038/s41586-024-07421-0

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines, 30*, 411–437. https://doi.org/10.1007/s11023-020-09539-2

Gubelmann, R. (2024). Large language models, agency, and why speech acts are beyond them (for now): A Kantian-cum-pragmatist case. *Philosophy & Technology, 37*, Article 32. https://doi.org/10.1007/s13347-024-00696-1

Jiang, B., Qi, Y., Hong, X., Fu, J., Cheng, Z., Meng, Y., Yu, D., Zhou, Y., et al. (2024). On large language models hallucination with regard to known facts. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1041–1053). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.60

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, F., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lai, H. (2020). Is knowing that P identical with knowing that P is true? *Philosophia, 48*, 1075–1092. https://doi.org/10.1007/s11406-019-00113-4

Li, J., Chen, G., Ren, Z., Cheng, Y., Zhao, Y., Nie, L., & Wen, J.-R. (2024). The dawn after the dark: An empirical study on factuality hallucination in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 10879–10899). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.586

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1906–1919). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.173

Menczer, F., et al. (2024). Factuality challenges in the era of large language models and opportunities for fact-checking. *Nature Machine Intelligence, 6*, 852–863. https://doi.org/10.1038/s42256-024-00881-z

Pataranutaporn, P., Liu, R., Finn, E., & Maes, P. (2023). Influencing human-AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence, 5*, 1076–1086. https://doi.org/10.1038/s42256-023-00720-7

Rapp, A., Di Lodovico, A., & Di Caro, L. (2025). How do people react to ChatGPT's unpredictable behavior? Anthropomorphism, uncanniness, and fear of AI: A qualitative study on individuals' perceptions and understandings of LLMs nonsensical hallucinations. *International Journal of Human-Computer Studies, 198*, Article 103471. https://doi.org/10.1016/j.ijhcs.2025.103471

Scharp, K. (2021). Conceptual engineering for truth: Aletheic properties and new aletheic concepts. *Synthese, 198*, 647–688.

Steyvers, M., Tejeda, H., Kumar, A., Belem, C., Karny, S., Hu, X., Mayer, L. W., & Smyth, P. (2025). What large language models know and what people think they know. *Nature Machine Intelligence, 7*, 221–231. https://doi.org/10.1038/s42256-024-00976-7

Suzgun, M., Gur, T., Bianchi, F., Ho, D. E., Icard, T., Jurafsky, D., & Zou, J. (2025). Language models cannot reliably distinguish belief from knowledge and fact. *Nature Machine Intelligence, 7*, 178–190. https://doi.org/10.1038/s42256-025-01113-8

Trott, S., Jones, C., Chang, T., Michaelov, J. A., & Bergen, B. (2023). Do large language models know what humans know? *Cognitive Science, 47*(7), Article e13309. https://doi.org/10.1111/cogs.13309

van Deemter, K. (2024). The pitfalls of defining hallucination. *Computational Linguistics, 50*(2), 807–816. https://doi.org/10.1162/coli_a_00509

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports, 13*, Article 14045. https://doi.org/10.1038/s41598-023-41032-5

Xu, R., Lin, B., Yang, S., Zhang, T., Shi, W., Zhang, T., Fang, Z., Xu, W., & Qiu, H. (2024). The Earth is flat because...: Investigating LLMs belief towards misinformation via persuasive conversation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 16259–16303). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.858

P-ISSN : 2654-5721
E-ISSN : 2654-7546
DOI : https://doi.org/10.37817/ikra-ithabdimas.v10i2

Zhao, W., Yoshinaga, N., & Oba, T. (2024). What matters in memorizing and recalling facts? Multifaceted benchmarks for knowledge probing in language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 13186–13214).