



THE CONCEPT OF TRUTH IN THE ERA OF INFORMATION TECHNOLOGY AND ARTIFICIAL INTELLIGENCE

**Li Ekaterina^{1*}, Izzetova Emine², Naumenko Oleg³, Tillavoldiyeva Matlubaxon⁴,
Mamadjonova Marvarid⁵**

Authors' affiliations:

¹ National Pedagogical University of Uzbekistan named after Nizami, 27 Bunyodkor Street, Tashkent, 100185, Uzbekistan. Email: liekaterina86@gmail.com
ORCID: <https://orcid.org/0000-0002-6624-9277>

² Tashkent State University of Oriental Studies, 20 Amir Temur Street, Tashkent, 100060, Uzbekistan. Email: izzetova@bk.ru
ORCID: <https://orcid.org/0000-0002-4989-2455>

³ National University of Science and Technology MISIS, 20 Lenin Ave., Moscow, 119049, Russia. Email: naumenko06@mail.ru
ORCID: <https://orcid.org/0000-0002-5789-8806>

⁴ National Pedagogical University of Uzbekistan named after Nizami, 27 Bunyodkor Street, Tashkent, 100185, Uzbekistan. Email: mtillavoldiyeva@mail.ru

⁵ Yangi Asr University, 5 Qatortol Street, Tashkent, 100113, Uzbekistan.
Email: ms.marvarid@inbox.ru
OrcID: <https://orcid.org/0000-0003-2708-1069>

ABSTRACT

This study analyzes religious and secular truths, as well as the transformation of truth in the context of AI-generated information and its assessment as either true or unreliable. While humans recognize truth as a value, the question of the significance of truth for AI remains open and requires philosophical analysis.

This article presents a comparative analysis of several AI programs, highlighting similarities and differences in their responses.

Key words: truth, artificial intelligence, false information, reliable information, private information, epistemic agent, digitalization, privacy

Introduction

The system of fundamental doctrines and beliefs known as religious truth helps explain the meaning of life, humanity's place in the universe, and the nature of the divine. Religious truth is attained through faith, personal experience, and revelation, unlike scientific truth, which is based on empirical data and observation. For a believer, religious truth is absolute, above doubt and speculation.

In Buddhism, there are four noble truths, the attainment of which is associated with inner enlightenment and liberation from illusions. In Christianity, Truth is identified with God: I am the Truth, I am God. Sufism, a branch of Islam, has a unique vision of Truth. Sufism is an ascetic and spiritual-mystical practice. Sufis understand Truth as God Himself; therefore, everything connected with the divine is Truth. Sufis set the goal of life as the knowledge of Truth (God), understanding this as the act of merging the Sufi's soul with God. Such an act is within the power of a person who has completed the stages of Sharia, Tariqah, Marifat, and Haqiqah, i.e., is physically and spiritually prepared. Remarkably, knowledge of Truth is possible through the heart.



Truth has been a subject of philosophical reflection since ancient times. The ancient Greeks believed that truth is born in debate. The Sophists preferred victory in debate to truth. Francis Bacon distinguished between secular and religious truth, a position that helps to defuse the tension between believers and non-believers. In the age of information technology and digitalization, as well as the rapid development of AI, the problem of Truth in the information it produces is relevant.

In a broad sense, the concept of truth refers to the correspondence of our knowledge to reality. Secular truth is a system of knowledge and principles based on reason, experience, observation, and the scientific method, without recourse to religious dogma or supernatural sources. Secular truth is constantly changing and pluralistic, unlike religious truth, which is often considered absolute, eternal, and claims exclusivity.

Artificial intelligence (AI) has become deeply ingrained in all aspects of human and social activity. AI processes vast amounts of information quickly, diversifiedly, and within a specific field of knowledge. AI has evolved from a simple information analyzer into not only a systematizer but also a producer of new information. The question of the reliability of this information naturally arises. What constitutes truth for AI, and are there any criteria for defining it?

Issue of truth as reliable knowledge has long intrigued philosophers and scientists. Today, this problem is presented in a new light due to the widespread and intensive use of artificial intelligence. Can the concept of truth be interpreted in the same way for humans and artificial intelligence? Is truth possible for artificial intelligence?

Understanding truth is represented in philosophy by a wide variety of concepts, but the authors of this study analyzed how most people—that is, non-experts—understand truth. They found that this understanding is not only multilayered but also individually structured. The article also examines the coherence of truth (Huber et al., 2026). The co-authors examined the relationship between certain emotions and false and true information, demonstrating different emotional dynamics (Kauk et al., 2025). The possibility of transforming myths into truth is examined in (Krahé, 2025), where the author concludes that the repetition of false information is remembered as truth. The role of truth in the political sphere of society (Sokolova, 2025) and how authorities should tell the truth are discussed by the next author.

Ground truth is the foundation of trustworthy AI information. Researchers have used 16 eXplainable Artificial Intelligence (XAI) methods to validate ground truth for AI (Miró-Nicolau et al., 2024). An innovative approach to assessing AI privacy and preserving sensitive information is proposed (Islam et al., 2025) as a privacy-preserving tool (Pilán et al., 2025).

The issue of information reliability arises from the use of LED technology (Avinadav and Shamir, 2025), which ensures the reliability of information. The possibility of creating lies based on facial microexpressions is being explored (De Marsico et al., 2024). The use of graph neural networks to detect fake information is being explored by a group of authors (Kontek et al., 2025). One of the key issues associated with AI is the ethical aspect of maintaining privacy, which can be addressed by the new text sanitization method INTACT (Pilán et al., 2025). A lie recognition system based on facial expression analysis has been developed as a human aid (De Marsico et al., 2024). AI is applied in various human and social activities, including education. The development of entrepreneurship in students based on two cognitive factors—global thinking and the perceived value of AI—is examined in a subsequent study (Duong, 2025).



Methods

This article utilizes dialectical, synergetic, and comparative methods. Researching types of truth, as well as analyzing their transformation through the use of AI in all spheres of society, requires a special approach and a thorough analysis of the subject matter of this research. The problem of truth has been considered in religious, philosophical, and scientific systems and therefore requires a comprehensive, objective, and systematic study. The article also compares AI applications and the changes occurring in the understanding of the essence of truth.

Results and discussion

Understanding truth, whether secular or scientific, differs greatly from understanding truth for AI. Since artificial intelligence lacks consciousness, values, or personal experience, its "truth" depends primarily on the quality of the data it is trained on and the algorithms it uses. This raises the question: is truth possible for artificial intelligence? Artificial intelligence operates with information. Comparing the concepts of knowledge and information, A.S. Karmin and G.G. Bernatsky argue that information cannot be considered false simply because it does not correspond to reality.

Truth is valued only by humans, and only humans determine the criteria for its determination. The results of artificial intelligence must be verified and interpreted within a broader ethical and social framework, without blindly trusting its "accuracy." The reliability of AI depends on the transparency of its algorithms and constant human supervision. Serious criticism of AI as an epistemic agent, that is, a source of knowledge, has challenged traditional notions of knowledge, truth, and cognitive activity. When analyzing this conversation, we must consider several important points that demonstrate why artificial intelligence cannot be considered a fully-fledged epistemic agent in the human sense.

Let's consider the problem of truth in philosophy. Truth is recognized as the goal of all knowledge and is recognized as a universally valid value. The main interpretations of truth in philosophy are correspondence, coherence, pragmatism, and constructivism. Correspondent truth represents the correspondence of our knowledge to reality, this is how Plato defines truth in the dialogue "Cratylus". For Aristotle, knowledge about reality, knowledge about the root causes, is true.

The coherence of truth is expressed in its consistency with other assertions of a given doctrine. It is also believed that coherence can be expressed in the willingness of the majority of representatives of a given science or theory to accept a given proposition. The fundamental condition of this view is consistency. A survey conducted by a group of researchers revealed that 54.7% of participants believed that truth equaled facts. Moreover, many demonstrated a pluralistic understanding: they were not limited to one theory, but combined two (most often, correspondence and authenticity). The authors also noted individual structure in their understanding of truth (Huber et al., 2026).

Scientific knowledge, or science, in the constructivist interpretation is viewed not as a reflection of objective reality, but as the result of "reflexive production," a product of social factors. Truth is a construct formed in the human mind.

Praxeologism demonstrates the relationship between the categories of "truth" and "usefulness" in the philosophy of technology. Both are used to document and evaluate necessary scientific, technological, and research activities. Truth and usefulness differ in that usefulness characterizes activity, while truth characterizes cognition. Truth, as a receptive characteristic of knowledge, is the condition for the possibility of usefulness as a projective characteristic of



knowledge, since the schema of activity is the inversion of the schema of cognition. For praxeologism, this condition is usefulness—not only truth, but also usefulness.

In its version of the truth-conditional conception of meaning, the correspondence theory of truth develops the view of M. Schlick and L. Wittgenstein that the method of verifying a proposition determines its meaning. The term "truth" pertains to epistemology and describes the goal of the cognitive process: when truth is achieved, that is, when the meaning of the desired proposition is determined, knowledge is complete. The procedure for determining meaning is a reflection that identifies and fixes in a rational concept the historically concrete (and thus the contingent), since by "sentence" we mean a grammatically correct sequence of words of a particular intersubjective language, a configuration of objects of reason, and "by conditions of truth" we mean the referent designated by this configuration. Since a "sentence" is a sequence of grammatically correct words of a particular intersubjective language, i.e. a particular configuration of objects of the understanding, and "under the conditions of truth" is the referent designated by this configuration, since the process of determining meaning becomes more complex than in any other case.

Reflection on the protocol sentence, which discovers and captures the historically specific (and therefore syntactically contingent) way in which the language of reason connects with sensory perception. When it comes to more complex situations and in real communicative practice in general, it is important to remember that linguistic sentences capture and convey both the sensory and the supersensory, as well as the syntactic rules of reason itself, as well as information about memory, fiction, and imagination. Language can also designate and express itself in metalinguistic constructions that are difficult to imagine. In the positive truth-conditional version of meaning, it is important that the correspondence theory of truth allows for the establishment of a semantic rule for the extremely complex designation relations realized by the intellect; it also allows for the establishment of perception and intellect as levels of representation realized by consciousness.

Language can also design and express itself in metalinguistic constructions that are difficult to imagine. In the positive truth-conditional version of meaning, it is important that the correspondence theory of truth allows for the establishment of a semantic rule for the extremely complex designation relations realized by the intellect; it also allows for the establishment of perception and intellect as levels of representation realized by consciousness. AI will learn from errors if the labeling is erroneous.

For AI, truth is what corresponds to the data. For example, if an AI predicts rain, and it actually begins, the forecast can be considered correct. However, AI predictions fail, indicating that its understanding of truth is limited. Ground Truth requires privacy preservation, and one way to achieve this goal is through the introduction of an additive version of case-based reasoning—AddCBR—which serves as an interpretable and model-consistent benchmark for evaluating privacy assessment methods for sensitive information (Islam et al., 2025).

Truth as coherence is also inherent to AI. Artificial intelligence can generate consistent internal responses, but these responses don't always correspond to objective reality. For example, a chatbot can create persuasive texts that appear logical but contain errors or "hallucinations." In many AI systems, truth is a probabilistic concept, not a binary "yes" or "no" answer. Each statement has a certain degree of truth. This means that the system can be 95% "confident" about one answer and 60% "confident" about another. AI has no intentions; it cannot intentionally lie, hence intentionality. When AI produces false information, it is not a lie; it could be the result of faulty data, bias in the training set, or flaws in the algorithm.

The modern information landscape is rife with fake information. On the one hand, how is fake information perceived by humans? On the other, how can it be countered using AI? Kelvin K. King found that modified disaster information spreads faster than unmodified information, regardless of its credibility. The more details, especially negative ones, the more in-demand it



becomes (King, 2025). This suggests that such information can be an effective means of manipulating people's consciousness and behavior. A team of authors also notes the negative connotations of fake information. Their study found that negative news increases negative emotions, while positive news has no effect, although positive emotions increase (Kauk et al., 2025). The significance of these studies could form the basis for understanding the dynamics of emotions and developing methods to combat manipulation.

Developing the idea of countering fake information, it's worth mentioning the following study, in which the authors presented the Face Truth Machine (FTM) system—an intelligent tool for automatic lie detection based on the analysis of facial microexpressions (De Marsico et al., 2024). The experiment was conducted using two datasets: Bag-of-Lies and Real-Life Trial Dataset, and, according to the authors, demonstrated high accuracy, especially in cross-testing conditions, where other methods often lose effectiveness. It is also noted that this is merely an auxiliary tool and cannot replace human expertise. Polish scientists propose an innovative approach to detecting fake news based on graph neural networks (GNNs) and multivariate content credibility assessment (Kątek et al., 2025). Unlike binary classification (fake/real), their method takes into account contextual, semantic, and structural relationships between text elements: author, headline, and content. The authors claim that the results of experiments on the SWAROG dataset showed a significant improvement in accuracy compared to baseline models, noting the high potential of GNNs for combating disinformation.

The functioning of AI is associated with certain problems and conditions.

The quality and completeness of the training data determine the truth of the AI. AI will repeat and amplify biases in the data. AI is good at understanding objective facts, but it lacks subjective truths that depend on personal experience and cultural context. Dangerous consequences can arise if artificial intelligence is not supervised by humans. In the case of deep neural networks, it is sometimes difficult to understand how the AI arrived at a particular conclusion. This makes it difficult to verify its accuracy.

Artificial intelligence lacks consciousness, intention, or understanding as can be explained by human language. Rather than understanding their meanings, it operates by manipulating symbols and patterns based on statistical regularities contained in data. This problem can be illustrated by John Searle's classic "Chinese Room" argument, which posits that a human in the room can correctly translate Chinese characters without understanding their meaning. AGI does the same, simulating intelligent behavior without actually possessing it.

The quality, volume, and representativeness of training data determine the volume and quality of AI. AI will not only replicate but also amplify biases in data containing biases (e.g., racial or gender). This leads to epistemic injustice when AI creates knowledge that reflects and perpetuates social inequality.

Inaccuracy or falsity in AI-generated information can be caused by insufficient or outdated data. The problem of opacity, known as the "black box," often makes it impossible to understand how AI arrived at a particular conclusion, especially in complex neural networks operating on the principle of deep learning. This opacity makes its judgments impossible to critically evaluate and assess their reliability, undermining its credibility as a source of knowledge.

AI cannot be held accountable for its "beliefs" or "conclusions," unlike humans. Humans bear moral and ethical responsibility for disseminating inaccurate information. AI simply produces results based on its algorithm without understanding the consequences.

According to some studies, users' ability to think critically and independently evaluate information is impaired by overreliance on AI tools. People stop asking questions and verifying information when AI offers ready-made answers, leading to epistemic passivity.

When it comes to AI as an epistemic agent, this could lead to the creation of "cognitive castes" in which epistemic power and privilege are determined by access to high-quality data and



advanced AI systems. People who blindly trust algorithms will lose out to people who are skilled in working with AI and critically evaluating its results.

Artificial intelligence is widely used in social media to create customized information flows. This leads to the creation of "epistemic bubbles" and "echo chambers," where people receive only what they like, hindering constructive conversation and the formation of informed opinions.

Rethinking the Role of AI: A critical analysis emphasizes that AI is an epistemic technology that extends and modifies our cognitive capacities, but does not replace them with fully-fledged epistemic agents. While artificial intelligence is a powerful tool, its epistemic status is very different from that of humans. It functions as a mediator, transforming data into information for humans.

Ultimately, when discussing AI as an equal epistemic agent, fundamental differences between machine and human cognition are ignored. Instead of entrusting epistemic responsibility to artificial intelligence, we should focus on developing technologies that will help humans improve their cognitive and critical capacities, as well as on creating moral and legal frameworks that support the use of artificial intelligence as an epistemic tool.

	ChatGPT	Capilot	Deepseek	YandexGPT
What is truth?	Truth is one of the most fundamental concepts in philosophy, logic, and everyday life. Generally speaking, truth is the correspondence of knowledge or statements to reality.	In philosophy, truth is the correct reflection of reality in human consciousness, the correspondence of knowledge to its object. It is viewed as a fundamental value of culture and science, uniting knowledge, morality, and practice.	The short answer: truth is the correspondence of our knowledge or statements to objective reality. This is the classic definition, dating back to Aristotle: "Truth is the correspondence of a thing to the intellect."	Truth is reliable knowledge that adequately reflects objective reality and corresponds to the properties of the object of knowledge. This is a fundamental concept in philosophy and epistemology, around which numerous debates have been and continue to be conducted.
The concept of truth for AI.	For artificial intelligence, truth is not a "sense of reality" or personal belief. AI doesn't know the world directly, like humans. It works with data, probabilities, and models.	For artificial intelligence, the concept of truth is interpreted not as a metaphysical category, but as the correspondence of the model's output to specified criteria of correctness, facts,	The concept of truth in AI depends on which aspect of its operation we're talking about. For modern neural networks (like me), it's not truth, but "plausibility."	In the context of AI, "truth" is a formal concept that determines the correspondence of a system's conclusions to: *the actual state of affairs (if verifiable); *given axioms and rules;



		or the user's goals.	I'm not conscious and can't "believe" or "know." For me, truth is statistical consistency with patterns in the training data.	*statistical patterns in the data.
How will the disappearance of the Moon affect the Earth?	If the Moon were to suddenly disappear, the consequences for Earth would be very serious—though the planet itself would not be immediately destroyed. Here are the main effects. Major consequences: • 1. Weak tides, • 2. Climate instability, • Ecosystem changes, • Darker nights, • 3. Long-term climate instability.	The disappearance of the Moon would have a catastrophic impact on Earth—not only physically, but also culturally. Physical Consequences Tides; 1. Stability of the Earth's axis; 2. Length of day; 3. Evolution of life. Cultural Consequences • The Moon is an important symbol in mythology, religion, and art. • Its disappearance would be a cultural shock: the loss of calendars, holidays, and symbols. • • Humanity would lose one of its main landmarks in the night sky.	The disappearance of the Moon would be a catastrophe of planetary proportions. Without the satellite, Earth would change radically in a relatively short (by geological standards) time. Here are the main consequences: 1. The Earth's climate and rotation would become chaotic; 2. Days would become shorter (approximately 3-4 times shorter); 3. The oceans would cease to move normally; 4. The Earth would begin to "wobble" with a range of tens of degrees;	• The disappearance of the Moon would lead to massive changes on Earth, affecting the climate, oceans, biosphere, and even the planet's rotation. These changes could make Earth significantly less habitable. • 1. Tides; • 2. The tilt of the Earth's axis and climate; • 3. The rotation of the Earth and the length of the day; • 4. Biological rhythms; • 5. Nighttime illumination; • 6. Protection from meteorites.



			5. Nights would become completely dark, and meteorites would become frequent; 6. Life would likely not survive (except for the most primitive forms).	
Can AI lie?	Yes, AI can produce false information – but the question is whether this should be considered a “lie.”	AI can create false statements, but this is not the same as human lying.	The short answer is no, in the human sense it cannot, but it can imitate lies or be a tool for deception.	Strictly speaking, AI doesn't lie in the human sense—it has no consciousness, no intention to deceive, or no personal gain. But it can produce false information—either intentionally (as directed) or accidentally.
Human lies vs AI errors.	This idea can be expressed almost as an aphorism: "A human lie is a mistake of conscience. An AI mistake is a lie without intent."	Human Lies • Intent: A person deliberately distorts the truth for profit, protection, or manipulation. AI Errors • Intent: None – AI has no will or goals.	Comparing human lies with AI errors is key to understanding why AI isn't a "liar" in the human sense. The difference is fundamental and runs along the lines of intention and belief.	Human Lies Definition: A deliberate distortion of the truth with the intent to mislead another person. AI Errors Definition: An unintentional presentation of incorrect information due to limitations of a model, data, or algorithm.

In this study, we analyzed four AI apps: **ChatGPT, Capilot, Deepseek, and YandexGPT**. An analysis of the responses to the questions posed reveals similarities, but also some differences. To the question "What is truth?", all apps provided similar answers, focusing on the key concepts of reality and consciousness. The answers focused on how truth is viewed in philosophy. To the question "The concept of truth for AI," the answers were also similar,



namely: data, statistics, axioms, facts. In responses to the question "How would the disappearance of the Moon impact Earth?", the main anticipated consequences were outlined. **Capilot** and **Deepseek** characterized the consequences as catastrophic, **ChatGPT** as serious, and **YandexGPT** as widespread. Only **Capilot** also noted cultural implications. In response to the question "Can AI lie?", **ChatGPT** and **Capilot** answered yes, but it is necessary to clarify what exactly is meant by "lying." **Deepseek** and **YandexGPT** answered no, since lying is a human product, and in their responses to the final question, all apps acknowledge the intentionality of human lies, which is not inherent in AI.

Conclusion

Truth is important and sought after, possessing value only for humans. Truth has been the goal of knowledge and philosophical reflection. Information produced by AI may be reliable or unreliable, but could it be intentional? Could AI deliberately create disinformation? A simple query to find an article on a specific topic published in a peer-reviewed journal yielded the following response: author, article title, and journal title. An attempt to find the article was unsuccessful, as no such author, article, or journal exists. Humans are clearly held responsible for using AI to create fake information. In the case of AI-based disinformation, who is held responsible?

Unexplored Areas: Is it possible to develop truth criteria for AI and requirements for the information it produces?

Is AI capable of "conscious" production? Fabricated photos, videos, and voice messages are created and used to blackmail and tarnish people's reputations. In such activities, AI is a tool in the hands of humans.

What is the role of critical thinking as an epistemic virtue in digital education?

How is truth formed in collective practices, online communities, and social media?

The aesthetic dimension of truth—truth in art and culture in the age of algorithms (for example, generative art).

Is there a connection between truth and action, utility, and ethics in a technological society?

References:

- Avinadav, T., Shamir, N., 2025. Truth, trust, and trade-offs: When blockchain in supply chains backfires. *Eur. J. Oper. Res.* 326, 467–480. <https://doi.org/10.1016/j.ejor.2025.05.011>
- De Marsico, M., Dionisi, G., Stanco, D.F.P., 2024. FTM: The Face Truth Machine—Hand-crafted features from micro-expressions to support lie detection. *Comput. Vis. Image Underst.* 249, 104188. <https://doi.org/10.1016/j.cviu.2024.104188>
- Duong, C.D., 2025. Human vs AI: What constitutes true intelligence for international entrepreneurship. *Res. Glob.* 11, 100320. <https://doi.org/10.1016/j.resglo.2025.100320>
- Huber, L.S., Künstle, D.-E., Reuter, K., 2026. Tracing truth through conceptual scaling. *Cognition* 266, 106321. <https://doi.org/10.1016/j.cognition.2025.106321>
- Islam, M.R., Weber, R.O., Ahmed, M.U., Begum, S., 2025. Privacy-preserving ground-truth data for evaluating additive feature attribution in regression models with additive CBR and CQV. *Knowl.-Based Syst.* 330, 114599. <https://doi.org/10.1016/j.knosys.2025.114599>



- Kątek, G., Kozik, R., Pawlicka, A., Pawlicki, M., Choraś, M., 2025. In depth analysis for securing the truth: Addressing the fake news challenge with graph neural networks. *Neurocomputing* 654, 131327. <https://doi.org/10.1016/j.neucom.2025.131327>
- Kauk, J., Humprecht, E., Kreysa, H., Schweinberger, S.R., 2025. Large-scale analysis of online social data on the long-term sentiment and content dynamics of online (mis)information. *Comput. Hum. Behav.* 165, 108546. <https://doi.org/10.1016/j.chb.2024.108546>
- King, K.K., 2025. How manipulating information affects information diffusion during disasters: The effects of modifying falsehoods versus corrections. *Decis. Support Syst.* 197, 114523. <https://doi.org/10.1016/j.dss.2025.114523>
- Krahé, B., 2025. Beware the truth effect: Why efforts to debunk rape myths may backfire. *J. Crim. Justice* 98, 102409. <https://doi.org/10.1016/j.jcrimjus.2025.102409>
- Miró-Nicolau, M., Jaume-i-Capó, A., Moyà-Alcover, G., 2024. Assessing fidelity in XAI post-hoc techniques: A comparative study with ground truth explanations datasets. *Artif. Intell.* 335, 104179. <https://doi.org/10.1016/j.artint.2024.104179>
- Pilán, I., Manzanares-Salor, B., Sánchez, D., Lison, P., 2025. Truthful text sanitization guided by inference attacks. *Appl. Soft Comput.* 185, 114013. <https://doi.org/10.1016/j.asoc.2025.114013>
- Sokolova, T., 2025. To speak truth as, with, and through power: Co-producing knowledge politics of a just transition with Swedish citizens and trade unions. *Environ. Sci. Policy* 171, 104166. <https://doi.org/10.1016/j.envsci.2025.104166>