reflektif
SOSYAL BİLİMLER DERGİSİ
JOURNAL OF SOCIAL SCIENCES

Reflektif Journal of Social Sciences, 2026, Vol. 7(1)

2026, Vol. 7(1), 109-121

DOI: 10.47613/reflektif.2026.265
Article type: Research Article

Received: 14.01.2026
Accepted: 08.02.2026
Published Online: 18.03.2026

Mahmut Özer*

The Epistemic Crisis of Artificial Intelligence as an Agent Without Responsibility

Sorumluluğu Olmayan Bir Fail Olarak Yapay Zekânın Epistemik Krizi

Abstract

This study argues that the increasing use of generative artificial intelligence and large language models (LLMs) gives rise not only to technical, ethical, or governance-related problems, but also points to a deeper and more structural epistemic crisis. Although the risks initially attributed to artificial intelligence have largely been discussed under headings such as data security, bias, and hallucinations, this study emphasizes that these issues are not isolated malfunctions but structural consequences arising from the way AI produces knowledge. While LLMs can generate fluent and persuasive outputs by mimicking human cognitive processes, these outputs emerge independently of core human cognitive dimensions such as meaning, intention, causality, and responsibility. Accordingly, the study underlines that artificial intelligence is built upon an architecture that imitates human cognition without assuming the epistemic and moral burdens intrinsic to those processes. This condition indicates that, despite functioning as an agent capable of acting successfully, artificial intelligence should not be regarded as an epistemic subject. From this perspective, the reproduction of biases, the generation of fabricated content, and a high degree of compliance with morally problematic decisions should be understood as natural and inevitable outcomes of this architecture. Thus, the core problem lies not in models' accuracy rates or performance levels, but in the processes through which their responses are generated. The study further discusses how generative artificial intelligence is transforming the human–machine relationship and examines the effects of the gradual transfer of cognitive load to machines on critical thinking, memory, patience, and independent problem-solving abilities. Existing findings suggest that rather than supporting humans, these tools may foster a dependency that leads to passivity in human judgment formation. In conclusion, the article maintains that the crisis surrounding artificial intelligence is not primarily rooted in technical inadequacy, but in an epistemic rupture caused by substituting humans with systems that bear no responsibility, and it argues that the solution lies not in more advanced models, but in rethinking the boundaries that safeguard human judgment, responsibility, and decision-making processes.

Öz

Bu çalışma, üretken yapay zekâ ve büyük dil modellerinin (LLM) giderek artan kullanımının yalnızca teknik, etik ya da yönetişimle ilgili sorunlar doğurmadığını; daha derin ve yapısal bir epistemik krize işaret ettiğini ileri sürmektedir. Başlangıçta yapay zekâya atfedilen riskler çoğunlukla veri güvenliği, yanlılık ve halüsinasyon gibi başlıklarda ele alınmış olsa da, bu sorunların tekil arızalar değil, yapay zekânın bilgi üretme biçiminden kaynaklanan yapısal sonuçlar olduğu vurgulanmaktadır. LLM'ler insanın bilişsel süreçlerini taklit ederek akıcı ve ikna edici çıktılar üretebilse de bu çıktılar, anlam, niyet, nedensellik ve sorumluluk gibi insani bilişsel süreçlerden bağımsız olarak ortaya çıkmaktadır. Bu nedenle çalışmada, yapay zekânın insan bilişsel süreçlerini taklit eden ancak bu süreçlerin taşıdığı epistemik ve ahlaki yükleri üstlenmeyen bir mimariye sahip olduğunun altı çizilmektedir. Bu durum, yapay zekânın başarılı biçimde eyleyebilen bir fail gibi işlev görmesine rağmen, epistemik bir özne olarak değerlendirilememesi gerektiğine işaret etmektedir. Bu nedenle yanlılıkların yeniden üretilmesi, uydurma içeriklerin ortaya çıkması ve ahlaki açıdan sorunlu kararlara yüksek uyum gösterilmesi, bu mimarinin doğal ve kaçınılmaz çıktıları olarak ele alınmalıdır. Dolayısıyla sorun, modellerin doğruluk oranlarından ya da performans seviyelerinden ziyade, cevapların hangi süreçlerle üretildiğiyle ilişkilidir. Çalışma ayrıca, üretken yapay zekânın insan–makine ilişkisini nasıl dönüştürdüğünü ve bilişsel yükün giderek makineye devredilmesinin eleştirel düşünme, hafıza, sabır ve bağımsız problem çözme becerileri üzerindeki etkilerini de bu bağlamda tartışmaktadır. Mevcut bulgular, bu araçların insanı desteklemekten ziyade, insanın yargı üretme süreçlerinde pasifleşmesine yol açabilecek bir bağımlılık ilişkisi doğurduğunu göstermektedir. Sonuç olarak makale, yapay zekâ bağlamında yaşanan krizin esasen teknik bir yetersizlikten değil, sorumluluğu olmayan sistemlerin insan yerine ikame edilmesinden kaynaklanan epistemik bir kırılma olduğunu savunmakta; çözümün daha gelişmiş modellerden ziyade insan yargısını, sorumluluğunu ve karar süreçlerini koruyacak sınırların yeniden düşünülmesinde yattığını ileri sürmektedir.

* Turkish Grand National Assembly, mahmutozer2002@yahoo.com, ORCID: 0000-0001-8722-8670.

Keywords
Large language models, epistemology, hallucination, confabulation, responsibility

Anahtar Kelimeler
Büyük dil modelleri, epistemoloji, halüsinasyon, konfabülasyon, sorumluluk

Introduction

As artificial intelligence technologies and large language models (LLMs) become widely used across many domains, the structural problems they embody have also begun to surface gradually (Suleyman, 2023; Perc et al., 2019; Perc and Özer, 2025). In this context, the risks initially associated with artificial intelligence were predominantly assessed in relation to data security, bias, and hallucination-related issues. In particular, the reproduction of social biases based on religion, culture, ethnicity, race, and gender by LLMs poses serious risks across a wide range of fields such as education, healthcare, and justice (Angwin et al., 2016; İlikhan et al., 2024; 2025; Lum and Isaac, 2016; Özer, 2024a; 2024b; Özer et al., 2024a; Suna and Özer, 2025; Tanberkan et al., 2024). Because biases linked to social power relations are reproduced through artificial intelligence technologies, the risk of entrenching social inequalities becomes far more pronounced (Özer et al., 2024b; Ulnicane and Aden, 2023). To mitigate these risks, a participatory governance approach that seeks to maximize human involvement in the development of artificial intelligence platforms is often proposed (Acemoglu et al., 2023; İlikhan et al., 2025; Özer and Perc, 2024).

On the other hand, hallucination behavior in artificial intelligence constitutes a major area of concern (Özer, 2025c). LLMs are capable of generating content that appears plausible but is factually incorrect. Initially visible through references to non-existent articles in scientific publications, this problem has been shown to be far deeper and more widespread (Ji et al., 2023). LLMs do not naturally stop themselves; once detached from a genuine epistemic grounding, they can produce content that does not correspond to reality (Sui et al., 2024). Given that LLMs are now widely used across many domains, particularly for text generation, this behavior of artificial intelligence is therefore rightly regarded as a serious risk.

Moreover, while the human–machine relationship had historically functioned in a complementary manner—enhancing human capabilities prior to the development of advanced artificial intelligence—an increasing body of research now seeks to determine how the balance of this relationship is evolving with the rise of AI technologies (Ahmad et al., 2023; Gerlich, 2025; Kosmyna et al., 2025; Özer et al., 2025; Özer and Perc, 2025; Stadler et al., 2024). In particular, the growing shift of cognitive load from humans to machines—that is, the increasing dependence on machines—has been accompanied by a rising number of descriptive and experimental findings indicating a weakening of critical thinking skills and memory.

These findings effectively constitute an early warning about the dangers we are likely to face (Kosmyna et al., 2025; Perc and Özer, 2025; Stadler et al., 2024).

In this context, studies examining the effects of the human–machine relationship typically employ three groups. The first group consists of participants who use no external tools in the designed experiment. The second group corresponds to participants who use more conventional tools, such as search engines, whose features do not substantially take over human cognitive load but rather provide support to human cognition. In other words, the tools used in the second group are not capable of fully assuming human cognitive tasks. In the third group, however, tools such as generative artificial intelligence, which can largely or entirely take over human cognitive load, are employed. Studies conducted within this framework to date demonstrate that as individuals use tools that allow them to offload cognitive tasks, they increasingly transfer these cognitive loads to external systems. As a result, the level of human involvement in learning and decision-making processes declines, leading to reduced brain activity and a weakening of critical thinking skills and memory (Kosmyna et al., 2025; Stadler et al., 2024). Moreover, individuals' confidence and motivation to solve problems independently also diminish over time (Ahmad et al., 2023). If this relationship, which is increasingly turning into a form of dependency, continues to develop along this trajectory, machines will continue to advance, while humans will gradually lose the very capacities that make them human.

In fact, the issues partially discussed above point beyond a focus on isolated problems and indicate the need to question the epistemic plane on which artificial intelligence operates. In this respect, the epistemic status of artificial intelligence remains uncertain (Hauswald, 2025). Within this framework, artificial intelligence is generally regarded not as an agent that acts with values and thus possesses epistemic agency, but rather as an element that contributes to the flow of information (Simbolon et al., 2025). Approaching the issue from a different perspective, Özer (2025a) examines the relationship between artificial intelligence and truth through the concept of nafs al-amr, which occupies a central place in Islamic thought, and emphasizes that AI's relation to truth is deeply problematic, giving rise to an epistemic and ontological crisis. LLMs may construct correct sentences, generate accurate information, and produce outputs that are difficult to distinguish from human ones; however, this correctness does not stem from a conscious orientation toward nafs al-amr (the epistemic ground). Artificial intelligence does not know the truth, does not orient itself toward it, and therefore does not enter into a relationship with it; yet despite this, it produces outputs that reflect a form of statistical convergence based solely on textual patterns.

Therefore, the problem arises from the fact that, unlike humans, LLMs do not generate content through meaning, but instead produce it through formal (statistical) structures (Floridi, 2023). Most of the observable problems stem from this structural difference. In this context, a recent and highly comprehensive study similarly addresses the issue on epistemic

grounds (Quattrociocchi et al., 2026). The study argues that the core problem lies far deeper than accuracy rates, emphasizing that the issue is not whether LLMs make errors, but rather the processes through which they generate responses, and that these processes fundamentally differ from those of human epistemology.

A thorough investigation of the relationship between artificial intelligence and epistemology—a field that is still relatively new—will not only make it possible to assess the costs of the current trajectory of the human–machine relationship, but will also contribute to defining the framework within which artificial intelligence should be used. Accordingly, the perspective presented in this study expands existing evaluations in this area, examines the problems generated by artificial intelligence on an epistemic ground, addresses how each issue is situated within this plane, and offers warnings against the emerging epistemic crisis.

Reflections of the Crisis

The first problem encountered in the outputs generated by LLMs is bias. LLMs learn from vast amounts of real-world data without questioning its validity and produce content through algorithms that use this data as memory (Özer et al., 2024a). Consequently, because LLMs are unable to interrogate either the data or the algorithms in terms of their accuracy or the extent to which they represent or reflect reality, they generate content that is effectively left to the mercy of both data and algorithm. As a result, inequalities reflecting social power structures—embedded within the data itself—can be reproduced (O'Neil, 2016). In this context, for example, recidivism prediction algorithms widely used in the United States have been shown to misclassify Black individuals at nearly twice the rate of white offenders, thereby producing biased decisions against Black individuals (Angwin et al., 2016). Moreover, this bias leads the software to classify white offenders as having a lower likelihood of reoffending, even when they do in fact reoffend.

Similarly, in the United States and Europe, artificial intelligence algorithms are widely used in security software—particularly for identifying potential crime hotspots and suspects—and these algorithms are largely trained on crime data obtained from police departments. The problem emerges precisely at this point. Because groups that are disadvantaged in terms of race, ethnicity, and socioeconomic status (SES) are initially more likely to appear more frequently in the data on which these algorithms are trained, the algorithms tend to select areas with higher concentrations of these groups when determining which regions to monitor. For example, Lum and Isaac (2016), who examined the outputs of a commonly used algorithm designed to identify areas with high levels of drug use in a U.S. state, found that despite the existence of other regions in the state with similarly high drug-use rates, the algorithm predominantly flagged neighborhoods inhabited by non-white populations and individuals with low socioeconomic status. As a result, areas where individuals initially recorded as offenders reside are subjected to more frequent surveillance, the likelihood of detection increases, recidivism

algorithms predict harsher penalties, and ultimately a negative feedback loop—resembling a self-fulfilling prophecy—is continuously reinforced through these algorithms (Özer, 2025b).

On the other hand, bias does not arise solely from biases embedded in data; it is also well known that the assumptions and priorities adopted during algorithm design can themselves generate bias. For instance, studies have shown that algorithms which use healthcare expenditures as a proxy for identifying patients in need of advanced care disproportionately recommend such care to white patients. When corrections were introduced into the algorithm, the rate of additional healthcare support provided to Black patients increased from 17.7% to 46.5% (Obermeyer et al., 2019). The core problem here is that socioeconomically disadvantaged groups often already lack adequate access to healthcare; as a result, their lower healthcare expenditures lead algorithms to exclude them from the category of patients deemed eligible for advanced care. In this way, the algorithm largely overlooks the needs of individuals who receive fewer services (Mittermaier, Raza, & Kvedar, 2023).

In short, artificial intelligence does not recognize the biases arising from the prioritizations made during both data collection and algorithm design, and instead reproduces these biases. Moreover, this is not a matter of AI's inadequacy; rather, it is something it is inherently incapable of recognizing. Consequently, AI systems neither detect this faulty behavior nor bear responsibility for it. Given that the architecture of LLMs is designed along these lines, the content they generate and the decisions they produce are not required to enter into a necessary relationship with truth (Özer, 2025a).

Precisely at this point, it becomes necessary to take a closer, more in-depth look at AI's so-called hallucinatory behavior As is well known, hallucination in artificial intelligence is commonly defined as the model's production of persuasive yet fabricated content after departing from the user's input or the preceding context. In a previous article, we conceptually interrogated the phenomenon of misinformation generation in generative AI systems—often labeled "hallucination"—and argued that this term is inadequate for explaining how large language models operate (Özer, 2024c). In clinical terms, hallucination refers to a sensory perception that arises without an external stimulus and is typically associated with conditions such as schizophrenia, bipolar disorder, and Parkinson's disease (Tamminga, 2009). LLMs, however, do not possess sensory perception, conscious experience, or subjective awareness (Berberette et al., 2024; Smith et al., 2023). For this reason, calling the unrealistic outputs produced by AI "hallucinations" carries the risk of attributing human-like perceptual and cognitive processes to these systems. Yet this behavior is triggered by prompts and emerges in ways that are tied to the data on which the model was trained (Østergaard & Nielbo, 2023).

For this reason, we argued that the term hallucination is not appropriate for describing this phenomenon, since artificial intelligence does not "see" something that is not there but rather fabricates it. Instead, we proposed that this behavior can be more accurately explained by confabulation, a well-defined concept in psychiatry referring to the production of narrative details that are incorrect yet not recognized as false. Confabulation denotes the filling of narrative

gaps on the basis of existing knowledge, experience, and contextual cues, without awareness of the inaccuracy. In LLMs, similarly, gaps, contradictions, and outdated information in the training data, when combined with the model's probabilistic structure, lead to the generation of erroneous yet seemingly coherent content (Smith et al., 2023). Recent studies also indicate that much of the information produced by LLMs can be interpreted within the framework of confabulation (Berberette et al., 2024; Sui et al., 2024). In this context, the problem is not that the model perceives something unreal, but that it reconstructs existing information incorrectly. The fact that fabricated outputs produced by LLMs tend to occur more frequently in texts characterized by narrative structure further supports this argument (Sui et al., 2024).

Therefore, the production of non-existent content by artificial intelligence—commonly referred to as hallucination—should not be seen as an exception, but as an outward manifestation of a structural property. In other words, what is labeled as hallucination is not a simple technical error or a temporary software flaw. Rather, this behavior offers an opportunity to better understand the structural limitations of these systems. It reveals that LLMs operate in relation to language rather than the world. The system fabricates not because it fails to reference reality, but because it fundamentally lacks the very concept of reference from the outset—and because it does not know when to stop. Sometimes it coincides with truth, sometimes it does not; however, it has no internal mechanism for distinguishing between true and false. An answer that appears correct is merely a statistically plausible continuation of text.

In this context, the findings of another study pointing to an epistemic crisis indicate that while immoral requests may give rise to hesitation—even momentarily—in humans, this possibility effectively disappears in machines (Köbis et al., 2025). According to the study's results, even when human agents are given instructions involving outright cheating, the rate of compliance ranges between 25% and 40%, whereas in machines this rate is far higher—between 60% and 90%, depending on the model used. In other words, machines exhibit a remarkably high capacity to comply with deceit. Even when safeguards against such moral corruption are incorporated into algorithms, the authors note that machines' propensity to comply with unethical decisions remains high. In short, the irresponsibility of artificial intelligence functions in a way that increases compliance with immoral decisions. Put differently, because artificial intelligence does not have to confront the costs generated by its decisions, the likelihood of its "bending toward wrongdoing" is extremely high. For this reason, the authors suggest that measures to prevent unethical machine agency should be directed primarily at human principals rather than machine agents themselves.

The Background of the Crisis

When examining the background of the crisis, we must first consider how the data used by artificial intelligence are obtained, and then how these data are evaluated. As is well known,

the process of data acquisition begins with measurement, which requires the prioritization of the domain being measured and, subsequently, the collection of data through measurement procedures aligned with those priorities. In this process, proxy features capable of representing the target domain are determined in advance (Erdi, 2020). Measurement conducted in this manner provides only an approximation of the domain in question, since it is impossible to measure the domain in its entirety. For this reason, predefined proxy features are used to approach the domain, and their respective weights are assigned during this process. In short, no data collection process fully captures a domain; it merely achieves a form of convergence toward it. The most critical problem here is that, over time, what cannot be measured—that is, what cannot be assigned proxy features—comes to be treated as insignificant. Yet features that are excluded as immeasurable, and thus dismissed as uncertainty, complexity, or ambiguity, are in fact integral components of the domain and inevitably exert influence on what is being measured under different conditions. Therefore, even when mathematical accuracy is achieved, the cost of omission (error) can be exceedingly high (O'Neil, 2016; Özer, 2025b).

Working with data that substitute for reality has an important consequence for LLMs: responses to prompts tend to lack depth in most cases. For example, in a study comparing two groups asked to research a topic and develop recommendations—one using only conventional search engines and the other using LLMs—it was shown that the group using LLMs produced a significantly smaller number of valid arguments than the group relying on traditional search engines (Stadler et al., 2024). A similar finding emerged in another study comparing three groups: participants using no tools, those using conventional search engines, and those using LLMs (Kosmyna et al., 2025). While essays produced by participants in the brain-only group exhibited diverse and original approaches to each topic, the texts written by participants in the LLM group were statistically more homogeneous and displayed significantly less variation across topics.

In other words, from the human perspective, deeper thinking leads to greater validity and diversity in the arguments produced; independently of this, however, a separate problem emerges with respect to LLMs: their inability to grasp the diversity that exists in the real world, resulting in standardized and homogenized responses (world). Put differently, the limitations inherent in the measurement and data production processes discussed above, as well as in the algorithms themselves, are directly reflected in the quality of the outputs generated and in their capacity to encompass real-world complexity.

On the other hand, there are structural differences between humans and LLMs that mimic human cognitive processes. Beyond the representational limitations of data discussed above, the distinct nature of cognitive processes in LLMs points to a separate and structural problem. As is well known, humans establish a causal relationship with their environment; in developing this relationship, they see, hear, and feel, and they make sense of the world through their bodies, emotions, and the bonds they form with others. When humans arrive

at a judgment, that judgment is shaped by cause-and-effect relations, experiences, memory, values, and a sense of responsibility. While this process may appear similar in LLMs, it is in fact fundamentally different (Quattrociocchi et al., 2026).

Moreover, for humans, knowing is always an activity that carries costs, entails risk, and generates responsibility, whereas LLMs occupy a position detached from responsibility and values. The relationship of LLMs to the world—and to reality—begins with data produced through a process marked by a form of certainty that avoids uncertainty and ambiguity, despite its problematic relation to reality. However, the nature of the data and the level of its representativeness are of no concern to LLMs. Precisely in this context, Innerarity (2024), in evaluating the relationship between politics and artificial intelligence, argues that humans cannot be replaced by AI in politics because humans are capable of understanding and assessing uncertainty and ambiguity. He emphasizes that the problem is epistemic rather than technical, and that it is the architecture of artificial intelligence itself that makes such replacement impossible. In other words, while politics is a human activity that requires decision-making under conditions of uncertainty, ambiguity, and contingency, LLMs are fundamentally oriented toward certainty and closed to uncertainty, making it impossible for them to grasp context, meaning, and the plurality of interpretations inherent in situations. Accordingly, while causality that encompasses uncertainty, ambiguity, and complexity plays a central role as a human characteristic, what determines processes in LLMs is statistical correlation (Quattrociocchi et al., 2026).

For this reason, whether a statement or decision is true or false has no counterpart within the epistemology of LLMs; instead, LLMs are concerned with which expressions occur more frequently in similar contexts—that is, which are more probable. This is because the model lacks an epistemic loop that can verify its outputs by reference to the external world, recognize its own errors, and stop itself accordingly. Fluent and persuasive inaccuracies are therefore a natural consequence of this architecture. Perhaps AI's confabulatory behavior is best understood as a symptom of its lack of capacity to evaluate uncertain, complex, and ambiguous situations in the real world.

Quattrociocchi et al. (2026) address the manifestations of epistemological ruptures observed in LLMs through a more general framework, identifying seven fundamental epistemic fault lines: the perceptual bond humans establish with the world, the way situations are interpreted, experience and memory, motivation and values, causal reasoning, the monitoring of uncertainty, and finally the moral and social responsibility of judgment. Across each of these domains, a structural disjunction between humans and LLMs is demonstrated. Human judgment begins with perception, deepens through experience, gains meaning through causal explanations, weighs uncertainty, and at times refrains from issuing judgment. In LLMs, by contrast, each of these stages is replaced by formal counterparts that are weaker in substance. Perception is replaced by textual input, experience by embedding spaces, purpose and value by optimization functions, and metacognition by the obligation to produce an answer. While

humans can hesitate and stop, LLMs cannot; they merely continue by attempting to predict the next word.

As the authors themselves emphasize, LLMs fail to replace human judgment not because they are deficient in these domains, but because they were never designed to operate within them in the first place. To make sense of the problem arising from this difference, Quattrociocchi et al. (2026) propose the concept they term epistemia. In this context, epistemia refers to a condition in which linguistic plausibility is substituted for epistemic evaluation, giving rise to a feeling in the user of having arrived at a ready-made conclusion without passing through a process of producing a justified belief. Accordingly, the danger lies not merely in the production of false information, but in the systematic disabling of verification, hesitation, and responsibility. Most critically, whereas successful human action (being an agent) has always been grounded in intelligence, understanding, thinking, comprehension, consciousness, and intention, artificial intelligence marks the first instance in which this connection is severed: objects that lack consciousness, understanding, and responsibility are nevertheless introduced as entities that act successfully—as agents (Floridi, 2023). Yet artificial intelligence is not an epistemic subject.

Discussion

With the widespread adoption of artificial intelligence, unlike previous technological disruptions, machines now reduce human cognitive intensity in learning, optimization, and decision-making processes by virtue of their capacity to generate responses that mimic human cognitive processes. Findings from recent studies indicate that a decline in human cognitive engagement leads to weaknesses in critical thinking skills and in memory related to the outcomes of these processes. The primary risk here is that while the core promise of artificial intelligence is to reduce human cognitive load and thereby free up time for more productive and value-generating activities, the gradual transformation of the human–machine relationship into one of dependency undermines critical thinking capacities and ultimately deprives humans of precisely this promised benefit.

On the other hand, it is evident that the advantages provided by LLMs come with significant costs in terms of human cognitive load. In human–machine communication and interaction, users tend to remain relatively passive when engaging with AI applications, and as structural cognitive load is reduced, learning carries the risk of becoming superficial. The research findings discussed in this study also indicate a decline in the diversity of content produced by LLMs, suggesting that, in the long run, humans' critical thinking and independent problem-solving abilities may weaken. Moreover, although both intrinsic and extraneous cognitive loads decrease with the use of these tools, the surplus time generated in this process has not enabled users—compared to those employing traditional methods—to produce more

qualified justifications, for example in developing recommendations. In other words, although LLMs appear to relieve individuals from demanding knowledge-building and experiential processes that deepen and solidify learning, if this trajectory continues, they may negatively affect the development of these human capacities in the long term.

Therefore, approaches should be developed that ensure both the use of LLMs and the production of justifications that are more diverse and of higher quality than those generated through traditional methods. In particular, AI applications—especially within education systems—should be employed in ways that encourage cognitive engagement and active interaction with complex content. Put differently, generative AI tools should be evaluated not only in terms of the convenience they provide, but also in terms of how they are used and what kinds of cognitive processes they activate.

Perhaps the most overlooked issue is the gradual loss of the virtue of patience, which is one of the most important supports in human intellectual work, in confronting difficult problems, and in overcoming challenges—and which, in turn, fosters human development. Especially under conditions of uncertainty and ambiguity, humans can deepen their engagement with processes through patience and thereby overcome such situations. Yet the irresistible allure of having a tool at hand that can constantly produce answers and content on any subject confines individuals to the outputs generated by these tools without exercising patience. Without patience, however, deep thinking cannot take place, and genuine understanding cannot emerge. With the widespread use of artificial intelligence, what is being lost or weakened is not only critical thinking and memory, but also patience, deep reflection, and ultimately the capacity for comprehension. In other words, the increasing use of artificial intelligence—particularly in ways that substitute for human judgment—accelerates the erosion of the very qualities that make humans human. Consequently, it is not only artificial intelligence that faces an epistemic crisis; humans themselves are confronting one as well.

On the other hand, the arguments presented in this study also demonstrate that artificial intelligence is experiencing an epistemic crisis of its own. This crisis, however, is not one that artificial intelligence itself undergoes intrinsically. Rather, it arises from a misinterpretation of the nature of a machine that, for the first time in history, imitates human cognitive processes, and from its substitution for humans. Humans relate to the world on the plane of truth, with values and responsibilities shaping their actions, whereas artificial intelligence neither claims a relationship with truth nor bears responsibility. LLMs can produce outputs that superficially resemble human judgment; they can speak fluently, persuasively, and with apparent confidence. Yet this resemblance exists at the linguistic level, not at the epistemic one. Human judgment is the product of a multilayered process extending from perception and causality to memory and values, whereas the "judgment" of LLMs consists merely of selecting the next word within a high-dimensional probability space. Humans can distinguish between what they know and what they do not know, and can suspend judgment when necessary. LLMs, by contrast, cannot stop; they are compelled to produce an answer in every case.

Therefore, artificial intelligence is not grounded in human values, but in data and algorithms. Evaluating data through algorithms, LLMs operate not within a space of truth, but within a space of probabilities. Consequently, it is not possible for LLMs to concern themselves with whether their data or algorithms represent truth. Even the emergence of bias or confabulation behavior is identified as a problem not by artificial intelligence itself, but through human evaluation. Moreover, particularly in domains such as finance, the tendency of AI systems to produce morally objectionable decisions is linked to this very architecture—that is, to the fact that artificial intelligence can act as an agent despite lacking consciousness, responsibility, and intention (Floridi, 2023). This is where the core crisis emerges: outputs generated without understanding, intention, or responsibility begin to replace relationships grounded in truth, and an object without responsibility starts to behave as if it were an epistemic subject.

In short, the substitution of artificial intelligence for humans and the growing pervasiveness of this practice not only erode the qualities that make humans human, but also accelerate detachment from the real world by constructing a new world simulation through artificial intelligence—one that confines humans to a narrower representational reality. Within this new world, a novel ecosystem is being formed through the actions of an entity that is not an epistemic subject. In this sense, the existential crisis at stake is not one facing artificial intelligence, but humanity itself.

Quattrociocchi et al. (2026) are also unequivocal in their warnings in this regard. The problem will not be resolved by better training of models or by achieving higher accuracy rates, because the issue is not one of performance but of structure. No matter how persuasive they may be, these systems are not epistemic subjects. Positioning them as if they were erodes both individual judgment and institutional responsibility. What is therefore required is a rethinking of the epistemic, institutional, and moral boundaries that safeguard human judgment. In the age of generative artificial intelligence, the central issue is not the quality of answers, but who retains judgment and who the agent truly is. For it is possible to live with fluent yet unjustified answers; however, it is not possible to build a just, authentic, and humane world on the basis of knowledge that bears no responsibility.

References

Acemoğlu, D., Autor, D., & Johnson, S. (2023). Can we have pro-worker- AI? Choosing a path of machines in service of minds. *CEPR Policy Insight*, No.123, 1-12.

Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M. K., Irshad, M., Arraño-Muñoz, M. and Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10(311), doi: 10.1057/s41599-023-01787-8.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*, Online Edition.

Berberette, E., Hutchins, J., & Sadovnik A. (2024). Redefining "Hallucination"in LLMs:Toward a psychology-informed framework for mitigating misinformation. arXiv.2402–01769v1.

Erdi, P. (2020). Ranking: The Unwritten Rules of the Social Game We All Play. Oxford University Press.

Floridi, L. (2023). AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models. *Philosophy & Technology*, 36, 15.

Gerlich, M. (2025). AI Tools in Society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6. doi: 10.3390/ soc15010006

Hauswald, R. (2025). Artificial Epistemic Authorities. *Social Epistemology*, 39(6), 716–725.

Innerarity, D. (2024). The epistemic impossibility of an artificial intelligence takeover of democracy. *AI & Society*, 39, 1667–1671.

İlikhan, S., Özer, M., Tanberkan, H., & Bozkurt, V. (2024). How to mitigate the risks of deployment of artificial intelligence in medicine? *Turkish Journal of Medical Science*, 54(3), 483-492.

İlikhan, S., Özer, M., Perc, M., Tanberkan, H., & Ayhan, Y. (2025). Complementary use of artificial intelligence in healthcare. *Medical Journal of Western Black Sea*, 9(1), 7-17.

Ji, Z., Lee, N., & Frieske R. et al (2023). Survey of hallucination in natural language generation. *ACM Computing Survey*, 55, 1–38.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J. et al (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for an essay writing task. arXiv:2506.08872. doi: 10.48550/ arXiv.2506.08872

Köbis, N., Rahwan, Z., Rilla, R., Supriyatno, B. I., Bersch, C., Ajaj, T., Bonnefon, J. F., & Rahwan, I. (2025). Delegation to artificial intelligence can increase dishonest behavior. *Nature*, 646, 126-134.

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, 13(5), 14–19.

Mittermaier, M., Raza, M. M., & Kvedar, J. C. (2023). Bias in AI-based models for medical applications: Challenges and mitigation strategies. *NPJ Digital Medicine*, 6, 113.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366, 447–453.

O'Neil, C. (2016). Weapons of math destruction: How big data increase inequality and threatens democracy. Crown Books.

Quattrociocchi, W., Capraro, V., & Perc, M. (2026). Epistemological Fault Lines Between Human and Artificial Intelligence. arXiv: 2512.19466.

Østergaard, S. D., &Nielbo, K. L. (2023) False Responses from Artificial Intelligence Models Are Not Hallucinations. *Schizophr Bull*, 49, 1105–7.

Özer, M. (2024a). Potential benefits and risks of artificial intelligence in education. *Bartın University Journal of Faculty of Education*, 13(2), 232-244.

Özer, M. (2024b). Impact of ChatCPT on Scientific Writing. İnsan ve Toplum Dergisi, 14(3), 210-217.

Özer, M. (2024c). Is artificial intelligence hallucinating? *Turkish Journal of Psychiatry*, 35(4), 333-335.

Özer, M., & Perc, M. (2024). Human complementation must aid automation to mitigate the unemployment effects due to AI technologies in the labor market. *Reflektif Journal of Social Sciences*, 5(2), 503-514.

Özer, M., Perc, M., & Suna, H. E. (2024a). Artificial intelligence bias and the amplification of inequalities in the labor market. *Journal of Economy, Culture and Society*, 69, 159-168.

Özer, M., Perc, M., & Suna, H. E. (2024b). Participatory management can help AI ethics adhere to the social consensus. Istanbul University Journal of Sociology, 44(1), 221-238. doi: 10.26650/ SJ.2024.44.1.0001

Özer, M. (2025a). The epistemic crisis of artificial intelligence in the context of Nafs al-Amr. *Journal of Economy Culture and Society*, 72, 254-264.

Özer, M. (2025b). Can mathematical models be weapons of mass destruction? *Reflektif Journal of Social Sciences*, 6(1), 259-268.

Özer, M., & Perc, M. (2025). Creative alineation in art due to artificial intelligence. *Reflektif Journal of Social Sciences*, 6(3), 1105-1118.

Özer, M., Tanberkan, H., & Perc, M., (2025). Artificial intelligence threatens critical thinking in education systems. *Journal of Higher Education and Science*, 15(2), 157-164.

Perc, M., Özer, M., & Hojnik, J. (2019). Social and juristic challenges of artificial intelligence. *Palgrave Communications*, 5(61).

Perc, M., & Özer, M., (2025). Disappearing Minds in the Age of Artificial Intelligence. İnsan ve Toplum Dergisi, 15(3), 1-9.

Simbolon, L., Manugeren, M., & Barus, E. (2025). Does AI Know Things? An Epistemological Perspective on Artificial Intelligence. *Journal of English Language and Education*, 10(5), 1022-1028.

Smith, A. L., Greaves, F., Panch, T. (2023) Hallucination or Confabulation? Neuroanatomy as metaphor in Large Language Models. *PLOS Digit Health*, 2, e0000388.

Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 160.

Sui, P., Duede, E., & Wu, S. et al. (2024) Confabulation: The Surprising Value of Large Language Model Hallucinations. arxiV:2406.04175v2.

Suleyman, M. (2023). The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma. New York: Crown.

Suna, H. E., & Özer, M. (2025). The human complimentary usage of AI and ML for fair and unbiased educational assessments. *Chinese/ English Journal of Educational Measurement and Evaluation*, 6(1), 1-21.

Tamminga, C. A. (2009). Schizophrenia and other psychotic disorders: Introduction and overview. In: Sadock BJ, Sadock VA, Ruiz P, eds. Kaplan and Sadock's Comprehensive Textbook of Psychiatry. 9th edition. Philadelphia: Lippincott Williams and Wilkins; p. 1432.

Tanberkan, H., Özer, M., & Gelbal, S. (2024). Impact of artificial intelligence on assessment and evaluation approaches in education. *International Journal of Educational Studies and Policy*, 5(2), 139-152.

Ulnicane, I., & Aden, A. (2023). Power and politics in framing bias in artificial intelligence policy. *Review of Policy Research*, 40 (5), 665– 687.