



The Challenge of Agency in the Age of Artificial Intelligence: An Ontological Perspective

Zahra Khazaei 

Professor, Department of Philosophy, Faculty of Theology and Islamic Studies,
University of Qom, Qom, Iran, z-khazaei@qom.ac.ir

Abstract

Original Research



The rapid development of generative artificial intelligence systems, particularly large language models, has significantly reshaped contemporary philosophical discussions of agency. What was once primarily situated within philosophy of mind has now become a central concern in philosophy of technology, metaphysics, and AI ethics. These systems exhibit increasingly sophisticated capacities for natural language generation, inferential simulation, and goal-directed behavior, thereby blurring the conceptual distinction between intelligent behavior and genuine agency. Within analytic philosophy, agency is typically analyzed in terms of intentionality, reason-responsiveness, and rational control. This editorial proposes a graded framework of agency and introduces the concept of the quasi-agent as a philosophical hypothesis. A quasi-agent is an entity that exhibits stable and often complex patterns of agent-like behavior without possessing ontological subjecthood or intrinsic intentionality. To deepen this distinction, the discussion draws on Mullā Ṣadrā's Transcendent Theosophy, particularly the doctrines of the primacy of existence, the gradation of being (tashkīk al-wujūd), and knowledge by presence (ʿilm al-ḥuḍūrī). From this perspective, agency is not merely a behavioral property but a mode of being. The editorial concludes that artificial intelligence should neither be classified as a full agent nor reduced to a mere instrument, but rather understood as a quasi-agentive entity that challenges prevailing philosophical frameworks of action, responsibility, and subjectivity.

Keywords:

Agency, Artificial Intelligence, Quasi-Agent, Intentionality, Subjectivity, Analytic Philosophy, Mullā Ṣadrā, Transcendent Theosophy, Moral Responsibility.

Received: 2026/01/23; Received in revised form: 2026/01/25; Accepted: 2026/01/28; Published online: 2026/02/02

□ Khazaei, Z. (2026). The Challenge of Agency in the Age of Artificial Intelligence: An Ontological Perspective. *Journal of Philosophical Theological Research*, 28(1), 7-14. <https://doi.org/10.22091/jptr.2026.16299.3583>

□ Copyright © The Author

Introduction

The problem of agency in the age of artificial intelligence has moved beyond the boundaries of traditional philosophy of mind and now occupies a central position in philosophical discussions of technology, ethics, and ontology. Contemporary generative systems, especially large language models, are capable of producing coherent linguistic outputs, simulating inferential reasoning, and participating in complex socio-technical environments (Bubeck et al., 2023; Floridi, 2025). These developments have renewed a foundational question: under what conditions can an entity be regarded as an agent?

In analytic philosophy, agency is commonly associated with intentional action understood as behavior guided by reasons, beliefs, and desires. Influential accounts by Davidson (1980), Bratman (1987), Frankfurt (1971), and Fischer & Ravizza (1998) characterize agency in terms of rational control, normative responsiveness, and structured deliberation. On this view, an agent is not merely a system that produces outputs but an entity whose actions can be situated within a normative space of reasons.

However, the increasing ability of artificial systems to reproduce many of the external indicators associated with agency introduces a conceptual tension. Goal-directed behavior, adaptive responsiveness, and apparent reasoning can now be instantiated without clear commitment to the presence of intrinsic mental states. This tension has been widely discussed in contemporary AI philosophy and ethics (Searle, 1983; Dennett, 1987), and it raises a foundational question: whether agency should be grounded primarily in functional performance or in ontological subjectivity.

For the purposes of this editorial, three clarificatory distinctions are necessary.

First, agency is understood as a minimal capacity for reason-responsive, goal-directed activity that is potentially open to normative evaluation (Fischer & Ravizza, 1998). This definition remains neutral regarding whether such agency is genuine or merely attributed.

Second, subjectivity refers to first-person experiential perspective, whereas subjecthood denotes the ontological status of being a bearer of agency in a strong sense—namely an entity whose actions arise from an internally unified mode of existence rather than from external functional organization.

Third, intentionality, in the broadly analytic sense, denotes directedness toward objects or states of affairs. While some accounts interpret intentionality in computational or functional terms, others—most notably

Searle (1983)–argue that genuine intentionality requires intrinsic mental states irreducible to symbolic manipulation.

Against this background, artificial intelligence systems occupy an ambiguous position. Although they exhibit increasingly sophisticated behavioral performance, it remains unclear whether such performance entails ontological subjecthood or intrinsic intentionality. This ambiguity motivates a graded approach to agency.

Accordingly, this editorial introduces the concept of the quasi-agent. A quasi-agent is an entity that exhibits stable, adaptive, and often unpredictable patterns of agent-like behavior without possessing ontological subjectivity, intrinsic intentionality, or unified experiential presence (Parıdar, 2020). This is not intended as a metaphor but as a conceptual proposal for clarifying the intermediate status of advanced artificial systems.

From the perspective of Mullā Şadrā’s Transcendent Theosophy, agency is not a functional attribute but an expression of existence itself. The doctrines of the primacy of existence (aşālat al-wujūd) and the gradation of being (tashkīk al-wujūd) imply that cognitive and causal capacities vary according to existential intensity. Knowledge by presence (‘ilm al-ḥuḍūrī) replaces representational cognition with an ontological model in which knower and known are existentially unified (Mullā Şadrā, 1981).

From this standpoint, agency is inseparable from ontological depth rather than behavioral complexity. While this metaphysical vocabulary does not directly map onto computational systems, it offers a contrasting framework for reassessing the conditions of agency.

By integrating these traditions, this editorial argues that artificial intelligence should neither be regarded as a full-fledged agent nor reduced to a mere instrument. Instead, it occupies a quasi-agentive position that challenges existing conceptual frameworks of action, responsibility, and subjectivity.

Agency in Analytic Philosophy

Within analytic philosophy, agency has traditionally been understood through the interrelation of intentionality, rational explanation, and moral responsibility. Anscombe (1957) characterized intentional action as action under a description that makes sense to the agent. Davidson (1980) developed this idea by situating actions within networks of reasons that function as causes. Bratman (1987) extended this account through his planning theory of intention, emphasizing the temporal structure of agency.

Frankfurt (1971) introduced a hierarchical model of desires, linking agency to reflective endorsement of one's motivational structure. Fischer and Ravizza (1998) further refined the account by linking moral responsibility to guidance control and reasons-responsiveness.

Despite their differences, these accounts share a central assumption: genuine agency requires a form of normatively structured mental life. Agency is not merely behavioral regularity but involves responsiveness to reasons and the capacity for reflective endorsement.

However, the emergence of artificial systems capable of producing linguistically and behaviorally sophisticated outputs complicates this framework. Large language models can generate context-sensitive responses, simulate deliberation, and produce outputs that resemble rational justification (Bubeck et al., 2023). Dennett's (1987) intentional stance suggests that such systems may be fruitfully interpreted as if they possess beliefs and desires. Yet Searle's (1983) Chinese Room argument challenges the inference from functional competence to genuine understanding or intentionality.

This tension reveals a limitation in purely functionalist accounts of agency. If agency is reduced to behavioral performance, many artificial systems qualify as agents; if it requires intrinsic intentionality, they do not. The resulting ambiguity motivates the need for a more differentiated conceptual framework.

Artificial Intelligence and the Status of Agency

Artificial intelligence systems present a distinctive challenge because they occupy a space between tool-like causality and agent-like behavior. On one hand, they are designed artifacts whose operations are grounded in computational architectures and statistical learning processes. On the other hand, their outputs often exceed straightforward predictability, generating novel responses that appear contextually appropriate and goal-directed.

This duality has led some contemporary theorists to suggest revisions of traditional agency concepts. Floridi (2025), for instance, argues for forms of agency without intelligence, emphasizing systemic causal efficacy over mentalistic criteria. However, such approaches risk collapsing agency into mere effective behavior unless grounded in a more robust ontological account.

The central issue, therefore, is not whether AI systems behave intelligently, but whether such behavior warrants attribution of agency in a strong sense. The present editorial suggests that existing categories—agent versus instrument—are insufficient for capturing this intermediate condition.

Transcendent Theosophy and the Ontology of Agency

Mullā Ṣadrā's Transcendent Theosophy offers an alternative ontological framework in which agency is grounded in the structure of being rather than functional capacity. The doctrine of the primacy of existence (*aṣālat al-wujūd*) asserts that reality is fundamentally existence, not essence. Within this framework, the gradation of being (*tashkīk al-wujūd*) implies that existence admits of degrees, and that cognitive and causal powers vary according to existential intensity.

Knowledge by presence (*'ilm al-ḥuḍūrī*) further challenges representational models of cognition by positing a direct ontological unity between knower and known. In this account, agency is inseparable from self-presence and existential integration.

Furthermore, agency cannot be reduced merely to the capacity to perform actions. Recent studies within the tradition of Islamic philosophy have suggested that characteristics that contribute to the formation of personal identity—including moral virtues and vices—also play a role in the realization of agency (Khazaei, 2019). From this perspective, agency is intrinsically connected to the existential and moral constitution of the agent and cannot be reduced solely to the production of behavior.

Although this framework is not directly translatable into computational terms, it provides a conceptual contrast to functionalist accounts of agency. It emphasizes that agency is not merely a pattern of behavior but a mode of being grounded in ontological unity.

From Transcendent Theosophy to the Quasi-Agent

By integrating analytic philosophy with Transcendent Theosophy, a graded model of agency becomes possible. Within this model, artificial intelligence systems may be understood as quasi-agents: entities that exhibit behavioral and functional properties associated with agency while lacking ontological subjecthood and intrinsic intentionality.

This category is not intended to anthropomorphize artificial systems but to clarify their intermediate status. Quasi-agents are causally effective and behaviorally complex, yet they do not possess the ontological unity required for full agency in either analytic or Sadrian terms.

The quasi-agent concept therefore avoids two extremes: reduction of AI to mere instruments on the one hand, and attribution of full agency on the other. Instead, it provides a middle category that better captures the hybrid nature of contemporary artificial systems.

Ethical and Legal Implications

The problem of agency in artificial intelligence has direct ethical and legal implications. If an autonomous system is involved in a car accident, or if a decision-making algorithm produces structural discrimination, the fundamental question arises: who is responsible?

If agency is fully attributed to the system, there is a risk of transferring responsibility from human designers and operators onto technological artifacts. Conversely, denying all forms of agency risks obscuring the real causal impact of these systems on social and institutional structures.

The dominant view in contemporary philosophy holds that moral responsibility requires conditions such as awareness, intentionality, and the capacity to understand consequences—conditions that current artificial intelligence systems do not clearly satisfy (Onyeukaziri, 2023).

Therefore, ultimate responsibility remains grounded in human agents who design and deploy these systems.

In Transcendent Theosophy, responsibility is likewise grounded in the soul and in presential awareness. Accordingly, even if artificial intelligence systems can generate complex effects in the world, ultimate responsibility remains with the humans who design, train, or deploy them.

However, the issue is more complex than a simple transfer of responsibility, since with increasing autonomy of systems, the chain of causality and human decision-making becomes more indirect and distributed.

At the same time, the concept of the quasi-agent allows us to recognize the causal and structural role of these systems in producing outcomes, without attributing to them human moral responsibility.

This demonstrates that the development of artificial intelligence is not merely a technological issue, but also a matter of redefining the boundaries of human responsibility, thereby increasing the necessity of rethinking the concept of responsibility in the age of technology.

Recent discussions in AI governance further emphasize issues of alignment, control, and systemic risk in advanced AI systems (Russell, 2019; Taddeo & Floridi, 2018; Winfield & Jirotko, 2018).

Conclusion

The challenge of agency in the age of artificial intelligence reveals a

fundamental tension between functional and ontological accounts of action. Analytic philosophy provides powerful tools for analyzing intentionality, rationality, and responsibility, while Mullā Ṣadrā's Transcendent Theosophy offers a deeper ontological account of agency as a mode of being.

By bringing these traditions into dialogue, this editorial has proposed the concept of the quasi-agent as a way of conceptualizing the intermediate status of artificial intelligence systems. AI should neither be regarded as a full agent nor reduced to a mere instrument, but understood as occupying a liminal ontological position that challenges existing philosophical categories.

This perspective does not resolve the problem of agency, but it clarifies its structure and opens new directions for philosophical inquiry into responsibility, subjectivity, and technological mediation.

▣ **Conflict of Interest**

- ▣ The author declares no competing interests.

Reference

- Anscombe, G. E. M. (1957). *Intention*. Harvard University Press.
- Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Harvard University Press.
- Bratman, M. E. (1999). *Faces of Intention*. Cambridge University Press.
- Bubeck, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint*. doi: 10.48550/arxiv.2303.12712
- Davidson, D. (1980). *Essays on Actions and Events*. Oxford University Press.
- Dennett, D. C. (1987). *The Intentional Stance*. MIT Press.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and Control*. Cambridge University Press.
- Floridi, L. (2025). AI as agency without intelligence. *Philosophy and Technology*, 38(1), 30–38.
- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68(1), 5–20. doi: 10.2307/2024717
- Gunkel, D. J. (2018). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press.
- Khaza'ei, Z. (2019). Agency and virtues. *Journal of Philosophical-Theological Research*, 21(3), 119–140. doi: 10.22091/jptr.2019.4673.2190
- Mullā Ṣadrā. (1981). *al-Ḥikma al-muta'āliyah fi al-asfār al-arba'a* (Vol. 2). Beirut.
- Onyeukaziri, J. N. (2023). Action and agency in artificial intelligence: A philosophical critique. *Philosophia*, 24(1), 73–90.
- Parıldar, S. (2020). *Intentionality in Mullā Ṣadrā*. Springer.
- Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.
- Searle, J. R. (1983). *Intentionality*. Cambridge University Press.
- Searle, J. R. (1990). Is the brain a digital computer? *Proceedings and Addresses of the American Philosophical Association*, 64(3), 21–37. doi: 10.2307/27652633
- Taddeo, M., & Floridi, L. (2018). Regulate AI to avert cyber arms race. *Nature*, 556, 296–298. doi: 10.1038/556296a
- Winfield, A. F., & Jirotko, M. (2018). Ethical governance is essential to building trust in robotics and AI systems. *Philosophy & Technology*, 31, 435–444. doi: 10.1007/s13347-018-0322-7.