Journal of Human Cognition
2024, VOL. 8, NO. 2, 5-7
DOI: 10.47297/wspjhcWSP2515-469901.20240802

Agency and Intentionality for Artificial Agents

Yidong Wei

School of Philosophy, Shanxi University

In this paper, the author will explore the relationship between agency and intentionality of the artificial agent in the following seven ways.

1 Agent vs. Subject

An agent (actor) as an entity includes physical systems, animals, and humans. In philosophy, subject is a concept relative to object and refers to a human being. In this way, subject and object are a pair of categories whose interaction is used to describe various human activities such as productive and scientific activities.

For human activities, the subject is the person and the object is the goat of his or her action or cognition; whereas agent and action are a pair of categories that are mainly used to describe entities and their actions, e. g., the behavior of robots, reactive agents in chemistry, and agency (intermediaries) in society. Thus, agent and subject are two different concepts, with the former encompassing the latter, e. g., an agent can also refer to a human being. But an agent may be unconscious, whereas a subject must be conscious.

2 Agency vs. Subjectivity

In philosophy and logic, subjectivity exhibits agency, so agency is sometimes called autonomy. In Computer Science and Artificial Intelligence, agent usually refers to intelligent agent, and agency refers to intelligent groups (combinations of intelligences, i. e., multi-intelligent systems, which are capable of acting). In the author's opinion, there is still a difference between these two concepts, subjectivity implies the presence of consciousness, while agency may not require consciousness, but requires action, for example, physical self-organizing systems also exhibit agency, not subjectivity.

3 Artificial agents vs. Human beings(subjects)

Similar to agent and subject, the concept of an intelligent agent is different from, but encompasses, humans, since humans are also an intelligent agent. However, in AI, the term subject is avoided because intelligences may be inanimate and unconscious. This is an obvious fact. Therefore, the environment in which agent and subject are located is very important; the environment in which agent is located is often artificial, such as physical systems and artificial intelligence systems; while the environment in which subject is located is usually natural and social, such as nature and social practices, and its contextualization is especially obvious. In other words, agent is context-free, while subject is contextualized.

Therefore, the presence or absence of contextuality is an intrinsic criterion for judging or distinguishing agent and subject.

4 Interagency vs. Intersubjectivity

Interaction between agents, which I call interagency, or interagentivity (a word coined by myself), means that different agents interact to produce the functional properties of multi-agents (that is, interaction between agents). Intersubjectivity is the overall properties produced by interactions between different subjects (people), such as communication and teamwork behaviors. Given that single agent is usually unconscious, only interagency is used in AI, which is a type of swarm or group intelligence resulting from collaborative behavior. This is similar to the collective behavior of swarms of bees, ant colonies, and flocks of birds, but different from the collective consciousness of human societies (the difference between individual consciousness and individual unconsciousness).

5 Subjectivity and Intentionality

In philosophy, subjectivity implies intentionality, i. e., the very notion of a subject implies the existence of intentions, expectations, beliefs, and purposes, because intentionality means aboutness or referring to something. In phenomenology, intentionality is a description of the nature of consciousness; humans are conscious, and certainly have intentionality. Does an unconscious artificial agent have no intentionality? The answer is not necessarily. Agents are directional, purposeful, and autonomous, so of course they have intentionality, but this kind of intentionality is not "machine intention" or "artificial intention" based on life and biological consciousness---it is a kind of "as if intention". Therefore, the intentionality of a human being is "real intentionality" and the intentionality of an artificial agent is "false intentionality".

6 Agency and Intentionality

In artificial intelligence or robot society, since agents also have directionality or aboutness, being goal or task oriented, the combination of agents---agency (a multi-agent system)---must also have intentionality (directionality). Thus, at the level of aboutness, both humans and intelligent robots have intentionality, but why is it said that humans are conscious and artificial agents are not? In this way, it seems that intentionality is not a criterion of judgment that distinguishes AI from human intelligence. There must be another criterion, which I call "reflexivity". Reflexivity indicates that the subject knows who he or she is. Thanks to this criterion, humans know who they are, AIs do not. In this sense, an artificial agent has no consciousness, no matter how intelligent it appears to be.

7 Adaptive Representation vs. Agency and Intentionality

The author is a staunch adaptive representationalist, arguing that cognition and intelligence are adaptive and representational. Adaptability itself implies intentionality, and representativeness itself implies agency. The artificial agent, as a self-organizing system or entity, must have the capacity for adaptive representation, which necessarily entails agency and intentionality. In terms of the representational relationship, the agent or the subject exhibits a different representational relationship between agency and intentionality in different disciplines.

In science, the subject is the scientist, the object of his or her study is usually a natural object or phenomenon, the research methods is carried out by means of scientific instruments (experiments), the mode of representation is modeling, and the representational relation between subjectivity and intentionality is:

Subject (scientist) → {model} → target (object) → scientific knowledge

In philosophy, where the subject is the philosopher, the interaction of the subject with the object can be direct, and in most cases indirect, i.e., carried out through some intermediary (e.g., thinking), and the representational relation between subjectivity and intentionality is:

Subject (Philosopher) → {Mediator (Thinking)} → Object (Concept) → Insights or ideas

In AI, the designer (human) is absent, the actor is a sub-subject (i.e., agent) whose task is goal-directed action, and the representational relation between agency and intentionality is:

{designer/user} → [agent or agency] → action → task

Here, the actor is an artificial agent, not a human subject. Therefore, for artificial agent, its autonomy is agency, not subjectivity.

Funding

"Philosophical Study on the Challenge of Artificial Cognition to Natural Cognition" (21&ZD061), a major project of the National Social Science Foundation of China.

References

[1] Barbaras R.Life and perceptual intentionality.Res Phenomenol 2003,33:157-166.

[2] Beim G.P. Contextual emergence of intentionality.J Conscious Stud, 2014,21(5/6):75–96.

[3] Ciardo,F.et al.Attribution of intentional agency towards robots reduces one's own sense of agency. Cognition,2020, 194:104109.

[4] Gallese V.The inner sense of action: agency and motor representations. Journal of Consciousness Studies, 2000, 7: 23-40.

[5] Haggard P.Sense of agency in the human brain.Nature Reviews Neuroscience,2017,18(4):196.

[6] Horst S.W.*Symbols,Computation,and Intentionality:A Critique of the Computational Theory of Mind.*Berkeley:University of California Press,1996.

[7] Obhi S.S.,& Hall P.Sense of agency in joint action: Influence of human and computer co-actors. Experimental Brain Research,2011,211(3-4): 663-670.

[8] Sheets-Johnstone M. Phenomenology and agency: methodological and theoretical issues in Strawson's 'The self'. In: Gallagher, S., Shear,J. (eds) .Models of the self. Imprint Academic,Thorverton, UK,1999,pp 231–252.

[9] Strasser,A.Social Cognition and Artificial Agents.In V C.Müller (Ed.): PT-AI 2017,SAPERE 44,pp.106–114, 2018.https://doi.org/10.1007/978-3-319-96448-5_12.Springer Nature Switzerland AG 2018.

[10] Schillaci G.,Hafner V.V.,and Lara B.2016.Exploration Behaviors, Body Representations, and Simulation Processes for the Development of Cognition in Artificial Agents. Frontiers in Robotics and AI 3: 1-39. www. frontiersin. org. doi: 10.3389/frobt.2016.00039

[11] Thompson E. Sensorimotor subjectivity and the enactive approach to experience. Phenomenology and the Cognitive Sciences,2005,4: 407-427.

[12] Wykowska A., Chaminade, T. and Cheng G. 2016. Embodied artificial agents for understanding human social cognition.The Royal Society Pubilishing.http://rstb.royalsocietypublishing.org/on April 11,2016.

[13] 魏屹东.适应性表征:架构自然认知与人工认知的统一范畴.哲学研究,2019(9):114-124.

[14] 魏屹东.论人工认知实现的方法论策略.哲学研究,2024(3):102-113.