Perspective

Is ChatGPT the way toward artificial general intelligence

Frank Emmert-Streib[1]

Abstract

The success of the conversational AI system ChatGPT has triggered an avalanche of studies that explore its applications in research and education. There are also high hopes that, in addition to such particular usages, it could lead to artificial general intelligence (AGI) that means to human-level intelligence. Such aspirations, however, need to be grounded by actual scientific means to ensure faithful statements and evaluations of the current situation. The purpose of this article is to put ChatGPT into perspective and to outline a way forward that might instead lead to an artificial special intelligence (ASI), a notion we introduce. The underlying idea of ASI is based on an environment that consists only of text. We will show that this avoids the problem of embodiment of an agent and leads to a system with restricted capabilities compared to AGI. Furthermore, we discuss gated actions as a means of large language models to moderate ethical concerns.

Keywords Artificial general intelligence · Conversational AI · Question-answering system · Natural language processing · Data science

1 Introduction

It is rare that the introduction of a new tool generates similar excitement in academia and the general public but ChatGPT is such a tool. This is even more astonishing considering that ChatGPT is the result of intense research in deep learning (DL) and natural language processing (NLP) which are both intricate fields by themselves. However, their combination via large language models (LLMs) [1] created a tool with truely remarkable capabilities.

Essentially, ChatGPT is an online chatbot based on a question-answering system [2, 3]. The suffix "GPT" means *generative pre-trained transformer* [4] that is a type of large language model (LLM) consisting of billions of parameters trained via transfer learning [5]. Since its introduction in November 2022, ChatGPT has generated significant attention, resulting in the publication of many articles addressing its impact on research and education as well as raising ethical concerns [6–11]. Additionally, there are studies discussing possibilities for ChatGPT to reach artificial general intelligence (AGI) [12, 13]. This fundamental perspective is also the focus of this paper.

The term artificial general intelligence (AGI) refers to a learning paradigm that results in an agent whose level of intelligence is comparable to human intelligence and that is capable of performing any intellectual task a human can perform [14–16]. Sometimes AGI is referred to as "narrow AI" or "strong AI" [17, 18] but the terminology for this is not consistent [19]. To date, the goal of designing such an agent hasn't been reached yet but there are speculations that we could achieve AGI within the next few decades [17, 20].

✉ Frank Emmert-Streib, frank.emmert-streib@tuni.fi | [1]Predictive Society and Data Analytics Lab Department, Tampere University, 33720 Tampere, Finland.

Discover

In this paper, we explore the relationship between ChatGPT and artificial general intelligence (AGI). To provide a comprehensive assessment, we first delve into the fundamentals of AGI, including its key components such as the perception-action cycle (PAC). This analysis helps uncover inherent limitations in ChatGPT that hinder its progression towards AGI. Subsequently, we introduce the concept of "artificial special intelligence" (ASI), which serves as the basis for a modified version of ChatGPT featuring a private input channel and an open PAC, termed LLM-PI. We then enhance LLM-PI with a gating mechanism for actions, effectively moderating the PAC to allow only selected actions. This model, referred to as gLLM-PI (gated large language model with private information), addresses ethical concerns associated with AGI. Ultimately, we argue that gLLM-PI has the potential to achieve ASI, offering a more attainable goal compared to AGI.

2 Artificial general intelligence

Artificial general intelligence (AGI) is a complex and multifaceted learning paradigm encompassing several key components that are considered essential for achieving human-level general intelligence in artificial systems. These components vary considerably depending on the specific approach or theory, but generally, the following aspects are considered integral [19, 21]:

- Learning and adaptation: AGI systems should possess the ability to learn from data and adapt their knowledge and behavior based on new information. This includes techniques such as reinforcement learning, active inference, and continual learning.
- Reasoning and problem-solving: AGI systems should exhibit advanced reasoning capabilities to analyze complex situations, make logical deductions, solve problems, and engage in higher-level thinking. This involves techniques such as logical reasoning, symbolic manipulation, and planning.
- Perception and sensing: AGI systems should be able to perceive and understand the environment using sensory inputs, including visual, auditory, and tactile information. This involves computer vision, natural language processing, speech recognition, and other sensory processing techniques.
- Knowledge representation and understanding: AGI systems should have the ability to represent and comprehend knowledge in a structured manner, enabling them to grasp and utilize information effectively. This involves techniques such as semantic networks, ontologies, and knowledge graphs.
- Cognitive flexibility and adaptability: AGI systems should demonstrate cognitive flexibility, enabling them to transfer knowledge and skills across domains, generalize from limited data, and exhibit creativity and innovation.

One particular theory of AGI was developed in the early 2000s by Hutter. He introduced a universal mathematical theory of AGI called AIXI [14]. Its basic idea is to combine Solomonoff induction with sequential decision theory where the former provides an optimal solution for induction and prediction problems while the latter addresses optimal sequential decisions. Since AIXI combines an approach that is optimal for prediction in unknown worlds with an approach that is optimal for decision making in known worlds Hutter argued that the resulting agent within this reinforcement learning model is an optimal rational agent. Unfortunately the resulting AIXI framework requires infinite computational power and is for this reason uncomputable. Overall, this theory gives a definition of intelligence [22] and, if one starts from this definition, it provides the construction of an intelligent agent reaching AGI.

A key element of AIXI and essentially any approach to AGI is the ability of an agent to interact with an environment and to make decisions by selecting actions. The neurobiological motivation for this cyclic process is discussed in the next section [23].

3 Perception-action cycle

It has been realized early that the foundation for understand the organizational behavior of an organism and learning is the relation between perception and action [24, 25] called the perception-action cycle (PAC). The concept of the perception-action cycle describes the continuous interaction loop between an agent and its environment, where the agent perceives information from the environment through its sensors, processes that information, and then takes actions based on its internal state and goals. Due to the intertwined interactions of the whole system, the perception-action cycle is a systems theoretical approach to understand the neuronal organization of animals rather than a reductionistic one [26].

For AI, the perception-action cycle is important because it has been used as a motivation and conceptual framework for reinforcement learning (RL) [27, 28]. That means depending on the action of the agent, a reward is received that reflects the quality of an action. In the field of robotics, reinforcement learning and, hence, the perception-action cycle is an integral part for building autonomous robots that learn from interacting with an environment where a reward is either obtained via a human, environmental or self-supervised feedback [29, 30]. In Fig. 1, we show an overview of the principal components of RL. Importantly, the perceptions and actions form iteratively a closed loop that enables a continuous learning cycle.

A problem in robotics is that a robot is actually a physical entity that interacts with a physical environment. This is generally referred to as embodiment [31, 32]. Specifically, embodiment refers to the idea that a robot is physically situated and interacts with the world through its body. Embodiment is essential for robotics as it enables the robot to gather sensory information from its environment, manipulate objects, and engage in physical interactions, leading to a more grounded and adaptive form of intelligence that is tightly coupled with the physical world. However, a main problem with embodiment in robotics is the challenge of designing and implementing a physical body that can effectively interact with the environment. This involves addressing issues related to mechanical constraints, sensor limitations, and the complexity of physical interactions. Another problem is the potential mismatch between the physical embodiment and the desired tasks or environments, as certain tasks may require specific morphologies or capabilities that are not easily achievable in a physical robot.

4 Shortcomings of ChatGPT for AGI

For the following discussion, we assume that ChatGPT takes the position of the agent in Fig. 1. Based on this configuration, we are now able to identify three important shortcomings of the current version of ChatGPT preventing it from reaching AGI.

1. Currently, ChatGPT cannot interact with the environment via an action.
2. Currently, ChatGPT does not have the inner structure (policy function) of an agent.
3. The environment consists only of text.

We would like to remark that the first two limitations are related because an agent without policy function cannot select an action but even with a policy function one could disallow the execution of an action which would prevent the interaction with an environment. While ChatGPT generates an output, this output cannot directly affect the environment, making it fundamentally different from the actions of an agent.

The third shortcoming is related to the fact that, in contrast to a robot, ChatGPT is a natural language processing system that reads text and generates text. That means ChatGPT lives per se in a virtual environment. This removes, in a natural way, the embodiment problem from robotics discussed above without the need for any approximation. However, this is a severe simplification and entirely different to the situation as faced by humans, animals or robots. We would also

Fig. 1 Key components of reinforcement learning based on the perception-action cycle of an agent. This is also a main element for theories of AGI
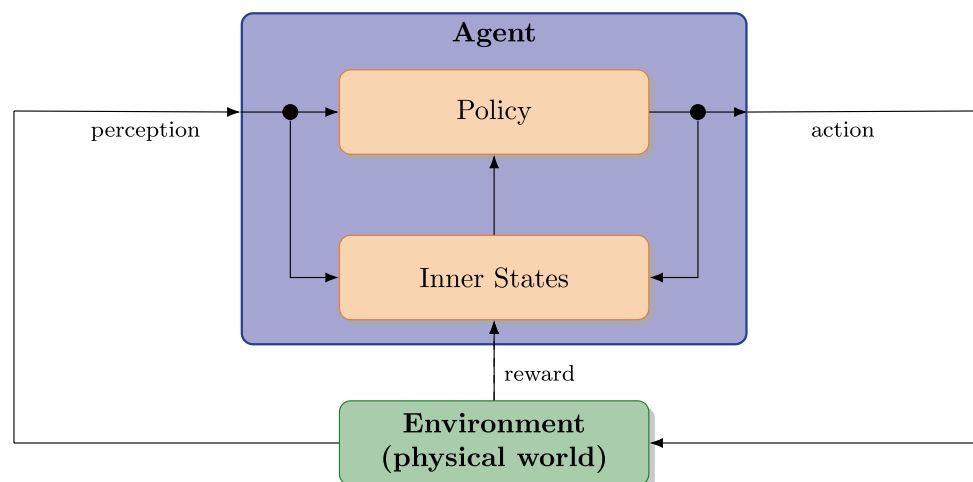
like to highlight that the text-based nature of the environment is a property of the environment rather than a technical limitation. As such (3) is different to (1) and (2) which can be changed by a modification of ChatGPT.

This indicates that, as it stands, ChatGPT fails to meet the fundamental criteria of an agent necessary for achieving AGI. However, this does not diminish its ability to amaze its users. Nonetheless, it's important to recognize that ChatGPT, in its current state, is not and cannot evolve into an AGI. Therefore, in the subsequent sections, we explore potential extensions of ChatGPT to gain insights into the most realistically achievable objectives.

5 GPT-in-the-loop as artificial special intelligence

For the following discussion, we will accept shortcoming (3) because it reflects the situation as is (corresponding to a property of the environment) but improve upon shortcomings (1) and (2). This will lead to the definition of artificial special intelligence (ASI).

While the goal of AGI is the development of autonomous systems that possess the cognitive capabilities and general intelligence similar to that of human beings one can limit this ambitious goal. Specifically, let's content ourselves with tasks only involving text. That means in this case, the environment consists only of text and no images, audio signals etc can occur. Hence, in this case the environment would be a text world.

Assuming the environment is a text world, this has two important implications.

1. An agent interacting with such an environment does not need physical sensors nor actors because the text world is virtual.
2. An agent living in a text world does not require embodiment because no "physical" entities are involved in this perception-action cycle.

These two implications enable an agent with more limited capabilities due to a simpler environment than required for AGI. For this reason we call this restricted case artificial special intelligence (ASI) because an optimal rational agent lives in a text world.

Definition 1 (*Artificial special intelligence [ASI]*) Artificial special intelligence (ASI) is an agent's capability to perform any intellectual task that a human being can, based on text data.

Regarding the chosen name, we would like to remark that we use artificial special intelligence (ASI) to mirror the same sentiment as in artificial general intelligence (AGI) but emphasizing a particular or "special" situation. More precisely it could be called artificial text intelligence (ATI), however, if considered in isolation, this asymmetry between AGI and ATI could be overlooked. For this reason, we prefer the name ASI over ATI, as it maintains consistency with the sentiment of AGI and underscores the specificity of an application.

Next, we address the need for an agent to act on the environment. Since ChatGPT is currently not an agent as in Fig. 1, we need to extend its structure to convert it into such an agent. That means, we need to place ChatGPT in-the-loop (ITL) and for this reason we call the resulting system GPT-ITL. Specifically, this would require the following modifications.

1. GPT-ITL would need access to the environment to collect new input (perception).
2. GPT-ITL would need to receive a reward from the environment.
3. The inner structure of GPT-ITL would need to be adjusted by including a policy.
4. GPT-ITL would need access to the environment by writing its output into the text world (action).

At the moment, all steps seem feasible from a technical point but ethically there are issues, especially with the last step. The problem with the last step is that it would give autonomy to the system to publish its output publicly. However, this would allow that GPT-ITL could spread misinformation because false output/answers could enter the public domain, e.g., on social media platforms. Even more severely, GPT-ITL could manipulate social media users and even persuade them to commit certain doings including crimes.

Looking from a fundamental perspective onto GPT-ITL, it is clear that due to our assumption of a simplified environment consisting of a text-world, GPT-ITL could at best reach an ASI. Hence, even GPT-ITL would not be capable of reaching AGI.

6 Modified version of ChatGPT with private input: LLM-PI

Now we turn from fundamental considerations to practical ones by outlining a way forward. For the moment, this approach will avoid ethical concerns encountered for GPT-ITL, instead, this will be considered in Sect. 7. For this approach, we start again from ChatGPT and augment it with a private input (PI). We call this model, LLM-PI where LLM (large language model) indicates that it could be GPT or any other LLM.

Figure 2 outlines the main components of LLM-PI. Importantly, this system is not allowed to act on the environment (violet box and its connecting lines are not part of this system; see below). That means the perception-action loop is open and, hence, the resulting agent, corresponding to LLM-PI, is no longer autonomous. Aside from this, LLM-PI has a separate private input (PI) channel allowing the user to supply text data, e.g., corresponding to documents, notes or files that are not available from the environment. In this way, problem specific or confidential information can be supplied that allows LLM-PI the answering of questions requiring either private or specific information that is not part of the public text world (environment).

We would like to emphasize that the crucial difference between LLM-PI and the current form of ChatGPT is the presence of the private input channel. While this may appear straightforward technically, it entails more than just a static input. In order to make this additional information actionable, the LLM must undergo another round of fine-tuning for learning from the provided input. While theoretically feasible, this process is time and resource-intensive in practice. Consequently, users may experience delays as they await the completion of this fine-tuning process before receiving answers to their questions, potentially rendering real-time conversations impractical at present.

In addition to the computational overhead, implementing a private input channel offers a distinct advantage for the personalization of a LLM. Over time, this feature would enable the system to tailor responses to individual users, providing more authentic interactions that align with personal preferences. However, this personalized approach raises ethical concerns surrounding user privacy, as safeguards would be necessary to prevent potential abuse by third parties. Nonetheless, these concerns are manageable and can be addressed through appropriate policies, representing a different set of challenges compared to those encountered with AGI.

7 Gating the loop for LLM-PI: gLLM-PI

The last extension we want to discuss provides an approach for the implementation of such a policy. Specifically, we suggest a gating mechanism for actions. Such a gating mechanism for actions allows to close the loop to the environment but in a gated manner. Hence, the gating would allow a moderation of actions in order to control ethical implications potentially incurred by the actions. While this controls the autonomy of the agent it mediates ethical concerns.
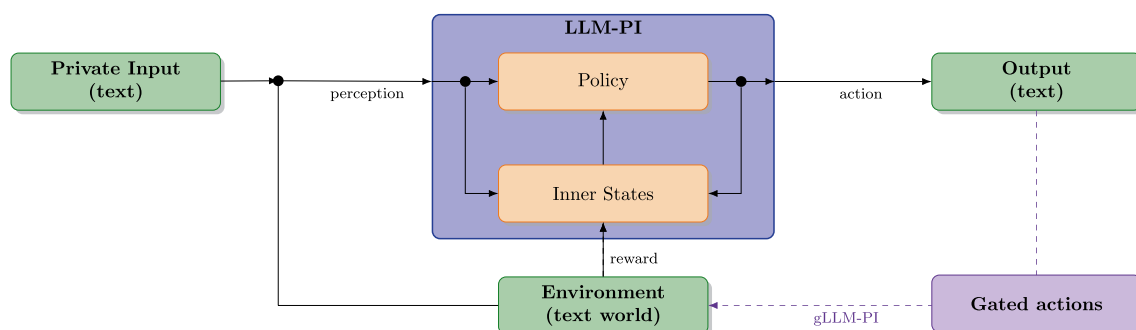

Fig. 2 Modified version of ChatGPT with a private input channel called LLM-PI (the violet box and its connecting lines are not part of this system). If the perception-action cycle is closed but gated by adding gated actions (now the violet box and its connecting lines are part of the system), the resulting new system is called gLLM-PI

With regard to the LLM-PI discussed in the previous section, one could start with such an optimized model allowing private input but augment it by an additional functionality enabled via the closed but gated perception-action-cycle. Overall, this brings us back to similar capabilities as discussed earlier in Sect. 5 and shows how to approach an artificial special intelligence (ASI) system that is limited to a text environment.

To distinguish this model from others, we call it gLLM-PI (gated large language model with private information). In Fig. 2, we show gLLM-PI that is obtained by adding gated actions (violet box) and its connecting lines to LLM-PI. Importantly, by using different gating mechanisms for selectively permitting actions different forms of ASI are enabled. Potentially, this allows us to further sub-categorize ASI into more refined classes that represent different types or severities of ethical concerns. In order to get a better understanding about this, let's consider a few examples.

Example 1 The following list gives hierarchical restrictions applied to a ASI via gated actions. We start at the top of such an ASI corresponding to the most liberal form and work our way down to the most restricted form.

Gated actions:

1. Politeness and courtesy
2. 1 + political correctness
3. 2 + falsifiability
4. 3 + expert in medicine

The first restriction (1) enforces a polite and courteous communication, thereby prohibiting the use of inappropriate language. This results in a formal rather than colloquial discussion. The second restriction (2) additionally (1 + means the first restriction and the second) requires that the communication be politically correct. Unfortunately, this restriction would be country- and time-specific and not universal, as the definition of political correctness varies widely among countries and changes over time. This is in contrast to restriction (3), which requires in addition (2 + means the second and the third) the falsifiability of statements, limiting communication to a factual basis with a clear distinction between facts and hypotheses. Finally, restriction (4) requires in addition that the communication is limited to the field of medicine. Overall, this results in a particular ASI where the gLLM-PI is similar to a medical expert. We would like to note that by replacing restriction (4) with, e.g., "3 + expert in economics" we obtain another ASI but now as an expert in economics.

Example 2 Another instance of a ASI for other gated actions is given by the following restrictions.

Gated actions:

1. Politeness and courtesy
2. 1 + falsifiability
3. 2 + expert in medicine

While this looks very similar to the gated actions for Example 1, it does not contain restrictions about political correctness (see above). However, as discussed above, such restrictions are not only country-specific but change also over time as political legislation can change. That means the gLLM-PI for Example 2 would be less restricted than the gLLM-PI for Example 1. As a result, one could study the effect different gated actions have on the resulting communication and the obtained ASI.

We would like to point out that our description of a gLLM-PI is unlike, e.g., standard agent types for ordinary artificial intelligence as described in [33]. The reason for this difference is due to our specific situation given by the text environment with which GPT-ITL can interact.

In summary, these examples illustrate various versions of a gLLM-PI, including ethical forms that can be studied and cross-investigated to enhance our understanding of an ASI.

8 Discussion

While there is currently a great deal of excitement surrounding ChatGPT, and it certainly has the potential to contribute significantly to research and education, the hope (or fear) that it will achieve AGI status is not justified. Instead, as discussed in this paper, several extensions are necessary. Even with these extensions, ChatGPT would at best achieve artificial special intelligence (ASI) status, primarily due to its limitation to a text-based environment. Still, the idea of studying such an ASI is intriguing because of the following reasons.

1. No embodiment problem: Usually, embodiment requires, e.g., a robot to be situated within a (real) environment to perceive and act correspondingly. However, this requires various approximations that all rely on assumptions. Instead, gLLM-PI lives in a space where only text exists. Hence, the difference between a physical world and a virtual word does not exist and no approximations to bridge both worlds are needed. Still, such a world is not a toy example but represents real texts as exchanged on all levels of society.
2. Categories of actions: The most problematic aspect of a closed loop of gLLM-PI would certainly be its possibility to performing actions. However, there are different categories of actions having different consequences. For instance, there is a difference if gLLM-PI would be allowed to upload a report to a public repository, or if it would be allowed to make an online purchase. While the former might lead to misinformation the latter triggers with certainty a direct action by a third party. That means by carefully categorizing actions of gLLM-PI one could moderate ethical implications and real-world consequences.
3. Declaration of gLLM-PI generated text: Another problems that could result from generated text by gLLM-PI is the spread of misinformation. This problem could be mitigated by the declaration of gLLM-PI generated text. Generally, it is always important to consider the source of information or text and a declaration would eliminate anonymity and promote source transparency. This would allow for a more informed assessment by readers of such a text. Technically, by using blockchain technology one could ensure that there is no tampering with such a declaration.

In summary, these are just a few aspects and approaches that could be studied to allow for a safe transition from ChatGPT in its current form to an ASI via a number of intermediate steps provided by different versions of gLLM-PI. If this paves also the way to a genuine AGI remains to be seen because it requires substantial extensions that would lead us beyond a text environment.

9 Conclusion

We showed in this paper that, for principle reasons, ChatGPT is neither an artificial general intelligence (AGI) nor can it naturally evolve into one. That means without significant modifications of ChatGPT this is not feasible.

A main problem for this is the confinement into a "text world" that does not require or enable embodiment to interact via sensors and actors with the environment. However, by introducing the notion of an artificial special intelligence (ASI) that is limited to an environment consisting only of text, a modified version of ChatGPT, we called gLLM-PI (gated large language model with private information), could transition into an ASI. In fact, the limitation to a "text world" seems to open new avenues for the exploration of the perception-action cycle that is an eminent component for the designing of an optimal rational agent in any setting. Furthermore, we showed that the gating mechanism for actions that closes the PAC in a controlled manner allows the implementation of policies for preventing ethical concerns that are inherently present for AGI. Hence, the gLLM-PI model has not only the potential to enhance our understanding of intelligent systems but to do this by moderating ethical considerations.

Data availability No data were used in this study.

Discover

Declarations

Competing interests The authors declare no conflict of interest.

References

1. Tay Y, Dehghani M, Bahri D, Metzler D. Efficient transformers: a survey. ACM Comput Surv. 2022;55(6):1–28.
2. Abbasiantaeb Z, Momtazi S. Text-based question answering from information retrieval and deep neural network perspectives: a survey. Wiley Interdiscip Rev Data Min Knowl Discov. 2021;11(6):1412.
3. Farea A, Yang Z, Duong K, Perera N, Emmert-Streib F. Evaluation of question answering systems: complexity of judging a natural language. arXiv preprint. 2022. arXiv:2209.12617.
4. Radford A, Narasimhan K, Salimans T, Sutskever I, et al. Improving language understanding by generative pre-training. 2018.
5. Bashath S, Perera N, Tripathi S, Manjang K, Dehmer M, Emmert-Streib F. A data-centric review of deep transfer learning with applications to text data. Inf Sci. 2022;585:498–528.
6. Biswas SS. Potential use of chat GPT in global warming. Ann Biomed Eng. 2023;51(6):1126–7.
7. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare. 2023;11(6):887. https://doi.org/10.3390/healthcare11060887.
8. Liebrenz M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. Lancet Digit Health. 2023;5(3):105–6.
9. Xue VW, Lei P, Cho WC. The potential impact of ChatGPT in clinical and translational medicine. Clin Transl Med. 2023;13(3): e1216.
10. Zhai X. ChatGPT for next generation science learning. XRDS Crossroads ACM Mag Stud. 2023;29(3):42–6.
11. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med. 2023;388(13):1233–9.
12. Zhou J, Ke P, Qiu X, Huang M, Zhang J. ChatGPT: potential, prospects, and limitations. Front Inf Technol Electron Eng. 2023. https://doi.org/10.1631/FITEE.2300089.
13. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, Lee P, Lee YT, Li Y, Lundberg S, et al. Sparks of artificial general intelligence: early experiments with GPT-4. arXiv preprint. 2023. arXiv:2303.12712.
14. Hutter M. Universal artificial intelligence: sequential decisions based on algorithmic probability. Berlin: Springer; 2005.
15. Goertzel B, Pennachin C. Artificial general intelligence. New York: Springer; 2007.
16. Yampolskiy RV, Fox J. Artificial general intelligence and the human mental model. In: Singularity hypotheses. Berlin: Springer; 2012. p. 129–45.
17. Kurzweil R. The singularity is near: when humans transcend biology. New York: Penguin; 2005.
18. Searle JR. Mind, language and society: philosophy in the real world. New York: Basic Books; 2008.
19. Goertzel B. Artificial general intelligence: concept, state of the art, and future prospects. J Artif General Intell. 2014;5(1):1.
20. Chalmers DJ. The singularity: a philosophical analysis. In: Science fiction and philosophy: from time travel to superintelligence. Hoboken: Wiley; 2016. p. 171–224.
21. Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. Behav Brain Sci. 2017;40: e253.
22. Legg S, Hutter M. Universal intelligence: a definition of machine intelligence. Mind Mach. 2007;17(4):391–444.
23. Freeman WJ. The physiology of perception. Sci Am. 1991;264(2):78–87.
24. Sperry RW. Neurology and the mind-brain problem. Am Sci. 1952;40(2):291–312.
25. Warren Jr WH. The perception-action coupling. In: Sensory-motor organizations and development in infancy and early childhood: proceedings of the NATO advanced research workshop on sensory-motor organizations and development in infancy and early childhood Chateu de Rosey, France. Springer; 1990. p. 23–37.
26. Emmert-Streib F. Active learning in recurrent neural networks facilitated by an hebb-like learning rule with memory. Neural Inf Process Lett Rev. 2005;9(2):31–40.
27. Barto AG, Sutton RS, Anderson CW. Neuronlike adaptive elements that can solve difficult learning control problems. IEEE Trans Syst Man Cybern. 1983;13:834–46.
28. Whitehead SD, Ballard DH. Learning to perceive and act by trial and error. Mach Learn. 1991;7:45–83.
29. Knox WB, Stone P. Tamer: training an agent manually via evaluative reinforcement. In: 2008 7th IEEE international conference on development and learning. IEEE; 2008. p. 292–7.
30. Sekar R, Rybkin O, Daniilidis K, Abbeel P, Hafner D, Pathak D. Planning to explore via self-supervised world models. In: International conference on machine learning. PMLR; 2020. p. 8583–92.
31. Pfeifer R, Lungarella M, Iida F. Self-organization, embodiment, and biologically inspired robotics. Science. 2007;318(5853):1088–93.
32. Clark A. An embodied cognitive science? Trends Cogn Sci. 1999;3(9):345–51.
33. Russell SJ, Norvig P. Artificial intelligence: a modern approach. Harlow: Pearson; 2016.