Psychology and Artificial Intelligence: Reconceptualization, Integration, and Platonic Intelligence

Yingrui Yang * and Hongbin Wang

Posted Date: 16 May 2025

doi: 10.20944/preprints202505.1277.v1

Keywords: machine; intelligence; sensation; perception; attention; mind; Searle; turing test; embedding; embodiment; platonic representation

 doi:10.20944/preprints202505.1277.v1

Article

Psychology and Artificial Intelligence: Reconceptualization, Integration, and Platonic Intelligence

Yingrui Yang [1,*] and Hongbin Wang [2]

[1] Department of Cognitive Science, Rensselaer Polytechnic Institute

[2] College of Medicine, Texas A&M University; hongbinwang@tamu.edu

* Correspondence: yangyri@rpi.edu

Abstract: The present paper first reviews those psychological concepts used in artificial intelligence. The embedding of psychological properties and mind in artificial intelligence has advantages and challenges. Instead of doing philosophical enquiries, we assume as our working hypothesis that machine intelligence is a kind of platonic reality which needs working definitions in AI modeling. It argues that following Gödel and Tarski, Turing test enriched Artificial intelligence, resulting a new kind of independent results, which refers to machine intelligence. Machine intelligence is dual with human intelligence.

Keywords: machine; intelligence; sensation; perception; attention; mind; Searle; turing test; embedding; embodiment; platonic representation

---

Artificial intelligence is not only a technology, but also a science. Psychology and artificial intelligence (AI) are clearly related in that they share a lot of concepts. This paper aims to clarify a number of psychological subdomains that are most relevant to the current research in artificial intelligence, including, sensation, attention, perception, intention, integration, consciousness, reasoning, and decision making.

1. Shared Concepts

AI researchers frequently "borrow" terminology from psychology and cognitive science – both to better understand how their systems operate and to communicate their ideas more effectively to the broader scientific community. The reasons and mechanisms behind this borrowing are fascinating in their own right, and we do not fault researchers for doing so. This trend dates back to the early days of AI, when *knowledge representations* and *planning/reasoning* were central topics, continues through to more recent breakthroughs such as the seminal paper *"Attention Is All You Need"* (Vaswani et al., 2017), which revitalized the field and catalyzed the development of transformers and large language models (LLMs).

Due to the familiarity of these terms, AI researchers often take them for granted and use them freely – frequently without acknowledging the depth of investigation and theoretical grounding these concepts have in psychology.

1.1. Sensation

In artificial intelligence, particularly in the context of LLMs, a verbal task is characterized by a sequence of tokens, which are then embedded into a high-dimensional feature space and processed by deep neural networks such as transformers to generate a response. As one can easily imagine, a complex compound task can have many intertwined features. In practice, only a set of these features prove relevant and will become sensitive for performance. Thus, for a given task, the design of neural

 doi:10.20944/preprints202505.1277.v1

network starts with a sensation-like process. From a psychological perspective, this process involves discourse processing and text comprehension.

1.2. Attention

Attention is one of the most extensively studied topics in psychology. William James famously defined it as "the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought" (James, 1890). Anne Treisman (1980) argued that the integration of task-relevant features calls for attention. Contemporary neuropsychology views attention not as a single, unified mechanism, but as a collection of distinct processes involving multiple brain networks responsible for distinct functions of alerting, orienting, and executive control (Posner & Petersen, 1990). This psychological perspective is, to some extent, reflected in the design of the transformer architecture, where the self-attention mechanism dynamically adapts and integrates features across a long context window. While this represents a prototype of how psychological constructs can be reconceptualized and implemented in AI, it is clear that transformer-based attention captures only a narrow slice of the full complexity of human attention.

1.3. Perception

Recently, Yingrui Yang (2025) has applied differential geometry to mathematical modeling of artificial intelligence. He resolved one issue involving how the bottom manifold is characterized. Making each point stand for a task and introducing the use of a task variable. The task variable can be multi-dimensional depending on the number of features that are taken into consideration. The bottom manifold can be a Hausdorff space, in which points are separable from each other by their open neighborhoods. For different tasks, they can be perceived from angels and scopes to understand. Thus, it is called the perceived neighborhood. This psychological treatment is not new. Perception is also used by Einstein in general theory of relativity. In the principle of equivalence, he states that one cannot perceive the difference between the local inertial mass and the gravitational mass.

1.4. Reasoning

In philosophy, logic as a formal language is regarded as the structure of thought or the language of thought. In psychology of reasoning, mental logic theory claims that people reasoning by applying inference schemas akin to formal rules in logic. Mental models theory claims that people reason by constructing mental models based on their understanding of the meanings of premises. It makes the distinction of cognitive routines and illusory inferences.

In a sense, artificial intelligence started from machine reasoning. The early expert systems are mostly on how to program or train machines to reason. The traditional approach to artificial general intelligence (AGI) follows this trend. Despite the debate on what could be counted as machine intelligence, reasoning is regarded as a major capacity of not only human intelligence but also artificial intelligence.

From meta-theoretical perspectives, logic permeates artificial intelligence. The traditional AGI approach is based on deductive reasoning. The LLM approach is based on abductive reasoning. In addition, artificial intelligence as an empirical science is based on inductive reasoning.

Recent efforts to endow LLMs with general reasoning capabilities have increasingly focused on reinforcement learning (RL), which enables neural networks to discover solutions – i.e., to "reason" – by actively and heuristically exploring the solution space, guided by feedback signals. For tasks such as playing Go, solving math problems, or writing computer code, the feedback signals (e.g., winning or losing a game of Go) are clear and objectively verifiable. However, in many other domains – such as writing a beautiful poem – the feedback is inherently subjective and difficult to define and

 doi:10.20944/preprints202505.1277.v1

obtain. The assumption is that the logical reasoning abilities developed through training on tasks with verifiable outcomes can generalize and transfer to broader, less well-defined domains.

1.5. Decision Making

Decision making is an everyday task to human life. Artificial intelligence also needs to make decisions all the time. For example, a testing item in standard educational tests has multiple options to choose but with only one best answer as the correct answer. Thus, solving a testing item involves decision making. In artificial intelligence, the training process of a neural network involves decision making. For example, for a given threshold, inverse testing is a decision process.

1.6. Learning

Just as humans learn from experience, enabling machines to learn from training data is a cornerstone of modern AI. Techniques such as supervised learning, unsupervised learning, and reinforcement learning are commonly used to train AI models, including LLMs. However, human learning and machine learning differ fundamentally in both structure and function. Even when performing the same task, humans and machines tend to understand it differently, follow different solution paths, and employ different strategies to integrate information. Learning, in both cases, involves generalizing from memory of trained problems to solve new problems. Yet the ways in which knowledge is stored and recalled—and the speed at which it is acquired—are vastly different between humans and machines. For instance, the scaling laws observed in LLMs (Kaplan et al., 2020), which suggest that performance improves predictably with more data and compute, do not clearly map onto human learning. It has been shown that humans can often learn far more efficiently from very limited data (Tenenbaum et al., 2011).

2. Embodiment: Efforts and Challenges

2.1. Gödel, Tarski, and Turing

In 1900, the 2nd World Congress of Mathematicians, Hilbert proposed 23 open conjectures. His second conjecture is on the logic foundation of the fast-growing mathematical architecture. In 1931, Gödel proved the incompleteness theorem for the first order theory. The theorem says that if a mathematical system is rich enough, then it is incomplete. Here, the richness of the first order theory means the integration of two components, first order logic and Peano arithmetic. Gödel created a numbering method to bridge the two components and defined the notion of expressibility, which enabled him to construct a self-referential statement. This self-referential statement is proven to be independent of the first order theory. This is the so-called independent result. The incompleteness theorem is a direct corollary. This story tells us that richness sometimes is the precondition of finding the limitation. Gödel's independent result is a purely syntactic result, which sets up the foundation of Proof Theory in mathematical logic.

In logic, the logical meaning of a statement is its truth value, being either true or false. For propositional logic, its semantics is the truth semantics. In 1933, Tarski proved that truth cannot be defined as a predicate. This is the well-known undefinibility theorem, which says that if define truth as a predicate, its model would be null. which states that arithmetical truth cannot be proved in arithmetic.

Like Gödel and Tarski, Turing was motivated to understand the fundamental limitations of computation. In 1936, he introduced the concept of universal Turing machine, and asked a foundational question: What cannot the machine do? One answer was provided by the halting problem, which is closely related to Gödel incompleteness theorem. Despite such theoretical limits, the boundaries of computation have continued to be pushed, as researchers pursue capabilities that go beyond the constraints implied by the halting problem. In 1950, Turing revisited his inquiry and posed a new question: Can machines think? In that context, he proposed the now-famous Turing test

 doi:10.20944/preprints202505.1277.v1

as a practical substitute for the original question, which he argued was too vague and ill-defined to address directly (Turing, 1950). Today, from many perspectives, artificial intelligence has moved beyond the scope of the Turing test, raising new challenges regarding what it means for machines to exhibit intelligent behavior.

2.2. Understanding and Memory

The psychological concepts such as attention, sensation, perception, and reasoning discussed in the last section have been, to some extent, construed in the structure of artificial intelligence; thus, these concepts can be regarded as having structural definitions in AI. Some concepts have functional definitions but with different structures. For example, memory is well-defined in psychology and AI. In psychology, there are well documented distinctions between working memory, short memory, and long memory. However, obviously, there are no need to make these distinctions in AI. Similarly, discourse processing and text comprehension also belong to this category. Some concepts in psychology are commonly available, such as understanding, but controversial in AI. Whether AI has the capacity to understand what it is doing is currently a topic being debated. Currently, machine understanding is still a working hypothesis with an existential definition, but without a functional definition. Some even argue for that the machine can be emotional. This is a kind of preferable definition. It is fine to assume the above psychological states to be modified in artificial intelligence as our working hypotheses, but currently they are lacking clear distinctions

2.3. What Is Mind

As John Searle (1984) points out, if one claims he is talking about human mind, there are four properties must be discussed, that are subjectivity, intentionality, causality, and consciousness. In artificial intelligence, currently subjectivity is discussed mostly as a philosophical enquiry. There are many discussions about causality, which is treated as an emerging property. This is a kind of phenomenological and existential definition. Note that the term subjectivity, causality, and intentionality are with suffix “ty”, which means these are intrinsic properties of mind. Differently, consciousness is with the suffix “ness”, which means it is a philosophical metaproperty of mind. It is interesting to look at the difference between intentionality and consciousness.

Some author uses the terms intentionality and consciousness interchangeably (Dennett, 1991). However, there is deeper distinction between the two terms. Intentionality always carries contents. In mathematical set theory, intension defines the membership of a set. In epistemology, the notion of intentionality is associated with certain contents and carries the contents over during a particular mental process. There many definitions about consciousness. In epistemology, one definition is by Menong (Aquila, 1977), which reads: Consciousness is a kind of irreducible directedness being through some intentional contents, toward some possible object without requiring the existence of that possible object. Here, consciousness is characterized by three metaproperties: namely, directedness, throughness, and towardness. Note that to go through some content does not mean to carry the content over. In this sense, consciousness is directional but contentless (massless). This property indicates that, consciousness can be regarded as a mental force akin to light, which enables consciousness to travel with the highest speed in the mental world. This is a functional as well as structural definition. Consider to build a QED model for artificial intelligence, it needs to introduce the invariance skin to the light (photon). Consciousness would be able to serve that structural requirement.

Here, the new issue is, can Searle’s four conditions of mind be embodied in artificial intelligence? Another way to address is, can these cognitive capacities be embedded in computer? There are many controversies and different philosophical enquiries. If we answer yes, the question is, are human intelligence and machine intelligence the same? If they are different, how to clarify the distinction between the two?

 doi:10.20944/preprints202505.1277.v1

3. Working Definitions vs Philosophical Enquiry

Instead of making philosophical enquiries, as an alternative approch, we propose the following hypothesis:

Postulate 1. We assume as our working hypothesis that machine intelligence exists. Artificial intelligence consists of human intelligence and machine intelligence. We call the former the proper intelligence, and the latter the anti-intelligence.

Machine intelligence can be seen as a special kind of Platonic reality. In neural network, Platonic reality happens in hidden layers, called Platonic representation. Yang (2025) applied this idea in artificial intelligence (LLM) modeling. We made a clear distinction of the human intelligence (α) and the machine intelligence (β). Further, we make a mathematical trivialization treatment such that α and β are orthogonal:

$$\delta_{\alpha\beta} = \begin{cases} 1, \alpha = \beta \\ 0, \alpha \neq \beta \end{cases}$$

Now we define α as the intelligence demand (D_α) and β as the intelligence supply (S_β). For an any given task φ, we have

Definition. $e^- = [D_\alpha, \varphi^-]$, and $e^+ = [S_\beta, \varphi^+]$.

The above definition can be characterized as a Wyle spinor below,

$$\psi = \begin{pmatrix} \psi_1 (= e^-) \\ \psi_2 (= e^+) \end{pmatrix}.$$

4. Conclusion

Artificial intelligence is not only some technology but a science. Current theories of artificial intelligence are mostly phenomenological, which can be characterized as statistical AI. As a science, it is lacking a basic theory. To build a scientific theory needs hypothetical working definitions. For example, in the early age of quantum chromodynamics (QCD), Gell-Mann introduced the working definition of the color charge as an intrinsic property of quarks. His motivation is purely theoretical to make QCD consistent with the Pauli exclusion principle. In quantum field theory, there are many non-physical parameters (Wang, 2008). These parameters are hypothetical but with clear definitions. This part of scientific theories is named theoretical scaffold, which is necessary in order to build a great theoretical architecture.

Simmel (1978/2004) once wrote: Money never lacks energy; when people think of money, they converted their energy into money. Money never lacks intelligence; when people spend money, they construed their intelligence into money. In the AI world, machine intelligence is the money.

When we are trying to embed psychology into artificial intelligence, we need to be careful about several things. First, from logic perspectives, conception theory tells us that the reconceptualization needs to go through a process from intuitions to ideas, concepts, notions and definitions. Second, from cognitive perspectives, machine cognition has different characters that are different from human cognation. These need to be identified. Third, from philosophical perspectives, it is crucial to disclose the epistemological structures and mental presentations as well as mental processes in artificial intelligence before making ontological commitments too quickly. Finally, if we want to embody psychology into artificial intelligence, it had better to embody psychological method together. From methodological perspectives, AI is in nature an empirical science and induction is its basic logic. Computing is not only digital, but also involves inductive inference, observation, judgement, decision making, etc. Induction follows the low of marginal decay. Thus, over statements would mislead to illusory predictions.

References

1. Aquila, R. (1977). Intentionality: A Study of Mental Acts. Pennsylvania State University Press.
2. Dennett, D. (1991). Consciousness Explained. Little Brown Co.

 doi:10.20944/preprints202505.1277.v1

3. James, W. (1890). *Principles of psychology*. Holt.
4. Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.
5. Searle, J. (1984). Minds, Brains, and Programs. Harvard University Press.
6. Simmel, G. (1978/2004). The Philosophy of Money. Routledge.
7. Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460.
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*, S0140525X16001837.
9. Wang, Z. X. (2008). Elementary Quantum Field Theory. Peking University Press (China).
10. Yang, Y. (2025) Maxwell and Artificial Intelligence: Preliminary QED Models of AI (LLM) Dynamics. Preprint, DOI: https://doi.org/10.20944/preprints202504.2624.v1