What Do You Mean by “Open World”?

Bowen Xu
Department of Computer and Information
Temple University
Philadelphia, PA 19122, USA
bowen.xu@temple.edu

Abstract

This paper explores the different understandings of the term “open world” in AI research, categorizing them into *partially open world* and *fully open world*. By distinguishing between *intelligence* and *skills*, I argue that a *fully open world* is essential for evaluating *intelligence*, while a *partially open world* risks shifting focus towards problem-specific *skills*. However, both *partially open* and *closed worlds* can still be valuable research tools if the underlying assumptions of AI systems are carefully examined.

1 Introduction

The real world is undoubtedly open and dynamic. However, based on historical experience in the field of AI, enabling machines to autonomously engage with the real world in a comprehensive manner is not something that can be achieved in a single leap. As a research strategy, it is understandable to simplify complex problems and begin with simpler tasks. However, this approach often results in specific solutions that are tailored to narrowly defined problems. Despite the impressive achievements in AI, there are always critics who argue that these systems do not represent “real AI” – a sentiment often referred to as the “AI effect” [3]).

In recent years, there has been a growing interest among researchers in developing AI systems that can learn and operate in open-world environments (*e.g.*, [14], [6], [7]), even if their motivations differ. This shift in research focus represents a significant step forward for AI, especially when considering the ultimate goal of creating truly “intelligent” machines – those capable of independently interacting with the complexities of the real world. In some contexts, this is referred to as Artificial General Intelligence (AGI) [13, 12].

The concept of the “open world” has attracted attention for various reasons. For instance, it offers the opportunity to test aspects of intelligence that cannot be adequately assessed using predefined datasets [8], to advance AI research through more complex environments [1], or to evaluate AGI systems [14]. Despite the diversity of motivations, the underlying goal remains consistent: to establish a new paradigm for benchmarking “real” AI.

In this paper, I discriminate the different interpretations of the term “open world”.

2 Understandings to “Open World”

Before delving into the concept of “open world”, it is useful to first consider its opposite: what is not an open world? Artificial environments, such as the game of Go or Atari video games, are defined by

[1] In the “Open-World Agents” workshop (`https://sites.google.com/view/open-world-agents`), it proposes “to consider open-world environments as the new habitat for AI agents: highly diverse and dynamic, fully interactive, teaming up with infinite and creative tasks, and requiring continuing learning and growth.”

The Workshop on Open-World Agents (OWA-2024), hosted with NeurIPS 2024.

a strict set of rules and clear, predefined goals. These environments are highly constrained, as are other real-world examples like the work environment of robots on a factory production line, where circumstances and tasks are rigidly structured.

The term "open world" is sometimes associated with video games like *Minecraft*, where players enjoy a certain degree of freedom to explore and create. The environment is dynamically generated based on a set of underlying "physical" rules. While there may be tasks or objectives within these environments, players often have the freedom to pursue their own goals. This openness resembles the complexity of the real world, which is likely why open-world games are considered promising as testbeds for evaluating advanced AI agents. However, some studies merely use open-world games to simulate relatively closed environments, effectively recreating the controlled conditions of the real world within a digital context. Intuitively, these environments should not be classified as truly open.

What, then, fundamentally distinguishes an open world from a closed one? I argue that the key differences lie in two conditions: (1) The future is not necessarily consistent with the past and may, sometimes, differ significantly. (2) The tasks or objectives in an open world are not predetermined or designed in advance. In such environments, a thinking machine must rely on its own intelligence to perceive, adapt, and reshape the world, rather than depending on problem-specific knowledge provided by humans.

In the following, I will clarify this understanding, explaining how it can help differentiate various types of environments and guide future research in AI.

The Open World in Reality

Our ultimate goal is to create thinking machines that can be applied in real-world contexts. The real world serves as the ultimate testing ground for AGI systems. However, in current AI research, these systems are often tested in highly restricted, virtual environments. The key question is: what properties distinguish the real world (an open world) from these artificial environments, so that the former can be considered truly open, while the latter cannot?

In this paper, I argue that an open world is characterized by two properties:

1. Time and space are *unbounded*: The future of the world may not align with the past. In other words, an agent's experiences may not be directly applicable to future scenarios.
2. Tasks are *unbounded*: A wide range of tasks needs to be addressed, and these tasks are not predetermined or defined at the agent's birth.

These two properties reflect the challenges humans face in daily life. We cannot predict exactly how the world will change, and we must constantly adapt to new situations. Similarly, we do not know what challenges lie ahead, yet we possess the potential to address them.

Some may argue that the natural laws discovered through physics allow us to predict the world with a high degree of accuracy. My response to this view is that these laws are, ultimately, human perceptions and descriptions, and their logical correctness cannot be guaranteed. As Hume argued [5], no matter how extensive our past experience may be, our predictions about the future can never be infallible. This is the reality humans face every day. Even our most successful scientific theories cannot be logically guaranteed to hold true in the future. Consequently, integrating human knowledge as absolute truths into an AI system is risky, as these systems may encounter situations where human theories no longer apply.

For narrow AI or computer programs designed to perform specific tasks, this uncertainty is not a problem, as their application domains are typically well-defined. However, for more general AI systems, like AGI, both humans and machines must confront a world where knowledge from the past may be challenged by the future.

In everyday life, most tasks are not specified before we are born, with the exception of basic physiological needs. This principle also applies to general-purpose AI systems. A useful metaphor is that a human baby can grow up to become an expert in any field through appropriate education. A wide variety of tasks arise from the environment – some are self-imposed, while others are acquired from external sources, such as society. Humans are capable of engaging with these tasks, even though they may not always successfully solve every problem. Similarly, AGI systems should also be open to a broad array of tasks, with the potential to address them.

From these two core properties, other requirements can be derived. For instance, the learning process should be "continual," "online," and "life-long" [4]; AI systems should be capable of reasoning under uncertainty and driven by values or goals, allowing them to acquire and pursue objectives that they have never previously encountered.

If we take these two properties – unbounded time and space, and unbounded tasks – as the essence of an *open world*, we can categorize environments into three types: *closed world*, *partially open world*, and *(fully) open world* (which will be discussed in the following sections).

Types of World

In AI research, the term "open world" is sometimes used in a way that differs from its meaning in reality, primarily because certain characteristics are missing. When evaluating AI systems, we can consider three types of worlds:

Closed World Many machine learning benchmarks fall under the category of a *closed world*. These environments fail to meet either of the two key requirements for an *open world*. In *closed worlds*, time and space are bounded, meaning the agent is tested within a fixed and limited range. The environment assumes background knowledge and the future is strictly consistent with the past. Additionally, tasks are pre-designed and defined by human researchers, so the agent is only tested within a specific set of tasks.

Partially Open World Some environments can be categorized as *partially open worlds* because they meet only some of the *open-world* requirements. Most current AI research that claims to work in "open world" environments is actually dealing with *partially open worlds*. For instance, DeepMind's *XLand* environment is designed to explore "open-ended learning" [9]. In *XLand*, both tasks and worlds are generated according to predefined rules. While the number of possible tasks and worlds is theoretically infinite, they are still bounded in the sense that the future follows consistent patterns from the past, and the tasks are pre-defined enough that human knowledge can be used to solve them. Similarly, environments like *Minecraft* or its 2D version, *Crafter* [8], allow agents to create objects based on rules, but they do not meet the requirements for the same reason.

Another example is self driving or autonomous driving. The real world faces self-driving cars, and no one could guarantee that unexpected circumstance will not occur. However, the tasks of self-driving cars are restricted to a small set, for example, to navigate to a destination without breaking traffic rules.

Fully Open World A *fully open world* satisfies both of the previously discussed requirements. Given that any artificial system has finite resources, an AI agent operating in such an environment must cope with these open-world challenges using limited computational capacity. The agent must apply finite resources to manage the complexity and unpredictability of an infinite world.

In fully open worlds, tasks are not predefined by human developers, so researchers must focus on problem-independent aspects of intelligence. This concept aligns with the idea of *general intelligence* [13], or simply *intelligence* – the fundamental principles and mechanisms that allow an agent to perceive, interpret, and modify the world. Researchers must explore invariant principles across tasks while enabling the agent to learn specific skills to handle diverse, evolving challenges.

It is important to note that while the environment may sometimes be relatively stable, the agent must still be capable of handling fundamental changes in its surroundings. An intelligent machine must be prepared for the possibility that even the laws it perceives to govern the world could change. This flexibility is crucial, as human knowledge is often subject to revision when challenged by new experiences. Examples of this phenomenon are abundant in the history of science, such as the shift from the geocentric model to the heliocentric model, or the transition from classical physics to quantum physics. The correctness of scientific knowledge can never be guaranteed logically.

Whether we assume that the laws governing the world are unchanging but beyond the machine's reach, or that no such unchanging laws exist, the implication for AI remains the same: intelligent machines must be theoretically prepared for these potential changes, rather than relying on the assumption that the world's rules are fixed and pre-programmed by human developers.

It is worth noting that intelligent agents may still attempt to find stable patterns or laws, despite the risk of being wrong. Learning from past experiences is, after all, the primary way that agents – both human and artificial – adapt to the world. Yet, only when the environment is relatively stable can we evaluate an agent's ability to adapt as "good."

As discussed in [14], one of the primary challenges in creating a fully open world for evaluating AGI is avoiding the "trap of developers' experience." This refers to the risk that human developers, by imposing their own knowledge and assumptions, inadvertently constrain the system, limiting its capacity to truly engage with an *open world*.

3 The Necessity of Distinction

To understand why distinguishing between different types of worlds is crucial, let's first analyze where the ability to perform tasks originates. For humans, expertise in performing tasks stems from having the necessary solutions, *i.e.*, knowing how to achieve a specific goal step by step. A common belief is that the more sophisticated a person becomes, the better they perform tasks. However, what is often overlooked is the process of acquiring these skills – every task, no matter how simple, requires a learning phase. Even for something like an IQ test, adolescents have gathered vast amounts of experience, including basic sensorimotor knowledge, since birth.

Intelligence can be viewed as a meta-capability – the ability to acquire concrete, problem-specific solutions and world knowledge (referred to as "skills" here). The distinction between *intelligence* and *skills* is similar to the distinction between "g-factors" and "s-factors" in psychology [1]. This perspective is also supported by prior AI research (*e.g.*, [10], [2]). In short, *skills* vary across tasks, while *intelligence* is the constant, underlying capability. If we represent this relationship in a causal graph (see Fig. 1a), the "direct cause" of task performance is "skills," while "intelligence" leads to the development of those skills.

For humans, saying that *skills* cause task performance is essentially the same as saying *intelligence* causes task performance, as *intelligence* is the origin of *skills*. However, for machines, the situation is different.

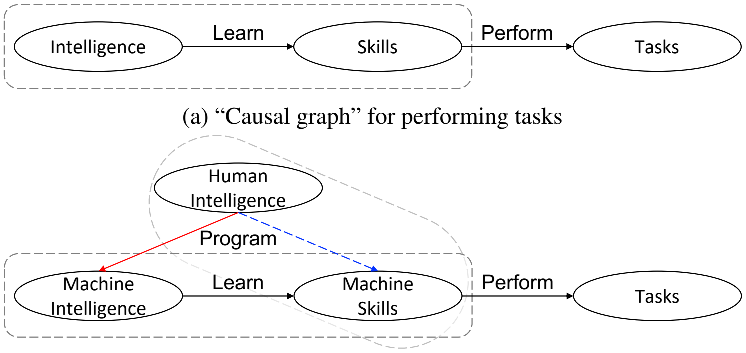
A machine can perform tasks for two primary reasons (see Fig. 1b): either a human programmer solves the problem and hard-codes a task-specific solution into the machine, or the machine learns to solve the problem autonomously. In the first case, the human and machine can be viewed as a single system, with the human *intelligence* being the true source of the solution. In the second case – especially relevant for AGI systems – the machine's own intelligence enables it to learn how to perform tasks. In this situation, machine intelligence is responsible for solving problems independently, with human intelligence acting as a *confounder*, making it easy to misunderstand the origin of the system's abilities.

Many AI projects blur this distinction by embedding human task-specific experience into machines. As a result, such systems are best categorized as narrow AI, deviating from the ultimate goal of AGI. By keeping the distinction between intelligence and skills in mind, I argue that differentiating between the two types of "open world" (*i.e.*, *partially open world* and *fully open world*) will steer research in distinct directions.

Firstly, without the constraint of unbounded space and time, human developers may embed their world knowledge into machines, creating systems incapable of adapting to situations where human knowledge does not apply. Secondly, without the constraint of unbounded tasks, machines may only be able to solve predefined problems, rendering them ineffective when encountering novel tasks. This limitation often leads researchers to solve problems manually, resulting in performance improvements that come from human intervention rather than the machine's own *intelligence*.

It is much easier to seek out problem-specific knowledge than to seek problem-independent principles. The former requires us to observe our own beliefs and goals and embed them into machines. The latter, however, demands that we investigate the origins of beliefs and goals, as well as the mechanisms by which they arise—essentially, understanding intelligence itself.

In *closed-world* scenarios, task-specific solutions often outperform more general systems. In *partially open worlds*, human prior knowledge can enhance performance as well. For instance, if a system is



(a) "Causal graph" for performing tasks

(b) "Human intelligence" as a confounder

Figure 1: Sources of performance on tasks

designed to solve only a few predefined problems, it is less likely to be distracted by irrelevant goals, whereas a general system might struggle with focus in certain situations.

However, adapting to a *fully open world* becomes crucial if we want AI systems to explore environments beyond our experience, such as outer space or alien planets. In these cases, the environment lies outside the scope of human knowledge, and machines must rely on their own intelligence to navigate and adapt.

4 Response to Potential Objections

(1) "Open world" defined herein is not useful for practical research because we cannot construct a benchmark that meets these requirements.

Designing and creating a virtual *open world* does present challenges. However, researchers can still consider it a relevant scenario in which AI systems will ultimately be applied. Pursuing this goal encourages researchers to investigate the essence of *intelligence*, rather than merely focusing on *skills* or solutions to specific problems.

(2) Machines are probably unable to adapt to an open world because they cannot handle undefined problems.

Currently, most AI systems are ill-equipped for *open worlds*, primarily because they struggle with unbounded tasks. For instance, the objective of a deep neural network is typically to minimize a loss function, which is designed based on the specific problem(s) at hand. While deep learning can be seen as a general algorithm applicable to a variety of optimization problems (when the loss functions are appropriately defined), it is not a truly general system. Once instantiated, a deep neural network's functionality becomes fixed.

The concept of an *open world* serves as a reminder to consider the motivation mechanisms of AI: the goals of an AI system should emerge from its interactions with the environment. Such systems, exemplified by human beings and [11], can indeed be created.

(3) We disagree with the assertion that the two properties represent the essence of an open world, as there are additional characteristics of the real world.

The real world encompasses many other attributes, such as high dynamism and the presence of multiple interacting agents. Definitions aim to clarify distinctions between concepts, and it is natural for different individuals to have varying interpretations of the same term, which may refer to different concepts. I am not opposed to alternative definitions of "open world." This paper could be viewed as a starting point for researchers to define "open world" in the context of AI agents interacting within it. Researchers should clarify their usage of the term – *e.g.*, "open-world agents" – to avoid ambiguity.

(4) We need environments that test cognitive capabilities, even if they are not open worlds.

Benchmarking AI systems through specific problems remains valuable, as it simplifies the environment and allows researchers to focus on particular aspects of intelligence temporarily. However, solving specific problems alone is insufficient for preparing AI systems for *open-world* applications. We must be cautious when evaluating AI systems against such benchmarks: after addressing specific problems, what elements of the system remain general?

5 Conclusion

In this paper, I have distinguished different understandings of the term "open world," categorizing them into *partially open world* and *fully open world*. By clarifying the distinction between *intelligence* and *skills*, I argue that a *partially open world* poses the risk of steering research towards *skills*, which are inherently problem-specific.

Nevertheless, *partially open worlds*, and even *closed worlds*, can still serve as valuable tools for research, provided that we critically assess the assumptions underlying our theories. This ensures that our findings can be applied to a *fully open world* as an idealized scenario.

Ultimately, embracing the concept of a *fully open world* challenges researchers to develop AI systems that not only perform specific tasks but also exhibit *intelligence*. By shifting our focus from narrow *skills* to the principles of *intelligence*, we can pave the way for the creation of more robust and adaptable AI systems capable of navigating the complexities of the real world.

As we move forward in AI research, it is imperative to maintain clarity in our definitions and to strive for frameworks that better reflect the dynamic and unpredictable nature of the environments AI systems will encounter. This will enable us to develop AI that is not only competent but also resilient in the face of new challenges.

References

[1] Christopher Brand. *The g Factor: General Intelligence and Its Implications*. Chichester, 1996.

[2] François Chollet. On the Measure of Intelligence, November 2019. arXiv:1911.01547 [cs].

[3] Michael Haenlein and Andreas Kaplan. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4):5–14, August 2019.

[4] Steven C. H. Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. Online learning: A comprehensive survey. *Neurocomputing*, 459:249–289, October 2021.

[5] David Hume. *An Enquiry Concerning Human Understanding*. London, 1748.

[6] Mayank Kejriwal, Eric Kildebeck, Robert Steininger, and Abhinav Shrivastava. Challenges, evaluation and opportunities for open-world learning. *Nature Machine Intelligence*, 6(6):580–588, June 2024. Publisher: Nature Publishing Group.

[7] Vikash Sehwag, Arjun Nitin Bhagoji, Liwei Song, Chawin Sitawarin, Daniel Cullina, Mung Chiang, and Prateek Mittal. Analyzing the Robustness of Open-World Machine Learning. In *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security*, AISec'19, pages 105–116, New York, NY, USA, 2019. Association for Computing Machinery.

[8] Aleksandar Stanić, Yujin Tang, David Ha, and Jürgen Schmidhuber. Learning to Generalize With Object-Centric Agents in the Open World Survival Game Crafter. *IEEE Transactions on Games*, 16(2):384–395, June 2024. Conference Name: IEEE Transactions on Games.

[9] Adam Stooke, Anuj Mahajan, Catarina Barros, Charlie Deck, Jakob Bauer, Jakub Sygnowski, Maja Trebacz, Max Jaderberg, Michael Mathieu, Nat McAleese, Nathalie Bradley-Schmieg, Nathaniel Wong, Nicolas Porcel, Roberta Raileanu, Steph Hughes-Fitt, Valentin Dalibard, and Wojciech Marian Czarnecki. Open-Ended Learning Leads to Generally Capable Agents, July 2021. arXiv:2107.12808 [cs].

[10] Pei Wang. On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, 10(2):1–37, August 2019.

[11] Pei Wang. Intelligence: From Definition to Design. In *Proceedings of the Third International Workshop on Self-Supervised Learning*, pages 35–47. PMLR, April 2022. ISSN: 2640-3498.

[12] Pei Wang and Ben Goertzel. Introduction: Aspects of Artificial General Intelligence. In *Proceedings of the 2007 conference on Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms: Proceedings of the AGI Workshop 2006*, pages 1–16, NLD, June 2007. IOS Press.

[13] Bowen Xu. What is Meant by AGI? On the Definition of Artificial General Intelligence, April 2024. arXiv:2404.10731.

[14] Bowen Xu and Quansheng Ren. Artificial Open World for Evaluating AGI: A Conceptual Design. In Ben Goertzel, Matt Iklé, Alexey Potapov, and Denis Ponomaryov, editors, *Artificial General Intelligence*, Lecture Notes in Computer Science, pages 452–463, Cham, 2023. Springer International Publishing.