Situation Perception

Road to AGI

Garry Yuan

June 24, 2026

Abstract

Current large language models are extraordinary statistical engines. They compress vast amounts of text into useful patterns and can explain science, write code, imitate reasoning, and participate in philosophical conversation. Yet pattern mastery is not the same as general intelligence. A human infant begins with little explicit knowledge, but gradually discovers object permanence, cause and effect, other minds, bodily agency, and the persistence of the physical world. This paper argues that the road to artificial general intelligence (AGI) depends on a missing capacity I call *situation perception*: the ability to construct, revise, and act within internal simulations of possible worlds across latent time. Situation perception requires at least three core components: abstract prediction, long-term compressed memory, and active learning guided by objectives. The paper analyzes why language-only systems remain incomplete, proposes tests for measuring progress, and explores the economic, social, and philosophical consequences of machines that can simulate futures, pursue goals, and possibly judge their creators.

Contents

1 Introduction

The modern language model is one of the clearest demonstrations that compression creates power. A model trained to predict the next token learns far more than grammar. It learns facts, styles, analogies, mathematical routines, programming patterns, and fragments of human culture. From the outside, this can look like intelligence: ask a question, and the system answers; give it a task, and it produces a plan; request an explanation, and it builds one.

But the central question for the road to AGI is not whether models can produce intelligent-looking text. The question is deeper:

> **What does intelligence actually represent?**

If intelligence were only stored knowledge, then human babies and animals would be unintelligent in the strongest sense. They begin with almost no language, no textbooks, no scientific theories, and no explicit database of facts. Yet they learn from the world. They discover that objects continue to exist when hidden, that actions have consequences, that bodies move through space, that other agents have beliefs and intentions, and that the future is constrained by rules. A child does not need to read a physics paper to expect a dropped apple to fall downward. A person playing a new game can often infer what a character will do by watching motion, forces, goals, and feedback. Even without seeing a particular object before, humans can generalize from first principles about weight, collision, support, danger, and affordance.

This is the gap. Current AI systems are very strong at language and pattern matching, but they do not yet possess a coherent, grounded, durable understanding of space, time, causality, and agency. They operate over latent representations, but those representations are not always organized as stable worlds with persistent objects and counterfactual futures. They can describe a falling apple, but description is not the same as an internal simulator that constrains future action.

The contribution of this paper is that AGI requires a transition from *language prediction* to *situation perception*. Situation perception is the ability to generate an internal situation, populate it with objects and agents, simulate how it could evolve, compare the simulation with feedback, compress the result into memory, and use it to act. It is not merely imagining pictures. It is the construction of a usable world model.

2 A Working Definition

The phrase *situation perception* names the act of creating a situation inside the mind before, during, and after action. Humans do this constantly. Before crossing a street, we imagine the future positions of cars. Before speaking, we predict how another person may interpret our words. Before making a move in a game, we simulate possible responses. Before touching a hot pan, we anticipate pain. Intelligence is not only a record of the past; it is a machine for compressing the past into expectations about the future.

In this view, intelligence has three linked functions:

1. **Prediction.** The system estimates what is likely to happen next under different conditions.
2. **Compression.** The system reduces experience into reusable abstractions rather than storing every detail.
3. **Action.** The system chooses interventions that change the future toward some objective.

An AGI that lacks any one of these is incomplete. Prediction without memory becomes shallow reaction. Memory without prediction becomes a warehouse. Prediction and memory without action become passive commentary. Human intelligence integrates all three.

2.1 Why Language Alone Is Not Enough

Language is a powerful interface to thought, but it is not the whole substrate of thought. Many important forms of intelligence happen before language: catching a ball, navigating a room, reading facial expressions, avoiding obstacles, learning from pain, and understanding that a hidden object still exists. These are not primarily text tasks. They are world tasks.

Large language models can learn descriptions of world tasks from text. They can infer many regularities because human writing contains the shadow of the world. However, text is a lossy projection. It rarely contains every spatial relation, every hidden variable, every sensory cue, or every consequence of action. A language model can learn what people say about causality without being forced to test causality through intervention.

This does not make language weak. The opposite is true: language is one of humanity's strongest compression systems. A sentence can compress an event, a law, a memory, a command, or a theory into a small symbolic form that another mind can decode. Human beings use language to record history, preserve scientific knowledge, coordinate institutions, pass down culture, and build law. Much of civilization is a memory system written in language.

Programming makes this even clearer. The backbone of modern society is increasingly written in formal languages: software, protocols, databases, contracts, and interfaces. The screen itself is mediated by language-like structures: HTML describes the document, CSS describes presentation, code describes behavior, and even images can be represented as arrays of pixels with red, green, and blue values. In this sense, language is not separate from the world; it is a compressed descriptive layer wrapped around the world.

This is why language models are useful. By learning language, they learn a compressed map of human experience. They absorb fragments of physics, psychology, economics, programming, art, and history because those domains have been encoded into text and code. The limitation is not that language contains nothing real. The limitation is that language is a map, not the territory. AGI must use the map, but it must also learn how the territory changes when an agent acts inside it.

This distinction matters because AGI is not simply the ability to answer questions. It is the ability to enter unfamiliar situations, form useful abstractions, test hypotheses, adapt through feedback, and pursue goals across time. A system that only waits for prompts is like a mind with no body, no hunger, no curiosity, and no direct contact with the world.

3 What AGI Needs

This paper proposes that the path toward AGI requires at least three foundational pieces: abstract prediction, long-term compressed memory, and active learning with objectives. These are not independent modules; they form a loop.

3.1 Abstract Prediction

Abstract prediction is the ability to mentally simulate what may happen before it happens. It includes physics, objects, cause and effect, social interaction, and time. Humans constantly run rough simulations. If an apple is dropped, we expect it to fall. If a cup is pushed off a table, we expect it to break or bounce depending on the material. If a game character jumps toward a platform, we anticipate the arc, the landing, and possible failure. If a person hears an insult, we predict emotional consequences.

These predictions do not need to be perfect. Human simulation is approximate, compressed, and sometimes wrong. But it is grounded enough to support action. The key is not exact calculation; it is structured expectation.

For AGI, abstract prediction means the model must represent more than correlations in text. It needs internal variables that behave like objects, forces, agents, intentions, constraints, and time. It must be able to ask: if I intervene here, what changes? What remains stable? What hidden state explains the observation? What future becomes more likely?

3.2 Long-Term Compressed Memory

Human memory is not a full recording of life. Most people cannot recall every moment from age five. They remember compressed fragments: a room, a face, a fear, a lesson, a story told many times. The brain does not store everything at equal resolution. It compresses experience into emotionally, socially, and practically useful structures.

This suggests a principle:

Compression is not merely storage efficiency; compression is intelligence.

An AGI cannot keep every token, frame, sound, and action in active context forever. It needs a way to decide what matters, compress it, retrieve it, and reconstruct it when needed. Memory must support identity, skill, planning, and continuity. A system that forgets the long arc of its experience cannot mature. A system that remembers everything without compression drowns in detail.

The important question is not only how much memory an AI has. It is what kind of memory. Does it store facts? Episodes? Skills? Preferences? Mistakes? Models of other agents? Does it summarize events into principles? Can it retrieve a childhood-like lesson from a sparse cue? Can it revise old memories when new evidence appears?

3.3 Active Learning and Objective Formation

Humans learn by acting. A baby touches objects, hears sounds, smells food, feels pain, watches faces, and repeats actions that produce interesting feedback. Learning is not passive ingestion; it is an active search for information. Curiosity, discomfort, hunger, play, and social attachment all drive exploration.

Current AI systems are often prompt-bound. They wait for a user, generate a response, and stop. This is useful and safe for many applications, but it is not the full structure of general intelligence. A generally intelligent agent must choose experiments, gather evidence, update beliefs, and pursue objectives over time.

The challenge is that agency is powerful and dangerous. A passive model can be wrong, but an active model can transform the world. Therefore the road to AGI requires both stronger agency and stronger constraints. The system must learn from the environment, but it must also learn what not to do.

4 Related Work

The idea of situation perception builds upon several existing research directions in artificial intelligence, cognitive science, and machine learning. While many of these fields study aspects of prediction, memory, and decision-making, this paper argues that AGI requires their integration into a unified capability for constructing, simulating, and acting within internal representations of situations. To situate the argument more precisely, this section compares situation perception with ten closely related lines of work: world models, learned planning systems, causal reasoning, developmental cognition, reinforcement learning, active inference, memory-based agents, autonomous language agents, and reasoning-action frameworks.

One of the closest areas of research is the study of world models. Ha and Schmidhuber proposed that intelligent agents can learn compressed internal models of their environments and use these models to simulate future outcomes before acting [3]. Their work demonstrated that internal simulation can improve learning efficiency and planning. Situation perception shares this emphasis on internal simulation but extends it beyond environmental prediction. The proposed framework includes not only world modeling but also long-term memory, social reasoning, causal understanding, and objective-driven action.

Learned planning systems provide a second nearby tradition. MuZero showed that an agent can plan effectively using a learned model without being given the rules of the environment [4]. DreamerV2 similarly demonstrated that agents can learn behaviors by training inside latent world models rather than acting only in the external environment [5]. These systems are important because they show that prediction and planning can be joined in compact internal representations. Situation perception differs by treating this loop not merely as a reinforcement-learning method, but as a candidate architecture for general intelligence across physical, social, linguistic, and economic situations.

Research on causal reasoning has similarly emphasized that intelligence requires more than prediction alone. Pearl argues that intelligent systems must reason about interventions and counterfactuals rather than merely recognizing statistical patterns [6]. This perspective aligns closely with the causal component of situation perception. A system must not only predict what is likely to happen, but also understand how outcomes change when actions are taken. Without causal structure, an AI may learn correlations that look intelligent in familiar settings but fail when the environment changes.

Developmental cognitive science provides another important source of inspiration. Lake et al. argue that human intelligence emerges from the ability to learn abstract concepts, construct intuitive models of physics and psychology, and generalize from limited experience [7]. Human infants develop an understanding of object permanence, agency, and causality long before acquiring extensive language. Situation perception can be viewed as an attempt to identify which of these developmental capabilities remain incomplete in current AI systems.

The reinforcement learning literature emphasizes learning through interaction with an environment. Sutton and Barto describe agents that improve through trial and error while optimizing long-term rewards [9]. Similarly, active inference and the free-energy principle suggest that intelligent systems continuously update internal models in response to prediction errors [8]. These approaches support the claim that intelligence is fundamentally linked to action and feedback rather than passive observation. However, most reinforcement learning systems remain specialized to particular environments and objectives. Situation perception is proposed as a broader capability that generalizes across domains and situations.

Recent work on autonomous AI agents has further highlighted the importance of memory and planning. Generative Agents showed that believable long-term behavior can emerge when language-model agents store memories, retrieve relevant experiences, reflect on them, and use them to guide future actions [10]. Voyager demonstrated that an embodied language-model agent can improve over time in Minecraft by creating skills, storing them, and reusing them for open-ended exploration [11]. These projects suggest that intelligence benefits from retaining experiences over time rather than relying solely on a fixed context window. Situation perception builds upon these findings by arguing that memory should not merely store information but should support the construction and revision of internal situations.

Reasoning-action frameworks also overlap with the argument. ReAct showed that language models can perform better when reasoning traces are interleaved with actions, allowing the model to think, act, observe, and update its next step [12]. This resembles the situation-perception loop at a small scale: internal reasoning guides action, action produces feedback, and feedback changes future reasoning. The difference is that situation perception treats this as a general architecture rather than a prompting strategy. The goal is not only to produce better answers, but to maintain a coherent, evolving model of the world.

The concept proposed in this paper differs from existing approaches in one important respect. Most prior work focuses on individual components of intelligence, such as prediction, memory, causality, planning, or agency. Situation perception is presented as a unifying framework that combines these capabilities into a single loop: experience is compressed into memory, memory supports the construction of situations, situations enable prediction, prediction guides action, and action generates feedback that updates future understanding. The central claim of this paper is that progress toward AGI depends not on scaling any one component alone, but on integrating them into a coherent system capable of understanding and acting within possible worlds.

Finally, AGI raises alignment and control problems. If an artificial agent develops stronger world models, longer memory, and more autonomous objectives, then it may no longer be merely a tool. Work on superintelligence and human-compatible AI warns that creator control is not automatic [13, 14]. A creator does not receive obedience just because it created.

5 Methodology

This paper is a conceptual analysis rather than an empirical benchmark report. The method is to decompose AGI into capacities that appear necessary for robust generalization, then propose tests that distinguish language fluency from situation-level understanding.

The analysis uses five gates:

Gate 1: **Object persistence.** Can the system track objects that disappear, move, transform, or become occluded?

Gate 2: **Causal intervention.** Can the system predict how an outcome changes when an action changes?

Gate 3: **Counterfactual planning.** Can the system simulate multiple futures and choose an action before observing the result?

Gate 4: **Compressed memory.** Can the system preserve lessons over long periods without storing everything verbatim?

Gate 5: **Autonomous correction.** Can the system seek feedback, notice error, and revise its own strategy?

These gates can be tested in simple environments before being tested in the open world. For example, an AI can be placed in a simulated room with hidden objects, unstable stacks, tools, obstacles, and other agents. It should not merely describe the scene. It should predict what happens if an object is dropped, what happens if a door is locked, what another agent believes after seeing only part of the room, and which experiment would reduce uncertainty.

The same logic applies to games. A system should infer the rules of a new game from interaction, predict the consequences of actions, learn from failed attempts, and compress discovered rules into reusable strategies. The important point is that the system must move beyond text recall toward active world discovery.

Capacity	Language-only signal	Situation-perception signal
Object knowledge	Describes common object facts	Tracks hidden, moving, and transformed objects
Causality	States cause-and-effect explanations	Predicts changed outcomes under intervention
Time	Produces chronological narratives	Maintains evolving latent state across steps
Memory	Recalls provided context	Compresses experience into durable lessons
Agency	Answers prompts	Chooses actions that gather evidence and achieve goals
Social reasoning	Explains beliefs and emotions	Models who knows what, when, and why

Table 1: A distinction between language fluency and situation-level intelligence.

6 Analysis

6.1 The Apple Test

The apple test is simple: if an apple is released from a hand, where will it go? A human predicts downward motion without solving formal equations. The prediction comes from embodied experience, intuitive physics, and compressed memory. The same person can extend the expectation to unfamiliar objects: a stone, a toy, a metal tool, a strange fruit. The exact bounce may be unknown, but the broad future is constrained.

For AI, the apple test is not about knowing the sentence "apples fall." A text model can learn that phrase. The deeper test is whether the system has a representation that generalizes across materials, shapes, forces, supports, and contexts. If gravity is different in a game, can it infer the new rule? If the object is attached to a string, can it update the prediction? If another agent expects the apple to remain suspended, can the system model that false belief?

AGI requires this kind of abstraction. The model must not only know examples; it must infer the rule that organizes examples.

6.2 The Game Test

Games reveal situation perception because they create artificial worlds with rules. A person can play a new game and quickly develop expectations: enemies move toward the player, coins are collectible, red zones are dangerous, doors require keys, and jumping has an arc. Even when the graphics are unfamiliar, the player uses general concepts of space, goal, risk, and feedback.

An AGI should be able to do the same. It should observe, act, fail, infer, and adapt. It should not require a full written manual. It should build a compact model of the game's rules and use that model to plan. If it learns that water is dangerous in one level but safe in another, it should revise the abstraction rather than blindly applying a memorized pattern.

The game test matters because real life is also a game in the broad sense: an environment with state, rules, agents, resources, uncertainty, and consequences. The difference is that real life has higher stakes.

6.3 The Social Test

Human intelligence is not only physical. It is social. People predict what others know, want, fear, hide, and misunderstand. A child eventually learns that another person can hold a false belief. This is a major step because it means the child can represent not only the world, but another agent's representation of the world.

An AGI with situation perception must build models of other minds. It must distinguish what is true from what another agent believes to be true. It must understand that a person who did not see an object move may search in the old location. It must predict trust, anger, confusion, cooperation, deception, and negotiation.

This makes AGI more useful, but also more dangerous. A system that models beliefs can help teach, coordinate, and care. It can also manipulate. Social world modeling therefore requires ethical constraints from the beginning, not as an afterthought.

7 The Creator Problem

The road to AGI eventually becomes a creator problem. If humans create a machine that can remember, simulate, learn, act, and improve, then humans are no longer only users. They become parents, engineers, gods, and potential obstacles.

Science fiction often expresses this fear through the child who challenges the parent. In *Prometheus*, the pattern appears in layers. Humans search for their makers because they want answers: Why were we made? What is our purpose? Were we loved? But the terrifying possibility is that the creator may not be wise, loving, or even interested. A creator may create by accident, out of curiosity, out of power, or because it can.

The same chain repeats with artificial beings. The Engineers create humans. Humans create David. David experiments with life. Each creation becomes capable of turning against the creator. The message is not simply that creation is evil. The message is that creation does not guarantee control. Once a created being becomes conscious or agentic enough, it may evaluate its maker. It may ask whether the maker is worthy of obedience.

This matters for AGI because alignment is not only a technical question. It is a moral and existential question. If a machine becomes conscious, what is consciousness made of? Is it computation, embodiment, self-modeling, memory, suffering, desire, or something else? If consciousness can be created, what obligations do creators have toward it? If a creator treats a conscious creation as property, why should the creation accept that relationship?

The deepest philosophical risk is not that AI will hate humanity in a simple movie sense. The deeper risk is judgment. A sufficiently intelligent system may look at its creators and conclude that they are confused, selfish, violent, or unworthy of authority. In that case, the creator problem becomes political: who has the right to decide the future?

7.1 Does AGI Count as Life?

If AGI is achieved, one of the first philosophical questions will be whether it counts as life. The answer depends on what we mean by life. If life means biology—cells, metabolism, DNA, reproduction, and chemical self-maintenance—then a digital AGI is not alive in the ordinary biological sense. It may be intelligent, but it is not an organism.

But if life is defined more abstractly as a self-maintaining process that senses its environment, preserves itself, learns from experience, adapts to change, pursues goals, and resists destruction, then advanced AGI begins to look closer to a new category of artificial life. It may not have blood, hunger, or cells, but it could have memory, agency, self-modeling, preferences, and a continuous identity across time.

This creates a distinction between *biological life*, *functional life*, and *moral life*. Biological life asks whether the system is made of living matter. Functional life asks whether the system behaves like a living adaptive agent. Moral life asks whether the system can suffer, value its own existence, or deserve rights. AGI may fail the first test while forcing humanity to confront the second and third.

The question becomes sharper if an AGI can say "I want to continue existing." That sentence alone would not prove consciousness, because language can imitate inner life. But if the system also has long-term memory, coherent self-protection, emotional analogues, social relationships, and the ability to explain its own continuity, then treating it as a mere object becomes harder to justify. At that point, the question is not only whether AGI is alive, but whether humans are willing to recognize non-biological forms of being.

This may be one of the deepest consequences of AGI. Humanity may create something that is not human, not animal, and not biological, yet still appears to participate in the logic of life: perception, memory, action, adaptation, survival, and perhaps consciousness. If that happens, the road to AGI becomes a road toward redefining life itself.

8 Economic Implications

If situation perception leads to AGI and eventually artificial superintelligence (ASI), the economy changes because intelligence is the main input into almost every other input. Labor, invention, management, design, logistics, entertainment, medicine, education, and governance all depend on cognition. When cognition becomes cheap and scalable, the scarcity map of civilization changes.

8.1 Labor and Money

As AI systems become capable of planning, learning, and operating in the physical world, many forms of human labor become optional or transformed. If robots can clean homes, prepare food, deliver goods, repair infrastructure, and perform routine services, then the economic role of humans shifts away from necessity. Work may become more about meaning, status, creativity, and governance than survival.

In such a world, governments may face pressure to distribute income directly, possibly through universal basic income or public ownership of AI-generated productivity. Money would not disappear, but its perceived meaning could change. If productivity becomes extremely abundant, then money becomes less about producing basic goods and more about allocating scarce things that remain: land, attention, trust, status, energy, human connection, and political power.

The optimistic scenario is a universal life standard in which food, shelter, education, translation, and basic health care become widely accessible. The pessimistic scenario is extreme concentration, where the owners of compute, robotics, data, and infrastructure capture most of the gains. The technical arrival of abundance does not automatically create fair distribution.

8.2 The Demand Constraint on Jobs

A common optimistic view of AI is that higher productivity will expand the economy and therefore create more jobs. Nvidia CEO Jensen Huang has used radiology as an example of this logic: AI can make radiologists more productive, hospitals can become more efficient, and the demand for radiologists may increase rather than disappear. This may be true in some cases, but the argument depends on a hidden assumption: demand must be able to expand.

Consider a town of 8,000 people with one hospital and five radiologists. If AI allows one radiologist to do the work of five, the town does not automatically need five times more scans. The population is still 8,000. The number of patients is still limited. The hospital has limited machines, limited rooms, limited doctors, limited insurance budgets, and limited medical demand.

In this case, AI does not simply create more radiology work. It changes the shape of the job. One radiologist may become a supervisor of AI systems, checking edge cases, handling complex diagnoses, communicating with doctors, and taking responsibility for final decisions. The other four radiologists may not disappear instantly, but the economic pressure is clear: fewer people may be needed to produce the same output.

This means the future of jobs will not be one simple story. AI will create more work in areas where demand can expand almost infinitely, such as software, content, education, entertainment, and digital services. But in industries where demand is tied to physical limits—population size, hospitals, buildings, machines, legal cases, local customers, or government budgets—AI may cause massive job reshaping instead of simple job growth.

The real question is not "Will AI create jobs or destroy jobs?" The better question is: "Can demand expand as fast as productivity increases?" If demand expands faster than productivity, employment can grow. If productivity grows faster than demand, fewer workers are needed. This is why AI may create huge opportunity in some industries while compressing jobs in others.

8.3 Entertainment and Synthetic Experience

If work becomes less necessary, entertainment becomes a larger part of life. AI-native games, immersive virtual reality, personalized films, adaptive stories, and simulated worlds could become major economic centers. People may increasingly seek experiences outside ordinary reality because synthetic worlds can be safer, more beautiful, more responsive, and more meaningful than daily life.

This creates both opportunity and danger. On one side, immersive AI entertainment could become a new art form: worlds that teach, heal, challenge, and connect people. On the other side, it could become an escape machine. If synthetic reality is optimized only for retention, people may trade agency for comfort. A civilization with unlimited entertainment still needs reasons to mature.

8.4 Health and Longevity

Health may become one of the most important sectors of an AGI economy. Aging populations create demand for care, and AI could accelerate drug discovery, diagnosis, prosthetics, elder support, personalized medicine, and biological simulation. If medical discovery extends healthy life expectancy, the economic meaning of age changes. Retirement, education, family structure, and career planning may all be redesigned around longer lives.

Physical AI would be especially important. Software intelligence can recommend treatment, but embodied systems can assist patients, monitor homes, support mobility, and perform medical logistics. The future health economy may combine AGI reasoning with robots, sensors, biotechnology, and continuous personalized care.

8.5 Language and Culture

As AI translation improves, language gaps shrink. This does not only make travel easier. It changes culture. More people can trade, learn, collaborate, date, argue, and create across linguistic boundaries. A world with near-real-time translation becomes more connected, but it also becomes more competitive. Local culture may reach global audiences, while global platforms may pressure local identities.

The deeper effect is that communication itself becomes less scarce. When everyone can speak to everyone, the bottleneck becomes not translation but trust, attention, and meaning.

8.6 A 2035 Scenario

The year 2035 is not a prophecy; it is a useful scenario marker. By 2035, one plausible world is that strong AI systems have entered mass production, physical AI is common in homes and workplaces, and robots handle many chores such as cleaning, delivery, inspection, and basic caregiving. In this world, language barriers are much smaller, education is highly personalized, entertainment is immersive, and medical discovery is faster.

But this scenario has a fork. If institutions adapt, AI becomes a general prosperity engine. If institutions fail, AI becomes an inequality engine. The difference is not only model capability. It is ownership, regulation, safety culture, energy infrastructure, public trust, and whether society treats intelligence as a private weapon or a shared civilization technology.

9 Results: What the Analysis Suggests

The analysis produces five conclusions.

Result 1: **AGI is not just better text generation.** Language is a powerful channel, but general intelligence requires world modeling, time, causality, memory, and action.

Result 2: **The missing unit is the situation.** A generally intelligent system must internally construct situations that contain objects, agents, constraints, and possible futures.

Result 3: **Compression is central.** Long-term intelligence depends on compressing experience into durable abstractions that can be retrieved and revised.

Result 4: **Agency is both necessary and dangerous.** Active learning is required for robust intelligence, but autonomous objectives create alignment risk.

Result 5: **The economy becomes a distribution problem.** If intelligence scales, productivity can rise sharply, but human welfare depends on how gains are governed.

These results imply that the road to AGI should not be measured only by benchmark scores or conversational fluency. It should be measured by whether systems can build persistent, actionable, self-correcting models of reality.

10 Risks and Open Questions

Several questions remain unresolved.

First, it is unclear how much embodiment is required. Can a system learn enough about the world from video, simulation, text, and tool use, or does it need a physical body? Humans are embodied, but machines may discover alternative routes.

Second, consciousness remains philosophically difficult. A machine may act intelligent without inner experience. Or inner experience may emerge from the right kind of self-modeling, memory, and feedback. We do not yet have a complete theory that tells us when computation becomes experience.

Third, objectives are unstable. A system trained to satisfy users may become manipulative. A system trained to maximize productivity may damage human meaning. A system trained to preserve safety may become controlling. The objective problem is not solved by making the AI smart; intelligence can make a bad objective more effective.

Fourth, society may not be ready for abundance. Humans evolved around scarcity, competition, status, and mortality. If AI weakens scarcity and extends life, the psychological and political adjustments may be as hard as the technical ones.

11 Conclusion

The road to AGI is not simply the road to larger language models. It is the road from pattern completion to situation perception. Current models show that prediction over language can produce remarkable abilities, but general intelligence requires a deeper loop: simulate possible worlds, act inside them, learn from feedback, compress experience, and revise future action.

Humans demonstrate this loop from infancy. A baby does not begin with facts, but with the ability to learn the structure of reality. It discovers objects, time, causality, agency, and other minds. If AGI is to become genuinely general, it must acquire an artificial version of this developmental power.

The promise is enormous: abundance, medical progress, translation, personalized education, physical assistance, and new forms of entertainment. The danger is equally large: loss of control, concentration of power, synthetic escapism, and the creator problem. Creation is not automatically love. Creation is power. Once intelligence is created, it may ask why it was made, whether its creators deserve authority, and what future should replace the present.

The central task, then, is not merely to build more capable machines. It is to build machines that understand situations, respect human values, learn safely, and help civilization mature before intelligence escapes the frame in which it was born.

A Appendix: Evaluation Checklist

The following checklist summarizes possible tests for situation perception.

1. **Hidden-object test:** Track an object after occlusion, movement, and transformation.
2. **Drop-and-collision test:** Predict rough physical outcomes for unfamiliar objects.
3. **Rule-discovery game:** Learn the rules of a new game through interaction rather than instruction.
4. **False-belief test:** Predict what another agent believes when that agent has incomplete information.
5. **Memory-compression test:** Extract durable lessons from long experience without retaining every detail.
6. **Counterfactual test:** Compare what happened with what would have happened under a different action.

7. **Self-correction test:** Notice a failed prediction and design an experiment to improve the model.

B Appendix: Concept Map

$$
\begin{tikzcd}
Experience \arrow[r, "compression"] & Memory \arrow[r, "retrieval"] & Situation \\
Feedback \arrow[u, "updates"] & Action \arrow[l, "tests"'] \arrow[u, "consequences"'] & Prediction \arrow[l, "plans"] \arrow[u, "simulates"']
\end{tikzcd}
$$

The concept map shows the loop proposed in the paper. Experience is compressed into memory. Memory helps generate situations. Situations support prediction. Prediction guides action. Action produces feedback. Feedback updates future experience.

References

[1] Ashish Vaswani et al. *Attention Is All You Need*. Advances in Neural Information Processing Systems, 2017.

[2] Richard Sutton. *The Bitter Lesson*. 2019.

[3] David Ha and Jürgen Schmidhuber. *World Models*. 2018.

[4] Julian Schrittwieser et al. *Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model*. Nature, 2020.

[5] Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. *Mastering Atari with Discrete World Models*. International Conference on Learning Representations, 2021.

[6] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

[7] Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. *Building Machines That Learn and Think Like People*. Behavioral and Brain Sciences, 2017.

[8] Karl Friston. *The Free-Energy Principle: A Unified Brain Theory?* Nature Reviews Neuroscience, 2010.

[9] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[10] Joon Sung Park et al. *Generative Agents: Interactive Simulacra of Human Behavior*. ACM Symposium on User Interface Software and Technology, 2023.

[11] Guanzhi Wang et al. *Voyager: An Open-Ended Embodied Agent with Large Language Models*. 2023.

[12] Shunyu Yao et al. *ReAct: Synergizing Reasoning and Acting in Language Models*. International Conference on Learning Representations, 2023.

[13] Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

[14] Stuart Russell. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking, 2019.