

Three approaches to help AI begin to understand the physical world.

AI is shifting to a world model to understand the real world, opening new avenues for robotics and autonomous vehicles.

Large-scale programming languages (LLMs) are gradually reaching their limits as they move beyond the digital environment to handle problems related to the physical world, such as robotics, autonomous vehicles, and manufacturing.

This led investors to shift their focus to a new concept: *world models* —models capable of simulating and understanding how the world operates. Proof of this is AMI Labs raising over \$1 billion, shortly after World Labs also reached a similar milestone.

Why do LLM programs struggle in real-world situations?

LLMs are very good at processing abstract knowledge by predicting the next word. But they lack a core element: understanding causal relationships in the physical world.

In other words, AI can write very well, but it doesn't really understand what will happen when an action takes place in real life.

Scientist Richard Sutton once stated that LLMs are merely 'mimicking how humans speak,' rather than truly modeling the world. This makes it difficult for them to learn from experience and adapt to change.

Meanwhile, Google DeepMind CEO Demis Hassabis calls this phenomenon 'jagged intelligence' — AI that can solve extremely difficult math problems but fails at basic physics problems.

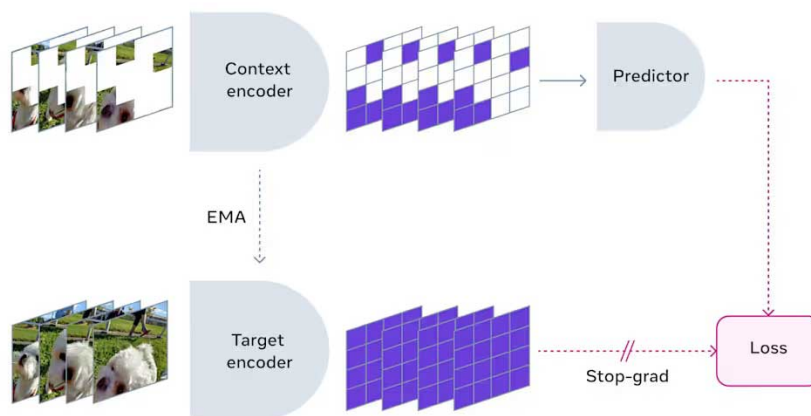
To overcome this weakness, researchers are building world models—systems that can simulate the inner world, allowing AI to 'experiment' before acting. Currently, there are three main approaches, each addressing a different aspect of the problem.

JEPA: Focus on the essence, ignore the details.

One important approach is to learn abstract representations instead of trying to simulate the entire world at the pixel level. This approach is driven by the JEPA architecture.

Instead of memorizing every minute detail, JEPA works similarly to how humans observe the world. When watching a car drive by, we focus on its direction and speed, not on the light reflecting off individual leaves.

JEPA does something similar: it ignores unnecessary details and focuses on core rules. As a result, the model becomes more stable against small changes and doesn't 'break' when the input data fluctuates.



A major advantage is its high computational efficiency, lower resource consumption, and low latency. This makes JEPA suitable for applications requiring real-time response, such as robotics, autonomous vehicles, or enterprise systems.

Scientist Yann LeCun stated that these types of world models can be 'controlled' according to specific goals, meaning they focus solely on completing assigned tasks.

Gaussian splats: Build complete 3D worlds

The second approach focuses on creating entire 3D spaces from scratch using a generative model.

This method utilizes the 'Gaussian splatting' technique, where a 3D scene is created from millions of tiny points that describe geometry and lighting. Unlike flat video, these scenes can be directly imported into engines like Unreal for interaction from multiple perspectives.

This approach significantly reduces the cost and time required to create complex virtual environments. According to Fei-Fei Li, current LLMs are like 'writers in the dark'—skilled in language but lacking spatial awareness. This type of world model helps fill that gap.

Although not suitable for tasks requiring immediate response, this method holds great potential in industrial design, interactive entertainment, and robot training in simulated environments.

End-to-end: Real-time world simulation

The third approach is to use an end-to-end generation model, where AI itself acts as the 'physical engine'.

Instead of creating an environment and then processing it through another system, the model will directly generate images, physics, and reactions in real time based on user actions.

Systems like DeepMind's Genie or NVIDIA's Cosmos follow this approach. They can create a continuously interactive environment while maintaining object consistency and the laws of physics.

The biggest advantage is the ability to generate massive amounts of simulation data. For example, autonomous vehicle companies can test rare and dangerous scenarios without needing real-world testing. However, this comes at a very high computational cost, as the system must process both images and physics simultaneously.

In the future, LLM will continue to play a role in communication and reasoning. However, world models will become the crucial infrastructure for AI systems operating in the real world. The current trend is to combine different architectures to leverage the strengths of each type. For example, the startup DeepTempo has developed a model that combines LLM and JEPA to detect anomalies in cybersecurity systems.

You finished reading the article "**Three approaches to help AI begin to understand the physical world.**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.