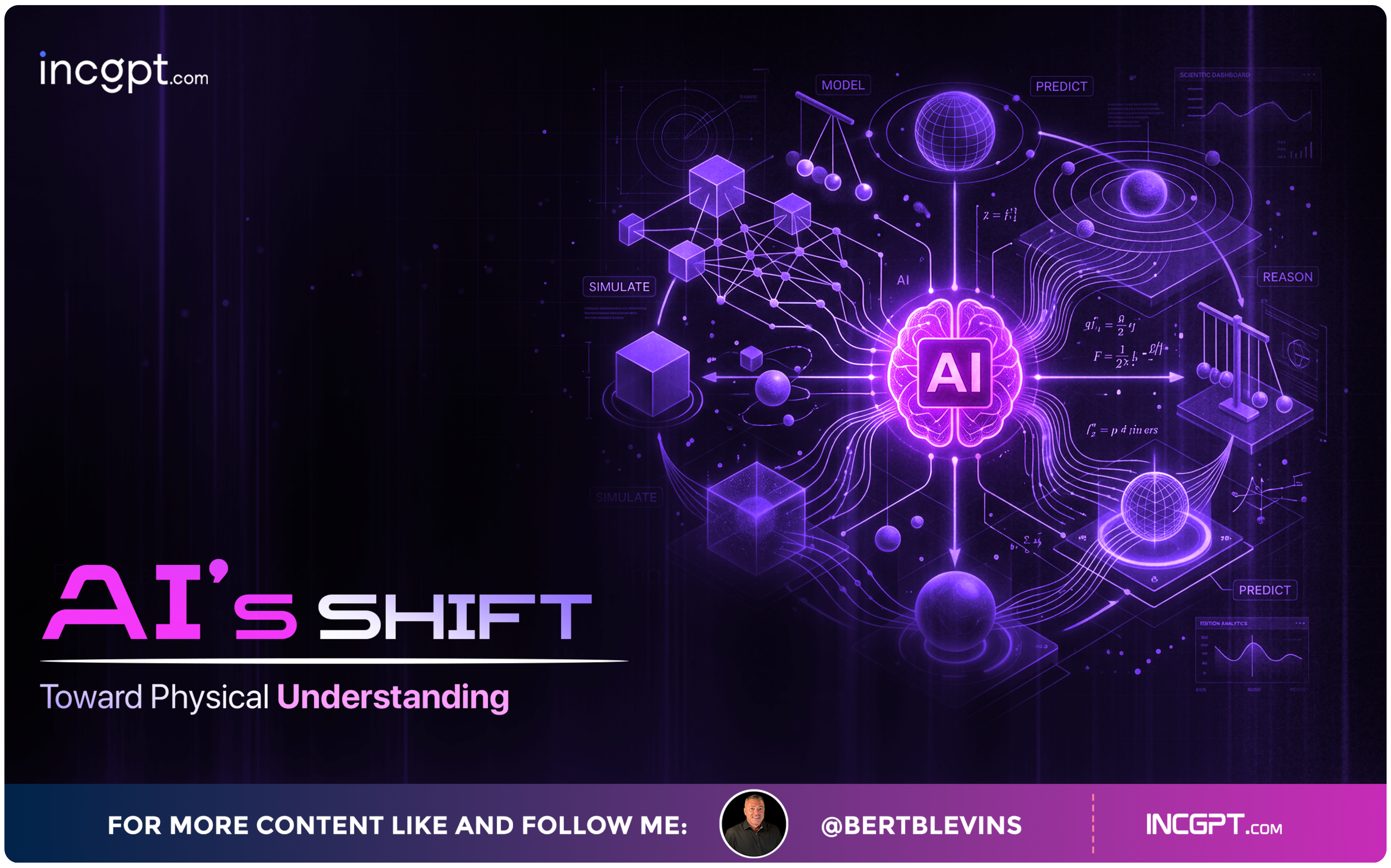
AI’s Shift Toward Physical Understanding

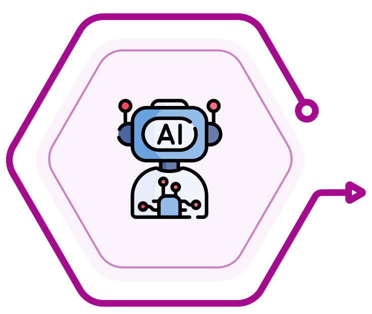


The Rise of World Models

AI has made tremendous strides in recent years, particularly in processing abstract knowledge through large language models (LLMs). However, these models face significant limitations in domains requiring a deeper understanding of the physical world, such as robotics, autonomous driving, and manufacturing. This challenge is pushing researchers and investors toward building better systems, known as world models, that can simulate and predict the physical consequences of actions in the real world. In response to this need, several architectural approaches to world models have emerged, each with unique strengths and tradeoffs.

The Limitations of Large Language Models (LLMs) in Physical Understanding

LLMs excel at abstract reasoning, such as processing and generating text based on patterns and correlations found in large datasets. Yet, their abilities fall short in areas that involve understanding physical causality—the way actions affect the real world. For example, LLMs cannot predict how a physical object will behave in a dynamic environment or accurately simulate the interaction of objects under real-world physics.



This gap in understanding is particularly evident in applications such as robotics, where a robot must respond to real-time environmental changes, or in autonomous driving, where vehicles must predict the consequences of their actions in complex, unpredictable environments. As Turing Award winner Richard Sutton has pointed out, LLMs mimic human language but fail to model the world in a way that allows them to adapt to real-world dynamics. This inability to interact with the physical world in a meaningful way leads to brittle behavior when small changes in inputs occur.

The Push for World Models: A New Approach

To address these shortcomings, AI researchers are increasingly focusing on world models—systems that act as internal simulators of the physical world. These models allow AI systems to test hypotheses, plan actions, and predict outcomes before taking any real-world action. World models are not limited to text generation or abstract reasoning; instead, they create representations of the physical world, allowing AI systems to understand and predict how objects and environments will behave.

Researchers have developed several distinct architectural approaches to world models. These approaches focus on different aspects of physical understanding, from real-time efficiency to spatial computing. Below, we explore three of the most prominent world model architectures.

JEPA: Latent Representations for Real-Time Efficiency

One of the primary approaches to world modeling focuses on learning latent representations of the world. Rather than trying to predict detailed changes in the physical world at a pixel level, models based on the Joint Embedding Predictive Architecture (JEPA) learn a compact set of abstract features that capture the core rules governing how objects interact in a scene.

01 In contrast to traditional models that focus on every detail in an environment, JEPA mimics how humans process the world. For example, when we observe a car driving down the street, we track its trajectory and speed but ignore irrelevant details, such as the reflections on the leaves of nearby trees. JEPA models replicate this cognitive shortcut, focusing on essential features while ignoring distractions.

02 The key benefit of this approach is its computational efficiency. Because the model doesn't need to process irrelevant details, it requires fewer training examples and operates with lower latency. This makes JEPA suitable for real-time applications such as robotics, autonomous vehicles, and dynamic enterprise workflows, where fast, responsive AI is crucial. The model is already being used in high-stakes settings like healthcare to simulate complex operations and reduce cognitive load for professionals in fast-paced environments.

Gaussian Splats: Building Complex Spatial Environments

A second architectural approach leverages generative models to construct complete 3D environments. This method is championed by companies like World Labs, which have developed techniques for generating 3D Gaussian splats—mathematical particles that define geometry and lighting in a 3D space.

01 The primary advantage of this approach is that it reduces the cost and time required to create interactive 3D environments. For example, starting from an initial prompt—whether it's an image or a textual description—generative models create a 3D scene that can be imported into standard physics and 3D engines. This enables users and AI agents to interact with the environment from any angle, making it a valuable tool for spatial computing, interactive entertainment, industrial design, and robotics training.

02 Unlike the first approach, which focuses on real-time processing, the Gaussian splats approach is ideal for creating static or pre-generated environments. Its use cases include virtual environments for training robots or building interactive experiences in design and entertainment. Autodesk's investment in World Labs is a clear indication of the commercial potential of this approach, which enables users to generate spatial data without the limitations of traditional modeling methods.

End-to-End Generative Models: Real-Time Simulation of the Physical World

The third approach to world modeling involves end-to-end generative models that can simulate the entire physical world in real time. Rather than relying on external physics engines to generate environments, these models calculate physical dynamics, object reactions, and lighting natively, producing real-time simulations.

01

DeepMind's Genie 3 and Nvidia's Cosmos are two examples of such models. These AI systems are capable of maintaining object permanence and consistent physics while generating dynamic environments on the fly. For instance, Genie 3 can simulate complex environments at 24 frames per second, keeping track of how objects interact within the space without relying on a separate memory module.

02

The end-to-end generative approach is particularly useful in applications where continuous data generation and physical simulation are essential. It powers large-scale synthetic data factories for autonomous vehicle and robotics development, where rare and dangerous edge-case scenarios can be simulated without the need for physical testing. Waymo, for example, uses this approach to train self-driving cars, creating simulated conditions that would be too costly or risky to test in the real world.

Hybrid Approaches: Combining Strengths of Each Architecture

As the field of world modeling advances, researchers are beginning to explore hybrid approaches that combine the strengths of different models. For example, the cybersecurity startup DeepTempo developed a model called LogLM that integrates elements of both LLMs and JEPA. This hybrid model is designed to detect anomalies and cyber threats by analyzing security and network logs, demonstrating how different world model architectures can complement each other in real-world applications.

The emergence of hybrid models represents the next stage in AI's evolution, blending the power of language models for reasoning with the physical understanding provided by world models. As these hybrid models mature, they will provide a more comprehensive framework for AI to operate effectively in both digital and physical spaces.

Conclusion

The Future of World Models and AI's Physical Understanding

While LLMs will continue to play a crucial role in AI's ability to reason, communicate, and process abstract information, the rise of world models marks a critical step in AI's ability to understand and interact with the physical world. These models are being developed to simulate the consequences of actions, predict real-world behaviors, and enable AI systems to adapt to the physical environment.

The three architectural approaches outlined—JEPA, Gaussian splats, and end-to-end generative models—each offer unique strengths, from real-time efficiency to spatial awareness to comprehensive physical simulation. As AI continues to evolve, hybrid models that combine the strengths of each approach will become increasingly important, laying the groundwork for the next generation of AI systems that can seamlessly navigate both the digital and physical worlds.