EPISODE #336

The World Model Revolution: Beyond LLM Token Prediction

Published January 28, 2026 • Runtime: 28:45

https://myweirdprompts.com/episode/ai-world-models-reasoning-evolution/

EPISODE SYNOPSIS

In this episode of My Weird Prompts, Herman and Corn tackle a growing frustration in the AI community: the "reasoning wall" hit by traditional large language models. As users notice coding assistants collapsing under the weight of complex architectural changes, the brothers discuss why statistical token prediction is no longer enough. They explore the emergence of world models—AI systems designed to internalize the laws of physics, causality, and 3D space. From Meta's JEPA architecture to the spatial intelligence breakthroughs at World Labs, this conversation maps out the transition from AI that merely "speaks" to AI that truly "understands" the environment it operates in. By examining the synergy between intuitive "System 1" language models and logical "System 2" world simulators, Herman and Corn provide a roadmap for the next stage of artificial general intelligence and what it means for the future of robotics, autonomous systems, and software development.

DANIEL'S PROMPT

Daniel

We've previously discussed criticisms of AI and the idea that Large Language Models (LLMs) may not lead us to Artificial General Intelligence (AGI) due to their fundamental limitations. I've experienced this firsthand when using AI for code generation; it often fails as soon as I request changes, possibly due to context window or reasoning degradation. While I previously thought world models were a theoretical concept, I recently came across a multimodal world model used for simulations. I'd like to discuss the current state of world models in 2026. Are they already here? What are some examples of world models that are currently available, and what are their use cases? Are LLMs a dead end, and are world models ready to take up the slack?

TRANSCRIPT

Corn

Welcome back to My Weird Prompts, everyone. I am Corn, and I am joined, as always, by my brother, who has likely spent the last seventy-two hours reading white papers on latent space dynamics.

Herman

Herman Poppleberry at your service, and you are not entirely wrong, Corn. Though, to be fair, I did take a break to eat some hummus. But today is a big one. Our housemate Daniel sent us a fascinating audio prompt that really cuts to the heart of where we are right now in January of twenty-six. He is hitting a wall that I think a lot of people are feeling lately.

Corn

Right, he was talking about that specific frustration when you are using an artificial intelligence to write code. It starts off brilliant, it builds the foundation, and then you ask for one small change, a little architectural pivot, and the whole thing just collapses. It is like the model loses its mind, or as Daniel put it, the reasoning just degrades.

Herman

It is the classic large language model trap. We have been living in this era where we treat language models like they are general purpose brains, but Daniel is seeing the cracks. He is asking if we are at a dead end with these token predictors and if these things called world models are finally ready to take over the heavy lifting.

Corn

It is a great question because for a long time, world models felt like this academic, almost philosophical concept. Something Yann LeCun would talk about on stage while everyone else was busy scaling up transformers. But now, we are seeing the term everywhere. So, Herman, let us start with the basics for our listeners. When we talk about a world model in twenty-six, what are we actually talking about, and how is it different from the models we have been using for the last few years?

Herman

That is the perfect place to start. Think about it this way. A traditional large language model is basically a master of the next word. It has read almost everything ever written, so it is incredibly good at predicting what comes next in a sequence of text. But it does not actually understand the physics of the world. It does not know that if you push a glass off a table, it will break. It just knows that the words glass and break often appear together in sentences about tables.

Corn

Right, it is a statistical map of human language, not a map of physical reality.

Herman

Exactly. A world model, on the other hand, is designed to understand the underlying rules of an environment. Instead of predicting the next token, it is trying to predict the next state of the world. If I am a robot and I move my arm three inches to the left, what does the room look like now? If I am a self-driving car and I turn the wheel, how does the geometry of the street change? It is about internalizing the causal links between actions and outcomes.

Corn

So, when Daniel says he is seeing reasoning degradation in code, is that because the model does not have a world model of the computer system? It is just guessing what the code should look like based on patterns, rather than understanding how the data actually flows through the logic?

Herman

Precisely. To an artificial intelligence without a world model, code is just text. It does not have a mental simulation of the program running. When you ask for a change, it tries to statistically blend your new request with the old code, and if the context window gets too messy, it loses the thread because it lacks a grounding in what the code is actually doing in a functional sense.

Corn

That makes sense. But Daniel mentioned he saw a multimodal world model recently that is being used for simulations. This feels like the big shift of twenty-five and twenty-six. We are moving past just text. Where are we seeing these actually deployed right now?

Herman

We are seeing them in three main areas. Robotics, autonomous vehicles, and what we call spatial intelligence. Let us look at robotics first because that is where the Joint-Embedding Predictive Architecture, or JEPA, has really changed the game. Meta has been pushing this hard. Unlike a generative model that tries to fill in every single pixel, a JEPA-style world model focuses on the high-level features. It ignores the flickering light or the dust on the floor and focuses on the fact that the door handle is a solid object that requires a specific force to turn.

Corn

I remember we talked a bit about that in episode three hundred and twenty-eight when we were looking at speaker identification, that idea of filtering out the noise to find the signal. But in this case, the signal is the physics of the room.

Herman

Right. And for Daniel's question about whether they are already here, the answer is a resounding yes, but they are often hidden under the hood. Take a company like Wayve or even the latest iterations of Tesla's end-to-end neural networks. They are not just following a set of if-then rules anymore. They are running internal simulations of what other drivers might do. They are predicting the future state of the intersection. That is a world model in action.

Corn

But here is the thing that confuses people. We have had simulators for decades. Video games are simulators. We have had physics engines like Havok or Unreal Engine. Is a world model just a fancy name for a neural physics engine?

Herman

That is a sharp distinction, Corn. A traditional physics engine is hard-coded by humans. We tell the computer that gravity is nine point eight meters per second squared. We define the friction of ice. A world model learns these rules from observation. It watches thousands of hours of video and figures out for itself that objects fall down and that round things roll. The reason this is more powerful is that it can handle the messy, unpredictable parts of reality that a programmer might forget to code.

Corn

So it learns the intuition of physics rather than just the math of it. But let us get to the big debate Daniel brought up. Are large language models a dead end for artificial general intelligence? There is this growing sentiment that we have reached the point of diminishing returns with just adding more layers and more data to transformers. Do you think world models are the missing piece that actually gets us to that next level?

Herman

I think the consensus in the research community this year is that language models are the front-end, the communicative layer, but they cannot be the whole brain. Think of the language model as the person talking to you, and the world model as the part of the brain that actually understands how to navigate a three-dimensional space. If you want a system that can actually do things in the physical world, or even just reason reliably about complex systems like a large software architecture, it needs that internal simulation. It needs to be able to test an idea in its head before it outputs the code.

Corn

It is like the difference between someone who has memorized a map and someone who has actually walked the streets. The person who memorized the map might get lost if a new building goes up, but the person with the internal model of the city can navigate around the obstacle.

Herman

Exactly. And this leads us to what Fei-Fei Li and her team at World Labs have been calling spatial intelligence. They are building models that understand three-dimensional geometry from a single image. They can take a photo of this living room and immediately understand where the space is, what is a solid object, and what is empty air. That is something a pure language model struggle with because language is inherently one-dimensional. It is a string of words. The world is three-dimensional and continuous.

Corn

So if I am Daniel, and I am frustrated with my coding assistant, am I waiting for a version of that assistant that is plugged into a world model of the operating system or the language's runtime environment?

Herman

That is exactly what is happening in the cutting-edge developer tools we are seeing enter beta this month. Instead of just a chat window, these tools are starting to use what are called persistent world models of the codebase. They don't just see the file you are working on. They have a latent representation of how every function in your entire repository interacts. When you ask for a change, the model simulates the impact of that change across the whole system in a hidden state before it ever writes a line of code. It is checking for those logic breaks that drive Daniel crazy.

Corn

That sounds like a massive jump in reliability. But let us talk about the generative side. Daniel mentioned world models for simulations. We have seen things like Google DeepMind's Genie or OpenAI's Sora. A lot of people see those as just video generators, but researchers call them world simulators. Why the distinction?

Herman

Because they aren't just stitching images together. To generate a minute of video where a character walks behind a tree and then reappears on the other side, the model has to understand that the character still exists even when they are not visible. It has to understand object permanence. It has to understand that the tree is a solid object. When Sora or Genie generates a scene, it is essentially running a neural simulation of reality. The video is just the output of that simulation.

Corn

But we have seen the bloopers, right? The videos where people merge into chairs or a glass of water disappears into someone's hand. If the world model is so great, why does it still make those basic physics mistakes?

Herman

Because we are still in the early days of grounding. Most of these models were trained on video data, which is just a two-dimensional projection of the world. They are trying to learn three-dimensional physics by watching a flat screen. It is like trying to learn how to swim by watching a movie of a pool. You get the idea, but you don't feel the buoyancy. The big breakthrough we are seeing now in twenty-six is training these models on multimodal data that includes depth information, tactile feedback from robotic sensors, and even synthetic data from those hard-coded physics engines we talked about.

Corn

That is the bridge, then. Using the perfect, albeit limited, math of a physics engine to teach the neural network the fundamentals, and then letting the neural network learn the nuances of the real world.

Herman

Precisely. It is called simulation to real transfer, or sim-to-real. And it is how we are getting robots that can walk over uneven terrain or fold laundry. They practice in a world model millions of times in a few hours, and then they apply that intuition to the physical world.

Corn

Let us pivot back to the LLM side of things. If world models are the future, does that mean the billions of dollars spent on language models was a waste? Or do they merge?

Herman

Oh, they definitely merge. We are seeing the rise of what people are calling the system one and system two architecture. This goes back to Daniel Kahneman's work. System one is fast, intuitive, and pattern-based. That is the LLM. It is great for quick conversation and creative brainstorming. System two is slow, deliberate, and logical. That is where the world model comes in. When you ask a question, the LLM generates a few possibilities, and the world model runs simulations to see which one actually works.

Corn

So the LLM is the imaginative storyteller, and the world model is the grumpy scientist who checks if the story is actually possible.

Herman

Haha, exactly. And that is why Daniel's coding assistant fails. It is all storyteller and no scientist. It is giving him code that looks right but doesn't actually work because it hasn't been verified against a model of reality.

Corn

I want to dig into some specific use cases that might surprise people. Beyond coding and robots, where is this spatial intelligence and world modeling going to hit the average person?

Herman

One of the most exciting areas is climate and weather. NVIDIA's Earth-two project is a great example. It is a digital twin of the entire planet. They aren't just using traditional meteorological equations. They have built a world model that has learned how weather patterns behave by looking at decades of satellite data. It can predict extreme weather events with a level of local specificity that traditional models just can't match because it understands the global context.

Corn

And what about something like gaming? We mentioned Genie earlier. Does this mean we are moving toward games that don't have pre-written scripts or levels?

Herman

We are already seeing the first experimental titles where the entire world is generated on the fly by a world model. You don't just follow a path. You can interact with anything, and the world model figures out what should happen. If you decide to build a fire in the middle of a wooden house, the model doesn't need a programmer to have written a fire script. It understands that wood is flammable and that fire spreads. It creates a truly emergent experience.

Corn

That sounds incredible, but also potentially terrifying for a game designer who wants to tell a specific story.

Herman

It is a total shift in how we think about digital environments. We are moving from built worlds to evolved worlds.

Corn

Okay, so let us address the elephant in the room. If world models are so powerful, what are the risks? If an artificial intelligence can simulate the world perfectly, does it become much better at manipulating it?

Herman

That is a deep rabbit hole, Corn. One of the concerns is that a model with a sophisticated world model can predict human behavior better than we can. It can run simulations of how we will react to certain information or incentives. This is why the alignment problem is moving away from just text and toward physical safety. We have to ensure that these models' internal simulations of a good outcome match our own.

Corn

It is like the difference between a child who doesn't know that hitting someone hurts and a person who knows exactly how to hurt someone but chooses not to. The world model provides the understanding of consequences, which makes the ethical framework even more critical.

Herman

That is a very thoughtful way to put it. Knowledge of the world is power over the world.

Corn

Let us take a break for a second and look at the practical side. If someone listening is a developer like Daniel, or just someone interested in the tech, what should they be looking at right now? What are the specific models or platforms that are leading this world model charge?

Herman

On the open source side, you definitely want to keep an eye on the JEPA releases from Meta. They have been very consistent about putting their research out there. For people interested in spatial intelligence, World Labs is the one to watch. They are building the infrastructure for what they call persistent, interactive three-dimensional worlds. And for the generative side, the research coming out of DeepMind regarding their world simulators is basically the gold standard for how to bridge the gap between video generation and physical reasoning.

Corn

And I think it is worth mentioning that this isn't just for the big players. We are seeing smaller, more specialized world models for things like medical simulations or urban planning. It is becoming a modular part of the AI stack.

Herman

Right. You might have a world model that is specifically an expert in fluid dynamics for a ship-building company, or a world model that is an expert in human anatomy for a surgical robot.

Corn

So, to go back to Daniel's question. Are LLMs a dead end?

Herman

I would say they are a dead end for AGI if they are left on their own. They are like a brain that is all talk and no action. But as a component of a larger system that includes world models, they are more relevant than ever. They are the interface. They allow us to communicate with these complex simulations.

Corn

It feels like we are moving from the era of the chatbot to the era of the agent. An agent that can actually see, move, and reason because it has a sense of where it is.

Herman

Exactly. We are giving the artificial intelligence a body, even if that body is just a virtual one for now. We are giving it a sense of place.

Corn

You know, it reminds me of how we learn as kids. We don't start by reading books. We start by dropping things, by crawling, by feeling the texture of the carpet. We build our world model through physical interaction long before we ever learn to speak or read. In a way, we have been building artificial intelligence backward. We taught it how to speak first, and now we are teaching it how to crawl.

Herman

That is such a great point, Corn. We started with the most complex human achievement, language, and now we are realizing that the foundation of intelligence is actually much more primal. It is about spatial awareness. It is about understanding that you are an entity in a world of other entities.

Corn

So, for Daniel, the reason his coding assistant is failing is that it never learned how to crawl. It never learned the physics of a computer program. It is just a very eloquent toddler who has read a lot of textbooks but has never actually tried to build a Lego tower.

Herman

Haha, I am definitely going to use that analogy. It is an eloquent toddler in a library.

Corn

I think we should look at the second-order effects here. If we get these world models right, what does that mean for the labor market? We have spent a lot of time worrying about writers and artists, but if world models enable truly capable robotics, are we looking at a much bigger shift in physical labor?

Herman

That is the big question for the second half of this decade. If a robot can navigate a construction site or a hospital wing because it has a robust world model, then we are talking about a revolution in blue-collar work that mirrors what we saw in white-collar work with the rise of GPT. But there is a silver lining. These models can also make those jobs much safer. They can simulate dangerous tasks and find the safest way to execute them.

Corn

And I imagine the training for those jobs will change too. Instead of a manual, you might have a world model that guides you in augmented reality, showing you the consequences of an action before you take it.

Herman

Absolutely. Imagine a mechanic wearing glasses that use a world model to identify a part and then show a transparent simulation of how that part fits into the engine. That is the power of spatial intelligence. It bridges the gap between digital information and physical reality.

Corn

We have covered a lot of ground here, Herman. From the limitations of LLMs to the rise of JEPA and spatial intelligence. Before we wrap up this section, is there anything else about the current state of world models in twenty-six that we missed?

Herman

I think it is important to mention that we are starting to see the first signs of world models that can reason across different scales. From the microscopic level of materials science to the macroscopic level of urban traffic. This multi-scale reasoning is something that was almost impossible with traditional simulations, but world models are beginning to find the patterns that connect them.

Corn

It is like they are finding the universal grammar of reality, not just the grammar of English.

Herman

Exactly. It is a very exciting, and slightly dizzying, time to be watching this space.

Corn

Well, I think Daniel's prompt really pushed us into a great deep dive. It is clear that while LLMs might be hitting a certain plateau, the field as a whole is just getting started on its most interesting chapter.

Herman

Definitely. And for anyone listening who is feeling that same frustration Daniel is, just know that the tools are evolving. We are moving toward a version of AI that is more grounded, more reliable, and ultimately, more useful because it finally knows what the world actually looks like.

Corn

That is a great place to take a breath. We have talked about the what and the why, but when we come back, I want to talk about the how. How do we, as humans, stay relevant in a world where AI can simulate reality better than we can?

Herman

That is the million-dollar question, Corn. Or, given inflation, maybe the billion-dollar question.

Corn

Let us get into that. But first, a quick reminder for everyone listening. If you are enjoying these deep dives into the weird and wonderful world of technology and beyond, we would really appreciate it if you could leave us a rating or a review on your favorite podcast app. It really does help other people find the show, and we love hearing from our community.

Herman

It really does. And don't forget, you can find all our past episodes and a way to get in touch with us at our website, myweirdprompts.com. We are also on Spotify, so make sure to follow us there so you never miss an episode.

Corn

Alright, let us talk about the human element. Herman, if we have these world models that can simulate and predict everything from weather to code to physical movement, what is left for the human brain?

Herman

This is where I think we need to look at the concept of intent. A world model can tell you how to get from point A to point B, or how to build a bridge that won't fall down. But it can't tell you why you should build the bridge in the first place. It doesn't have desires, values, or a sense of purpose. Those are uniquely human traits that come from our biological evolution and our social connections.

Corn

So the AI is the engine, but we are still the drivers.

Herman

Exactly. And even more than that, we are the ones who define what a good world looks like. A world model is a tool for understanding reality, but it doesn't define morality. It can simulate a world where everyone is perfectly efficient but miserable, or a world that is messy and inefficient but full of joy. We are the ones who have to provide the moral and aesthetic compass.

Corn

I like that. It suggests that as the technical barriers fall, the importance of our values and our creativity actually increases. We don't have to spend as much time worrying about the how, so we can spend more time on the what and the why.

Herman

That is the optimistic view, and I think it is a valid one. It is a shift from being a laborer to being an architect. From being a coder to being a system designer.

Corn

But it requires a different kind of education, doesn't it? We have spent decades teaching people how to be human calculators or human encyclopedias. If AI can do that, what should we be teaching the next generation?

Herman

We should be teaching them how to ask better questions. We should be teaching them critical thinking, ethics, and how to collaborate with these powerful systems. The most valuable skill in twenty-six isn't knowing the answer; it is knowing how to frame the problem so that the AI can help you solve it.

Corn

It is back to the prompt, right? Daniel's frustration came from a place of knowing what he wanted but not being able to get the tool to understand his intent. As the tools get better at understanding the world, our job is to get better at communicating our vision.

Herman

Precisely. And that requires a deep understanding of ourselves. The more powerful our technology becomes, the more we need to understand what it means to be human.

Corn

That is a profound thought to end on, Herman. I think we have given Daniel, and everyone else, a lot to chew on.

Herman

I hope so. It is a fascinating transition we are living through. From the world of words to the world of models.

Corn

Well, thank you all for joining us on this episode of My Weird Prompts. A big thanks to our housemate Daniel for sending in that prompt and for keeping the fridge stocked with hummus while we record these.

Herman

Yes, thank you, Daniel. And thank you, Corn, for the great questions. This was a fun one.

Corn

If you haven't already, head over to myweirdprompts.com to see our full archive. We have covered everything from subsea fiber optics to the secret history of spying, and we have many more weird prompts coming your way.

Herman

Until next time, stay curious and keep asking those weird questions.

Corn

Thanks for listening to My Weird Prompts. We will see you next week.

Herman

Goodbye, everyone!