From Words to Worlds: Spatial Intelligence is AI's Next Frontier

FEI-FEI LI

NOV 10, 2025

In 1950, when computing was little more than automated arithmetic and simple Alan Turing asked a question that still reverberates today: can machines think? I remarkable imagination to see what he saw: that intelligence might someday be rather than born. That insight later launched a relentless scientific quest called Artificial Intelligence (AI). Twenty-five years into my own career in AI, I still fin myself inspired by Turing's vision. But how close are we? The answer isn't simpl

Today, leading AI technology such as large language models (LLMs) have begun transform how we access and work with abstract knowledge. Yet they remain wordsmiths in the dark; eloquent but inexperienced, knowledgeable but ungroun **Spatial intelligence will transform how we create and interact with real and vir worlds—revolutionizing storytelling, creativity, robotics, scientific discovery, a beyond. This is AI's next frontier.**

The pursuit of visual and spatial intelligence has been the North Star guiding me I entered the field. It's why I spent years building ImageNet, the first large-scale learning and benchmarking dataset and one of three key elements enabling the b of modern AI, along with neural network algorithms and modern compute like graphics processing units (GPUs). It's why my academic lab at Stanford has spen last decade combining computer vision with robotic learning. And it's why my cofounders Justin Johnson, Christoph Lassner, Ben Mildenhall, and I created Wo Labs more than one year ago: to realize this possibility in full, for the first time.

In this essay, I'll explain what spatial intelligence is, why it matters, and how we' building the world models that will unlock it—with impact that will reshape cre embodied intelligence, and human progress.

Spatial Intelligence: The scaffolding of human cognitio

AI has never been more exciting. Generative AI models such as LLMs have move from research labs to everyday life, becoming tools of creativity, productivity, an communication for billions of people. They have demonstrated capabilities once thought impossible, producing coherent text, mountains of code, photorealistic images, and even short video clips with ease. It's no longer a question of whethe will change the world. By any reasonable definition, it already has.

Yet so much still lies beyond our reach. The vision of autonomous robots remain intriguing but speculative, far from the fixtures of daily life that futurists have lo promised. The dream of massively accelerated research in fields like disease cur new material discovery, and particle physics remains largely unfulfilled. And the promise of AI that truly understands and empowers human creators—whether students learning intricate concepts in molecular chemistry, architects visualizi spaces, filmmakers building worlds, or anyone seeking fully immersive virtual experiences—remains beyond reach.

To learn why these capabilities remain elusive, we need to examine how spatial intelligence evolved, and how it shapes our understanding of the world.

Vision has long been a cornerstone of human intelligence, but its power emerge something even more fundamental. Long before animals could nest, care for thei young, communicate with language, or build civilizations, the simple act of sensi quietly sparked an evolutionary journey toward intelligence.

This seemingly isolated ability to glean information from the external world, wh a glimmer of light or the feeling of texture, created a bridge between perception survival that only grew stronger and more elaborate as the generations passed. L upon layer of neurons grew from that bridge, forming nervous systems that inter the world and coordinate interactions between an organism and its surrounding Thus, many scientists have conjectured that **perception and action became the c loop driving the evolution of intelligence**, and the foundation on which nature c our species—the ultimate embodiment of perceiving, learning, thinking, and doi

Spatial intelligence plays a fundamental role in defining how we interact with th physical world. Every day, we rely on it for the most ordinary acts: parking a car l imagining the narrowing gap between bumper and curb, catching a set of keys to across the room, navigating a crowded sidewalk without collision, or sleepily pou coffee into a mug without looking. In more extreme circumstances, firefighters navigate collapsing buildings through shifting smoke, making split-second judgements about stability and survival, communicating through gestures, body language and a shared professional instinct for which there's no linguistic substi And children spend the entirety of their pre-verbal months or years learning the through playful interactions with their environments. All of this happens intuiti automatically—a fluency machines have yet to achieve.

Spatial Intelligence is also foundational to our imagination and creativity. Storyt create uniquely rich worlds in their minds and leverage many forms of visual me bring them to others, from ancient cave painting to modern cinema to immersive video games. Whether it's children building sandcastles on the beach or playing Minecraft on the computer, spatially-grounded imagination forms the basis for interactive experiences in real or virtual worlds. And in many industry applicatio simulations of objects, scenes and dynamic interactive environments power cour numbers of critical business use cases from industrial design to digital twins to r training.

History is full of civilization-defining moments where spatial intelligence played central roles. In ancient Greece, Eratosthenes transformed shadows into geomet measuring a 7-degree angle in Alexandria at the exact moment the sun cast no sh in Syene—to calculate the Earth's circumference. Hargreave's "Spinning Jenny" revolutionized textile manufacturing through a spatial insight: arranging multip spindles side-by-side in a single frame allowed one worker to spin multiple threa simultaneously, increasing productivity eightfold. Watson and Crick discovered structure by physically building 3D molecular models, manipulating metal plates wire until the spatial arrangement of base pairs clicked into place. In each case, intelligence drove civilization forward when scientists and inventors had to manipulate objects, visualize structures, and reason about physical spaces - none which can be captured in text alone.

Spatial Intelligence is the scaffolding upon which our cognition is built. It's at when we passively observe or actively seek to create. It drives our reasoning and planning, even on the most abstract topics. And it's essential to the way we inter verbally or physically, with our peers or with the environment itself. While most aren't revealing new truths on the level of Eratosthenes most days, we *routinely* tl in the same way—making sense of a complex world by perceiving it through our senses, then leveraging an intuitive understanding of how it works in physical, s terms.

Unfortunately, today's AI doesn't think like this yet.

Tremendous progress has indeed been made in the past few years. Multimodal L (MLLMs), trained with voluminous multimedia data in addition to textual data, h introduced some basics of spatial awareness, and today's AI can analyze pictures answer questions about them, and generate hyperrealistic images and short vide And through breakthroughs in sensors and haptics, our most advanced robots ca begin to manipulate objects and tools in highly constrained environments.

Yet the candid truth is that AI's spatial capabilities remain far from human level the limits reveal themselves quickly. State-of-the-art MLLM models rarely perfo better than chance on estimating distance, orientation, and size—or "mentally" rotating objects by regenerating them from new angles. They can't navigate maz recognize shortcuts, or predict basic physics. AI-generated videos—nascent and very cool—often lose coherence after a few seconds.

While current state-of-the-art AI can excel at reading, writing, research, and patt recognition in data, these same models bear fundamental limitations when representing or interacting with the physical world. Our view of the world is hol not just what we're looking at, but how everything relates spatially, what it mean why it matters. Understanding this through imagination, reasoning, creation, an interaction—not just descriptions—is the power of spatial intelligence. Without is disconnected from the physical reality it seeks to understand. It cannot effecti drive our cars, guide robots in our homes and hospitals, enable entirely new ways immersive and interactive experiences for learning and recreation, or accelerate discovery in materials science and medicine.

The philosopher Wittgenstein once wrote that "the limits of my language mean t limits of my world." I'm not a philosopher. But I know at least for AI, there is mo than just words. Spatial intelligence represents the frontier beyond language—th capability that links imagination, perception and action, and opens possibilities machines to truly enhance human life, from healthcare to creativity, from scienti discovery to everyday assistance.

The next decade of AI: Building truly spatially intelliger machines

So how do we build spatially-intelligent AI? What's the path to models capable c reasoning with the vision of Eratosthenes, engineering with the precision of an industrial designer, creating with the imagination of a storyteller, and interacting their environment with the fluency of a first responder?

Building spatially intelligent AI requires something even more ambitious than L world models, a new type of generative models whose capabilities of understand reasoning, generation and interaction with the semantically, physically, geometr and dynamically complex worlds - virtual or real - are far beyond the reach of toc LLMs. The field is nascent, with current methods ranging from abstract reasonii models to video generation systems. World Labs was founded in early 2024 on th conviction: that foundational approaches are still being established, making this defining challenge of the next decade.

In this emerging field, what matters most is establishing the principles that guid development. For spatial intelligence, I define world models through **three essen capabilities:**

1. Generative: World models can generate worlds with perceptual, geometrical, and physical consistency

World models that unlock spatial understanding and reasoning must also genera simulated worlds of their own. They must be capable of spawning endlessly varie diverse simulated worlds that follow semantic or perceptual instructions—*while* remaining geometrically, physically, and dynamically consistent—whether representing real or virtual spaces. The research community is actively exploring whether these worlds should be represented implicitly or explicitly in terms of th innate geometric structures. Furthermore, in addition to powerful latent representations, I believe the outputs of a universal world model must also allow generation of an explicit, observable state of the worlds for many different use ca In particular, its understanding of the present must be tied coherently to its past the previous states of the world that led to the current one.

2. Multimodal: World models are multimodal by design

Just as animals and humans do, a world model should be able to process inputs— known as "prompts" in the generative AI realm—in a wide range of forms. Giver partial information—whether images, videos, depth maps, text instructions, gest or actions—world models should predict or generate world states as *complete* as possible. This requires processing visual inputs with the fidelity of real vision wh interpreting semantic instructions with equal facility. This enables both agents a humans to communicate with the model about the world through diverse inputs receive diverse outputs in return.

3. Interactive: World models can output the next states based on in actions

Finally, if actions and/or goals are part of the prompt to a world model, its outpu must include the *next* state of the world, represented either implicitly or explicitl When given only an action with or without a goal state as the input, the world m should produce an output consistent with the world's previous state, the intende state if any, and its semantic meanings, physical laws, and dynamical behaviors. spatially intelligent world models become more powerful and robust in their rea and generation capabilities, it is conceivable that in the case of a given goal, the models themselves would be able to predict not only the next state of the world, also the next actions based on the new state.

The scope of this challenge exceeds anything AI has faced before.

While language is a purely generative phenomenon of human cognition, worlds by much more complex rules. Here on Earth, for instance, gravity governs motio atomic structures determine how light produces colors and brightness, and coun physical laws constrain every interaction. Even the most fanciful, creative world composed of spatial objects and agents that obey the physical laws and dynamic behaviors that define them. Reconciling all of this consistently—the semantic, th geometric, the dynamic, and physical—demands entirely new approaches. The dimensionality of representing a world is vastly more complex than that of a one dimensional, sequential signal like language. Achieving world models that delive kind of universal capabilities we enjoy as humans will require overcoming sever formidable technical barriers. At World Labs, our research teams are devoted to making fundamental progress toward that goal.

Here are some examples of our current research topics:

- **A new, universal task function for training:** Defining a universal task functi simple and elegant as next-token prediction in LLMs has long been a centra of world model research. The complexities of both their input and output sp make such a function inherently more difficult to formulate. But while much remains to be explored, this objective function and corresponding represent must reflect the laws of geometry and physics, honoring the fundamental na of world models as grounded representations of both imagination and realit
- **Large-scale training data**: Training world models requires far more comple than text curation. The promising news: massive data sources already exist. Internet-scale collections of images and videos represent abundant, accessib

training material—the challenge lies in developing algorithms that can extra deeper spatial information from these two-dimensional image or video fram based signals (i.e. RGB). Research over the past decade has shown the power scaling laws linking data volume and model size in language models; the key unlock for world models is building architectures that can leverage existing data at comparable scale. In addition, I would not underestimate the power high-quality synthetic data and additional modalities like depth and tactile information. They supplement the internet scale data in critical steps of the training process. But the path forward depends on better sensor systems, mo robust signal extraction algorithms, and far more powerful neural simulatio methods.

- **New model architecture and representational learning:** World model resea will inevitably drive advances in model architecture and learning algorithms particularly beyond the current MLLM and video diffusion paradigms. Both these typically tokenize data into 1D or 2D sequences, which makes simple s tasks unnecessarily difficult - like counting unique chairs in a short video, o remembering what a room looked like an hour ago. Alternative architecture help, such as 3D or 4D-aware methods for tokenization, context, and memor example, at World Labs, our recent work on a real-time generative frame-ba model called RTFM has demonstrated this shift, which uses spatially-groun frames as a form of spatial memory to achieve efficient real-time generation maintaining persistence in the generated world.

Clearly, we are still facing daunting challenges before we can fully unlock spatial intelligence through world modeling. This research isn't just a theoretical exerci is the core engine for a new class of creative and productivity tools. And the prog within World Labs has been encouraging. We recently shared with a limited num users a glimpse of Marble, the first ever world model that can be prompted by multimodal inputs to generate and maintain consistent 3D environments for use storytellers to explore, interact with, and build further in their creative workflow we are working hard to make it available to the public soon!

Marble is only our first step in creating a truly spatially intelligent world model. progress accelerates, researchers, engineers, users, and business leaders alike are beginning to recognize its extraordinary potential. The next generation of world models will enable machines to achieve spatial intelligence on an entirely new le an achievement that will unlock essential capabilities still largely absent from to AI systems.

Using world models to build a better world for people

It matters what motivates the development of AI. As one of the scientists who h usher in the era of modern AI, my motivation has always been clear: AI must aug human capability, not replace it. For years, I've worked to align AI development, deployment, and governance with human needs. Extreme narratives of techno-ut and apocalypse are abundant these days, but I continue to hold a more pragmatic AI is developed by people, used by people, and governed by people. It must alway respect the agency and dignity of people. Its magic lies in extending our capabili making us more creative, connected, productive, and fulfilled. Spatial intelligenc represents this vision—AI that empowers human creators, caregivers, scientists, dreamers to achieve what was once impossible. This belief is what drives my commitment to spatial intelligence as AI's next great frontier.

The applications of spatial intelligence span varying timelines. Creative tools are emerging now—World Labs' Marble already puts these capabilities in creators' a storytellers' hands. Robotics represents an ambitious mid-term horizon as we ref the loop between perception and action. The most transformative scientific applications will take longer but promise a profound impact on human flourishin

Across all these timelines, several domains stand out for their potential to reshap human capability. It will take significant collective effort, more than a single tean company can possibly achieve. It will require participation across the entire AI ecosystem—researchers, innovators, entrepreneurs, companies, and even policym —working toward a shared vision. But this vision is worth pursuing. Here's what future holds:

Creativity: Superpowering storytelling and immersive experiences

"Creativity is intelligence having fun." This is one of my favorite quotes by my personal hero Albert Einstein. Long before written language, humans told storie painted them on cave walls, passed them through generations, built entire cultur shared narratives. Stories are how we make sense of the world, connect across di and time, explore what it means to be human, and most importantly, find meanir life and love within ourselves. Today, spatial intelligence has the potential to transform how we create and experience narratives in ways that honor their fundamental importance, and extend their impacts from entertainment to educa from design to construction.

World Labs' Marble platform will be putting unprecedented spatial capabilities a editorial controllability in the hands of filmmakers, game designers, architects, a storytellers of all kinds, allowing them to rapidly create and iterate on fully explo 3D worlds without the overhead of conventional 3D design software. The creativ remains as vital and human as ever; the AI tools simply amplify and accelerate w creators can achieve. This includes:

- Narrative experiences in new dimensions: Filmmakers and game designers a using Marble to conjure entire worlds without the constraints of budget or geography, exploring varieties of scenes and perspectives that would have be intractable to explore within a traditional production pipeline. As the lines between different forms of media and entertainment blur, we're approaching fundamentally new kinds of interactive experiences that blend art, simulatio play—personalized worlds where anyone, not just studios, can create and inl their own stories. With the rise of newer, more rapid ways to lift concepts an storyboards into full experiences, narratives will no longer be bound to a sin medium, with creators free to build worlds with shared throughlines across myriad surfaces and platforms.
- Spatial narratives through design: Essentially every manufactured object or constructed space must be designed in virtual 3D before its physical creatior process is highly iterative and costly in terms of both time and money. With spatially intelligent models at their disposal, architects can quickly visualize

structures before investing months into designs, walking through spaces tha don't yet exist—essentially telling stories about how we might live, work, an gather. Industrial and fashion designers can translate imagination into form instantly, exploring how objects interact with human bodies and spaces.

- New immersive and interactive experiences: Experience itself is one of the d ways that we, as a species, create meaning. For the entirety of human history there has been one singular 3D world: the physical one we all share. Only in decades, through gaming and early virtual reality (VR), have we begun to gli what it means to share alternate worlds of our own creation. Now, spatial intelligence combined with new form factors, like VR and extended reality (headsets and immersive displays, elevates these experiences in unprecedent ways. We're approaching a future where stepping into fully realized multi-dimensional worlds becomes as natural as opening a book. Spatial intelligen makes world-building accessible not just to studios with professional produ teams but to individual creators, educators, and anyone with a vision to shar

Robotics: Embodied intelligence in action

Animals from insects to humans depend on spatial intelligence to understand, navigate and interact with their worlds. Robots will be no different. Spatially-aw machines have been the dream of the field since its inception, including my own with my students and collaborators at my Stanford research lab. This is also why so excited by the possibility of bringing them about using the kinds of models W Labs is building.

- **Scaling robotic learning via world models**: The progress of robotic learning hinges on a scalable solution of viable training data. Given the enormous sta spaces of possibilities that robots have to learn to understand, reason, plan, interact with, many have conjectured that a combination of internet data, synthetic simulation, and real-world capture of human demonstration are re to truly create generalizable robots. But unlike language models, training da scarce for today's robotic research. World models will play a defining role in As they increase their perceptual fidelity and computational efficiency, outp world models can rapidly close the gap between simulation and reality. This in turn help train robots across simulations of countless states, interactions environments.

- **Companions and collaborators:** Robots as human collaborators, whether aic scientists at the lab bench or assisting seniors living alone, can expand part workforce in dire need of more labour and productivity. But doing so deman spatial intelligence that perceives, reasons, plans, and acts while—and this i important—staying empathetically aligned with human goals and behaviors. instance, a lab robot might handle instruments so the scientist can focus on needing dexterity or reasoning, while a home assistant might help an elderly person cook without diminishing their joy or autonomy. Truly spatially intel world models that can predict the next state or possibly even actions consist with this expectation are critical for achieving this goal.
- **Expanding forms of embodiment:** Humanoid robots play a role in the world built for ourselves. But the full benefit of innovation will come from a far mc diverse range of designs: nanobots that deliver medicine, soft robots that na tight spaces, and machines built for the deep sea or outer space. Whatever tl form, future spatial intelligence models must integrate both the environmen these robots inhabit and their own embodied perception and movement. But challenge in developing these robots is the lack of training data in these wid varieties of embodied form factors. World models will play a critical role in simulation data, training environments, and benchmarking tasks for these e

The Longer Horizon: Science, Healthcare, and Education

In addition to creative and robotics applications, spatial intelligence’ profound i will also extend to fields where AI can enhance human capability in ways that sa lives and accelerate discovery. I highlight below three areas of applications that deeply transformative, though it goes without saying the use cases of spatial intelligence are truly expansive across many more industries.

In **scientific research**, spatially intelligent systems can simulate experiments, tes hypotheses in parallel, and explore environments inaccessible to humans—from oceans to distant planets. This technology can transform computational modelin fields like climate science and materials research. By integrating multi-dimensio simulation with real-world data collection, these tools can lower compute barrie extend what every laboratory can observe and understand.

In **healthcare**, spatial intelligence will reshape everything from laboratory to bec At Stanford, my students and collaborators have spent many years working with hospitals, elder care facilities, and patients at home. This experience has convinc of spatial intelligence's transformative potential here. AI can accelerate drug dis by modeling molecular interactions in multi-dimensions, enhance diagnostics by helping radiologists spot patterns in medical imaging, and enable ambient moni systems that support patients and caregivers without replacing the human conne that healing requires, not to mention the potential of robots in helping our healt workers and patients in many different settings.

In **education**, spatial intelligence can enable immersive learning that makes abst or complex concepts tangible, and create iterative experiences so essential to hov brains and bodies are wired in learning. In the age of AI, the need for faster and effective learning and reskilling is particularly important for both school-aged children and adults. Students can explore cellular machinery or walk through historical events in multi-dimenality. Teachers gain tools to personalize instruct through interactive environments. Professionals—from surgeons to engineers— safely practice complex skills in realistic simulations.

Across all these domains, the possibilities are boundless, but the goal remains constant: AI that augments human expertise, accelerates human discovery, and amplifies human care—not replacing the judgment, creativity, and empathy that central for being humans.

Conclusion

The last decade has seen AI become a global phenomenon and an inflection poir technology, the economy, and even geopolitics. But as a researcher, educator, anc entrepreneur, it's still the spirit behind Turing's 75-year-old question that inspire most. I still share his sense of wonder. It's what energizes me every day by the challenge of spatial intelligence.

For the first time in history, we're poised to build machines so in tune with the physical world that we can rely on them as true partners in the greatest challeng face. Whether accelerating how we understand diseases in the lab, revolutionizir we tell stories, or supporting us in our most vulnerable moments due to sickness injury, or age, we're on the cusp of technology that elevates the aspects of life we about most. This is a vision of deeper, richer, more empowered lives.

Almost a half billion years after nature unleashed the first glimmers of spatial intelligence in the ancestral animals, we're lucky enough to find ourselves amon generation of technologists who may soon endow machines with the same capab and privileged enough to harness those capabilities for the benefits of people everywhere. Our dreams of truly intelligent machines will not be complete withc spatial intelligence.

This quest is my North Star. Join me in pursuing it.