SIMLABS RESEARCH / POSITION PAPER

AI Needs a Place to Stand

World models, AI Terminals, AI Hosts, and the physical layer
for intelligence in real places.

AUTHORS	AFFILIATION	PUBLISHED	VERSION
simLabs	simulatingMachines	2026-08-24	2.5.0

MANIFESTO

Machines that understand the place.

AI has left the chat. For years, intelligence lived behind a browser: brilliant at answering, absent from the aisle, the concourse, the counter, the queue. We see a widening gap between rapidly improving online assistance and place-dependent help inside many venues. Our mission is to make helpful AI present wherever people need assistance: a walk-up AI Terminal first, embodied AI Hosts later, and the Operations Platform behind them, so that real people, in real places, in real situations, can simply speak and be helped.

1. Help is a physical problem.

The questions that remote help struggles to resolve are tied to place and changing state: where, which way, how far, is it available, can I make it in time.

The answer to those is not a paragraph. It is a route that holds, a shelf that is actually there, a person reached before the moment passes. We judge our work by what happens after the answer: the person acts, and it works.

2. A place speaks more than language.

Language models learn statistical structure in text. Physical places also require geometry, dynamics and current state that text alone does not determine. The shortfall this creates in the systems tested on spatial benchmarks is measured, not asserted.

So every agent we intend to deploy carries, by design, a versioned representation of its venue: the map as it stands, the inventory where it sits and the routes allowed by venue-approved constraints. Language is how our machines talk with people. Space is what they must understand.

3. Help is measured by what happens next.

Intelligence has not finished when it produces a sentence. The person still has to reach the gate, find the item, complete the task or connect with the right human. We build for that full arc: understand the need, ground the response in the place, support the next action and learn from the outcome.

Measured outcomes give the next release something concrete to improve, with people deciding what changes. The system gets better by learning from consequences, not from conversation alone.

4. Reality is more than a rendering.

A world can look convincing and still lack what action requires. Useful physical intelligence needs current state it can reason over, spatial relationships it can preserve and consequences it can anticipate. Rendering, simulation and planning serve different functions; visual realism alone cannot show that a system can route, predict or act.

Natural conversation and visual presentation make intelligence easier to use. They do not define the world behind the interface. The system must remain connected to what is true in the venue now and able to change when the venue changes.

5. Rehearse before deployment. Learn from reality.

Simulation gives physical AI a place to encounter difficult situations before visitors have to. It makes those situations repeatable, so the service can be tested, compared and improved before it faces the same conditions in a venue.

Reality then reveals what the model missed. A failure, surprise or changed condition can become a scenario the next release must handle. Simulation makes experience reusable; experience makes the simulation more useful. That exchange is the method behind our name.

6. The place is the dataset.

A useful record binds a versioned place, human intent, observation, action and physical outcome. The Terminal or Host earns the right to create that record by helping the person first. Simulation can turn authorized hard cases into repeatable tests. The service is primary; the dataset is a possible consequence of operating it well.

7. Venues learn with people in control.

Every question can signal what a place fails to communicate. A venue should be able to learn from those moments without turning a visitor into a profile. Each deployment names who controls and processes interaction data, gives appropriate notice, follows applicable law and respects people's data rights. Training, cross-venue learning and licensing each require purpose-specific permission. Learning expands only as trust and permission expand.

8. People stay the point.

We put agents where repeated questions gather so staff can be where judgment, care and exception live. Handing a moment to a person is a feature, not a failure. The machine succeeds when the place feels more attentive, not merely more automated.

9. Intelligence is a system, not a model.

Once AI enters a venue, its intelligence includes everything a person depends on: the microphone that hears, the state that stays current, the route that holds, the handoff that reaches a person and the operations that keep the service alive. A failure in any layer reaches the visitor as one failure.

We design, test and operate those layers as one service. The person asking for help should never have to diagnose the machine before the machine can help them.

10. AI needs a place to stand.

Language gave AI access to what people have written. World models are beginning to give it representations of what places contain, how they change and what follows from an action. But intelligence cannot help in the physical world until it is connected to current state, a physical interface, permission to act and consequences it can learn from.

That is what we are building. The Terminal is the first foothold; the Host is the embodied extension; venue models, simulation and operations form the learning loop around them. We begin with one person asking for help in one real place. If the person can act and the place can improve, the model has met the world.

ABSTRACT

Language models made intelligence conversational. World models are beginning to make it spatial, predictive and action-aware. Yet a model cannot help inside a venue unless it has a physical point of presence, current state, permission to act and feedback from what happened next. simulatingMachines is building that bridge: AI Terminals first, AI Hosts next, and the physical AI support infrastructure around them. The service creates immediate value by helping people navigate, find, act and reach the right person. Simulation and real-world outcomes form a learning loop; under appropriate agreements, aligned experience may become evaluation and training assets for world-model builders. This paper presents the thesis: make helpful AI present wherever people need assistance.

External claims link to selected evidence with visible evidence classes. Company statements are labeled by maturity, and the present proof is separated from the pilot and research roadmap.

SECTION

1. The world is still waiting for AI

AI's most important next interface may not be another app. It may be the place itself.

A person is standing in an unfamiliar building. They need a gate, a product, a service, an exit or someone who can help. The question may fit in one sentence. The answer depends on everything the sentence leaves out: where the person is, what is around them, what is open, what has changed and what they should do next.

When an answer changes where a person walks, whether they find what they need or who comes to help, AI is no longer only generating content. It is participating in the world. Fluency matters, but it is not enough. The intelligence needs a point of presence, current knowledge of the place, a path from language to action and feedback from what happened after the answer.

simulatingMachines exists to make helpful AI present wherever people need assistance: for real people, in real places, in real situations. We call the service architecture **Artificial Concierge Intelligence**. It connects an approachable physical interface to the knowledge, spatial state, supported actions and people of the place it serves.

The AI Terminal is the first product: fixed, walk-up and present in the venue. The AI Host is the embodied extension: the same service with a voice and visual presence, and eventually a greater ability to perceive and move through the environment. Around both is the physical AI support infrastructure that turns a capable model into a dependable service: venue state, operations, simulation and evaluation.

> World models are helping AI represent and predict physical environments, giving agents a basis for planning and action. simulatingMachines is building the place where that intelligence meets people, operations and consequences.

We begin deliberately: one place, one repeated assistance problem and one outcome that matters. The Terminal exists today as a working software prototype. The next milestone is a live venue deployment that tests whether place-aware AI can help people complete a real task at the moment they need it. That is the first step toward a much larger idea: places that can understand what is happening inside them and respond.

SECTION

2. AI is moving from words to worlds

Language models changed computing by learning the structure of what people have written. They made knowledge conversational and software easier to direct. But a place is not a paragraph. It has state, geometry, time, motion and consequences that text alone does not determine.

World models are the emerging attempt to cross that boundary. The term covers several technical approaches, but the common purpose is straightforward: build an internal representation of an environment that helps an agent understand the present, anticipate what could happen next and choose an action. A recent survey organizes the field around two closely related functions: representing the mechanisms of the external world and predicting future states to guide decisions [1].

Fei-Fei Li describes the same frontier in functional terms: world models should be generative, multimodal and interactive, able to produce next states in response to actions [2]. That final property is the decisive one. The goal is not only to make a plausible picture of a place, but to represent enough of its structure and dynamics for an agent to ask what happens if it acts.

The idea predates the current wave. In 2018, David Ha and Jürgen Schmidhuber trained a compressed spatial and temporal representation of an environment, trained an agent inside the model's generated "dream" and transferred the policy back to the original environment [3]. The research frontier has since expanded in several directions. Dreamer learns environment dynamics and improves behavior by imagining future scenarios [4]. Navigation World Models predicts future observations under candidate actions so an agent can plan a path [5]. V-JEPA 2 learns latent spatiotemporal representations from video and demonstrates action-conditioned planning on physical robots [6].

Several industrial research programs are pursuing the same transition through different representations. NVIDIA describes a world foundation model as a general model that can be adapted into customized world models for physical applications [7]. Google DeepMind describes world models as systems that simulate how an environment may evolve and how actions affect it [8]. World Labs separates renderers that produce observations, simulators that produce computable state and planners that produce actions [9]. These are different research programs, not evidence of one settled architecture. Their shared direction is what matters: from describing an observation to representing state, action and consequence.

That shift changes what becomes valuable. A model of language can learn from documents copied at internet scale. A model of a place needs grounded experience: what was present, what changed, what an agent did and what followed. Li argues that explicit three-dimensional and physically annotated data remains far scarcer than internet video [9]. The path from words to worlds therefore leads back to the world itself.

SECTION

3. A world model still needs a place to stand

A frontier model can know what an airport, store, hospital or arena is. It does not automatically know this gate, this shelf, this closure or this person's position right now. General intelligence holds priors about the world. Physical service depends on the state of a particular place.

That distinction is the opening for simulatingMachines. We do not need to out-train the laboratories building general world models. We need to make their progress useful inside real venues. That requires a layer the model cannot provide by itself:

- a physical interface present at the moment of need;
- a current representation of the venue and the systems that hold its facts;
- supported actions, routes and a path to a person;
- operations that keep the service alive as the venue changes;
- outcomes that show whether the answer worked in the world.

Together, these form **physical AI support infrastructure**. A chatbot has language without the place. A map has the place without natural intent or an operating loop. A conventional kiosk is a surface, not a system that can learn from consequences. A frontier model laboratory can supply remarkable capability without owning installation, venue integrations or the service on the floor. Artificial Concierge Intelligence joins those missing pieces around the person who needs help.

Technically, this is a **venue adaptation layer**. A general world prior meets venue-specific state and a bounded action space. A Terminal or Host turns a human need into an interaction with that state. The outcome returns through operations and matched simulation-and-reality evaluation. The model can change; the role of the layer remains: make intelligence useful and measurable in one exact place.

The representation must also be useful, not merely convincing. A beautiful three-dimensional scene may help a person recognize a route while still being wrong about a closure, a service or what an action will change. Spatial intelligence research makes the same point from another angle: systems tested on spatial reasoning still struggle with relationships people handle naturally [10]. The important venue model is the one that can support a correct decision now and grow toward predicting the effect of change later.

This is also why the work is larger than installing an interface. Even very capable AI remains constrained by physical time, available data, institutional adoption and human decisions [11]. Capability does not distribute itself. It has to be connected to the systems and places where life happens. A place to stand is both a coordinate and an operating relationship.

SECTION

4. Terminal first. Host next.

The Terminal is a deliberate first form of physical AI. Its location is known. It is persistently available. A person can approach it without downloading an app, finding the correct website or leaving the moment for a remote support queue. Because the interaction starts inside the place, the useful answer can end in the world: a route, a task, a service point or the right person.

For the visitor, the product promise is simple: ask naturally and receive a next step the place can support. For the venue, the Terminal creates a consistent service point that can connect approved knowledge, spatial guidance and staff handoff. It can also reveal repeated friction: what people cannot find, which questions remain unresolved and where the physical experience fails to explain itself.

The Host carries that service into a more embodied form. Voice, visual identity and presence are intended to make assistance more approachable; mobility and sensing can eventually let the system experience more of the place it represents. The surface changes, but the foundation remains: the same venue model, supported actions, operating layer, simulation and human handoff.

This sequence is strategically important. A fixed Terminal lets us test whether useful service can arrive before general embodied intelligence is mature. A successful deployment would establish physical distribution, connect the company to live venue state and create a recurring reality check for the simulation. The Host then extends an operating system that already understands the venue instead of beginning as an isolated robot or character.

The first customer is the team accountable for service and operations in a place where the same physical questions recur. They do not need to buy a world model. They need people to get where they are going, find what they came for and reach staff when the situation requires one. The initial commercial unit is therefore an operated Terminal service and the software that keeps it useful. The larger platform grows from value delivered on the floor.

SECTION

5. Simulation before consequence

A physical AI cannot treat deployment as its first serious test. When an answer sends a person down a route, announces what is available or hands a request to staff, an error leaves the screen and enters someone's day. Simulation gives us a place to discover those failures before visitors do.

The useful twin is not the one with the most beautiful pixels. It is the one that preserves what can change an outcome: the geometry, current state, available actions, interfaces and relevant failure modes. Then it can put the service under pressure. A route closes. State falls behind the venue. The network disappears. Speech is ambiguous. A person's constraints change. The question is not whether the scene looks real. It is whether the machine can still help, or recognize the moment to hand over.

The field offers credible reasons to build this way. Simulation-trained control has transferred to physical machines, though the sim-to-real gap remains an active problem rather than a solved step [12]. World Labs has reported policies trained in aligned simulated worlds transferring across several robot platforms; the result is a company report, not an accompanying peer-reviewed publication [13]. The relevant lesson is not that a robotics result proves a venue service. It is that a faithful environment can become both a learning ground and an evaluation instrument while the physical world remains the final judge.

More capable models increase the value of that discipline. Systems can behave differently in new conditions, and physical action raises the cost of a confident mistake. Evaluation therefore cannot end when a general model ships; it has to continue in the environment where the model is asked to act [14].

That creates the loop implied by our name. An important failure in a venue returns as a repeatable scenario. The next release meets it before another visitor does. Matched simulated and physical outcomes reveal which details matter and where the venue model's assumptions break. Reality corrects the simulation; simulation makes hard-won experience reusable.

Over time, this can turn deployment from a series of bespoke installations into a system that compounds. Venue patterns become reusable. Integrations become repeatable. Difficult situations become a shared evaluation library. The next Terminal or Host starts with more of the world already understood, while every local venue remains grounded in its own current state.

SECTION

6. The service is the value. Experience becomes an asset.

The primary business is the service delivered to the person and the venue. A visitor receives help at the physical moment. Staff can spend more attention on judgment, care and exceptions. The venue gains an operating view of what people need, where journeys break and what should improve next. The Terminal and, later, the Host are products people can use and venues can operate. Their first job is useful service.

A properly instrumented and authorized service could create a form of experience that is difficult to obtain from the open internet. A grounded episode can connect human intent to the state of a real place, the evidence consulted, the action proposed and what happened afterward. The outcome is outside the transcript: the destination was reached, the item was found, the handoff completed or the plan corrected.

That alignment is already important in adjacent research. HoloAssist synchronizes verbal guidance, sensor streams, actions, mistakes and interventions during real physical tasks [15]. DROID aligns language instructions, observations, robot state and actions across diverse manipulation trajectories [16]. Their relevance here is structural: world-model and embodied-AI research values records that bind language and perception to action over time.

The two product forms can contribute different kinds of experience. A Terminal episode can bind a human goal to the venue-state version, evidence, permitted route or action, correction and outcome. A future Host episode can add synchronized perception, pose, motion and the resulting state change. A matched simulation-and-reality episode can record where the same initial state and action diverge. These are three different research objects, not one generic interaction log. Together, they form a path from place-grounded evaluation toward richer world-model training material.

The order matters. Operational records first improve the service that produced them. Where customer agreements and appropriate permissions allow, selected experience can then support internal evaluation and training. Curated scenario libraries, matched simulation-and-reality episodes and place-grounded datasets could become a second commercial layer for spatial and world-model builders. Their value would come from the precise alignment between a real place, an action and a consequence: a structure directly useful for evaluation and training.

The compounding advantage is broader than data alone. Each deployment can strengthen venue-state schemas, integrations, simulation scenarios, operating practice and physical distribution. Those assets improve the service and reduce the distance to the next place. The installed service earns the experience; the experience improves the system; the system makes the next deployment stronger.

SECTION

7. Make every complex place understandable

The promise of powerful AI is not fulfilled when capability exists in a datacenter. It is fulfilled when people can feel its benefit in ordinary life. Realizing that promise will take sustained work by many companies and institutions, not intelligence alone [11].

simulatingMachines intends to do one part of that work: make helpful AI present wherever people need assistance. The Terminal gives intelligence its first persistent foothold in the venue. The Host gives the same service a more embodied presence. The operating layer keeps both connected to the place. Simulation lets them learn before and after deployment. Each layer serves one human moment: a person needs help, the place understands enough to respond and the person can act.

If that works once, it is a useful machine. If it can stay current, improve and repeat, it becomes physical AI support infrastructure. A network of such systems could make complex places easier to understand while creating a new bridge between frontier world models and the environments in which people actually live, travel, shop, work and gather.

The longer horizon is not a smarter kiosk. It is a place with a working model of itself: one that can explain what is true now, rehearse what might happen next and support increasingly capable agents without losing the people they are meant to serve. The Terminal is the first point of contact. The Host is the embodied extension. The venue model and learning loop are what let the whole system grow.

AI needs a place to stand. We are building it at the point where intelligence meets a person, a place and a consequence, one real situation at a time.

APPENDIX

Notes and references

1. PEER REVIEWED Jingtao Ding et al. 2025. *Understanding World or Predicting Future? A Comprehensive Survey of World Models*. ACM Computing Surveys 58(3), 1-38. DOI 10.1145/3746449.
2. ESSAY / POSITION Fei-Fei Li. 2025-11-10. *From Words to Worlds: Spatial Intelligence is AI's Next Frontier*. Substack (drfeifei.substack.com). Original essay. Accessed 2026-08-23.
3. PEER REVIEWED David Ha and Jürgen Schmidhuber. 2018. *Recurrent World Models Facilitate Policy Evolution*. Advances in Neural Information Processing Systems 31 (NeurIPS 2018). Primary source.
4. PEER REVIEWED Danijar Hafner, Jurgis Pasukonis, Jimmy Ba and Timothy Lillicrap. 2025. *Mastering diverse control tasks through world models*. Nature, volume 640, pages 647-653. DOI 10.1038/s41586-025-08744-2.
5. PEER REVIEWED Amir Bar, Gaoyue Zhou, Danny Tran, Trevor Darrell and Yann LeCun. 2025. *Navigation World Models*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 15791-15801. Primary source.
6. COMPANY REPORTED PREPRINT Mido Assran et al. 2025-06-11. *V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning*. arXiv preprint (Meta FAIR). Company source. Accessed 2026-08-23.
7. COMPANY REPORTED PREPRINT Niket Agarwal et al. (NVIDIA, 77 authors). 2025-01-07. *Cosmos World Foundation Model Platform for Physical AI*. arXiv preprint (NVIDIA). Company source. Accessed 2026-08-23.
8. COMPANY REPORTED Google DeepMind. 2025-08-05. *Genie 3: A new frontier for world models*. Google DeepMind. Company source. Accessed 2026-08-23.
9. ESSAY / POSITION Fei-Fei Li. 2026-06-03. *A Functional Taxonomy of World Models*. Substack (drfeifei.substack.com), mirrored on the World Labs blog. Original essay. Accessed 2026-08-23.
10. PEER REVIEWED Jihan Yang, Shusheng Yang, Anjali W. Gupta, Rilyn Han, Fei-Fei Li and Saining Xie. 2025. *Thinking in Space: How Multimodal Large Language Models See, Remember, and Recall Spaces*. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2025, pages 10632-10643. DOI 10.1109/CVPR52734.2025.00994.
11. ESSAY / POSITION Dario Amodei. 2024-10. *Machines of Loving Grace: How AI Could Transform the World for the Better*. darioamodei.com. Original essay. Accessed 2026-08-23.
12. PEER REVIEWED Wenshuai Zhao, Jorge Peña Queralta and Tomi Westerlund. 2020. *Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: a Survey*. 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pages 737-744. DOI 10.1109/SSCI47803.2020.9308468.
13. COMPANY REPORTED World Labs. 2026-07-28. *Building Worlds That Train Robots*. World Labs. Company source. Accessed 2026-08-23.
14. ESSAY / POSITION Dario Amodei. 2026-01. *The Adolescence of Technology: Confronting and Overcoming the Risks of Powerful AI*. darioamodei.com. Original essay. Accessed 2026-08-23.
15. PEER REVIEWED Xin Wang et al. 2023. *HoloAssist: an Egocentric Human Interaction Dataset for Interactive AI Assistants in the Real World*. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 20270-20281. Primary source.
16. PEER REVIEWED Alexander Khazatsky, Karl Pertsch, Suraj Nair, Ashwin Balakrishna, Sudeep Dasari, Siddharth Karamcheti et al. 2024. *DROID: A Large-Scale In-The-Wild Robot Manipulation Dataset*. Proceedings of Robotics: Science and Systems XX. DOI 10.15607/RSS.2024.XX.120.