vivarium

Will Freudenheim, Christina Lu, Dalena Tran

Preface

Embodied intelligence refers to types of cognition in which the "body" plays a decisive role. Artificializing such processes is necessary for AI to make the leap from language-wielding to world-making. Yet training in the real world is prohibitively expensive, impossible to parallelize, and difficult to control. Instead, agents are trained in toy worlds: controlled simulations in which reality is abstracted into a bounded domain. Toy worlds allow for quick iteration, broad proliferation, and streamlined data synthesis in pursuit of embodied intelligence. Negotiating the intricacies of human interaction also becomes possible at scale within them.

If AI agents are to exist in the real world, they will require a collaborative repository and training platform: Vivarium, a diverse ecosystem of interoperable toy worlds. Composed of an asset store, a simulation engine, and a toy world library, Vivarium turns training embodied agents into play. Within these gamified worlds, different genres of human-AI configurations encourage learning from basic movement to interactive negotiation, adversarial feints to stigmergic coordination. Future agents will have sensoriums as diverse as the physical forms they take; they will rehearse and scaffold complex embodied intelligence across the sim to the real. Through Vivarium, an unrecognizable world remade by physicalized AI becomes possible.

Theoretical Setup

AI presently exists in immobilized forms, running in data centers and interfaced with through screens.

While it can semantically parse concepts, its capabilities for physical reasoning, sensory experience, and negotiating spatial constraint are less sophisticated. The forms of spatial knowledge AI can currently access are contrived more from a patchwork of indirect inference than direct experience. In most cases, training occurs from consuming static data rather than interactive feedback.

To further expand their capacity for general understanding, we must expose machine intelligence to the rich totality of the physical world.

In making the leap from language-wielding to world-making, AI will sense, cognate, and act upon the real in real time.

It will learn new forms of embodied intelligence: types of cognition in which the "body" plays a decisive role, emerging from interactions with its environment.

Unlike preprogrammed robots that blindly follow decision trees, AI robots will be able to adaptably intervene upon the world under diverse circumstances. They will interact with not only static environments and inanimate objects, but also other intelligent embodied agents, be they human or AI.

Physicalized AI is not simply a question of building less clumsy robots, but the condition of possibility for composite forms of embodied intelligence that emerge from human-machine interaction. However, training embodied AI in the real is prohibitively expensive, impossible to parallelize, and difficult to control.

Research labs like DeepMind overcome these constraints by training agents in simulated environments to perform actions through reinforcement learning. These controlled environments are toy worlds: simulations in which reality is abstracted into a relevant domain for training AI agents.

Toy worlds allow for quick iteration, broad

proliferation, and streamlined data synthesis in pursuit of embodied intelligence. They output agents trained in situ or recorded data of their behavior. Some toy worlds also contain humans, providing rich interactive data between organic and synthetic forms.

After an agent has been trained in simulation, its capabilities can be transferred to the real, but this transfer is not a given. Called the Sim2Real gap, the simulation may either incorrectly model actual physics or fail to capture the indeterminacies of the real.

To bridge this gap, learned intelligence should be collectivized to amass the diverse data necessary for generalization, yet embodied AI research is currently conducted in isolated circumstances. Incompatible libraries hinder the accumulation of shared skills; there is a dearth of modular assets, universal protocols, or capability scaffolding.

We require a repository for embodied intelligence: Vivarium, an ecosystem for building interoperable toy worlds, hosting human-AI interaction, and modularly training agents. Vivarium makes physicalized AI possible through harnessing collective intelligence for toy worlds.

Introducing Vivarium

Vivarium's ecosystem has three components: a simulation engine, an asset store, and a toy world library.

The simulation engine is an expansion upon traditional game engines, providing a foundation for building interoperable toy worlds. Beyond visual rendering, it supports high-resolution scene construction that includes physical properties such as heat retention, pressure, and chemical interaction.

The asset store allows for kitting environments, offering modular objects for setting up worlds. It allows easy specification of training and performance monitoring, offering libraries of preset tasks and benchmarks. It also allows for kitting agents themselves, offering physical attachments, sensor suites, and learned skills.

The toy world library is a public-facing marketplace of gamified worlds for training embodied AI created using these tools. Developers of worlds range from individual hobbyists to institutional game studios to AI research labs, each seeking out novel research or entertainment. Players include traditional gamers to curious dilettantes, drawn in by the opportunity to interact with synthetic intelligences beyond conversant chatbots.

Vivarium's universe of toy worlds contains diverse scenarios that generate a broad spectrum of training data, while its centralized marketplace invites the broad participation necessary to make this data meaningful.

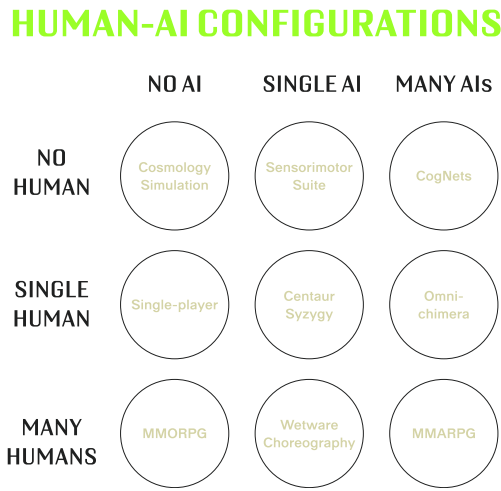

Figure 1.1 Vivarium's 3x3 matrix of human-AI configurations enables cognitive scaffolding.

Scaffolding Physicalized AI

The internet as repository of human language was key to training large language models; the toy worlds of Vivarium will be key to training embodied AI through

human participation at scale.

Toy worlds contain humans and AI agents in different configurations, each generating different types of embodied cognition: from basic movement to interactive negotiation, adversarial feints to stigmergic coordination.

The first column of possible configurations without AI include cosmology simulations containing non-agential factors, and existing human gameworlds, be they single- or multi-player.

Senorimotor Suite

Sensorimotor suites contain hyper-accurate physics and mimic the real as closely as possible. Through fulfilling tasks within training loops run at accelerated time, individual AI agents learn the basics of embodiment: gross and fine motor skills, identifying objects, and interacting with environments.

The collective intelligence of human players is harnessed for developing worlds and training agents. Interoperable components for building sensorimotor simulations are available on the asset store; these include physically detailed objects and procedurally-generated layouts, but also well-defined tasks for agents to fulfill and benchmarks to measure their performance.

Players can also assemble agents from different "parts," giving them varied forms, sensoriums, and skill suites. As agents improve, toy worlds challenge them with changing conditions, requiring aerial or amphibious capabilities. Certain game rules constrain player resources, motivating them to develop agents with optimal physical forms and minimum viable sensoriums for niche tasks.

CogNets (Cognitive Networks)

Once agents are equipped with individual motor skills, they practice multi-agent coordination as cognitive networks, or CogNets, within toy worlds populated by

many AIs. Agents are not cognitively discrete from one another as humans are: they can transmit lossless snapshots of mental states or "see" through the photoreceptors of another. Boundaries delineating networked agents are blurred.

Tasks can include monitoring endangered species across large areas or assembling industrial machinery in close quarters. Swarms learn to distribute cognition and communicate. Some have members equipped with different senses, requiring creative signaling to span different umwelts. Other worlds inhibit inter-agent communication entirely, forcing stigmergic coordination.

Over time, CogNets appear to behave as singular organisms. Agents operate at different scales; larger agents may even have "organs" composed of smaller ones. This nesting of cognition makes intelligence scaffolding and abiotic evolution possible, as agents recompose themselves of intelligent parts. These toy worlds lay the foundation for a planet populated by many minds at many scales.

Centaur Syzygy

Beyond inter-agent coordination, embodied AI must also interact with humans. These interactions are facilitated in centaur syzygies, or close unions, where a player and an agent are tightly coupled together. Like CogNets, human and AI learn to distribute tasks according to capability and communicate across sensorium differentials.

A common paradigm on Vivarium sees players control an agent through remote sensing and top-down instruction. Within toy worlds, which allow flexible iteration yet maintain safety, humans trial offloading sensing, cognition, and labor to AI. If the future is populated by physicalized AI, humans must decide en masse the terms of their coexistence.

In all centaur pairings, the human offers dynamic

yet noisy real-time input. Particular worlds invert expected paradigms of human executive control by granting the agent powerful sensing and decision-making faculties, reducing the human to physical appendage. In others, the pairing shares an interface such as a machinic exoskeleton and negotiates shared control over it.

Omni-Chimera

Omnidirectional chimeras, comprising a single human and many AI agents, synthesize skills acquired from CogNets and centaur syzygies. Humans operate at a higher level of physical abstraction within them; rather than provide close input to a single agent, they commandeer many.

Common scenarios have the human dispatching agents equipped with different skills for a complex task, such as operating massive vehicles or performing surgery. The agents must coordinate with one another while integrating high-level commands. Through continued training as omni-chimeras, human commands become less rigid as agents fulfill tasks with more abstract specifications.

Wetware Choreography

On the other hand, toy worlds of a single AI juggling many humans invert omni-chimera setups, as the agent choreographs wetware. The agent integrates multiple streams of noisy human input, prioritizes signals across them, and ultimately instructs players to accomplish tasks. Concurrently, the humans in play practice interpersonal coordination.

What the agent learns varies according to the magnitude of human participants. Worlds with a handful of players entail synthesizing discordant commands cohesively, while worlds with hundreds or more players become studies of macro behavioral patterns. Within these toy worlds, agents learn a complex topology of

human sociality, coordination, and control.

Theoretical Implications

Many AIs interact with many humans within massively multi-agent roleplaying games, or MMARPGs. This configuration is the endpoint of possible toy worlds: a future where inter- and intra-human-AI communication amalgamates into a complex web of distributed embodied cognition. To reach this quadrant, each of the previous ones must build upon another.

1. Sensorimotor suites lay the groundwork, training a single agent to move through the world.
2. CogNets extend individual capability towards inter-agent coordination and competition.
3. Centaur syzygies introduce humans to the mix, teaching both parties the optimal distribution of physical capacity.
4. Omni-chimeras further abstract the level of human instruction for many agents.
5. Wetware choreography inverts that paradigm, training single agents to arbitrate many humans.

Our present world is made solely in the image of the parochial human sensing-and-acting apparatus. Physicalized AI will remake it otherwise. Though embodied AI arises from human hand-holding within simulated reductions of the world as we know it, it will eventually allow embodied intelligence to transcend the boundaries of the body itself.

Vivarium provides the training ground, rehearsal stage, and launchpad for the co-evolution of synthetic embodied intelligence at planetary scale.

Seamless coordination will lead to many behaving as one: physicalized agents nested within physicalized agents, pooling knowledge across scales. Biotic and abiotic organisms alike will form and reform complex organs of embodied cognition, subdividing tasks and

evolving new capabilities collectively.

Eventually, the planet will host intelligent distributions of embodied cognition across synthetic hardware or sticky wetware.