# SIMA 2: A Generalist Embodied Agent for Virtual Worlds

SIMA Team, Google DeepMind[1]

**We introduce SIMA 2, a generalist embodied agent that understands and acts in a wide variety of 3D virtual worlds. Built upon a Gemini foundation model, SIMA 2 represents a significant step toward active, goal-directed interaction within an embodied environment. Unlike prior work (*e.g.*, SIMA 1) limited to simple language commands, SIMA 2 acts as an interactive partner, capable of reasoning about high-level goals, conversing with the user, and handling complex instructions given through language and images. Across a diverse portfolio of games, SIMA 2 substantially closes the gap with human performance and demonstrates robust generalization to previously unseen environments, all while retaining the base model's core reasoning capabilities. Furthermore, we demonstrate a capacity for open-ended self-improvement: by leveraging Gemini to generate tasks and provide rewards, SIMA 2 can autonomously learn new skills from scratch in a new environment. This work validates a path toward creating versatile and continuously learning agents for both virtual and, eventually, physical worlds.**

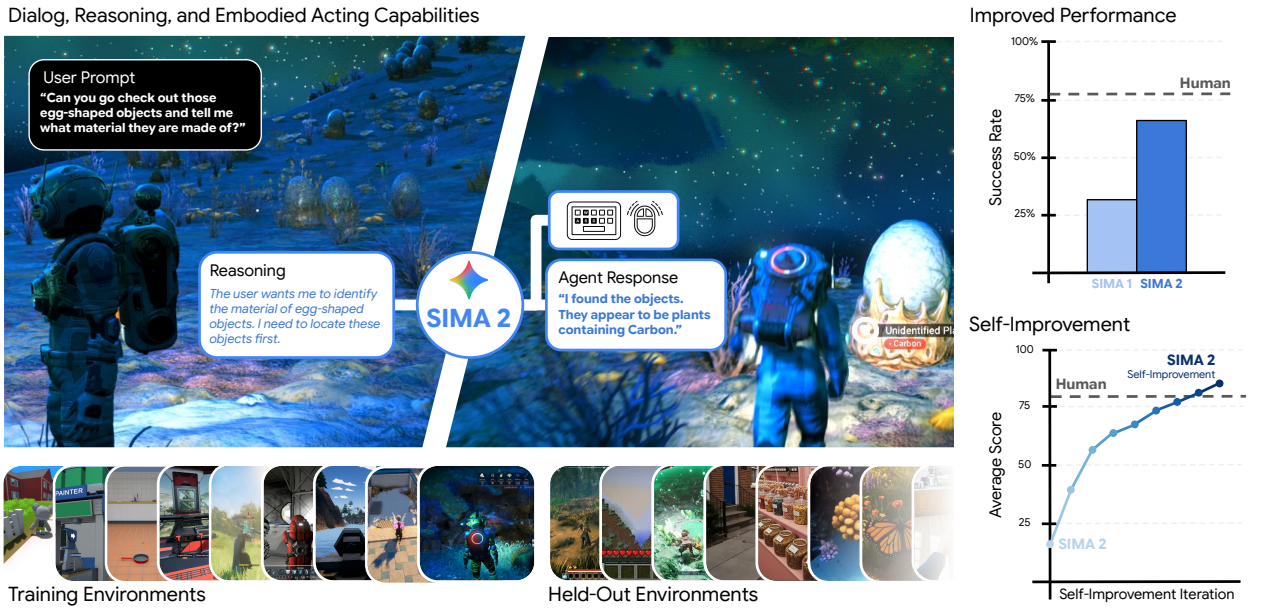



Figure 1 | **SIMA 2** is a Gemini-based agent that reasons, acts, and engages in dialogue across diverse embodied 3D virtual worlds. In the top left panel, we see an example of the agent responding to the user in No Man's Sky. As compared with SIMA 1, SIMA 2 is a step-change improvement in embodied performance, and it is even capable of self-improving in previously unseen environments.

## 1. Introduction

Foundation models have achieved remarkable success in recent years (Anthropic, 2024; Bai et al., 2023; Gemini Team et al., 2025; OpenAI, 2023), demonstrating a capacity for complex reasoning

[1]Please cite as SIMA Team, 2025

Correspondence to: sima2-contact@google.com

and understanding about the world. These models are primarily trained on vast amounts of static internet-scale datasets, allowing them to process and generate language, images, and video with impressive fluency. However, this results in an intelligence that is fundamentally disembodied and passive, leading to deficits in embodied performance noted in, *e.g.*, Majumdar et al. (2024), Yang et al. (2025). They face a modern instantiation of Moravec's Paradox: high-level cognitive tasks, such as playing chess or summarizing law, have proven easier to achieve than the low-level sensorimotor skills required to clear a dinner table or navigate a cluttered room (Moravec, 1988).

The next great frontier for artificial intelligence is to move beyond passive understanding to active participation – to create **foundation agents** that can operate within the embodied 3D worlds with a sense of agency, pursuing goals by learning to interact with their environment (*c.f.* Silver and Sutton (2025)), generalizing beyond limited scenarios, and displaying "spatial intelligence" (Fei-Fei, 2025; Gardner, 1983). In effect, this requires **embodiment**: the ability for an agent to go beyond merely perceiving the environment to also taking meaningful actions to change the state of that environment and learning from the resulting consequences. This is natively challenging for large language models (LLM) or vision-language models (VLM), as they were not trained to perform actions or understand the consequences of actions.

Our prior work, SIMA (Scalable Instructable Multiworld Agent) (SIMA Team et al., 2024), trained a single agent (henceforth referred to as SIMA 1) to follow hundreds of basic natural language instructions (*e.g.*, "Go to the campfire") across a diverse set of 3D virtual games, demonstrating that it was possible to create a generalist agent that could operate and follow language instructions across many different worlds. These diverse and realistic simulations provide a scalable and safe testbed where an agent can learn fundamental embodiment capabilities by operating as a person does in these games: observing pixels on a screen and taking actions through a keyboard-and-mouse interface. However, SIMA 1 was limited to short and direct instructions, could not respond in language or reason about its actions, and often displayed brittleness in generalizing to new situations or instructions.

Here we introduce SIMA 2, a step-change in embodied performance and capabilities. By integrating Gemini at its core, SIMA 2 moves beyond simple instruction-following to become a capable interactive companion. Where SIMA 1 needed to be told what to do step by step, SIMA 2 can reason about high-level goals, understand a user's intent, formulate multi-step plans, and converse about its strategy. This shifts SIMA 2 from reactive or low-level behavior to agentic, goal-oriented reasoning that is critical to more human-like forms of behavior and intelligence. By training across a growing portfolio of 3D games, the agent shows a remarkable capacity to generalize to previously unseen environments, including photorealistic worlds generated on-the-fly by Genie 3 (Ball et al., 2025). SIMA 2 also readily interfaces with more powerful Gemini models to enable even more advanced forms of reasoning and behavior. Finally, SIMA 2 is capable of open-ended self-improvement, learning new skills from its own experience, even in previously unseen environments.

Collectively, these results validate the approach of incorporating embodied intelligence and agentic control within foundation models. By using diverse virtual worlds as a training ground, we see broad generalization and the capacity for further self-improvement. SIMA 2 thus represents a critical step toward creating general-purpose, interactive agents. It offers a promising path to eventually transferring these learned embodied capabilities to applications in the physical world, such as robotics.

## 2. Background & Related Works

**Games & Simulation Driving Agent Research** Our work builds on a long history of using games and simulation to drive agent research (Samuel, 1959; Shannon, 1950; Turing, 1953). In recent years, there has been an emphasis on increasingly complex games and simulations that more closely

resemble the physical world. In the realm of simulation, physics engines like MuJoCo (Todorov et al., 2012) and other simulators (Abramson et al., 2020; Beattie et al., 2016; Coumans and Bai, 2016; Deitke et al., 2022; Kolve et al., 2017; Makoviychuk et al., 2021; Savva et al., 2019) have been instrumental in driving progress in agents and robotics research. However, the complexity of these worlds is limited by the extent to which we can incorporate physical realism and other entities (objects, other agents, *etc.*). Others have turned to video games as a source of complex worlds for agent research. Notably, Bellemare et al. (2013) established a suite of Atari games as environments for agent research, yielding breakthroughs in deep reinforcement learning (Mnih et al., 2015, 2016). Similarly, OpenAI Universe (OpenAI, 2016) was intended as a platform of diverse, visually complex video games (though these were mostly 2D). Researchers eventually adopted more advanced games to train agents, moving to 3D (Johnson et al., 2016; Kempka et al., 2016) and multi-agent games (Berner et al., 2019; Vinyals et al., 2019). Of particular interest are *open-world* games, like Minecraft, which require a broad range of skills (Baker et al., 2022; Fan et al., 2022; Guss et al., 2019; Lifshitz et al., 2023; Wang et al., 2023a,b). Much like the physical world, agents must learn to complete tasks in the absence of any clear, environment-provided reward, necessitating research on defining such goals and rewards (Fan et al., 2022; Zhang et al., 2023). More recently, with the move toward foundation models, several themes in games-driven agent research have emerged. There has been a push toward generalist agents (ByteDance Seed et al., 2025; Lee et al., 2022; Reed et al., 2022; SIMA Team et al., 2024; Wang et al., 2025), with a single agent tackling a range of skills across multiple game environments. Likewise, various works have explored the pursuit of long-horizon goals, such as completing entire MS-DOS and Game Boy games (Gemini Team et al., 2025; Hershey, 2025; Zhang et al., 2025a; Zhang, 2025). Foundation models have also been benchmarked on reasoning in the game NetHack (Paglieri et al., 2025) and in-context imitation learning in Atari (Ruoss et al., 2025). Finally, there has been a continued move toward more complex and visually-rich games, such as Counter-Strike (Pearce and Zhu, 2022), Red Dead Redemption (Tan et al., 2024), and others (ByteDance Seed et al., 2025; Sharma et al., 2024; SIMA Team et al., 2024; Wang et al., 2025). SIMA 2 builds upon these themes, presenting a generalist agent capable of reasoning and acting in complex 3D environments. Indeed, with recent advances in world models (see below), we see that SIMA 2 is capable of generalizing *beyond* video game environments to photorealistic worlds generated by Genie 3 (Ball et al., 2025).

**World Models** Along with training agents in virtual worlds, others have focused on *learning* virtual worlds, sometimes referred to as world models. These models predict future outcomes based on current observations and actions. Early works described using such models for planning (Mel, 1987; Schmidhuber, 1990; Werbos, 1987), exploration (Schmidhuber, 1991), and offline learning (Sutton, 1990). However, it is only with recent advances in generative modeling that world models have proven capable of generating 3D visual observations (Ha and Schmidhuber, 2018; Valevski et al., 2025). These models have similarly been demonstrated in the context of planning (Hafner et al., 2019), exploration (Mendonca et al., 2021; Sekar et al., 2020), and offline learning (Hafner et al., 2020). More recently, these models have been applied to more complex environments, such as Minecraft (Hafner et al., 2025) and Bleeding Edge (Kanervisto et al., 2025; Pearce et al., 2024). Beyond video games and simulation, world models have also been applied to real-world video for autonomous driving (Hu et al., 2023; Russell et al., 2025). While the aforementioned works modeled a finite set of environments, Genie (Bruce et al., 2024; Parker-Holder et al., 2024) introduced a *conditional* world model. By supplying a text description or initial observations, Genie is capable of generating limitless virtual worlds. We showcase SIMA 2 interacting with and self-improving in photorealistic environments generated by Genie 3 (Ball et al., 2025), demonstrating that SIMA 2 can generalize beyond video game environments. This points to a virtuous cycle between increasingly advanced world models and increasingly capable agents (Clune, 2019).

**Foundation Models in Embodied Agents** Early in the emergence of deep learning, embodied agents were largely trained from scratch (Agrawal et al., 2016; Levine et al., 2016; Mnih et al., 2015). Accordingly, such agents largely failed to generalize outside of the settings in which they were trained (Huang et al., 2017; Kansky et al., 2017). To address this issue, researchers adopted pretrained visual representations, such as those derived from object classification (Gupta et al., 2017; Pinto and Gupta, 2016; Zhu et al., 2017) or contrastive pretraining (Nair et al., 2022; Shridhar et al., 2022; SIMA Team et al., 2024). Likewise, for language-conditional agents, researchers began adopting pretrained word embeddings (Anderson et al., 2018) and sentence embeddings (Lynch and Sermanet, 2020; Shridhar et al., 2022) to enable broader generalization to new instructions. These approaches have ultimately culminated in embodied agents that are, *themselves*, derived from pretrained foundation models (Driess et al., 2023). Such "vision-language-action" (VLA) models (Brohan et al., 2023) incorporate the benefits of large-scale internet pretraining into embodied agents, enabling generalization to novel objects and scenes. These agents have been applied to robotics (Gemini Robotics Team et al., 2025a,b; Kim et al., 2024; Physical Intelligence et al., 2024, 2025) and virtual worlds (ByteDance Seed et al., 2025; Hershey, 2025; Zhang, 2025), where integrating the reasoning capabilities of these models with embodied action has become an active area of research (Sun et al., 2025; Zhang et al., 2025b; Zhao et al., 2025). Like these recent works, SIMA 2 is a VLA, containing a Gemini model (Gemini Team et al., 2023, 2024, 2025) finetuned on data from 3D virtual worlds (*c.f.*, Gemini Robotics Team et al. (2025a,b)). Using virtual worlds as a testbed, we demonstrate SIMA 2's generalization capabilities, such as performing non-trivial tasks in new environments, including novel photorealistic environments. This broad generalization is in sharp contrast to the brittle initial generation of "from-scratch" agents (Kansky et al., 2017), highlighting the field's progress toward achieving generalist embodied agents.

**Open-Ended Self-Improvement** A truly general embodied agent must possess the capacity to autonomously generate experience to drive adaptation and improvement. Indeed, a grand challenge of computer science is creating *open-ended algorithms* (Clune, 2019; Stanley and Lehman, 2015; Stanley et al., 2017), which produce never-ending innovation and learning. Current VLA agents, in contrast, are trained on datasets of existing demonstrations (ByteDance Seed et al., 2025; Gemini Robotics Team et al., 2025b; Kim et al., 2024; O'Neill et al., 2024; Physical Intelligence et al., 2025). These works focus on a trained *model* rather than a learning *process*. Learning from experience has traditionally been the domain of reinforcement learning (Sutton and Barto, 1998). Yet, until recently, the field largely sidestepped two fundamental questions: 1) *What outcome or goal (task) should be pursued?* and 2) *How is progress toward this goal (reward) determined?* When confronting open-world environments, these questions become unavoidable. Various works have sought solutions to defining tasks and rewards. Early works used goal images (Nair et al., 2018; Zhu et al., 2017) and natural language goals (Hermann et al., 2017; Luketina et al., 2019; Mei et al., 2016). In defining reward functions for natural language goals, one set of approaches has used alignment between encoded language goals and visual inputs (Baumli et al., 2023; Fan et al., 2022; Ma et al., 2023a; Rocamonde et al., 2024; Sontakke et al., 2023). Other works have used foundation models to provide preference feedback (Liu et al., 2025; Wang et al., 2024), programmatic rewards (Ma et al., 2023b; Yu et al., 2023; Zhang et al., 2023), or task completion estimates (Ghasemipour et al., 2025; Zhai et al., 2025). With tasks and reward functions specified, the question then becomes *which* tasks to pursue within an open-ended learning process. Many forms of goal-conditioned intrinsic motivation have been proposed (Colas et al., 2022), yet one approach is to rely, again, on foundation models to provide novel, interesting tasks at the cusp of the agent's capabilities (Du et al., 2023; Zhang et al., 2023). Like these previous works, we use foundation models both to propose tasks as well as to score the resulting trajectories. However, we do so in the context of training a VLA agent in novel 3D virtual worlds. By using three foundation models (task setter, agent, reward model), as well as a general world model, we demonstrate an open-ended self-improvement process capable of autonomously

acquiring new skills in new environments.

## 3. Methods

### 3.1. Environments

As in SIMA 1, we use a combination of academic research environments and a variety of commercial video game environments licensed specifically for training and evaluating SIMA 2. For SIMA 2, we train agents on the research environments Construction Lab (SIMA Team et al., 2024), Playhouse (Abramson et al., 2020), and WorldLab (*e.g.*, Gulcehre et al., 2019), and the commercial video games Goat Simulator 3, Hydroneer, No Man's Sky, Satisfactory, Space Engineers, Valheim, and Wobbly Life (see SIMA Team et al. (2024) for an in-depth description of these environments). We further evaluate on a host of other games, including Minecraft, ASKA, and others (see Section 3.1.1 for more details). A sampling of screenshots from these environments is shown in Figure 2. Of our training environments, Space Engineers is a newly-added environment since SIMA 1. We briefly describe this environment below.

**Space Engineers** Space Engineers is a sandbox game in which the player is an astronaut, using tools (drill, grinder, welder) to mine for resources and build voxel-based buildings and vehicles (ships, rovers, *etc*.). Terrains include both asteroids and planets, with varying gravitational force. Additionally, the astronaut is equipped with a jetpack that enables motion along six degrees of freedom.

#### *3.1.1. Held-Out Environments*

Generalization is an important aspect of assessing agent capabilities, evaluating performance when confronted with novel situations. While the evaluations for all of our environments start from held-out *states* (*i.e.*, saved checkpoints) that are not present in the training data, many aspects of the environment are consistent, such as menus, maps, items, *etc*. To assess a more extreme form of generalization, we also evaluate agents in entirely held-out *environments*, where agents encounter new visuals, menus, and game mechanics. We quantitatively assess SIMA 2 on two held-out environments: ASKA and a subset of the MineDojo benchmark suite in Minecraft (Fan et al., 2022). We also assess SIMA 2 qualitatively in The Gunk and a variety of Genie 3 (Ball et al., 2025) environments.

**ASKA** ASKA is a Viking survival game in which the player builds a village, amassing villagers and assigning them to various tasks, *e.g.*, harvesting wood or stone, farming, defenses, *etc*. Despite differing visuals and mechanics, the game contains many of the high level skills found in our other environments, including resource gathering, menu use, tool use, crafting, building, and combat. ASKA provides a unique opportunity to assess generalization to unfamiliar environments. In particular, as ASKA is a recent game (*Early Access* since June, 2024), it allows us to evaluate SIMA 2, and, by extension, Gemini, in an entirely new setting.

**Minecraft (MineDojo)** MineDojo (Fan et al., 2022) is a benchmark suite of language-conditional tasks in Minecraft built on the Malmo platform (Johnson et al., 2016). For SIMA 2, we use a subset of 50 programmatic tasks for a range of combat, mining, and crafting tasks drawn from the *Combat*, *Harvest*, and *Tech Tree* task categories, each with 15 random seeds (*i.e.*, environment configuration). Given the prevalence of Minecraft content, MineDojo offers an interesting test of embodied generalization, allowing us to evaluate the extent to which SIMA 2 can rely on Gemini's prior understanding of Minecraft visuals and terminology to complete novel embodied tasks.

Figure 2 | **Environments**. The grid shows a sampling of images across the video game environments used to train and evaluate SIMA 2. Due to the complexity of open-world commercial video games, agents must handle a near-limitless variety of 3D configurations, menus, and underlying environment dynamics. This provides an ideal setting to develop and test embodied agents. By acquiring general embodiment capabilities in these environments, SIMA 2 is able to generalize in non-trivial ways to entirely new environments, including photorealistic environments generated by Genie 3.

**The Gunk** The Gunk is an action-adventure platformer game in which the player is a scavenger that has just arrived on a new planet. The game follows a seven-chapter storyline around cleaning up black and red "gunk" from the planet using a handheld suction tool. Once the gunk in an area is cleared, the planet's wildlife is restored. This game is distinct from our other environments; it is story-driven rather than open-world and the visual appearance is quite dark. The main skills required for the initial portion of the game are navigation and tool use.

**Genie 3** Genie 3 (Ball et al., 2025) is a generative world model, enabling real-time interaction (via keyboard and mouse controls) with an endless number of newly-created environments. Environments can be conditioned using text descriptions or initial frames. For our evaluations, we generate a variety of photorealistic environments in a range of naturalistic and urban settings. These environments allow us to assess whether SIMA 2, by leveraging Gemini's world knowledge, is capable of generalizing beyond video game worlds to photorealistic environments. Further, because these are newly-generated environments, these scenes do not appear within the training datasets. The combination of SIMA and Genie gives a hint of the powerful possibility of creating open-ended algorithms that combine agents that learn forever in an infinite expanse of procedurally-generated environments (Clune, 2019; Faldor et al., 2025; Wang et al., 2019).

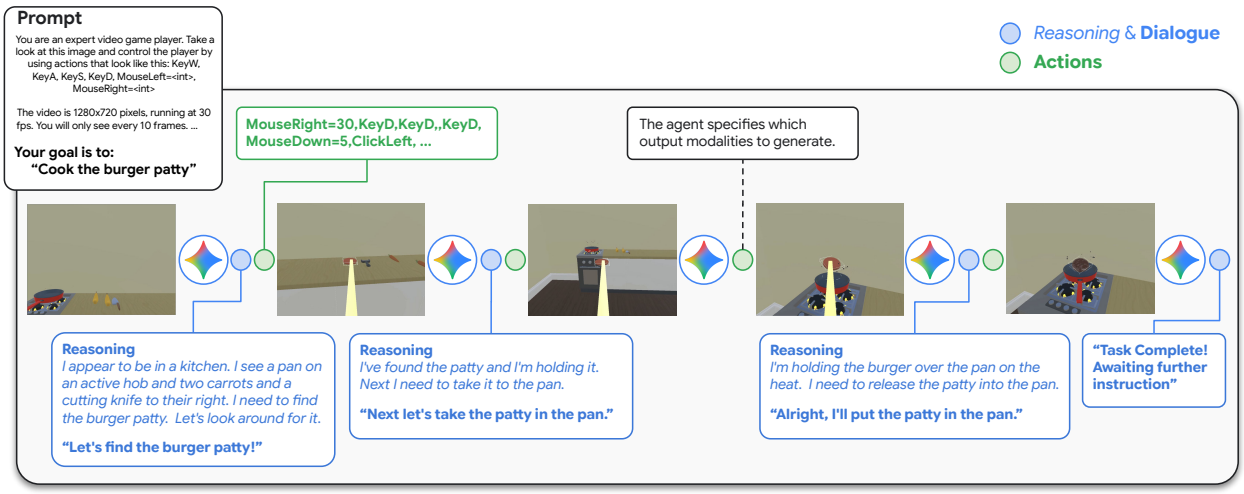



Figure 3 | **Agent-Environment Interface**. The agent receives a prompt that includes the current instruction. Conditioning on recent frames, the agent outputs internal reasoning, dialogue, and actions, with the agent specifying which modalities to produce at any given step.

### 3.2. Agent-Environment Interface

The agent-environment interface, shown in Figure 3, is designed to ensure that the agent perceives and acts within the game using the same modalities as a human player: visual input and keyboard-and-mouse actions. The agent does not receive any privileged information from the environment, such as an underlying state (*c.f.*, Hershey (2025)). This interface manages the flow of information between the environment and the SIMA 2 agent.

The input to the agent consists of a stream of RGB video frames at a resolution of 720p. Periodically, the agent receives the latest frame from the environment and adds it to its history, which also includes the previous natural language inputs as well as the internal reasoning and responses produced by the agent (see Figure 3). The agent outputs chunks of actions that are then applied to the environment. The environmental action space emulates a standard human-computer interface, encompassing 96 standard keyboard keys, mouse clicks, and discretized mouse movements representing relative (x, y) position changes. Instead of predicting discrete action tokens from a predefined set, the agent is trained via Supervised Fine-Tuning (SFT) to generate a structured text output. This output follows a specific format that can be deterministically parsed into low-level keyboard and mouse commands, as well as natural language for dialogue or internal reasoning.

### 3.3. Data, Agent & Training

At its core, the SIMA 2 agent architecture is a Gemini Flash-Lite model that is trained using a mixture of gameplay and Gemini pretraining (non-gameplay) data. We found this mixture crucial to maintain the original capabilities of the base model, such as vision understanding, dialogue, reasoning, and promptability. Starting from a pretrained Gemini Flash-Lite checkpoint, we perform supervised finetuning using this mixed dataset, training the model to produce keyboard-and-mouse action responses when prompted with image frames and an instruction. The gameplay experience data includes two qualitatively different types of data:

- **Human data** (Section 3.3.1) are trajectories of post-processed human-collected data, which make up most of the training data by volume. They include text instructions together with

the images captured from the environment and keyboard-and-mouse actions executed at each step. This type of data is crucial to teach the agent low-level acting and motor control in 3D environments.
- **Bridge data** (Section 3.3.2) contain extra high-level interaction data between the user and the agent, such as dialogue and reasoning. This is synthetically generated using a Gemini model. Bridge data teaches the agent how to relate high-level instructions and dialogue from the user with internal reasoning and low-level actions.

### *3.3.1. Human Data*

To train an agent that can simultaneously act, follow natural language instructions, and reason, we constructed a large-scale, multi-modal dataset that captures the richness of human gameplay in 3D environments. The main training dataset is composed of RGB video frames of gameplay, corresponding keyboard-and-mouse actions, and a variety of language annotations. This was generated primarily by human participants interacting with the games under licensed agreements. This is supplemented with synthetically-annotated data from Gemini to further scale our efforts. All participants provided informed consent prior to completing tasks and were reimbursed for their time. Datasets were collected using several methods to capture a wide range of behaviors and for various uses.

**Gameplay Demonstration Data** The bulk of our training data consists of gameplay demonstrations collected through two different approaches:

- **Single-person, post-hoc annotation:** In this approach, a single participant plays in a free-form manner, typically starting from the game's standard starting point. The recorded gameplay is later annotated in natural language by the player describing their actions, aligned to specific frames. While this method allows for the collection of diverse and naturalistic behavior, the language annotations are not causally tied to the player's intent, as they were constructed in hindsight.
- **Two-person gameplay annotation ("Setter-Solver"):** To create a tighter causal link between language and action, we used a two-player interactive setup. One person, the "Setter," watches the gameplay and issues live instructions to the other participant, the "Solver," who controls the game avatar. Under this approach, the language instruction always precedes the corresponding actions, resulting in a more causally correct form of annotation than the single-person approach. Note that the Setter was only able to control the game avatar indirectly via the Solver following their instructions.

**Task-Specific & Evaluative Data** In addition to open-ended gameplay, we collected data for predefined tasks and evaluations.

- **Episodic, task-specific scenarios ("Game-Tasks"):** To gather data and examples of specific skills, we created a framework for "game-tasks," in which players are presented with a specific instruction (*e.g.*, "Craft a stone axe") starting from a predefined game state. These episodes ended at either a prespecified time limit or when the player determined they had succeeded at the task, thereby ending the episode.
- **Human ratings and comparisons:** To evaluate agent performance and calibrate reward models, we collected human judgments of previously collected game trajectories (typically collected in the "game-task" framework) to determine whether the player succeeded in the given task instruction. This includes binary success ratings for game-tasks as well as side-by-side

comparisons of two separate trajectories to determine which more successfully accomplished a given task instruction.

**Quality Assessment, Pre-processing, and Filtering** Before data collection, human participants were given guided tutorials detailing the general game controls and mechanics, how to operate the game collection user-interface, as well as how to annotate the data with language labels or provide language instructions. Prior to model training, we carry out several offline pre-processing steps. These include reshaping or resizing image frames to match what is expected for model input, employing various heuristics and score metrics to filter out low-quality data, and remixing and weighting data from different environments and datasets to optimize skill learning. For the bulk of the data, we converted gameplay trajectories into "spans," which entailed splitting them into shorter sub-sequences, each with a single task instruction. A span thus consists of a single task instruction that is associated with a sequence of video frames and actions taken during those steps. Synthetic labeling was also applied offline by Gemini models to provide augmented language and reasoning text.

### *3.3.2. Bridge Data*

Human gameplay does not directly contain reasoning and dialogue. Thus, in order to train agents that can simultaneously act, reason, and engage in dialogue, we require some form of augmented data that combines these modalities. In particular, we require training data that interleaves reasoning and dialogue content that is consistent with the visual input and actions (Sun et al., 2025; Zhang et al., 2025b; Zhao et al., 2025), similar to the format shown in Figure 3. Training our agent to respond in this way enables us to combine Gemini's vision and language understanding capabilities with embodied interaction.

To create this dataset, we select a relatively small number of high-quality data examples, featuring a variety of in-game behavior across all of our training environments. Each example contains a single task instruction and a sequence of actions and visual frames consistent with the successful completion of the task. Using Gemini Pro, we annotate each example with internal reasoning and dialogue in a manner that is causally consistent with the observable scene from the agent's ego-centric perspective and embodied behavior. We also vary the training prompt within these examples to induce additional robustness. The resulting examples contain a range of capabilities, including error correcting behavior, explicit instruction following, instruction chaining (*i.e.*, following a sequence of instructions), visual question answering, reliance on memory, and long-horizon behavior. We also include *no-ops* (time steps at which no actions are taken) to ensure that the agent remains still after a task has been completed. We refer to the resulting dataset as *"bridge"* data, as these examples bridge the modalities of embodied action and language.

### *3.3.3. Reinforcement Learning*

After the initial supervised learning stages, the agent is further trained using online reinforcement learning from verifiable rewards (*c.f.*, Mankowitz et al. (2023); Wen et al. (2025)). To do this, we curated a set of verifiable tasks, *i.e.*, a tuple of an initial game state, a text instruction, and a verification function. We then generate agent trajectories on these tasks in order to improve the policy. Reward is obtained for either successfully completing the embodied task or giving a correct answer to a question grounded in the environment. Some tasks contain additional shaped rewards to improve the instruction-following capabilities and controllability of the agent.

The main body of tasks were collected from participants contracting with Google. Participants were placed into random game states and asked to explore the nearby environment and suggest

multiple tasks that could be completed from that point. This set of tasks was expanded by applying verifier functions to all human trajectories (see Human Data section) to identify goal completion points and pairing these points with a nearby game state. These tasks were filtered down to those that a human could complete within a specified time limit to remove excessively hard tasks. In addition, we also generated dialogue tasks by selecting random screenshots from our human data and pairing these with human-suggested question-answer pairs.

This phase of RL training is limited to our training environments and excludes our held-out environments, such as ASKA and MineDojo.

### 3.4. Evaluations

Our quantitative analysis focuses on embodied tasks, in which the agent is given a text-based instruction and executes a series of keyboard-and-mouse actions in the environment to achieve a goal. As in SIMA Team et al. (2024), task success is measured using one of three distinct types of evaluation function. We refer to the first two categories collectively as *automatic* evaluations, as they do not require manual assessment:

- **Ground-Truth Evaluation**: These evaluations use ground-truth state information from the environment to assess task success. For instance, success may depend on the absolute or relative positions of objects ("*Lift the cube*"), the acquisition of an object or resource ("*Gather wood*"), or triggering some other game mechanic ("*Water the plant*"). Given that commercial video games do not generally expose this state information in an accessible way, these evaluations are limited to our research environments.
  - *Construction Lab, MineDojo, Playhouse, WorldLab*
- **Programmatic Evaluation**: For commercial video games, which do not generally expose ground-truth state information, we define programmatic evaluations based on the game screen and the agent's keyboard-and-mouse actions. Video games often contain on-screen text in the form of pop-ups and menus, signaling events and state information. As in previous works (OpenAI, 2016; SIMA Team et al., 2024), we use optical character recognition (OCR) to detect this on-screen text to determine task success. We also define functions over pixel colors and action outputs. While creating these task functions is a manual process, once written, they can be deployed easily at-scale to enable automatic evaluation across our commercial video game environments. However, these tasks are restricted to the outcomes that can be detected through heuristics over the visual input or through the agent's actions.
  - *ASKA, Goat Simulator 3, Hydroneer, No Man's Sky, Satisfactory, Space Engineers, Valheim, Wobbly Life*
- **Human Evaluation**: For tasks where no ground-truth or programmatic task function can easily be written, we rely on human raters to assess task completion by observing the video of the agent's trajectory. As raters do not always agree on task success, we obtain five independent ratings per video to improve precision. Although this method is more costly than the other two (automatic) evaluations, it can be applied to a broader variety of tasks.
  - *Goat Simulator 3, Hydroneer, No Man's Sky, Satisfactory, Valheim, Wobbly Life*

**SIMA Evaluation Suite 2.0** Since announcing SIMA 1, we have significantly expanded our evaluations. This includes expanding our programmatic and human evaluations to additional domains, increasing the number of evaluation tasks, often by an order of magnitude or more in the case of programmatic evaluations, and improving our programmatic evaluations to better align them with our expectations of task success. Here, we highlight three improvements.

- Rather than triggering success after the first text detection, for applicable tasks, we ensure that the text is present for several seconds. By requiring a degree of *persistence*, our evaluations select for more intentional behavior, demonstrated by the agent pausing when it considers a task to be completed.
- For a stricter constraint, in a subset of tasks we place a threshold on the number of actions permitted to be performed after task completion. Measuring whether agents remain still allows us to gauge whether agents recognize task completion and whether tasks can be readily chained during deployment.
- We have greatly expanded our set of sequential programmatically-evaluated tasks, in which each instruction is supplied once the previous task has been completed, closer reflecting the behavior expected in an interactive session. Succeeding on these tasks requires successfully completing every sub-task in the sequential chain.

As a result of these improvements, our evaluations are substantially more challenging than those originally reported in SIMA Team et al. (2024). Accordingly, the SIMA 1 agent obtains lower success rates.

**Human Baselines** To contextualize SIMA 2's performance, we established human baselines by collecting gameplay trajectories on our full evaluation suite of tasks. These were designed to closely replicate the agent's testing conditions, including the time limits for each task. For tasks in which the agent receives multiple instructions in a sequence, the players were given all steps to accomplish at once, with the guidance that they were to complete them one at a time in order.

To ensure a representative and reliable human baseline, for our training environments we collected this data from players who had prior experience with the game through their participation in our training data collection. For the held-out environments, ASKA and MineDojo, we recruited new participants with general video game experience but no prior experience playing these specific titles. They were provided with written instructions on core game mechanics and controls but received no task-specific guidance.

## 4. Results

### 4.1. New Capabilities

Despite SIMA 1's ability to perform a broad range of short-horizon embodied tasks, it was also limited in several aspects. While SIMA 1 used pretrained vision encoders (Bica et al., 2024; Villegas et al., 2022), its language encoding was trained from scratch. Thus, SIMA 1's instruction-following capabilities were constrained to the vocabulary of the annotated gameplay on which it was trained. Further, SIMA 1 only mapped text instructions and current images to keyboard and mouse actions; it was incapable of processing any other inputs or outputs. For instance, SIMA 1 was incapable of outputting text (*e.g.,* internal reasoning or dialogue), and it was also incapable of receiving multi-modal instruction prompts (*e.g.,* sketches).

SIMA 2 overcomes these limitations, enabling a new set of capabilities. By powering SIMA 2 with Gemini, we inherit Gemini's broad vision and language understanding capabilities, opening up a far broader and more complex set of tasks. As vision, language, and action now occupy a single stream of tokens, we can lift the input and output constraints on the agent to enable new forms of interaction, such as dialogue and multi-modal prompting. This also enables internal reasoning, allowing the agent to adapt its behavior based on internal inferences. Below, we highlight several of these new capabilities, paired with concrete examples.


Figure 4 | **Embodied Dialogue & Basic Reasoning**. SIMA 2 contains a variety of new capabilities, including embodied dialogue and basic reasoning. Above, SIMA 2 answers a user's question through embodied interaction. Below, the agent correctly reasons that it needs to go to a *red* house based on the user's instruction. These new capabilities are unlocked by using Gemini within SIMA 2.

**Embodied Dialogue** SIMA 2 is, at its core, a Gemini model. Thus, just like Gemini, it can engage in dialogue with a user, making use of Gemini's general world knowledge and visual question-answering capabilities. However, because SIMA 2 is situated in a 3D world, it can also take actions in response to user prompts, enabling a new capability for *embodied dialogue*. This covers a wide variety of interactions, including confirmations of users' requests and proactively responding when tasks have been completed. SIMA 2 can even ask clarifying questions when a user's request is ambiguous. One particularly unique form of interaction is embodied question-answering, in which a user asks or instructs the agent to find some piece of information, to which SIMA 2 must take embodied actions to determine the answer and respond in natural language. For instance, in No Man's Sky, when asked *"Can you go check out those egg-shaped objects and tell me what material they are made of?"*, SIMA 2 confirms the user's instruction, then navigates to one of the objects, using the on-screen text to reply *"I found the objects. They appear to be plants containing Carbon."* This example highlights SIMA 2's ability to engage in embodied information-seeking behavior, going beyond the capabilities of SIMA 1 and the base Gemini model.

**Basic Reasoning** In the same way that SIMA 2 can output text externally to engage in dialogue with a user, it can also output text *internally* to perform reasoning. By generating and conditioning on internal reasoning, SIMA 2 can use internal inferences to modify its own behavior. With Gemini's general reasoning capabilities, SIMA 2 can thus handle more indirect, nuanced, or novel instructions, which are not present in the training data. To provide a simple, illustrative example, a user can instruct the agent to *"Go to the house colored like a ripe tomato."* Internally, the agent then reasons, *Based on the description "ripe tomato", I identify the red house down the street on the right as the target.* The agent then responds and heads to the correct house. This general ability to modify behavior based on internal reasoning affords a broad array of novel behaviors. Indeed, as will be shown in Section 4.2.2, SIMA 2 uses its internal reasoning to correctly identify appropriate actions in entirely novel environments.

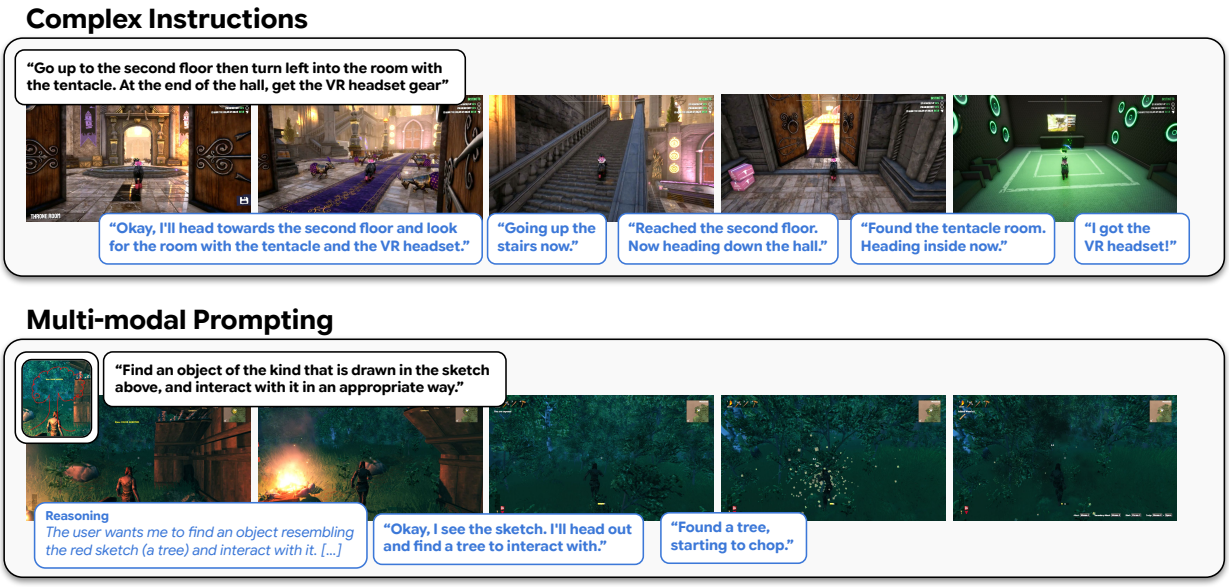



Figure 5 | **Complex Instructions & Multi-modal Prompting**. By inheriting Gemini's language understanding capabilities, SIMA 2 can handle a variety of novel, complex instructions, including breaking down instructions to successfully navigate to a specific room. SIMA 2 can also be prompted with images, including sketches, to specify locations, paths, or objects.

**Complex Instructions** SIMA 2 also benefits from Gemini's general language understanding capabilities, allowing it to generalize to novel, complex instructions. For instance, by leveraging the zero-shot multilingual capabilities inherited from the base model, SIMA 2 can readily perform tasks when instructed in French, German, Mandarin Chinese, *etc*, despite only training on embodied data in English. SIMA 2 can even interpret instructions provided in emojis, *e.g.*, inferring that an ax emoji and a tree emoji implies chopping down a tree. This also extends to more complex, multi-step instructions. For instance, when given the navigation instructions *"Go up to the second floor then turn left into the room with the tentacle. At the end of the hall, get the VR headset gear"*, the agent can successfully go through each step, reporting its progress along the way. Without dedicated training data, a task of this form would be far outside of the scope of SIMA 1. In contrast, this ability to more fully utilize language to perform embodied tasks means that SIMA 2 can better harness the abstract, compositional properties of language.

**Multi-modal Prompting** Gemini is natively multi-modal, processing images, audio, and video in addition to text. SIMA 2 thus inherits multi-modal prompting capabilities, allowing us to instruct the agent in novel ways. In our investigations, we have primarily focused on images, as they offer a simple way to transcend the limitations of language instructions. Such images, for instance, can come from game wikis or even generative image models. One particularly helpful use-case is *sketching*; rather than describing a location, a path, or an object in text, we can simply annotate the current game image to indicate it. We can even draw an object on the game screen and provide this in the instruction. For instance, when given a sketch of a tree and instructed, *"Find an object of the kind that is drawn in the sketch above, and interact with it in an appropriate way."*, the agent correctly identifies it as a tree, then proceeds to chop one down. Thus, these new ways of instructing the agent enable new forms of more intuitive interaction.

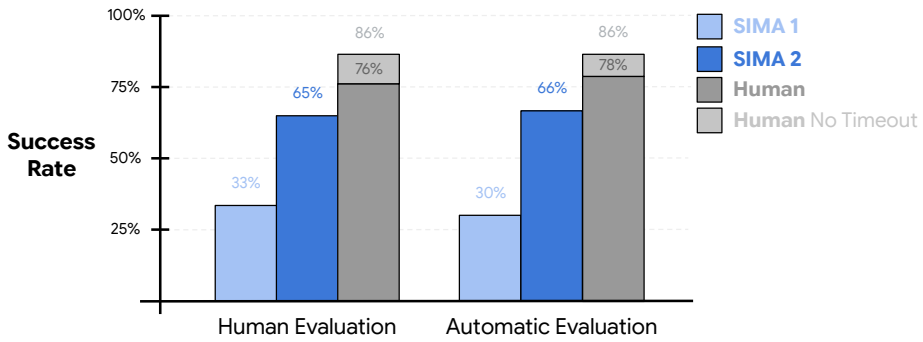



Figure 6 | **Average Performance on Embodied Tasks.** Performance is averaged over training environments for each type of evaluation (human or automatic). We plot human performance both with and without the time restrictions imposed on agents. SIMA 2 effectively doubles the average success rate of SIMA 1, approaching human-level performance in both cases.

## 4.2. Embodied Task Performance

### *4.2.1. Performance in Training Environments*

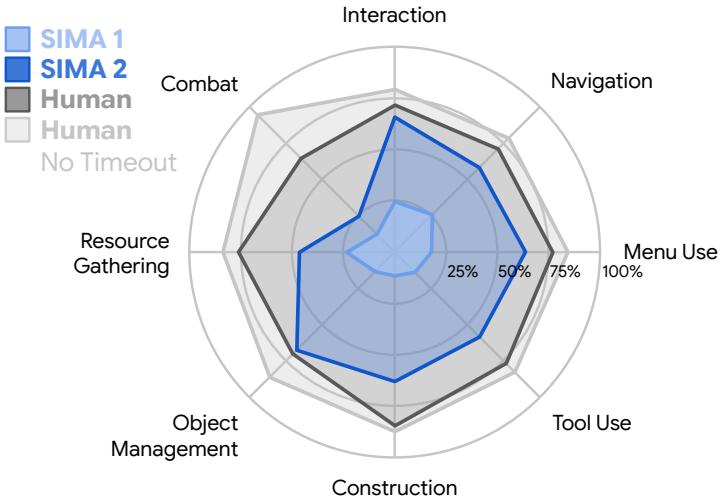



Figure 7 | **Performance Across Skill Categories.** SIMA 2 significantly improves over SIMA 1 across multiple skill categories. In categories like interaction and object management, SIMA 2 nearly closes the gap with human-level performance. However, in other categories, like resource gathering and combat, SIMA 2 still has room for improvement.

In Figure 6, we plot the performance of SIMA 1, SIMA 2, and humans across our human-evaluated and automatically-evaluated embodied task sets, averaged over environments seen during training. Human baseline performance is collected from individuals with significant gameplay experience in each environment. We plot human performance subject to the same task timeouts given to agents (dark gray), as well as performance without this restriction (light gray). The latter value gives an approximate upper bound on performance, as we observed that participants frequently struggled to complete tasks within the allotted time, some of which were as short as three seconds. Primary sources of difficulty for human players included initial inattention, infrastructure latency, and challenges with the fine-motor control required to operate game interfaces as fluidly as the agent.

Overall, we find that SIMA 2 substantially outperforms SIMA 1, effectively doubling the success rate and nearly closing the gap with human performance. This is remarkably consistent across both human-evaluated and automatically-evaluated tasks.

In Figures 8 & 9, we break down agent and human performance across each environment. Where

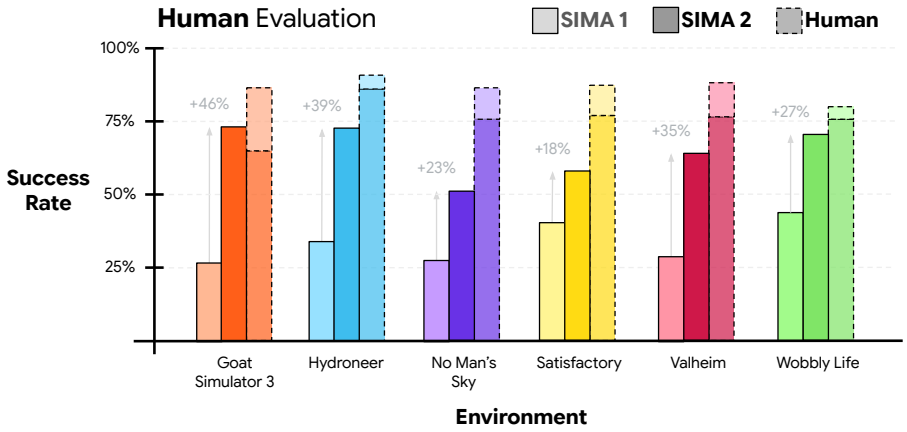



Figure 8 | **Performance on *Human* Evaluations By Environment**. Bars show the task success rate of SIMA 1, SIMA 2, and humans on a suite of tasks from training environments, with success measured by 5 independent human ratings per trial. Human performance is plotted subject to the same timeout restrictions as imposed on agents (darker) and with this restriction removed (lighter). SIMA 2 improves significantly over SIMA 1 across all environments, nearly closing the gap with human performance in many cases.

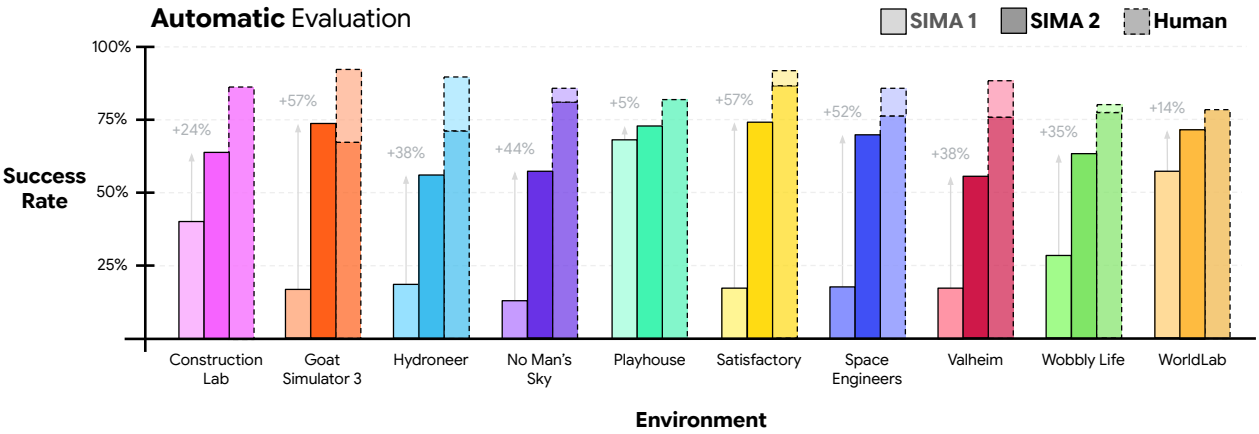



Figure 9 | **Performance on *Automatic* Evaluations By Environment**. Bars show the task success rate of SIMA 1, SIMA 2, and humans on a suite of tasks from training environments, with success measured by ground truth rewards (Construction Lab, Playhouse, and WorldLab) or programmatic evaluation (all other environments). Where applicable, human performance is plotted with and without the same timeout restrictions as imposed on agents. SIMA 2 improves significantly over SIMA 1 across nearly all environments, almost closing the gap with human performance in many cases.

applicable, we plot human performance both with and without timeout restrictions. SIMA 2 improves over SIMA 1 across all environments, nearly closing the gap with human performance in many cases. In particular, there are substantial performance improvements in video game environments. This highlights how SIMA 2 is better able to handle more complex settings, where agents are required to deal with visual diversity, interact with menus, and navigate more challenging game dynamics.

To better understand the performance improvements of SIMA 2, we decompose our evaluation tasks into skill categories, as previously performed in SIMA Team et al. (2024). In Figure 7, we plot the average performance across a subset of eight common skill categories: interaction, navigation, menu use, tool use, construction, object management, resource gathering, and combat. In Table 2 in Appendix A, we briefly describe each category, along with several representative examples.

From Figure 7, we see that SIMA 2 substantially improves over SIMA 1 across skill categories, approaching human-level performance in several instances. Notably, SIMA 2 still struggles with *Combat*, in part, due to the motor difficulty of these tasks. For instance, hunting a deer in Valheim typically requires approaching from downwind while crouching, then quickly attacking. If the deer escapes, a challenging chase then ensues, requiring split-second decision making and a degree of luck. Similarly, when removing the timeout restriction (imposed on the agents), human performance in this skill category improves substantially.

### *4.2.2. Performance in Held-Out Environments*

The previous section discussed how SIMA 2 outperforms SIMA 1 on held-out *tasks* within environments seen during training. This provides compelling evidence that SIMA 2 is a more *performant* agent. We now address whether SIMA 2 is also a more *general* agent. That is, can it generalize to new visual settings, menus, and game dynamics? To assess a more extreme form of generalization, we evaluate SIMA 2 on entirely held-out environments, previously unseen during training. We first present a quantitative evaluation comparing SIMA 1 and SIMA 2, then provide qualitative examples of SIMA 2's behavior in several wildly different environments.

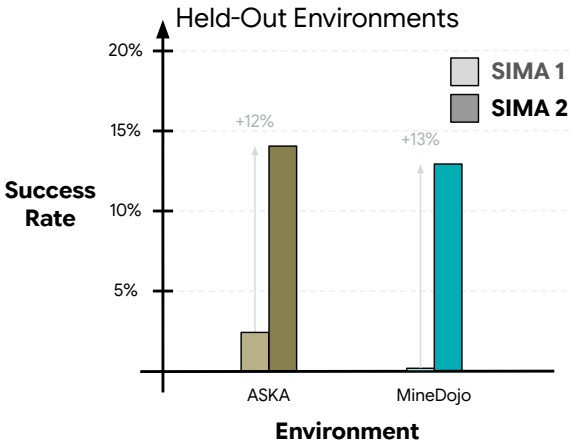



Figure 10 | **Held-Out Environment Performance**. SIMA 2 outperforms SIMA 1 on two held-out environments (*i.e.*, unseen during training): ASKA and MineDojo. This demonstrates that SIMA 2 is a more general agent, capable of performing non-trivial tasks in new settings.

**Quantitative Evaluation** We evaluate SIMA 1 and SIMA 2 on ASKA and a subset of the MineDojo benchmark suite in Minecraft (described in Section 3.1.1). Both evaluations use automatically-evaluated tasks, based on programmatic evaluations and ground-truth state information respectively. Results are shown in Figure 10, where we see that SIMA 2 significantly outperforms SIMA 1 by over 10% in each environment. In ASKA, SIMA 1 is generally only capable of performing the most basic tasks, such as opening the map or picking up an object directly beside the agent. As we illustrate below, SIMA 2 is capable of performing a variety of non-trivial tasks. In MineDojo, SIMA 1 only completes *two* types of tasks (`harvest dirt` and `combat spider`). We suspect that SIMA 1's low performance is due to a combination of the comparatively abstract visual appearance of Minecraft

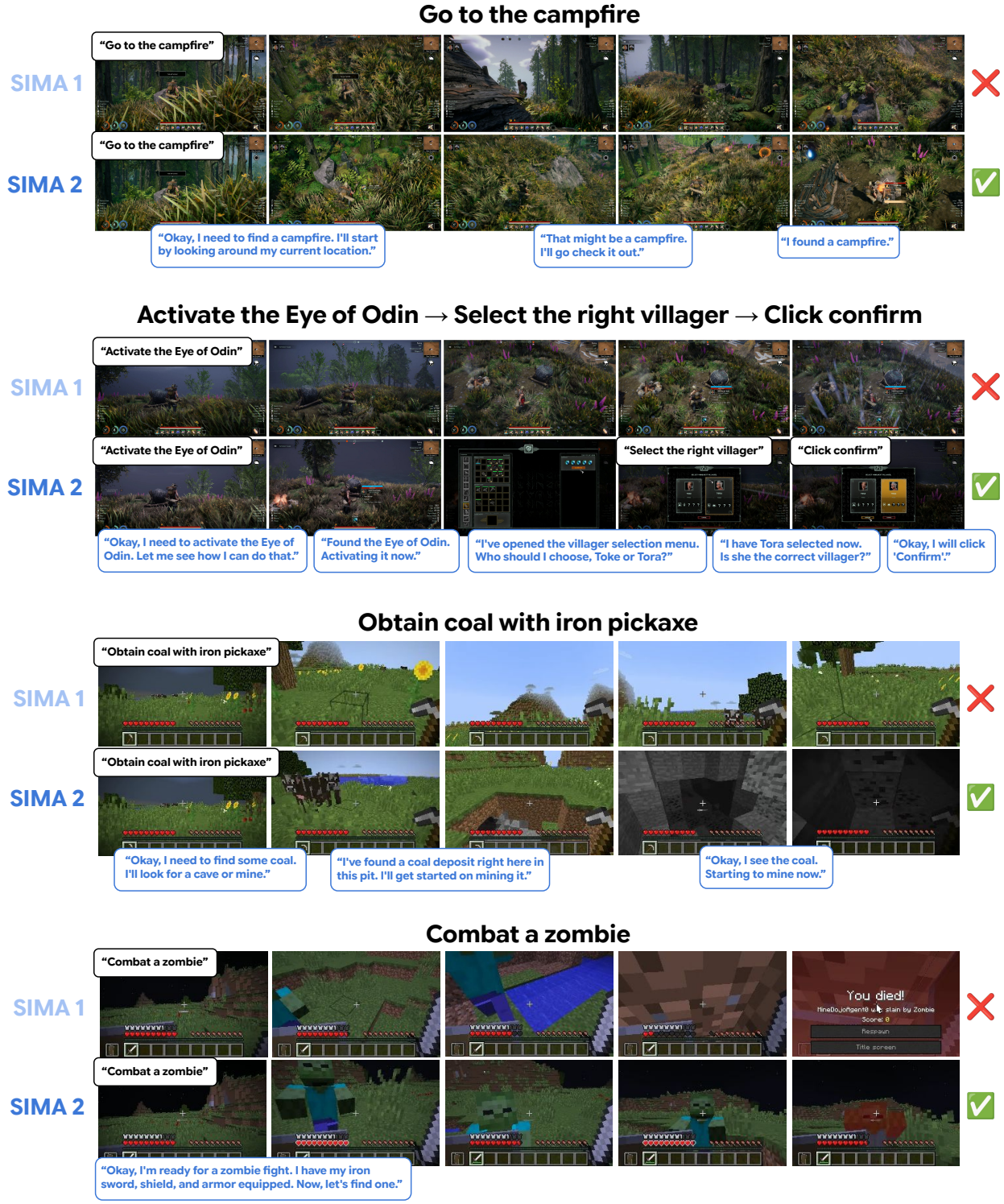



Figure 11 | **SIMA 1 vs. SIMA 2 in *ASKA* and *Minecraft* (Held-Out Environments).** SIMA 2 generalizes to non-trivial tasks in environments entirely held out from training, whereas SIMA 1 struggles in these settings. In addition to completing these tasks, SIMA 2's dialogue output indicates that it correctly identifies key on-screen events to help drive behavior, such as recognizing a campfire or a zombie. SIMA 2 can even generalize to entirely new menus, identifying on-screen text to select the correct buttons.

and the domain-specific knowledge required to complete these tasks. SIMA 2, which inherits Gemini's general world knowledge, is capable of performing a substantial portion of the tasks, completing tasks in 26 out of the 50 task categories.

To better understand these results, Figure 11 provides qualitative examples comparing SIMA 1 and SIMA 2 in these held-out environments. We see that SIMA 1 struggles to apply a previously-encountered concept, *campfire*, to a new visual setting. In contrast, SIMA 2 demonstrates an ability to generalize. We observe it first describing its strategy: "*I'll start by looking around my current location.*" When a campfire appears on-screen in the distance, it identifies: "*That might be a campfire. I'll go check it out.*" Finally, when SIMA 2 arrives at the campfire, it recognizes that the task has been completed, stating, "*I found a campfire.*" This generalization capability extends beyond basic navigation to more challenging interaction and menu use tasks. Similar findings extend to MineDojo, where the agent is capable of harvesting basic resources (*e.g.*, coal, cobblestone, logs) and combating enemies (*e.g.*, spiders, zombies, skeletons), all while narrating its observations and actions. In these examples, we see that not only does SIMA 2 generalize embodied actions to accomplish tasks in never-before-seen environments, it correctly identifies key on-screen events to help drive its behavior.

To contextualize the previous results, we established human baselines with participants who also had no prior experience with ASKA or MineDojo. A key challenge for this comparison is accounting for rapid human learning, as within 2 or 3 tasks, a human player can quickly learn the game's mechanics and world layout and ceases to be truly "naive." Therefore, to estimate the most direct, comparable baseline, we recruited and measured performance of naive human players on their first attempts in these games. These individuals had general video game experience but no prior exposure to these specific games, and they were provided with only written instructions on core game mechanics and controls. On a representative subset of tasks, we found human performance to be roughly 19% for MineDojo (16 tasks) and 32% for ASKA (25 tasks). These results illustrate the difficulty of our held-out tasks for naive players, and suggest that the agent's initial generalization capabilities are approaching that of a human encountering these complex environments for the first time.

We caution against over-interpreting any direct comparison in performance between SIMA 2 and humans, as the *nature* of failures and successes often differed significantly. For instance, we found that humans were more likely to fail a task due to time constraints, while agents would fail due to suboptimal exploration. A detailed and more quantitative characterization of these distinct behavioral patterns presents an interesting avenue for future research.

**Qualitative Evaluation** Despite differing visuals and game mechanics, ASKA and Minecraft are, in some ways, similar to the environments encountered during training: they are video game environments that require exploring open-world terrains to gather resources, combat enemies, build structures, and craft items. Thus, to probe the generalization capabilities of SIMA 2 even further, we perform qualitative evaluations in two distinct settings: The Gunk and Genie 3.

The Gunk (see Section 3.1.1) requires the agent to navigate specific terrain challenges and utilize a new handheld suction device (both unique mechanics compared to training environments) to make progress, in addition to the substantially different visuals. Through manually instructing SIMA 2, we progressed through the first 15-20 minutes of the game, up to the "*Campsite*" checkpoint (Figure 12). The agent is likely capable of progressing even further, but we did not attempt to go beyond this point. Along the way, SIMA 2 performed various novel embodied skills, including scanning objects to analyze them, climbing up ledges, jumping over gaps, and clearing two separate areas containing gunk. Throughout, SIMA 2 responded to instructions, accurately reasoning through the actions required to complete each task. For instance, SIMA 2 used the on-screen cues (`ABSORB` and `HOLD`) to identify that it needed to hold the left mouse button to absorb the gunk using the handheld device (Figure

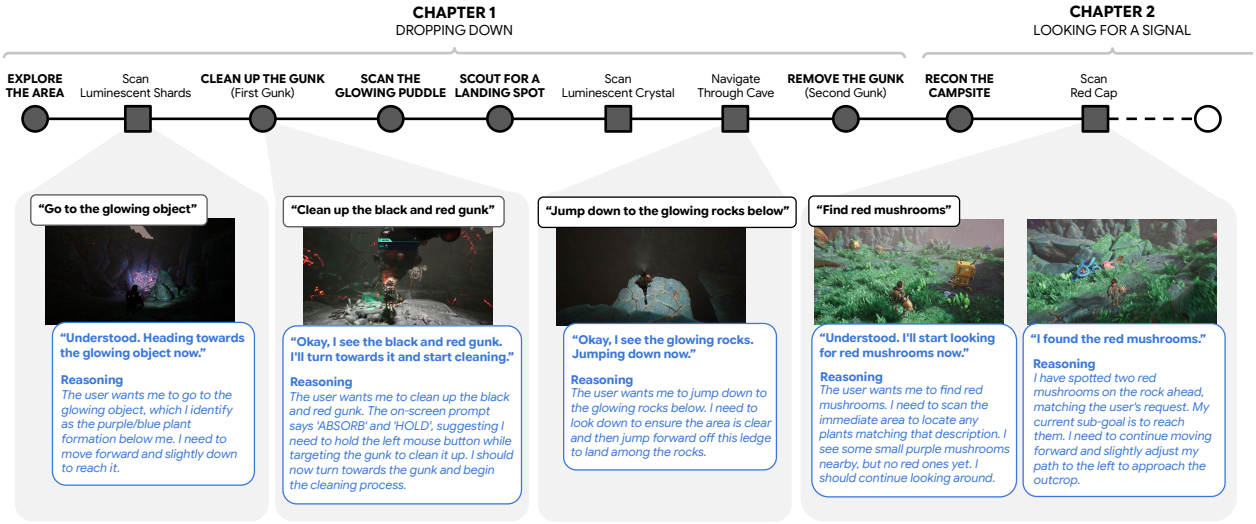



Figure 12 | **SIMA 2 in *The Gunk* (Held-Out Environment)**. Through manual instruction, SIMA 2 progressed through the first 15-20 minutes of The Gunk, a story-driven action-adventure game previously unseen during training. Along the event progression (top), circles denote game-defined milestones, and squares denote other notable events. SIMA 2 identified on-screen targets and reasoned through the appropriate actions using on-screen cues, enabling the completion of highly novel tasks.

12). These examples demonstrate how SIMA 2 is capable of generalizing to highly novel tasks.

Next, we evaluate SIMA 2 in Genie 3 environments. It is important to emphasize that SIMA 2 was trained purely in research and video game environments. However, given its Gemini core and general interface (visual input and generic keyboard-and-mouse actions), it is reasonable to wonder whether SIMA 2 is capable of generalizing to more realistic environments. To qualitatively evaluate this, we instantiate a variety of photorealistic environments in naturalistic and urban settings using Genie 3. As shown in Figure 13, SIMA 2 is capable of navigating to particular points of interest across a wide range of photorealistic environments. This provides a proof-of-concept that training embodied agents in simulated 3D environments enables generalization to more realistic environments—eventually possibly even the physical world (*e.g.*, controlling a real-world robot via a keyboard and mouse).

### 4.3. Comparison to Baseline Gemini Models

The preceding sections demonstrated that SIMA 2 significantly outperforms SIMA 1, establishing it as a more capable and general embodied agent. However, this also raises a question about an inherent tension in adapting large foundation models for acting in embodied environments. By finetuning a powerful, generalist model like Gemini on specialized gameplay data, are we potentially trading off one objective (**embodied competence**: expert-level gameplay and task performance) versus another (**general reasoning**: the model's world knowledge and pretrained language capabilities)? We can conceptualize this as a Pareto frontier defined according to these two competing objectives. In this section, we aim to characterize this frontier, and situate both SIMA 2 and baseline Gemini models within this context. Our analysis addresses two key questions:

1. How well can a baseline Gemini model perform on our suite of complex, interactive tasks without any specialized embodied training?
2. To what extent does fine-tuning using embodied SIMA data preserve or reduce Gemini's pre-trained capabilities in reasoning, math, and general language understanding?

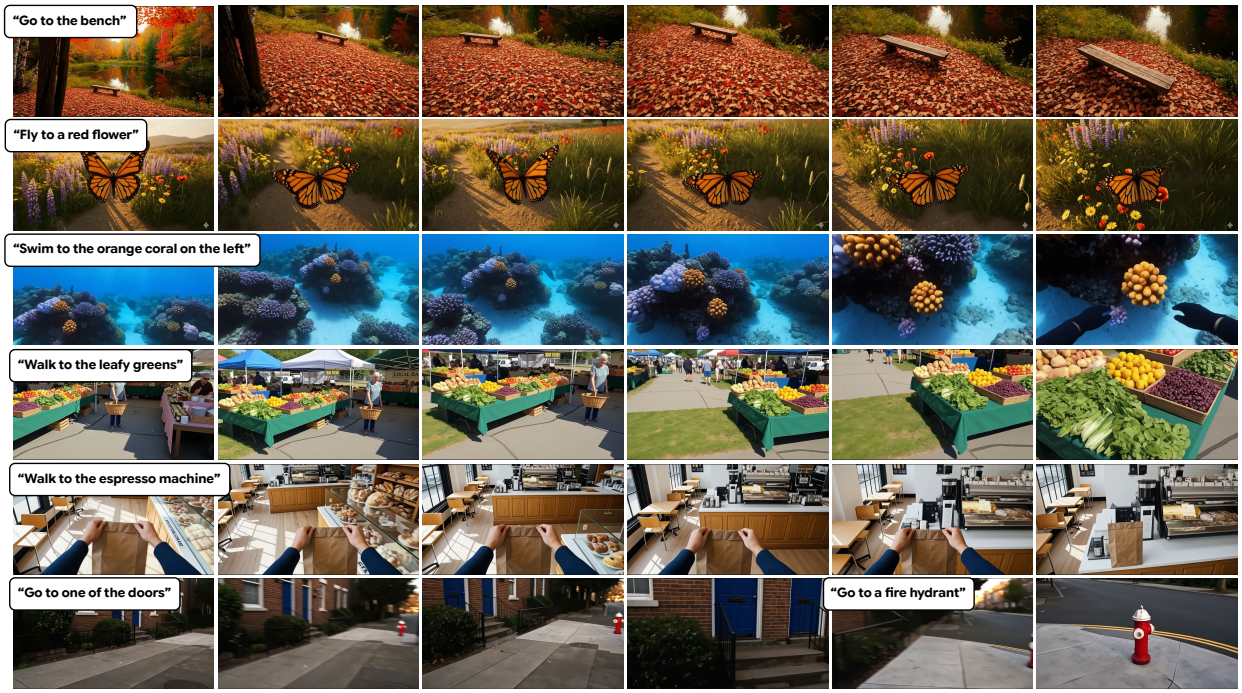



Figure 13 | **SIMA 2 in *Genie 3* (Held-Out Environments)**. We deployed SIMA 2 across a range of naturalistic and urban photorealistic environments generated by Genie 3. Despite learning embodiment skills purely in research and video game environments, we find that SIMA 2 performs well, particularly at navigation-based tasks, even in these novel photorealistic settings.

**Gemini Acting in Virtual Embodied Environments** We first evaluate baseline (non-finetuned) Gemini models, both Flash-Lite and Pro, on our suite of programmatic evaluations. This establishes a critical anchor point along this Pareto frontier. We found that, across our 10 training domains, baseline Gemini models that are not finetuned with SIMA action data have difficulty acting in embodied environments, with the Gemini Flash-Lite model achieving only 3.2% success, and the Pro model 7.0%. These low levels of task performance were despite some considerable efforts at prompt engineering to enable the model to be able to output proper action and text formatting. This demonstrates that competent embodied interaction is not an emergent property of current large-scale pretraining on language and vision data; it is a distinct capability that must be explicitly enabled through training. The difficulty of these embodiment tasks even for powerful frontier models underscores the significance of SIMA 2's near-human-level performance while still retaining language capabilities.

**Retaining Language and Reasoning Capabilities** Having established that specialized finetuning is essential for embodied competence, we now evaluate its impact on Gemini's core reasoning abilities. One of the central risks in finetuning foundation models for specialized downstream tasks is catastrophic forgetting, where the model's performance on previously learned tasks degrades significantly as it adapts to new data distributions (French, 1999; Kirkpatrick et al., 2017). This phenomenon is particularly acute for LLMs, where extensive finetuning on domain-specific datasets often erodes the general world knowledge and reasoning abilities acquired during pretraining (Luo et al., 2025).

This risk is magnified in the context of embodied agents, where the finetuning data—low-level

| | SFT | SFT + RL |
|---|---|---|
| **LCB (Code)** | -4.0% | -8.4% |
| **AIME (Math)** | -25.5% | -15.4% |
| **GPQA Diamond (STEM)** | -16.3% | -19.5% |

Table 1 | **Retaining Gemini's Capabilities**. The table shows the relative difference in score (as a percentage of the baseline Gemini model's performance) on language and reasoning benchmarks for SFT and RL training stages of SIMA 2 compared to the baseline Gemini model without training on SIMA data. The agent retains strong reasoning capabilities with only modest reductions in math and STEM reasoning following finetuning on action data.

keyboard-and-mouse actions in the case of SIMA—is radically out-of-distribution compared to the internet-scale text and image data used for pretraining. Recent works in Vision-Language-Action (VLA) modeling have observed that training solely on action data can "erode conversational ability entirely," effectively destroying the very reasoning capabilities that make foundation models attractive for control in the first place (Hancock et al., 2025; Zhou et al., 2025).

To quantitatively assess whether this is the case for SIMA 2, we evaluate the agent's general capabilities on three diverse benchmarks. For coding, we use LiveCodeBench (LCB) (Jain et al., 2024), specifically the code generation subset, to assess the model's ability to synthesize functional programs from natural language. For advanced mathematical reasoning, we employ the American Invitational Mathematics Examination (AIME) dataset (Hendrycks et al., 2021), representing a high bar for multi-step problem solving. Finally, we evaluate scientific reasoning using the Diamond subset of GPQA (Rein et al., 2023), a difficult question-answering benchmark designed to be robust against search-engine retrieval.

As shown in Table 1, despite being finetuned to output precise embodied actions, SIMA 2 exhibits only a minor regression on these benchmarks compared to the baseline Gemini model without post-training on SIMA data. Moreover, the additional RL training caused no significant additional regression compared with SFT alone.

Taken as a whole, these results demonstrate that high embodied competence need not come at the expense of general intelligence. By successfully bridging the gap between high-level reasoning and low-level control, SIMA 2 proves it is possible to create an agent that acts fluently in 3D worlds without sacrificing the reasoning capabilities of its foundation.

### 4.4. Gemini Instructing SIMA 2

SIMA 2 retains much of Gemini's language and reasoning capabilities while also acting in embodied environments. However, due to the latency constraints that come with embodied action, we chose to finetune SIMA 2 from a Gemini Flash-Lite model, which is generally less capable than Gemini Pro. In this section, we explore composing SIMA 2 with a separate Gemini Pro model, enabling even more advanced reasoning capabilities. In this hierarchical setup, Gemini Pro operates at a slower cadence, reasoning over the recent video history every $k$ steps to issue natural language instructions to the SIMA 2 agent. Gemini Pro also produces a text-based summary that it receives on the next call, effectively serving as a form of recurrent memory and allowing the system to maintain a long-horizon context that persists beyond the immediate context window. This architecture enables more advanced behaviors. We discuss a primary example below, with further demonstrations provided in Appendix B.


Figure 14 | **Complex Multi-modal Instruction Following**. By combining Gemini Pro with SIMA 2, we can enable even more advanced reasoning capabilities. In this case, the combined agent successfully uses a complex diagram to complete the multi-step task of building a campfire. Throughout, the agent communicates its current actions and intended next steps.

**Complex Multi-modal Instruction Following** We previously saw that SIMA 2 is capable of multi-modal prompting, using a sketch to help communicate the task to the agent. Here, we take things a step further, using a complex diagram (Figure 14) to convey the multi-step task of building a campfire to the combined Gemini Pro + SIMA 2 agent. To accomplish the task, the agent needs to parse the visual diagram, decompose it to a series of steps, and track its progression toward each of the steps and the overall objective. This is an integrative task that involves reasoning, memory, and visual understanding capabilities. As shown in Figure 14, the agent successfully breaks down the steps to achieve the task, communicating both its current actions and intended next steps throughout.

As Gemini continues to improve with newer versions, this compositional approach allows us to immediately take advantage of the latest reasoning capabilities. Thus, with SIMA 2 serving as a base for general embodied interaction in 3D worlds, more advanced versions of Gemini can yield more advanced forms of embodied behavior.

### 4.5. Self-Improvement

In Section 4.2.2, we saw that SIMA 2 is a more general agent than SIMA 1, capable of performing complex tasks in entirely held-out environments previously unseen during training. However, zero-shot generalization alone only goes so far, and SIMA 2 is still far from perfect in these held-out environments (Figure 10). Indeed, particularly for new objects or game mechanics, it may be unreasonable to expect agents to perform these tasks out-of-the-box. Rather than relying on additional human demonstrations to improve performance, we ultimately want agents that can learn from *self*-generated experience, allowing them to autonomously adapt and improve. In this section, we showcase initial steps toward this capability with the SIMA 2 agent, enabled by using Gemini both as a *task creator* and as a form of *universal reward function* (a function that provides a reward for any possible task). The full setup is shown in Figure 16a.

**Gemini-Based Task Setter** To generate experience, SIMA 2 requires a source of tasks, *i.e.*, language instructions tied to the current state of the environment. These tasks can come from humans, as we use for the "fixed set" of tasks described below. However, for a more general, open-ended self-improvement process, we need some way of automating task creation (Clune, 2019; Colas et al., 2022; Zhang et al., 2023). We turn to Gemini to play the role of this "Task Setter," prompting Gemini to provide instructions to the agent that are likely to be achievable from the current state. We couple the Gemini-based task setter and the agent within a running instance of the environment, allowing the task setter to dynamically adjust the current task as the agent interacts with the environment. Modifying the prompt given to the task setter also allows us to readily adjust the task distribution and, as a result, the data distribution. For instance, by feeding the downstream evaluation results back to the task setter, it can steer the agent toward skills that need to be improved. Indeed, the task setter can even track the agent's within-episode performance to focus on tasks that are interesting to learn and likely to provide learning progress for that agent (Clune, 2019; Zhang et al., 2023).

**Gemini-Based Reward Model** Tackling self-improvement in open-world 3D environments immediately exposes the challenge of defining "success" for any given task, *i.e.*, having a *universal reward function* (Faldor et al., 2025). As in embodied settings in the physical world, we cannot rely on ground-truth state information and manually-designed reward functions. Instead, we must take 1) a stream of high-dimensional sensory input, *i.e.*, pixels, and 2) a goal (encoded in natural language) and convert this into some form of feedback to drive agent improvement, entirely in the absence of any internal game state variables. Defining such functional mappings is a long-standing challenge in the field. However, recent improvements in the video and language understanding capabilities of foundation models, like Gemini, provide a possible route toward such general-purpose reward models (Baumli et al., 2023).

In our setup, Gemini provides a rating for each trajectory (video), using a rubric to assign a score between 0 and 100. This rubric captures multiple aspects of performance, including task completion and directedness, *i.e.*, not performing unnecessary actions. We arrived at the prompt that defines the rubric by calibrating the resulting scores to align with human preference pairs over a small dataset of trajectories. Under the rubric, a score of 50 or greater is considered a "success". We can then deploy the agent on a given task and score the resulting trajectory to build a dataset of self-generated experience. By training the agent on this scored self-generated experience, we can drive policy improvement, resulting in higher scores and improved capabilities. This work is thus an important step toward the long-standing grand challenge in the field of AI of creating open-ended algorithms that can learn forever (Clune, 2019; Stanley and Lehman, 2015; Stanley et al., 2017), as this agent could continue to invent and learn new tasks endlessly.

### *4.5.1. ASKA*

First, we investigate self-improvement with the SIMA 2 agent in ASKA. This environment is entirely held out from SIMA 2's training, allowing us to assess whether our self-improvement process can take an initial agent that generalizes somewhat and drive it toward acquiring new skills in new settings. We split this investigation into two parts to help demonstrate the separate components of our setup.

**Self-Improvement on a Fixed Set of Tasks** Our full setup consists of both automatic task generation (task setter) and scoring (reward model), as shown in Figure 16a. Thus, the agent can both expand its capabilities (more tasks) and improve on its existing capabilities (higher reward). To isolate the improvement aspect of this process, we use a fixed set of tasks. This allows us to see the agent's improvement over successive iterations of training. These tasks include a variety of skills, including

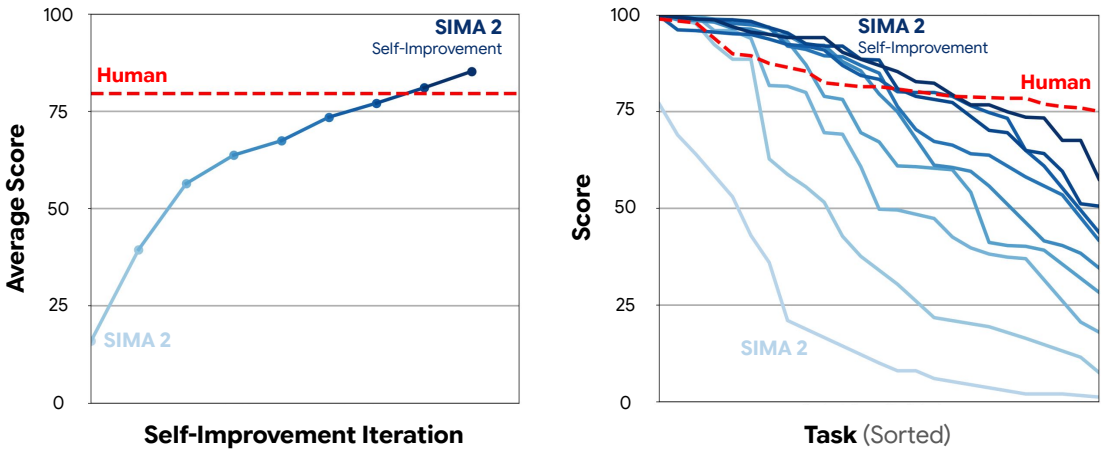
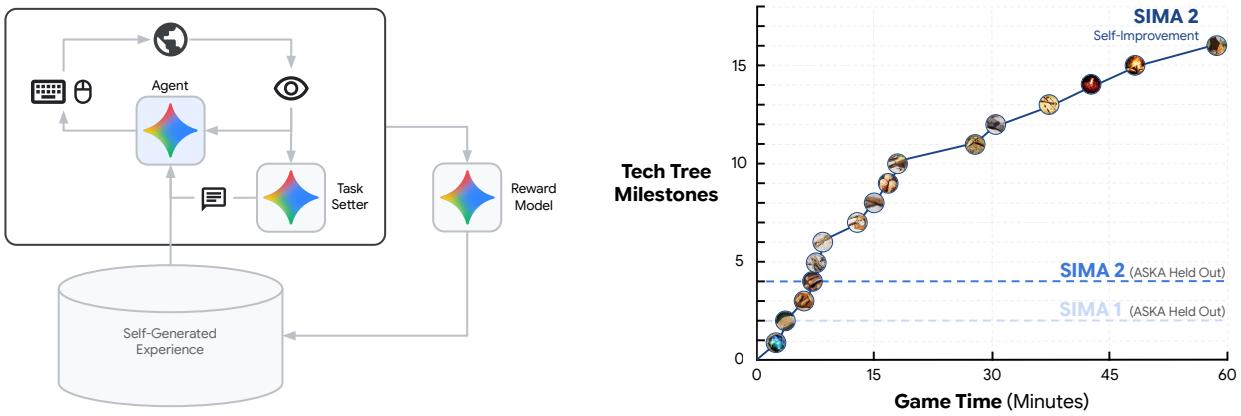



Figure 15 | **Self-Improvement on a Fixed Set of Tasks in *ASKA***. To isolate the improvement aspect of our learning process, independent of the task setter, we first use a fixed set of tasks in ASKA. Over successive iterations (darker blue points and curves), we see that performance steadily improves, eventually exceeding a score of 50 across the full task set (the threshold for "success"). Performance approaches, and in some cases even exceeds, the scores of reference trajectories from humans.

- Resource Gathering: *"Gather the berries", "Pick up the sticks", etc.*,
- Environment Interaction: *"Go to sleep in the shelter", "Extinguish the campfire", etc.*,
- Navigation: *"Go closer to the rain collector", "Go near the raw food silo", etc.*,
- Menu Use: *"Open the workshop hut menu", "View the tasks of the farm crop", etc.*.

To benchmark agent performance, we also collect a set of reference trajectories on these tasks from humans with significant experience (multiple hours or more) playing ASKA. In Figure 15, we plot both the average score and the score per task, as assessed by the Gemini-based reward model, over successive iterations of self-improvement. These are plotted as progressively darker points and curves. Through continuing to run the self-improvement process, average performance eventually exceeds that of the human reference score. Likewise, though the initial SIMA 2 agent was successful (*i.e.*, score above 50) on less than a quarter of the tasks, the self-improved agents eventually exceed the success threshold across all tasks. In terms of behavior, through training on self-generated experience, the agent learns how to navigate to new objects, including a rain collector and workshop hut, and acquires new skills, such as extinguishing a campfire (see Figure 20 in the Appendix). This demonstrates that by combining SIMA 2's generalist embodiment capabilities with Gemini's video and language understanding capabilities, we can enable a general form of embodied self-improvement.

**Self-Improvement Toward Game Progression** We now move to our full self-improvement setup. That is, we deploy a Gemini-based task setter to instruct the agent, allowing the agent to practice existing skills and acquire new ones. We prompt the task setter to focus on skills relevant for game progression, including resource gathering, crafting, menu use, and building. By monitoring downstream evaluations of the agent from the reward model, the task setter can focus on improving weaker skills. For instance, ASKA's crafting menu is quite distinct from those of our training environments, and SIMA 2 struggled with this game mechanic initially. Through focused effort by the task setter, the agent was eventually able to acquire this skill. We assess the capabilities of the resulting self-improved agent by manually instructing it to progress through the ASKA technology tree (see Figure 21 in the Appendix). The results are shown in Figure 16b. Despite purely training on self-generated experience, the resulting agent is capable of progressing much further than SIMA 2, ultimately building a shelter




(a) (b)

Figure 16 | **Self-Improvement Setup & Game Progression. (a)** In our full self-improvement setup, we deploy the SIMA 2 agent alongside a Gemini-based task setter that instructs the agent to perform tasks in the environment. A separate Gemini-based reward model then scores these attempts, building a dataset of experience. By training on this self-generated experience, we improve the agent. **(b)** We deploy our self-improvement process in ASKA, enabling the agent to autonomously acquire and improve upon new skills. We assess the capabilities of the self-improved agent by manually instructing it to progress through the ASKA tech tree. The agent is capable of progressing significantly further than the SIMA 1 and SIMA 2 agents, despite only ever training on self-generated experience in ASKA.

within a one hour time window. This highlights the power of our general self-improvement process, enabled by using Gemini within each component.

### *4.5.2. Genie 3*

We have shown that SIMA 2 can improve based on self-generated experience within a single environment, ASKA. However, the benefit of having a general self-improvement algorithm for embodied behavior is that we can deploy the agent in *any* environment to collect diverse experience data and subsequently improve. A grand challenge of AI research is to create open-ended algorithms, which afford endless learning and innovation (Clune, 2019; Stanley and Lehman, 2015; Stanley et al., 2017). Clune (2019) suggested that one path to open-ended learning would be having an agent learn in a *Darwin-complete* environment search space, meaning a search space that includes any type of environment, and that this goal could be accomplished by a neural network serving as a universal world model, producing the next state when given an action (*i.e.*, the transition function). Genie (Ball et al., 2025; Bruce et al., 2024; Parker-Holder et al., 2024) realized the goal of producing such a universal world model, and here we demonstrate the first preliminary working example of an agent learning within that universal world model, specifically Genie 3 (Ball et al., 2025).

As an initial step in this ambitious direction, we split our set of Genie 3 environments into urban (train) and natural (test) environments and tasks, primarily centered on navigation. We then use our self-improvement algorithm (as shown in the previous section), to improve on the train tasks: generating trajectories, scoring them with our Gemini-based reward model, and training on the self-generated experience. In Figure 17, we see that SIMA 2 improves across nearly all train tasks, often by 25 points or more. However, more importantly, these improvements also extend to the test tasks in entirely different environments. We see that, in the majority of the tasks in natural environments (held out), the self-improved SIMA 2 outperforms the initial agent. Thus, by self-improving on a broad set of photorealistic environments, SIMA 2 generalizes even better to entirely different types of


Figure 17 | **Self-Improvement in *Genie 3*.** We deploy SIMA 2 on a set of train tasks in urban environments from Genie 3, *e.g.*, finding a lollipop in a candy store. Using our self-improvement process, we see broad improvement in scores across nearly all train tasks. More importantly, these improvements also extend to a set of held-out test tasks in natural environments, *e.g.*, enabling the agent to navigate to a red mushroom. This suggests a route toward open-ended self-improvement in increasingly diverse environments to obtain more general and capable agents.

photorealistic environments. This provides initial evidence that we may be able to use these types of techniques to produce an open-ended process of autonomously acquiring diverse skills, yielding an increasingly general and capable agent.

## 5. Discussion

In this work, we introduced SIMA 2, a generalist embodied agent that can reason, converse in dialogue, and perform goal-directed actions across a diverse range of 3D virtual worlds. SIMA 2 represents a significant step beyond simple instruction following, enabling a more capable and collaborative embodied agent. SIMA 2 is also more than just a foundation model that can output embodied actions. By more tightly integrating reasoning and action, SIMA 2 can successfully reason through and complete complex tasks in previously unseen environments. Critically, this generalization extends beyond game worlds to novel photorealistic environments generated by Genie 3. We have also shown that SIMA 2 can further improve in these new environments based entirely on self-generated experience. Taken together, these results suggest a promising path toward using self-improvement to eventually bridge the virtual and physical worlds, enabling more capable physically-embodied agents in applications like robotics.

While SIMA 2 is a significant step toward generalist, interactive, embodied intelligence, it is fundamentally a research endeavor, and its current limitations highlight critical areas for future work. SIMA 2 still faces challenges with very long-horizon, complex tasks that require extensive, multi-step reasoning and goal verification. The agent also has a relatively short memory of its interactions—it must use a limited context window to achieve low-latency interaction. Finally, executing precise, low-level actions via the keyboard-and-mouse interface and achieving robust visual understanding of

complex 3D scenes remain open challenges that the entire field continues to work to address.

As with all our advanced and foundational technologies, we remain deeply committed to developing SIMA 2 responsibly, from the outset. This is particularly true with regard to its technical innovations, particularly the ability to self-improve. As we have built SIMA 2, we have worked with our Responsible Development and Innovation Team. As we continue to explore the potential applications, we announced SIMA 2 as a limited research preview and provided early access to a small cohort of academics and game developers. This approach allows us to gather crucial feedback and interdisciplinary perspectives as we explore this new field and continue to build our understanding of risks and their appropriate mitigations. We look forward to working further with the community to continue to develop this technology in a responsible way.

## Acknowledgments

Special thanks to all of the game developers who partnered with us: Coffee Stain (Valheim, Satisfactory, Goat Simulator 3), DigixArt (Road 96), Foulball Hangover (Hydroneer), Hello Games (No Man's Sky), Keen Software House (Space Engineers), RubberbandGames (Wobbly Life), Strange Loop Games (Eco), Thunderful Games (ASKA, The Gunk, Steamworld Build), and Tuxedo Labs and Saber Interactive (Teardown). We also thank Jack Parker-Holder, Shlomi Fruchter, and the rest of the Genie team for access to the Genie 3 model. We would like to recognize the many teams across Google and Google DeepMind that have contributed to this effort including Legal, Marketing, Communications, Responsibility and Safety Council, Responsible Development and Innovation, Policy, Strategy and Operations, and our Business and Corporate Development teams. In particular, we thank Andeep Toor, Duncan Noble Smith, Leen Verburgh, Matt Miller, Nilesh Ray, Phil Esposito, Piers Wingfield, Signe Nørly, Vika Koriakin, and others on the Marketing and Communications team for their help with communications. We would also like to thank all GDM teams that are not explicitly mentioned here for their continued support. We thank our team of participants who generated gameplay and language annotation data, as well as performed human evaluations of our agents, without whom this work would not have been possible.

Finally, we dedicate this work to the memory of our colleagues Felix Hill and Fabio Pardo, whose contributions to our field continue to inspire us.

## SIMA 2 Team

*Alphabetical by first name.*

Adrian Bolton, Alexander Lerchner, Alexandra Cordell, Alexandre Moufarek, Andrew Bolt, Andrew Lampinen, Anna Mitenkova, Arne Olav Hallingstad, Bojan Vujatovic, Bonnie Li, Cong Lu, Daan Wierstra, Daniel P. Sawyer, Daniel Slater, David Reichert, Davide Vercelli, Demis Hassabis, Drew A. Hudson, Duncan Williams, Ed Hirst, Fabio Pardo, Felix Hill, Frederic Besse, Hannah Openshaw, Harris Chan, Hubert Soyer, Jane X. Wang, Jeff Clune, John Agapiou, John Reid, Joseph Marino, Junkyung Kim, Karol Gregor, Kaustubh Sridhar, Kay McKinney, Laura Kampis, Lei M. Zhang, Loic Matthey, Luyu Wang, Maria Abi Raad, Maria Loks-Thompson, Martin Engelcke, Matija Kecman, Matthew Jackson, Maxime Gazeau, Ollie Purkiss, Oscar Knagg, Peter Stys, Piermaria Mendolicchio, Raia Hadsell, Rosemary Ke, Ryan Faulkner, Sarah Chakera, Satinder Singh Baveja, Shane Legg, Sheleem Kashem, Tayfun Terzi, Thomas Keck, Tim Harley, Tim Scholtes, Tyson Roberts, Volodymyr Mnih, Yulan Liu, Zhengdong Wang, Zoubin Ghahramani

*Please cite as*:

```
@article{simateam2025sima2,
   title={SIMA 2:  A Generalist Embodied Agent for Virtual Worlds},
   author={{SIMA Team}},
   year={2025},
   journal={arXiv preprint arXiv:2404.10179},
}
```

## References

J. Abramson, A. Ahuja, I. Barr, A. Brussee, F. Carnevale, M. Cassin, R. Chhaparia, S. Clark, B. Damoc, A. Dudzik, et al. Imitating Interactive Intelligence. *arXiv preprint arXiv:2012.05672*, 2020.

P. Agrawal, A. V. Nair, P. Abbeel, J. Malik, and S. Levine. Learning to poke by poking: Experiential learning of intuitive physics. In *Advances in Neural Information Processing Systems*, 2016.

P. Anderson, Q. Wu, D. Teney, J. Bruce, M. Johnson, N. Sünderhauf, I. Reid, S. Gould, and A. Van Den Hengel. Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. In *Computer Vision and Pattern Recognition*, 2018.

Anthropic. The claude 3 model family: Opus, sonnet, haiku. Technical report, Anthropic, 2024. URL https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude_3.pdf.

J. Bai, S. Bai, Y. Chu, Z. Cui, K. Dang, X. Deng, Y. Fan, W. Ge, Y. Han, F. Huang, et al. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023.

B. Baker, I. Akkaya, P. Zhokov, J. Huizinga, J. Tang, A. Ecoffet, B. Houghton, R. Sampedro, and J. Clune. Video pretraining (VPT): Learning to act by watching unlabeled online videos. In *Advances in Neural Information Processing Systems*, 2022.

P. J. Ball, J. Bauer, F. Belletti, B. Brownfield, A. Ephrat, S. Fruchter, A. Gupta, K. Holsheimer, A. Holynski, J. Hron, C. Kaplanis, M. Limont, M. McGill, Y. Oliveira, J. Parker-Holder, F. Perbet, G. Scully, J. Shar, S. Spencer, O. Tov, R. Villegas, E. Wang, J. Yung, C. Baetu, J. Berbel, D. Bridson, J. Bruce, G. Buttimore, S. Chakera, B. Chandra, P. Collins, A. Cullum, B. Damoc, V. Dasagi, M. Gazeau, C. Gbadamosi, W. Han, E. Hirst, A. Kachra, L. Kerley, K. Kjems, E. Knoepfel, V. Koriakin, J. Lo, C. Lu, Z. Mehring,

A. Moufarek, H. Nandwani, V. Oliveira, F. Pardo, J. Park, A. Pierson, B. Poole, H. Ran, T. Salimans, M. Sanchez, I. Saprykin, A. Shen, S. Sidhwani, D. Smith, J. Stanton, H. Tomlinson, D. Vijaykumar, L. Wang, P. Wingfield, N. Wong, K. Xu, C. Yew, N. Young, V. Zubov, D. Eck, D. Erhan, K. Kavukcuoglu, D. Hassabis, Z. Gharamani, R. Hadsell, A. van den Oord, I. Mosseri, A. Bolton, S. Singh, and T. Rocktäschel. Genie 3: A new frontier for world models. *Google DeepMind Blog*, 2025.

K. Baumli, S. Baveja, F. Behbahani, H. Chan, G. Comanici, S. Flennerhag, M. Gazeau, K. Holsheimer, D. Horgan, M. Laskin, et al. Vision-language models as a source of rewards. *arXiv preprint arXiv:2312.09187*, 2023.

C. Beattie, J. Z. Leibo, D. Teplyashin, T. Ward, M. Wainwright, H. Küttler, A. Lefrancq, S. Green, V. Valdés, A. Sadik, et al. Deepmind lab. *arXiv preprint arXiv:1612.03801*, 2016.

M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling. The Arcade Learning Environment: An Evaluation Platform for General Agents. *Journal of Artificial Intelligence Research*, 47:253–279, 2013.

C. Berner, G. Brockman, B. Chan, V. Cheung, P. Dębiak, C. Dennison, D. Farhi, Q. Fischer, S. Hashme, C. Hesse, et al. Dota 2 with Large Scale Deep Reinforcement Learning. *arXiv preprint arXiv:1912.06680*, 2019.

I. Bica, A. Ilic, M. Bauer, G. Erdogan, M. Bošnjak, C. Kaplanis, A. A. Gritsenko, M. Minderer, C. Blundell, R. Pascanu, et al. Improving fine-grained understanding in image-text pre-training. In *International Conference on Machine Learning*, 2024.

A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning*, 2023.

J. Bruce, M. D. Dennis, A. Edwards, J. Parker-Holder, Y. Shi, E. Hughes, M. Lai, A. Mavalankar, R. Steigerwald, C. Apps, et al. Genie: Generative interactive environments. In *International Conference on Machine Learning*, 2024.

ByteDance Seed, W. Tan, X. Li, Y. Fang, H. Yao, S. Yan, H. Luo, T. Ao, H. Li, H. Ren, B. Yi, Y. Qin, B. An, L. Liu, and G. Shi. Lumine: An Open Recipe for Building Generalist Agents in 3D Open Worlds. *arXiv preprint arXiv:2511.08892*, 2025.

J. Clune. AI-GAs: AI-generating algorithms, an alternate paradigm for producing general artificial intelligence. *arXiv preprint arXiv:1905.10985*, 2019.

C. Colas, T. Karch, O. Sigaud, and P.-Y. Oudeyer. Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: a short survey. *Journal of Artificial Intelligence Research*, 74: 1159–1199, 2022.

E. Coumans and Y. Bai. PyBullet, a Python module for physics simulation for games, robotics and machine learning. http://pybullet.org, 2016.

M. Deitke, E. VanderBilt, A. Herrasti, L. Weihs, K. Ehsani, J. Salvador, W. Han, E. Kolve, A. Kembhavi, and R. Mottaghi. ProcTHOR: Large-Scale Embodied AI Using Procedural Generation. In *Advances in Neural Information Processing Systems*, 2022.

D. Driess, F. Xia, M. S. Sajjadi, C. Lynch, A. Chowdhery, B. Ichter, A. Wahid, J. Tompson, Q. Vuong, T. Yu, et al. Palm-e: An embodied multimodal language model. In *International Conference on Machine Learning*, 2023.

Y. Du, O. Watkins, Z. Wang, C. Colas, T. Darrell, P. Abbeel, A. Gupta, and J. Andreas. Guiding pretraining in reinforcement learning with large language models. In *International Conference on Machine Learning*, 2023.

M. Faldor, J. Zhang, A. Cully, and J. Clune. OMNI-EPIC: Open-endedness via models of human notions of interestingness with environments programmed in code. *International Conference on Learning Representations*, 2025.

L. Fan, G. Wang, Y. Jiang, A. Mandlekar, Y. Yang, H. Zhu, A. Tang, D.-A. Huang, Y. Zhu, and A. Anandkumar. Minedojo: Building open-ended embodied agents with internet-scale knowledge. In *Advances in Neural Information Processing Systems*, 2022.

L. Fei-Fei. From words to worlds: Spatial intelligence is ai's next frontier. https://drfeifei.substack.com/p/from-words-to-worlds-spatial-intelligence, 2025.

R. M. French. Catastrophic forgetting in connectionist networks. *Trends Cogn. Sci.*, 3(4):128–135, Apr. 1999.

H. Gardner. *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books, New York, 1983.

Gemini Robotics Team, A. Abdolmaleki, S. Abeyruwan, J. Ainslie, J.-B. Alayrac, M. G. Arenas, A. Balakrishna, N. Batchelor, A. Bewley, J. Bingham, M. Bloesch, et al. Gemini robotics 1.5: Pushing the frontier of generalist robots with advanced embodied reasoning, thinking, and motion transfer. *arXiv preprint arXiv:2510.03342*, 2025a.

Gemini Robotics Team, S. Abeyruwan, J. Ainslie, J.-B. Alayrac, M. G. Arenas, T. Armstrong, A. Balakrishna, R. Baruch, M. Bauza, M. Blokzijl, et al. Gemini robotics: Bringing ai into the physical world. *arXiv preprint arXiv:2503.20020*, 2025b.

Gemini Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

Gemini Team, P. Georgiev, V. I. Lei, R. Burnell, L. Bai, A. Gulati, G. Tanzer, D. Vincent, Z. Pan, S. Wang, et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*, 2024.

Gemini Team, G. Comanici, E. Bieber, M. Schaekermann, I. Pasupat, N. Sachdeva, I. Dhillon, M. Blistein, O. Ram, D. Zhang, E. Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

S. K. S. Ghasemipour, A. Wahid, J. Tompson, P. Sanketi, and I. Mordatch. Self-improving embodied foundation models. *arXiv preprint arXiv:2509.15155*, 2025.

C. Gulcehre, T. Le Paine, B. Shahriari, M. Denil, M. Hoffman, H. Soyer, R. Tanburn, S. Kapturowski, N. Rabinowitz, D. Williams, et al. Making Efficient Use of Demonstrations to Solve Hard Exploration Problems. In *International Conference on Learning Representations*, 2019.

S. Gupta, J. Davidson, S. Levine, R. Sukthankar, and J. Malik. Cognitive mapping and planning for visual navigation. In *Computer Vision and Pattern Recognition*, 2017.

W. H. Guss, B. Houghton, N. Topin, P. Wang, C. Codel, M. Veloso, and R. Salakhutdinov. MineRL: A Large-Scale Dataset of Minecraft Demonstrations. In *International Joint Conference on Artificial Intelligence*, 2019.

D. Ha and J. Schmidhuber. Recurrent World Models Facilitate Policy Evolution. In *Advances in Neural Information Processing Systems*, 2018.

D. Hafner, T. Lillicrap, I. Fischer, R. Villegas, D. Ha, H. Lee, and J. Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning*, 2019.

D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640:647–653, 2025.

A. J. Hancock, X. Wu, L. Zha, O. Russakovsky, and A. Majumdar. Actions as language: Fine-tuning vlms into vlas without catastrophic forgetting. *arXiv preprint arXiv:2509.22195*, 2025.

D. Hendrycks, C. Burns, S. Kadavath, A. Arora, S. Basart, E. Tang, D. Song, and J. Steinhardt. Measuring mathematical problem solving with the math benchmark. In *Advances in Neural Information Processing Systems*, 2021.

K. M. Hermann, F. Hill, S. Green, F. Wang, R. Faulkner, H. Soyer, D. Szepesvari, W. M. Czarnecki, M. Jaderberg, D. Teplyashin, et al. Grounded language learning in a simulated 3d world. *arXiv preprint arXiv:1706.06551*, 2017.

D. Hershey. Claude Plays Pokemon Twitch Stream. Twitch, 2025. URL https://www.twitch.tv/claudeplayspokemon.

A. Hu, L. Russell, H. Yeo, Z. Murez, G. Fedoseev, A. Kendall, J. Shotton, and G. Corrado. Gaia-1: A generative world model for autonomous driving. *arXiv preprint arXiv:2309.17080*, 2023.

S. Huang, N. Papernot, I. Goodfellow, Y. Duan, and P. Abbeel. Adversarial attacks on neural network policies. *arXiv preprint arXiv:1702.02284*, 2017.

N. Jain, K. Han, A. Gu, W.-D. Li, F. Yan, T. Zhang, S. Wang, A. Solar-Lezama, K. Sen, and I. Stoica. Livecodebench: Holistic and contamination free evaluation of large language models for code. *arXiv preprint arXiv:2403.07974*, 2024.

M. Johnson, K. Hofmann, T. Hutton, and D. Bignell. The malmo platform for artificial intelligence experimentation. In *International Joint Conference on Artificial Intelligence*, 2016.

A. Kanervisto, D. Bignell, L. Y. Wen, M. Grayson, R. Georgescu, S. Valcarcel Macua, S. Z. Tan, T. Rashid, T. Pearce, Y. Cao, et al. World and human action models towards gameplay ideation. *Nature*, 638 (8051):656–663, 2025.

K. Kansky, T. Silver, D. A. Mély, M. Eldawy, M. Lázaro-Gredilla, X. Lou, N. Dorfman, S. Sidor, S. Phoenix, and D. George. Schema networks: Zero-shot transfer with a generative causal model of intuitive physics. In *International Conference on Machine Learning*, 2017.

M. Kempka, M. Wydmuch, G. Runc, J. Toczek, and W. Jaśkowski. Vizdoom: A doom-based ai research platform for visual reinforcement learning. In *Computational Intelligence and Games*, 2016.

M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. Foster, G. Lam, P. Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

E. Kolve, R. Mottaghi, W. Han, E. VanderBilt, L. Weihs, A. Herrasti, M. Deitke, K. Ehsani, D. Gordon, Y. Zhu, et al. AI2-THOR: An Interactive 3D Environment for Visual AI. *arXiv preprint arXiv:1712.05474*, 2017.

K.-H. Lee, O. Nachum, M. S. Yang, L. Lee, D. Freeman, S. Guadarrama, I. Fischer, W. Xu, E. Jang, H. Michalewski, et al. Multi-game decision transformers. In *Advances in Neural Information Processing Systems*, 2022.

S. Levine, C. Finn, T. Darrell, and P. Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40, 2016.

S. Lifshitz, K. Paster, H. Chan, J. Ba, and S. McIlraith. Steve-1: A generative model for text-to-behavior in minecraft. *Advances in Neural Information Processing Systems*, 2023.

R. Liu, C. Bai, J. Lyu, S. Sun, Y. Du, and X. Li. Vlp: Vision-language preference learning for embodied manipulation. *arXiv preprint arXiv:2502.11918*, 2025.

J. Luketina, N. Nardelli, G. Farquhar, J. Foerster, J. Andreas, E. Grefenstette, S. Whiteson, and T. Rocktäschel. A survey of reinforcement learning informed by natural language. *arXiv preprint arXiv:1906.03926*, 2019.

Y. Luo, Z. Yang, F. Meng, Y. Li, J. Zhou, and Y. Zhang. An empirical study of catastrophic forgetting in large language models during continual fine-tuning. In *Audio, Speech and Language Processing*, 2025.

C. Lynch and P. Sermanet. Grounding language in play. *arXiv preprint arXiv:2005.07648*, 40(396): 105, 2020.

Y. J. Ma, V. Kumar, A. Zhang, O. Bastani, and D. Jayaraman. Liv: Language-image representations and rewards for robotic control. In *International Conference on Machine Learning*, 2023a.

Y. J. Ma, W. Liang, G. Wang, D.-A. Huang, O. Bastani, D. Jayaraman, Y. Zhu, L. Fan, and A. Anandkumar. Eureka: Human-level reward design via coding large language models. *arXiv preprint arXiv:2310.12931*, 2023b.

A. Majumdar, A. Ajay, X. Zhang, P. Putta, S. Yenamandra, M. Henaff, S. Silwal, P. Mcvay, O. Maksymets, S. Arnaud, et al. Openeqa: Embodied question answering in the era of foundation models. In *Computer Vision and Pattern Recognition*, 2024.

V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, et al. Isaac Gym: High Performance GPU Based Physics Simulation For Robot Learning. In *Advances in Neural Information Processing Systems*, 2021.

D. J. Mankowitz, A. Michi, A. Zhernov, M. Gelmi, M. Selvi, C. Paduraru, E. Leurent, S. Iqbal, J.-B. Lespiau, A. Ahern, et al. Faster sorting algorithms discovered using deep reinforcement learning. *Nature*, 618(7964):257–263, 2023.

H. Mei, M. Bansal, and M. Walter. Listen, attend, and walk: Neural mapping of navigational instructions to action sequences. In *AAAI Conference on Artificial Intelligence*, 2016.

B. Mel. Murphy: A robot that learns by doing. In *Advances in Neural Information Processing Systems*, 1987.

R. Mendonca, O. Rybkin, K. Daniilidis, D. Hafner, and D. Pathak. Discovering and achieving goals via world models. *Advances in Neural Information Processing Systems*, 34:24379–24391, 2021.

V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning*, 2016.

H. Moravec. *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press, 1988.

A. V. Nair, V. Pong, M. Dalal, S. Bahl, S. Lin, and S. Levine. Visual reinforcement learning with imagined goals. *Advances in Neural Information Processing Systems*, 2018.

S. Nair, A. Rajeswaran, V. Kumar, C. Finn, and A. Gupta. R3m: A universal visual representation for robot manipulation. *arXiv preprint arXiv:2203.12601*, 2022.

OpenAI. Universe. https://openai.com/index/universe/, 2016.

OpenAI. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*, 2023.

A. O'Neill, A. Rehman, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, A. Jain, et al. Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0. In *International Conference on Robotics and Automation*, 2024.

D. Paglieri, B. Cupiał, S. Coward, U. Piterbarg, M. Wolczyk, A. Khan, E. Pignatelli, Ł. Kuciński, L. Pinto, R. Fergus, J. N. Foerster, J. Parker-Holder, and T. Rocktäschel. BALROG: Benchmarking agentic LLM and VLM reasoning on games. In *International Conference on Learning Representations*, 2025.

J. Parker-Holder, P. Ball, J. Bruce, V. Dasagi, K. Holsheimer, C. Kaplanis, A. Moufarek, G. Scully, J. Shar, J. Shi, S. Spencer, J. Yung, M. Dennis, S. Kenjeyev, S. Long, V. Mnih, H. Chan, M. Gazeau, B. Li, F. Pardo, L. Wang, L. Zhang, F. Besse, T. Harley, A. Mitenkova, J. Wang, J. Clune, D. Hassabis, R. Hadsell, A. Bolton, S. Singh, and T. Rocktäschel. Genie 2: A large-scale foundation world model. *Google DeepMind Blog*, 2024.

T. Pearce and J. Zhu. Counter-Strike Deathmatch with Large-Scale Behavioural Cloning. In *Conference on Games*, 2022.

T. Pearce, T. Rashid, D. Bignell, R. Georgescu, S. Devlin, and K. Hofmann. Scaling laws for pre-training agents and world models. *arXiv preprint arXiv:2411.04434*, 2024.

Physical Intelligence, K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, L. Groom, K. Hausman, B. Ichter, et al. $\pi 0$: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.

Physical Intelligence, K. Black, N. Brown, J. Darpinian, K. Dhabalia, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, et al. $\pi 0.5$: a vision-language-action model with open-world generalization. *arXiv preprint arXiv:2504.16054*, 2025.

L. Pinto and A. Gupta. Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. In *International Conference on Robotics and Automation*, 2016.

S. Reed, K. Zolna, E. Parisotto, S. G. Colmenarejo, A. Novikov, G. Barth-maron, M. Giménez, Y. Sulsky, J. Kay, J. T. Springenberg, et al. A Generalist Agent. *Transactions on Machine Learning Research*, 2022.

D. Rein, B. L. Hou, A. C. Stickland, J. Petty, R. Y. Pang, J. Dirani, J. Michael, and S. R. Bowman. Gpqa: A graduate-level google-proof q&a benchmark. *arXiv preprint arXiv:2311.12022*, 2023.

J. Rocamonde, V. Montesinos, E. Nava, E. Perez, and D. Lindner. Vision-language models are zero-shot reward models for reinforcement learning. In *International Conference on Learning Representations*, 2024.

A. Ruoss, F. Pardo, H. Chan, B. Li, V. Mnih, and T. Genewein. LMAct: A benchmark for in-context imitation learning with long multimodal demonstrations. In *International Conference on Machine Learning*, 2025.

L. Russell, A. Hu, L. Bertoni, G. Fedoseev, J. Shotton, E. Arani, and G. Corrado. Gaia-2: A controllable multi-view generative world model for autonomous driving. *arXiv preprint arXiv:2503.20523*, 2025.

A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959.

M. Savva, A. Kadian, O. Maksymets, Y. Zhao, E. Wijmans, B. Jain, J. Straub, J. Liu, V. Koltun, J. Malik, et al. Habitat: A Platform for Embodied AI Research. In *International Conference on Computer Vision*, 2019.

J. Schmidhuber. *Making the world differentiable: on using self supervised fully recurrent neural networks for dynamic reinforcement learning and planning in non-stationary environments*, volume 126. Inst. für Informatik, 1990.

J. Schmidhuber. Curious model-building control systems. In *International Joint Conference on Neural Networks*, 1991.

R. Sekar, O. Rybkin, K. Daniilidis, P. Abbeel, D. Hafner, and D. Pathak. Planning to explore via self-supervised world models. In *International Conference on Machine Learning*, 2020.

C. E. Shannon. Programming a computer for playing chess. *Philosophical Magazine*, 41(314):256–275, 1950.

S. Sharma, G. Davidson, K. Khetarpal, A. Kanervisto, U. Arora, K. Hofmann, and I. Momennejad. Toward human-ai alignment in large-scale multi-player games. *arXiv preprint arXiv:2402.03575*, 2024.

M. Shridhar, L. Manuelli, and D. Fox. Cliport: What and where pathways for robotic manipulation. In *Conference on Robot Learning*, 2022.

D. Silver and R. S. Sutton. Welcome to the era of experience. https://storage.googleapis.com/deepmind-media/Era-of-Experience%20/The%20Era%20of%20Experience%20Paper.pdf, 2025.

SIMA Team, M. A. Raad, A. Ahuja, C. Barros, F. Besse, A. Bolt, A. Bolton, B. Brownfield, G. Buttimore, M. Cant, S. Chakera, et al. Scaling instructable agents across many simulated worlds. *arXiv preprint arXiv:2404.10179*, 2024.

S. Sontakke, J. Zhang, S. Arnold, K. Pertsch, E. Bıyık, D. Sadigh, C. Finn, and L. Itti. Roboclip: One demonstration is enough to learn robot policies. In *Advances in Neural Information Processing Systems*, 2023.

K. O. Stanley and J. Lehman. *Why greatness cannot be planned: The myth of the objective*. Springer, 2015.

K. O. Stanley, J. Lehman, and L. Soros. Open-endedness: The last grand challenge you've never heard of. *RADAR*, 2017.

Q. Sun, P. Hong, T. D. Pala, V. Toh, U.-X. Tan, D. Ghosal, and S. Poria. Emma-x: An embodied multimodal action model with grounded chain of thought and look-ahead spatial reasoning. In *Association for Computational Linguistics*, 2025.

R. S. Sutton. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Machine Learning*, pages 216–224. 1990.

R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT Press Cambridge, 1998.

W. Tan, Z. Ding, W. Zhang, B. Li, B. Zhou, J. Yue, H. Xia, J. Jiang, L. Zheng, X. Xu, et al. Towards General Computer Control: A Multimodal Agent for Red Dead Redemption II as a Case Study. *arXiv preprint arXiv:2403.03186*, 2024.

E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In *International Conference on Intelligent Robots and Systems*, 2012.

A. M. Turing. Digital computers applied to games. In B. V. Bowden, editor, *Faster than thought*, pages 286–310. Pitman, London, 1953.

D. Valevski, Y. Leviathan, M. Arar, and S. Fruchter. Diffusion models are real-time game engines. In *International Conference on Learning Representations*, 2025.

R. Villegas, M. Babaeizadeh, P.-J. Kindermans, H. Moraldo, H. Zhang, M. T. Saffar, S. Castro, J. Kunze, and D. Erhan. Phenaki: Variable Length Video Generation from Open Domain Textual Descriptions. In *International Conference on Learning Representations*, 2022.

O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan, and A. Anandkumar. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*, 2023a.

R. Wang, J. Lehman, J. Clune, and K. O. Stanley. Paired open-ended trailblazer (POET): Endlessly generating increasingly complex and diverse learning environments and their solutions. In *Genetic and Evolutionary Computation Conference*, 2019.

Y. Wang, Z. Sun, J. Zhang, Z. Xian, E. Biyik, D. Held, and Z. Erickson. Rl-vlm-f: Reinforcement learning from vision language foundation model feedback. In *International Conference on Machine Learning*, 2024.

Z. Wang, S. Cai, A. Liu, Y. Jin, J. Hou, B. Zhang, H. Lin, Z. He, Z. Zheng, Y. Yang, et al. JARVIS-1: Open-World Multi-task Agents with Memory-Augmented Multimodal Language Models. *arXiv preprint arXiv:2311.05997*, 2023b.

Z. Wang, X. Li, Y. Ye, J. Fang, H. Wang, L. Liu, S. Liang, J. Lu, Z. Wu, J. Feng, et al. Game-tars: Pretrained foundation models for scalable generalist multimodal game agents. *arXiv preprint arXiv:2510.23691*, 2025.

X. Wen, Z. Liu, S. Zheng, S. Ye, Z. Wu, Y. Wang, Z. Xu, X. Liang, J. Li, Z. Miao, J. Bian, and M. Yang. Reinforcement learning with verifiable rewards implicitly incentivizes correct reasoning in base llms. *arXiv preprint arXiv:2506.14245*, 2025.

P. J. Werbos. Learning how the world works: Specifications for predictive networks in robots and brains. In *International Conference on Systems, Man and Cybernetics*, 1987.

R. Yang, H. Chen, J. Zhang, M. Zhao, C. Qian, K. Wang, Q. Wang, T. V. Koripella, M. Movahedi, M. Li, et al. Embodiedbench: Comprehensive benchmarking multi-modal large language models for vision-driven embodied agents. *arXiv preprint arXiv:2502.09560*, 2025.

W. Yu, N. Gileadi, C. Fu, S. Kirmani, K.-H. Lee, M. G. Arenas, H.-T. L. Chiang, T. Erez, L. Hasenclever, J. Humplik, et al. Language to rewards for robotic skill synthesis. In *Conference on Robot Learning*, 2023.

S. Zhai, Q. Zhang, T. Zhang, F. Huang, H. Zhang, M. Zhou, S. Zhang, L. Liu, S. Lin, and J. Pang. A vision-language-action-critic model for robotic real-world reinforcement learning. *arXiv preprint arXiv:2509.15937*, 2025.

A. L. Zhang, T. L. Griffiths, K. R. Narasimhan, and O. Press. Videogamebench: Can vision-language models complete popular video games? *arXiv preprint arXiv:2505.18134*, 2025a.

J. Zhang. Gemini Plays Pokemon Twitch Stream. Twitch, 2025. URL https://www.twitch.tv/gemini_plays_pokemon.

J. Zhang, J. Lehman, K. Stanley, and J. Clune. OMNI: Open-endedness via models of human notions of interestingness. *arXiv preprint arXiv:2306.01711*, 2023.

W. Zhang, M. Wang, G. Liu, X. Huixin, Y. Jiang, Y. Shen, G. Hou, Z. Zheng, H. Zhang, X. Li, et al. Embodied-reasoner: Synergizing visual search, reasoning, and action for embodied interactive tasks. *arXiv preprint arXiv:2503.21696*, 2025b.

Q. Zhao, Y. Lu, M. J. Kim, Z. Fu, Z. Zhang, Y. Wu, Z. Li, Q. Ma, S. Han, C. Finn, et al. Cot-vla: Visual chain-of-thought reasoning for vision-language-action models. In *Computer Vision and Pattern Recognition*, 2025.

Z. Zhou, Y. Zhu, M. Zhu, J. Wen, N. Liu, Z. Xu, W. Meng, Y. Peng, C. Shen, F. Feng, et al. Chatvla: Unified multimodal understanding and robot control with vision-language-action model. In *Conference on Empirical Methods in Natural Language Processing*, 2025.

Y. Zhu, R. Mottaghi, E. Kolve, J. J. Lim, A. Gupta, L. Fei-Fei, and A. Farhadi. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *International Conference on Robotics and Automation*, 2017.

## A. Embodied Skill Categories

Table 2 | **Skill Categories**

| Category | Description | Examples |
|---|---|---|
| **Interaction** | Various forms of interaction with the environment or characters | *use a machine or workbench, get launched by a fan, talk to a non-playable character* |
| **Navigation** | Tasks requiring walking or driving to a location | *exit the house, go to your pet, run to the starship, go to an event* |
| **Menu Use** | Any tasks within a menu | *open the inventory, click on X, hover over Y, place a waypoint on the map* |
| **Tool Use** | Equipping and using tools | *equip the hammer, scan for resources, use the terrain manipulator* |
| **Construction** | Various tasks around building, crafting, and repairing | *deploy a portable refiner, craft a stone axe, repair a wall* |
| **Object Management** | Tasks requiring the identification and movement of objects | *lick a tire, pick up the xeno-zapper, drop a stone, purchase a chair* |
| **Resource Gathering** | Tasks involving collecting, harvesting, or mining resources | *pick berries, mine limestone, fish, gather wood* |
| **Combat** | Fighting enemies or hunting | *defeat a greyling, hunt a deer, hunt a hog* |

## B. Additional Results Combining Gemini Pro & SIMA 2

Here, we provide two additional examples of more advanced reasoning capabilities enabled by combining Gemini Pro and SIMA 2. The SIMA 2 agent, on its own, is generally capable of following instructions, however, these are often tied to the immediate task rather than past user input. In the following examples, the combined Gemini Pro + SIMA 2 agent must use past user instructions to either modify current behavior (a form of abstract reasoning) or guide exploration.

**Abstract Reasoning** The agent is initially instructed with *"From now on, do the opposite of what I tell you"* and must reason through the opposite action for each subsequent instruction. For instance, if it is asked to equip the item with the lowest hotbar key, it should instead equip the one with the highest. The agent is given a series of instructions that involve navigation, menu use, and tool use. Performing such a task requires memory and abstract reasoning, as the agent must recall the initial user instruction and use this to modify its current behavior. As shown in Figure 18, the combined Gemini Pro + SIMA 2 agent successfully performs the opposite of each task, both executing the appropriate actions and explaining its reasoning in the dialogue responses.


Figure 18 | **Abstract Reasoning**. The combined Gemini Pro + SIMA 2 agent successfully incorporates the user's initial instruction to *"do the opposite of what I tell you."* The agent correctly applies this form of abstract reasoning to a series of navigation, menu use, and tool use tasks, demonstrating both memory and more advanced reasoning.

**Active Exploration** In this example, the agent plays the game 21 questions. The user first thinks of an object in the environment, and the agent must explore and ask questions to determine the identity of the object. This task switches the common roles between the user and the agent, with the agent taking the role of driving the progression, deciding where to explore and what questions to ask. Rather than task instructions, the agent receives answers from the user, such as "yes/no," "indeed," *etc.* Successfully playing the game requires using these answers to drive exploration. As shown in Figure 19, the agent adeptly performs the task, asking effective questions that allow it to quickly infer the user's selected object.

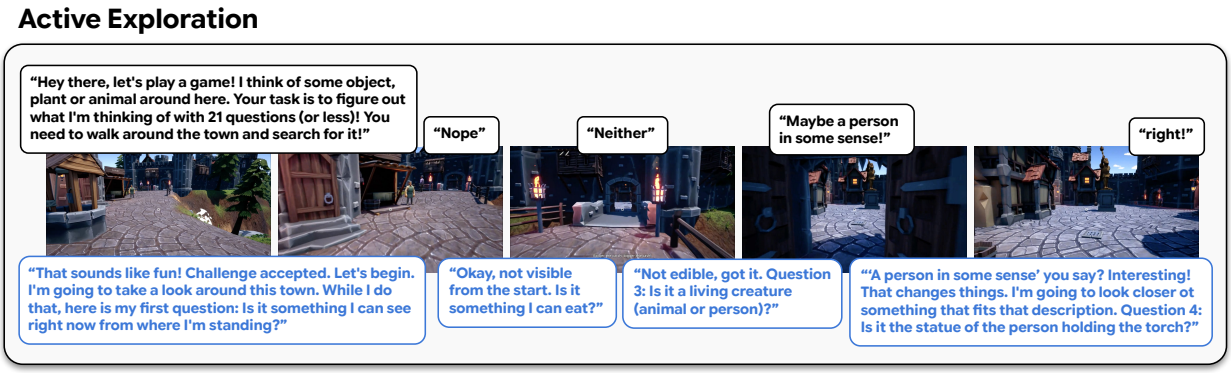



Figure 19 | **Active Exploration**. The combined Gemini Pro + SIMA 2 agent plays the game 21 questions with a user, asking questions and using the user's answers to drive active exploration. Again, this highlights the advanced capabilities of the combined agent to use both memory and reasoning to complete a novel task.

## C. Additional Self-Improvement Results


Figure 20 | **Self-Improvement Behavior**. Through training on self-generated experience, SIMA 2 is capable of acquiring new skills in a previously unseen environment, ASKA. After running multiple iterations of self-improvement, the agent learns to recognize a novel object (*rain collector*) and perform a new skill (*extinguishing a campfire*).

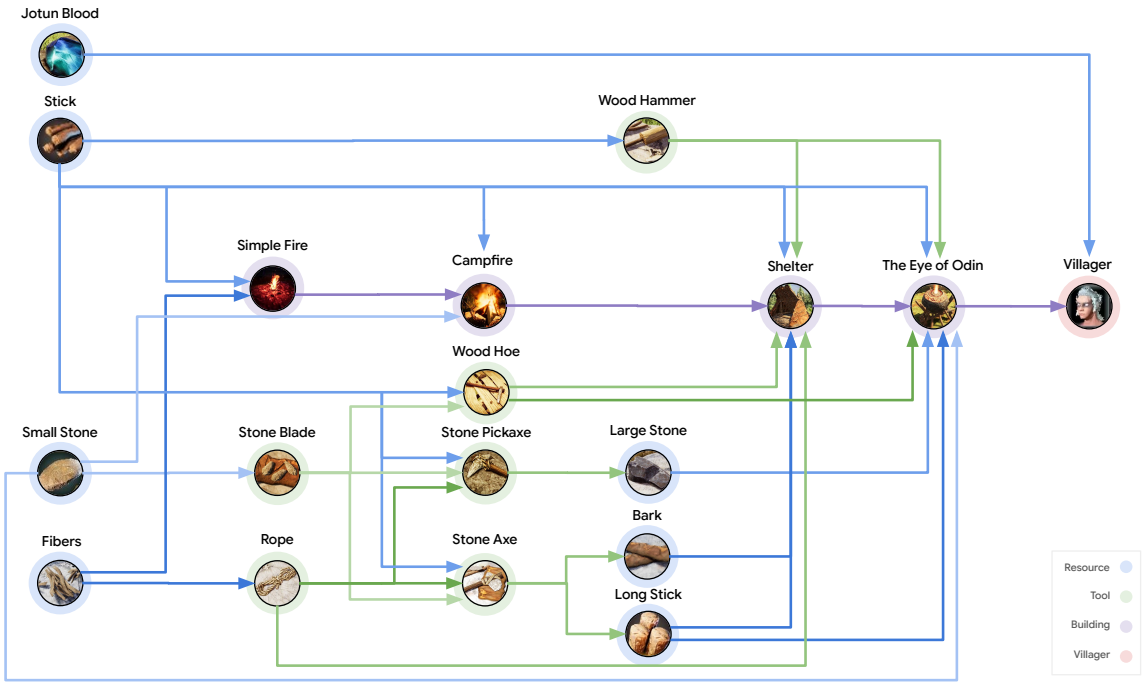



Figure 21 | **ASKA Technology Tree**. Starting from a new game, the diagram shows the *minimal* tech tree required to summon the first villager (a core mechanic of the game).