Google

Gemini Robotics On-Device 2 Model Card

Model Cards are intended to provide essential information about models, including known limitations, mitigation approaches, and safety performance. Model cards may be updated from time-to-time; for example, to include updated evaluations as the model is improved or revised.

Updated: July 12, 2026

Model Information

Description

Gemini Robotics On-Device is a state-of-the-art Vision-Language-Action (VLA) model based on our on-device Gemma models. It is designed for general-purpose robotic manipulation, operating efficiently on local devices. This model supports a wide range of tasks, scenes, and multiple robot types. It is currently available to a select group of trusted testers.

Inputs

Text (e.g., a question or instruction), images (e.g., robot's perspective of the environment), and robot proprioception as numerical values.

Outputs

Robot actions as numerical values.

Architecture

Gemini Robotics On-Device is an on-device VLA model based on Gemini Robotics 1.5 technology and our on-device Gemma models.

Model Data

Training Dataset

Gemini Robotics On-Device was trained on datasets consisting of images, text, and robot sensor and action data.

Training Data Processing

Data filtering and preprocessing included techniques such as deduplication, safety filtering in line with Google's commitment to advancing AI safely and responsibly and quality filtering to mitigate risks and improve training data reliability.

Implementation and Sustainability

Hardware

Gemini Robotics On-Device was trained using Google's Tensor Processing Units (TPUs). TPUs are specifically designed to handle the massive computations involved in training LLMs and can speed up training considerably compared to CPUs. TPUs come with large amounts of high-bandwidth memory, allowing for the handling of large models and batch sizes during training, which can lead to better model quality. TPU Pods (large clusters of TPUs) also provide a scalable solution for handling the growing complexity of large foundation models. Training can be distributed across multiple TPU devices for faster and more efficient processing.

The efficiencies gained through the use of TPUs are aligned with Google's commitment to operate sustainably.

Software

Training was done using JAX and ML Pathways.

Distribution

Gemini Robotics On-Device 2 is distributed only to Trusted Testers (signup link).

Evaluation

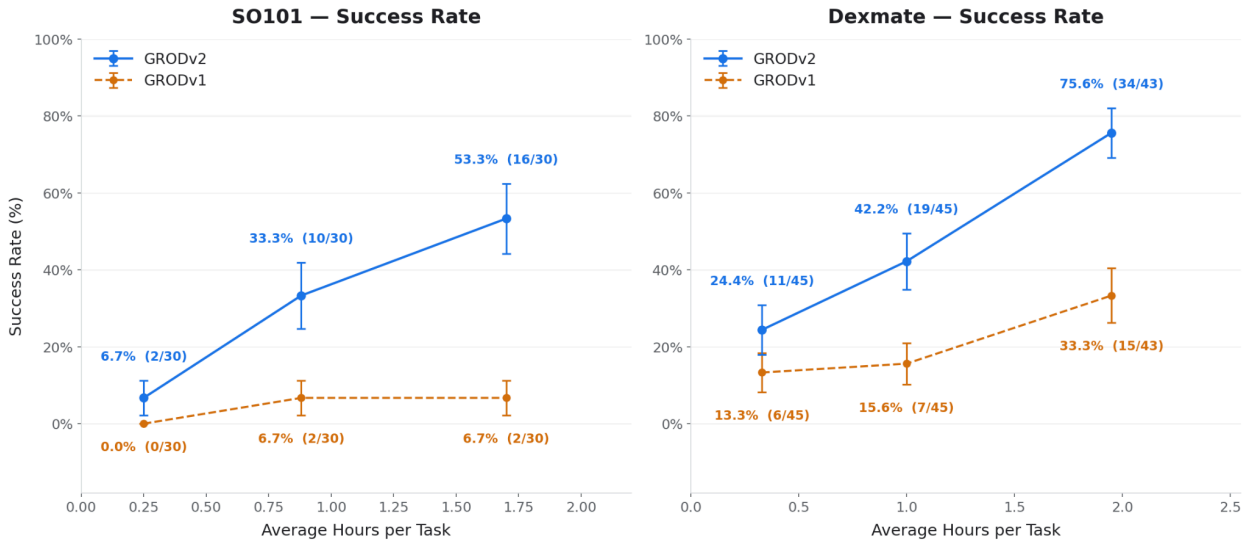
Approach

Gemini Robotics On-Device was evaluated on several simulation and on-robot benchmarks. This evaluation assessed scene, instruction, and action generalization, as well as instruction following for tasks outside the training data distribution.

Results

Gemini Robotics On-Device 2 demonstrates substantial performance advancements over Gemini Robotics On-Device 1 across diverse robotic embodiments. When evaluated on novel platforms introduced exclusively during post-training, Gemini Robotics On-Device 2 exhibits faster and more efficient learning and achieves a significantly higher final success rate.

Gemini Robotics On-Device 1 vs Gemini Robotics On-Device 2 - Data scaling evaluation



Intended Usage and Limitations

Intended Usage

The Gemini Robotics On-Device model is a VLA model designed for on-device deployment in robotics applications, offering competitive results to larger models (like Gemini Robotics VLA) on in-domain tasks. The model allows for efficient on-robot inference and is trained to provide a general base for a variety of tasks on bi-arm robots. Its intended use is to serve as a core component in robotics systems, enabling them to understand and respond to linguistic instructions, and to act in a given environment.

Known Limitations

The Gemini Robotics On-Device model is limited in its ability to generalise to out of distribution tasks as well as controlling high-degree-of-freedom robots.

Ethics and Safety

Safety Evaluations, Risks, and Mitigations: Gemini Robotics On-Device was developed in partnership with our internal safety, security, and responsibility teams. For specific physical embodiments, we conducted safety evaluations and red-teaming, tracking metrics such as collision rates, motion quality, and out-of-distribution robustness.

Gemini Robotics On-Device inherently minimizes standard generative content risks by restricting its outputs to physical actions rather than text. Because it has been evaluated primarily on standing bi-arm manipulation tasks, risks associated with mobile platforms or whole-body control are outside its current scope. We recommend a layered approach to safety: interfacing VLAs with Embodied Reasoning models for high-level semantic safety; using low-level controllers for collision-free motion planning, balance and force control, and implementing hardware-specific functional safety mechanisms.
