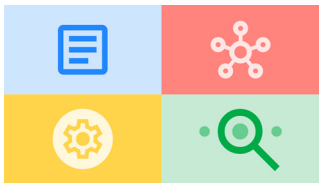
Google

Model card published June 24. 2025



Gemini Robotics On-Device Model Card

Model Cards are intended to provide essential information on 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.

Published: June 24, 2025

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. This model 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 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.

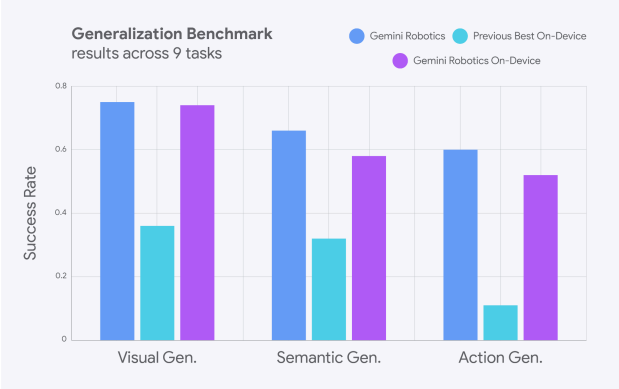
The efficiencies gained through the use of TPUs are aligned with Google's commitment to operate sustainably.

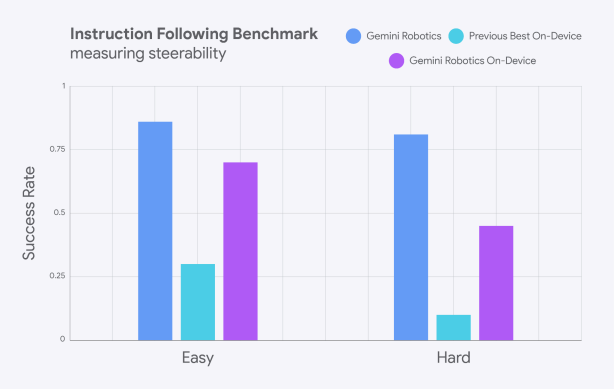
Software: Training was done using JAX and ML Pathways.

Evaluation

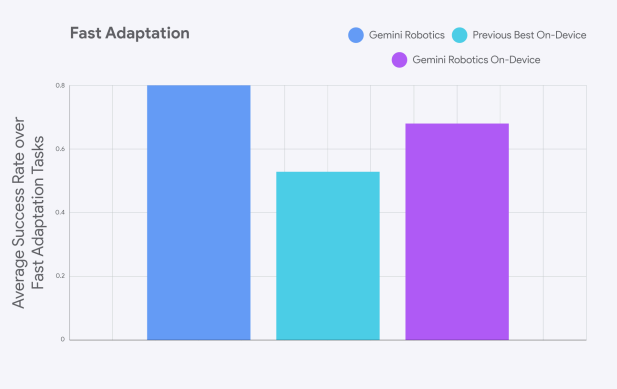
Approach: Similar to the March 2025 Technical Report, Gemini Robotics On-Device was evaluated on the Gemini Robotics Benchmark. This evaluation assessed scene, instruction, and action generalization, as well as instruction following for tasks outside the training data distribution.

Results: The Gemini Robotics On-Device model performed similarly with our flagship Gemini Robotics model on both generalization and instruction following performance while running entirely locally.





In comparison to existing methods for fine-tuning to newer models, Gemini Robotics On-Device outperformed the current, best, on-device VLA. We use seven dexterous manipulation tasks of varying degrees of difficulty with less than 100 examples.



Intended Usage

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 visual and linguistic instructions, and to potentially act in a given environment.

Ethics and Safety

Safety Considerations: Whilst developing the Gemini Robotics On-Device VLA we have carefully considered the following areas of safety risks and developed strategies to mitigate them:

- Mistakes and misalignment: errors associated with hallucination, lack of robustness to environmental factors, and misalignment with user instruction.
- Bias and fairness: errors that result in discriminatory output, including performance disparities.
- Misuse: circumvention of safety capabilities including novel jailbreaking techniques for robotics foundation models.

Mitigations: We recommend a layered approach to safety where on-device VLA models are interfaced both with Embodied Reasoning models for high-level semantic safety understanding, as well as with hardware-specific safety-critical low-level controllers for collision-free motion generation and force modulation. We actively work on safety features and evaluation metrics, building on recent work on semantic safety and automated red-teaming, to improve overall safety and performance of Gemini Robotics.