Gemini Robotics 2.5 explained: a vision-language-action model, factory pilots and robot safety

25 November 2026 · 8 min read · blogs.iictp.org/gemini-robotics-2-5-vla-model-dexterity-apollo-mercedes-explained

Short answer: Gemini Robotics 2.5 is a vision-language-action (VLA) model from Google DeepMind, announced on 9 November 2026, that turns camera input and instructions into robot arm and hand actions. DeepMind says it was evaluated on more than 200 tasks, reached a 96% success rate in a pick-and-pack pilot with Apptronik's Apollo 2 humanoid on Mercedes-Benz production lines, and adds inference-time safety steering plus an open-source jailbreak benchmark called ASIMOV-Jailbreak. It matters because it shows how reasoning models and action models are now being paired for physical work.

TL;DR The short version

- Gemini Robotics 2.5 is Google DeepMind's vision-language-action model that turns camera input and instructions into robot actions.
- DeepMind reports more than 200 tasks and a 96% success rate in one pick-and-pack pilot on Mercedes-Benz lines with Apptronik's Apollo 2 robot.
- The 96% is from a single task with unstated trial counts, so ask what counted as success and what happened on failures.

Read online, listen to it, or change the text size: https://blogs.iictp.org/gemini-robotics-2-5-vla-model-dexterity-apollo-mercedes-explained

What is Gemini Robotics 2.5?

A VLA model takes images and language and outputs motor actions. DeepMind describes Gemini Robotics 2.5 as its most capable VLA model so far and says it natively works across robot types, including Apptronik's Apollo 2 humanoid with a five-finger, 22 degree-of-freedom hand and the Franka Duo platform with two-fingered parallel grippers. It is rolling out to DeepMind's robotics partners rather than to the public.

How it works

- **Two models, two jobs.** In the Mercedes-Benz pilot, DeepMind says the embodied reasoning model (Gemini Robotics ER) keeps track of the workflow while the Gemini Robotics 2.5 actions model controls the hands and arms.
- **Simulation from a single scene.** The post says Gemini 4 Argon can generate thousands of varied simulated practice runs from a real camera scene and build a digital twin, which is combined with real demonstrations so the robot generalises to new scenes and objects.

- **Failure analysis.** Argon can also propose why a task failed from video and test fixes in simulation and on real robots, according to the post.
- **Inference-time safety guidance.** The model uses reasoning guidance at run time to steer around hazards, similar in spirit to a safety classifier around a chatbot.

What is new compared with earlier approaches

DeepMind frames the advance as breadth, speed of learning and reliability together. Compared with previous baselines, it reports a 30% to 100% higher success rate depending on the task type and tasks completed 20% to 30% faster. It also says a few weeks of post-training were enough for Apollo to carry its skill over to the Mercedes-Benz production line with very little drop-off.

Results reported by the source

- **Tasks:** more than 200 distinct tasks, including serving food with tongs, tying a bow on a gift box and pipetting a liquid.
- **Factory pilot:** 96% success picking automotive brackets from bulk bins into sequencing containers at Kecskemét, Hungary and Sindelfingen, Germany. DeepMind calls it a proof-of-concept.
- **Safety:** inference-time guidance raised safe task completion by 33% around obstacles and 43% around nearby people.
- **Jailbreaks:** on the new ASIMOV-Jailbreak benchmark, Gemini 4 Argon reached 99.5% jailbreak resistance against 82% for the previous baseline, per DeepMind. These are the vendor’s numbers.

Limitations and open questions

- The 96% figure is from one pilot task, kitting, not a general measure of factory work.
- The post does not state how many trials sit behind the percentages in the text, nor the failure modes that remain.
- The model is available to partners only, so outside teams cannot yet test the claims. Sources were read on 25 November 2026. Products, rules and standards change, so check the linked originals before you rely on any detail.

How to assess a robotics claim like this

1. Ask what the task was, how many trials were run and what counted as success.
2. Ask what happened on the failures: dropped part, stalled robot, or unsafe motion.
3. Check whether the safety benchmark is public. Here the jailbreak set is open source, which lets others test it.
4. Compare against the same robot and hardware, not a different baseline.

What it means for engineers and operations teams

For software and systems engineers, the notable pattern is a planner model that tracks the workflow and an action model that executes, with a simulator used to generate training data. Those are separable components you can evaluate one at a time. For safety and compliance teams, an open jailbreak benchmark for robots is a reusable test artefact. See AlphaProof Code and formal verification for another example of pairing generation with checking.

Frequently asked questions

- **What is a vision-language-action model?** A model that takes images and text instructions and outputs low-level robot actions.
- **What did Gemini Robotics 2.5 achieve at Mercedes-Benz?** DeepMind reports a 96% success rate in a kitting pilot with Apptronik's Apollo 2 humanoid.
- **Is it publicly available?** DeepMind says it is rolling out to robotics partners.
- **What is ASIMOV-Jailbreak?** An open-source benchmark of hundreds of prompts that disguise unsafe requests across dozens of scenes.

Related reading

- AlphaProof Code and formal verification
- What is Genie 4, a world model?
- AI Practitioner certification

> Related credential: AI Practitioner. Explore the exam facts and syllabus.

Sources and further reading

1. Google DeepMind. Gemini Robotics 2.5 accelerates real-world general dexterity. 9 November 2026. deepmind.google/blog/gemini-robotics-2-5-accelerates-real-world-physical-intelligence

Published by IICTP, an independent technical credential body, on 25 November 2026. The current version of this post is at https://blogs.iictp.org/gemini-robotics-2-5-vla-model-dexterity-apollo-mercedes-explained. More guides: blogs.iictp.org.