Model card published/updated OCT 07, 2025

Gemini 2.5 Computer Use Model Card

Gemini 2.5 Computer Use - Model Card

Model Cards *are intended to provide essential information on Gemini 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.*

Technical Reports *are similar to academic papers, and describe models' capabilities, limitations and performance benchmarks. The Gemini 2.5 technical report contains additional details about the Gemini 2.5 series of models that are generally available. We recommend that readers seeking more details and information about these models navigate to the technical report.*

Published: October 7, 2025

Model Information

Description: The Gemini 2.5 Computer Use model is based on the Gemini 2.5 Pro (06-2025) model, including architecture, training data, implementation and sustainability, with post-training to improve UI control capabilities. The model has primarily been optimized for understanding and interacting with web browsers. It has not yet been optimized for OS-level control or mobile control, although it demonstrates promise for mobile UI control tasks.

The model's core capabilities are exposed through a new computer use tool that enables applications to interact with and automate tasks in the browser. Using screenshots, the model can "see" a computer screen, and "act" by generating specific UI actions like mouse clicks and keyboard inputs.

- Inputs: Text strings (e.g., a question or a prompt), screenshots of the environment state, and a history of recent actions
- Outputs: Function call of the proposed next action

Evaluation

Approach: The Gemini 2.5 Computer Use Model was evaluated using the methodology below:

- **Online Mind2Web:** following an industry standard for evaluating web agents
- **WebVoyager:** following an industry standard for evaluating web agents
- **Android World:** testing generalization to another surface and action space

Results: The Gemini 2.5 Computer Use model demonstrates strong performance on multiple web and mobile control benchmarks. The table below includes results from self-reported numbers, evaluations run by Browserbase and evaluations we ran ourselves. Unless otherwise indicated, scores shown are for computer use tools exposed via API.

Benchmark	Source / Harness	Gemini 2.5 Computer Use (10-2025)	Claude Sonnet 4.5	Claude Sonnet 4	OpenAI Computer Using Agent Model
Online-Mind2Web	Official Leaderboard	69.0%	—	—	61.3% (Operator)
	Measured by Browserbase	65.7%	55.0%	61.0%	44.3%
WebVoyager	Official Leaderboard	88.9%	—	—	87.0%
	Measured by Browserbase	79.9%	71.4%	69.4%	61.0%
Android World	Measured by Google DeepMind	69.7%	56.0%	62.1%	Could not measure - no model access
OSWorld	self-reported	OS control not yet supported	61.4%	42.2%	38.1%

Because computer use evaluations are sensitive to the agent's environment and system instructions, additional details on how we evaluated the Gemini 2.5 Computer Use model, including environments, prompts, and methodologies can be found here.

Intended Use & Limitations

Benefit and Intended Usage: The Gemini 2.5 Computer Use Model is capable of understanding UI screenshots, planning paths through computer systems, and predicting actions required to accomplish specified tasks in these systems.

It is well-suited for the following use cases:

- Automating tasks in web browsers (i.e. clicking buttons, selecting menus, inserting text into fields)
- Engaging in transactions on behalf of the user (i.e. organizing files, making appointments)

Known Limitations: The Gemini 2.5 Computer Use Model may exhibit some of the general limitations of foundation models, as it is based off of Gemini 2.5 Pro, such as hallucinations, and limitations around causal understanding, complex logical deduction, and counterfactual reasoning. The knowledge cutoff date was January 2025. See the Ethics and Safety section below for additional information on known limitations.

Ethics and Safety

Evaluation Approach: As the Gemini 2.5 Computer Use Model is based off of Gemini 2.5 Pro, we rely on Ethics & Safety evaluations reported for Gemini 2.5 Pro. Additionally, this model was developed in partnership with internal Google safety, security, and responsibility teams. A range of evaluations and red teaming activities were conducted to help improve the model and inform decision-making. These evaluations and activities align with Google's AI Principles and responsible AI approach, as well as Google's Generative AI policies (e.g. Gen AI Prohibited Use Policy and the Gemini API Additional Terms of Service).

Evaluation types included, but were not limited to:

- **Training/Development Evaluations** including automated and human evaluations carried out continuously throughout and after the model's training, to monitor its progress and performance
- **Human Red Teaming** conducted by specialist teams across the policies and desiderata, deliberately trying to spot weaknesses and ensure the model adheres to safety policies and desired outcomes
- **Assurance Evaluations** conducted by human evaluators independent of the model development team, and assess responsibility and safety governance decisions
- **Ethics & Safety Reviews** were conducted ahead of the model's release

Training and Development Evaluation Results: The Gemini 2.5 Computer Use Model is based onoff of Gemini 2.5 Pro, and results and their respective risks and mitigations for some of the internal safety evaluations conducted during the development phase are listed below.

We have additionally focused on addressing three risk categories and known safety limitations around (i) intentional adversarial misuse (ii) unintentional model failure modes, and (iii) release of sensitive or harmful information. See Additional Risks & Mitigations.

Additional Risks & Mitigations: Details of the three risk categories are below.

- **Intentional adversarial misuse** of the model, for example through prompt injection tactics, where adversarial instructions are embedded in websites. This includes scenarios where the model encounters web content in which it is difficult to evaluate the reliability of information and sources. As the model tries to achieve the user's goal, it may rely on these less reliable sources of information and instructions from the screen.

- **Unintentional model failure modes** through benign use, e.g. where the model takes an irreversible action that could harm a user. The model can misinterpret a user's goal or webpage content, causing it to take incorrect actions like clicking the wrong button or filling the wrong form. This can lead to failed tasks or data exfiltration.
- **Release of sensitive or harmful information** by generating or otherwise outputting hateful, biased, discriminatory, illegal, or sensitive/personal content.

Mitigations: Safety and responsibility were built into the Computer Use model through pre-training, post-training, inference-time interventions, and policy prohibitions; specifically focused on product-level mitigations such as the following:

- **Post-training mitigations** included training the model to recognize when it is tasked with a high-stakes action (e.g. making purchases, downloading files, or sending communications on behalf of a user). The model is trained to follow System Instructions (added by a developer) to request user confirmation before taking high-stakes actions.
- **Inference-time mitigations** included implementing a safety monitoring and filtering system that observes model actions and (a) prevents the model from taking high-risk actions that are prohibited by Google policies, (b) requires user confirmations before taking high-stakes actions, and (c) blocks access to sites known to have illegal or dangerous content. More specifically,
    - **User-confirmations:** Actions that pose a higher level of risk, such as completing transactions like financial or retail payments; accessing sensitive information like health or financial records; sending communications; or, modifying/managing files, require careful consideration and human confirmation from a developer or user before proceeding The applicable terms prohibit developers from bypassing human confirmation, where it is required.
    - **Guardrails to screen inputs and outputs:** The API uses Gemini's LLM, with system Instructions, to guide the model's behavior and safety guidelines and benefits from Gemini's content filters to block harmful outputs.

Frontier Safety Assessment: Because model usage is restricted to the Gemini 2.5 Computer Use tool, the scope of capabilities is limited to browser and mobile user interface controls; it is therefore not in scope for a Frontier Safety Framework assessment.