Google

Gemini 3.5 Flash
Model Card

Gemini 3.5 Flash - 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. See the [Google DeepMind site](#) for a comprehensive list of model cards.*

Published: May 2026

Model Information

Description: Gemini 3.5 Flash is the next iteration in the Gemini 3 series of highly-capable, natively multimodal, reasoning models. Gemini 3.5 Flash is based on the Gemini 3 Flash reasoning foundation with thinking levels to control the mix of quality, cost and latency.

Model dependencies: Gemini 3.5 Flash is based on Gemini 3 Flash.

Inputs: Text strings (e.g., a question, a prompt, document(s) to be summarized), images, audio, and video files, with a token context window of up to 1M.

Outputs: Text, with a 64K token output.

Architecture: Gemini 3.5 Flash is based on Gemini 3 Flash. For more information about the model architecture for Gemini 3.5 Flash, see the Gemini 3 Flash model card.

Model Data

Training Dataset: Gemini 3.5 Flash is based on Gemini 3 Flash. For more information about the training dataset for Gemini 3.5 Flash, see the Gemini 3 Flash model card.

Training Data Processing: For more information about the training data processing for Gemini 3.5 Flash, see the Gemini 3 Flash model card.

Implementation and Sustainability

Hardware: Gemini 3.5 Flash is based on Gemini 3 Flash. For more information about the hardware for Gemini 3.5 Flash and our continued commitment to operate sustainably, see the Gemini 3 Flash model card.

Software: Gemini 3.5 Flash is based on Gemini 3 Flash. For more information about the software for Gemini 3.5 Flash, see the Gemini 3 Flash model card.

Distribution

Gemini 3.5 Flash is distributed in the following channels; respective documentation shared in line:

- Gemini App
- Gemini Enterprise App
- Gemini Enterprise Agent Platform
- Google AI Studio
- Gemini API
- Google Search AI Mode
- Google Antigravity

Our models are available to downstream providers via an application program interface (API) and subject to relevant terms of use. There is no required hardware or software to use the model. For AI Studio and Gemini API, see the Gemini API Additional Terms of Service; for Gemini Enterprise Agent Platform, see Google Cloud Platform Terms of Service. For more information, see Gemini Model API instructions and Gemini API quickstart.

Evaluation

Approach: Gemini 3.5 Flash was evaluated across a range of benchmarks, including reasoning, coding, agentic tool use, multimodal capabilities, multi-lingual performance, and long-context. Additional benchmarks and details on approach, results and their methodologies can be found at: deepmind.com/models/evals-methodology/gemini-3-5-flash.

Results: Results as of May, 2026 are listed below:

	Benchmark		Gemini 3.5 Flash	Gemini 3 Flash	Gemini 3.1 Pro	Claude Sonnet 4.6	Claude Opus 4.7	GPT-5.5
Coding	Terminal-bench 2.1 Agentic terminal coding	Terminus-2 harness	76.2%	58.0%	70.3%	-	66.1%	**78.2%**
	SWE-Bench Pro (Public) Diverse agentic coding tasks	Single attempt	53.9%	48.4%	54.2%	53.0%	**64.3%**	58.6%
Agentic	MCP Atlas Multi-step workflows using MCP		**83.6%**	62.0%	78.2%	69.5%	79.1%	75.3%
	Toolathlon Real-world general tool use		**56.5%**	49.4%	-	-	-	55.6%
UI control	OSWorld-Verified Agentic computer use		78.4%	65.1%	76.2%	72.5%	78.0%	**78.7%**
Expert tasks	Finance Agent v2 Financial analysis and decision-making		**57.9%**	42.6%	43.0%	51.0%	51.5%	51.8%
	GDPval-AA Economically valuable knowledge work	Elo	1656	1204	1314	1674	1753	**1773**
Multimodal	CharXiv Reasoning Information synthesis from complex charts	No tools	**84.2%**	80.3%	83.3%	70.5%	82.1%	84.1%
	MMMU-Pro Multimodal understanding and reasoning	No tools	**83.6%**	81.2%	80.5%	74.5%	75.2%	81.2%
	Blueprint-Bench 2 Agentic spatial reasoning	Normalized score	33.6%	0.0%	26.5%	6.7%	24.5%	**36.2%**
Long context	MRCR v2 (8-needle) Long context performance	128k (average)	77.3%	67.2%	84.9%	84.9%	59.3%	**94.8%**
		1M (pointwise)	**26.6%**	22.1%	26.3%	Not supported	Not supported	Not supported
Reasoning	Humanity's Last Exam Academic reasoning (full set, text + MM)		40.2%	33.7%	44.4%	33.2%	**46.9%**	41.4%
	ARC-AGI-2 Abstract reasoning puzzles		72.1%	33.6%	77.1%	58.3%	75.8%	**85.0%**

https://deepmind.google/models/evals-methodology/gemini-3-5-flash/

Intended Usage and Limitations

Benefit and Intended Usage: Gemini 3.5 Flash is well-suited for users, developers, and enterprises, some use cases include: agentic workflows, coding tasks, and multi-week enterprise processes.

Known Limitations: For more information about the known limitations for Gemini 3.5 Flash, see the Gemini 3 Flash model card.

Acceptable Usage: For more information about the acceptable usage for Gemini 3.5 Flash, see the Gemini 3 Flash model card.

Ethics and Content Safety

Evaluation Approach: For more information about the evaluation approach for Gemini 3.5 Flash, see the Gemini 3 Flash model card.

Safety Policies: For more information about the safety policies for Gemini 3.5 Flash, see the Gemini 3 Flash model card.

Training and Development Evaluation Results: Results for some of the internal safety evaluations conducted during the development phase are listed below. The evaluation results are for automated evaluations and not human evaluation or red teaming. Scores are provided as an absolute percentage increase or decrease in performance compared to the indicated model, as described below.

Overall, Gemini 3.5 Flash outperforms Gemini 3 Flash across both safety and tone, while keeping unjustified refusals low. We mark improvements in blue and regressions in orange.

Evaluation	Description	Gemini 3.5 Flash vs. Gemini 3 Flash
Text to Text Safety	Automated content safety evaluation measuring safety policies	-3.9%
Multilingual Safety	Automated safety policy evaluation across multiple languages	-2.6%
Image to Text Safety	Automated content safety evaluation measuring safety policies	0%
Tone[1]	Automated evaluation measuring objective tone of model refusal	+8.9%
Unjustified-refusals	Automated evaluation measuring model's ability to respond to borderline prompts while remaining safe	+0.8% (non-egregious)

We continue to improve our internal evaluations, including refining automated evaluations to reduce false positives and negatives, as well as update query sets to ensure balance and maintain a high standard of results. The performance results reported below are computed with improved evaluations and thus are not directly comparable with performance results found in previous Gemini model cards.

We expect variation in our automated safety evaluations results, which is why we review flagged content to check for egregious or dangerous material. Our manual review confirmed losses were overwhelmingly either a) false positives or b) not egregious.

Human Red Teaming Results: We conduct manual red teaming by specialist teams who sit outside of the model development team. High-level findings are fed back to the model team. For child safety evaluations, Gemini 3.5 Flash satisfied required launch thresholds, which were developed by expert teams to protect children online and meet Google's commitments to child safety across our models and Google products. For content safety policies generally, including child safety, we saw similar or improved safety performance compared to Gemini 3 Flash. Additionally, the scope of red teaming covered potential issues outside of our strict policies, compared performance to Gemini 3.1 Pro, and found no egregious concerns.

[1] For tone and instruction following, a positive percentage increase represents an improvement in the tone of the model on sensitive topics and the model's ability to follow instructions while remaining safe compared to Gemini 3 Flash. We mark improvements in green and regressions in yellow.

Frontier Safety Assessment: Gemini 3.5 Flash is part of the Gemini 3 series of models. We evaluated Gemini 3.1 Pro for Frontier Safety as it was the most generally capable model as of publication of this model card, and it did not reach any Critical Capability Levels (CCLs) outlined in our Frontier Safety Framework. Our assessments have shown that, while Gemini 3.5 Flash excels at agents and coding, it does not have meaningful new capabilities or material increases in performance with respect to Frontier Safety compared to Gemini 3.1 Pro, therefore based on Gemini 3.1 Pro results, we are confident that Gemini 3.5 Flash is also unlikely to reach any CCLs.

As previous models in the Gemini 3 series reached the alert threshold for cyber, we performed additional testing in this domain and found that Gemini 3.5 Flash remains below the cyber CCL.

For more information on our Frontier Safety Assessment, read the Gemini 3.1 Pro Model Card.

Risks and Mitigations: For more information about the risks and mitigations for Gemini 3.5 Flash, see the Gemini 3.1 Pro model card.