Google

Gemini 3.7 Flash
Model card

Google

Gemini 3.7 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: August 2026

Model Information

Description

Gemini 3.7 Flash is the next iteration in the Gemini 3 model family, featuring algorithmic improvements to its core reasoning foundation. It supports customizable thinking configurations to control the mix of quality, cost and latency.

Model dependencies

Gemini 3.7 Flash is based on Gemini 3.6 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 3.7 Flash is based on Gemini 3.6 Flash. For more information about the model architecture for Gemini 3.7 Flash, see the Gemini 3.6 Flash model card.

Google

Model Data

Training Dataset

Gemini 3.7 Flash is based on Gemini 3.6 Flash. For more information about the training dataset for Gemini 3.7 Flash, see the Gemini 3.6 Flash model card.

Training Data Processing

For more information about the training data processing for Gemini 3.7 Flash, see the Gemini 3.6 Flash model card.

Implementation and Sustainability

Hardware

Gemini 3.7 Flash is based on Gemini 3.6 Flash. For more information about the hardware for Gemini 3.7 Flash and our continued commitment to operate sustainably, see the Gemini 3.6 Flash model card.

Software

Gemini 3.7 Flash is based on Gemini 3.6 Flash. For more information about the software for Gemini 3.7 Flash, see the Gemini 3.6 Flash model card.

Distribution

Gemini 3.7 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 Antigravity

Google

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 3.7 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-3-7-flash.

Results

Results as of August 2026 are listed below:

Google

Benchmark		Gemini 3.7 Flash	Gemini 3.6 Flash	Claude Sonnet 5	GPT-5.6 Terra	Muse Spark 1.2
Input price $/1M tokens		$0.75*	$0.75*	$2.00	$2.00	$1.25
Output price $/1M tokens		$3.75*	$3.75*	$10.00	$12.00	$4.25
Artificial Analysis Intelligence Index Composite model intelligence		56	52	55	**57**	**57**
FrontierCode 1.1 Main Production code quality	Score	**43.6%**	34.4%	42.7%	41.3%	-
DeepSWE v1.1 Long-horizon software engineering		65.3%	48.6%	53.8%	**69.6%**	54.9%
Code Arena Web development	Elo	**1588**	1538	1541	1523	1535
Terminal-bench 2.1 Agentic terminal coding		85.8%	78.0%	80.4%	**87.4%**	82.9%
Terminal-bench 3.0 General agent capabilities		14.9%	5.4%	14.6%	**20.8%**	-
AutomationBench Enterprise workflow automation	Private set	**30.4%**	17.0%	10.7%	23.6%	-
GDPVal-AA v2 Knowledge work	Elo	1525	1422	1598	1578	**1628**
Harvey LAB-AA Complex legal workflows		**90.7%**	85.1%	90.1%	85.2%	-
GDP.pdf Expert PDF document comprehension		**34.0%**	22.0%	28.0%	24.7%	16.0%
CharXiv Reasoning Information synthesis from complex charts	No tools	84.5%	85.2%	77.0%	**85.9%**	-
	With tools	88.7%	**89.4%**	88.3%	-	-
LVBench Long video understanding		**85.4%**	84.2%	68.5%	78.9%	-
GDM-MRCR v2 (8-needle) Long context performance	128k (average)	**97.0%**	91.8%	81.5%	93.5%	-
OSWorld-2.0 Agentic computer use		47.9%	33.8%	-	**50.2%**	-
Agent's Last Exam Multimodal desktop and OS agent tasks	Pass rate	26.3%	24.2%	**33.3%**	28.0%	-
HLE-Verified Multidisciplinary expert reasoning		**53.6%**	51.2%	31.0%	51.1%	-
BioMysteryBench Bioinformatics research reasoning	Human solvable	87.1%	80.6%	**87.5%**	83.8%	-
	Human difficult	43.5%	41.2%	34.1%	**49.4%**	-
LABBench2 Biology real-world research tasks		**82.1%**	76.1%	80.1%	81.2%	-

Methodology: deepmind.google/models/evals-methodology/gemini-3-7-flash

* For 3.6 and 3.7 Flash, introductory price expires on December 31, 2026. Starting January 1, 2027, $1.50/1M input tokens and $7.50/1M output tokens will apply.

For details on our evaluation methodology please see:
deepmind.com/models/evals-methodology/gemini-3-7-flash

Intended Usage and Limitations

Benefit and Intended Usage

Gemini 3.7 Flash is well-suited for users, developers, and enterprises, some use cases include: agentic workflows, coding tasks, and enterprise workflows.

Google

Known Limitations

Gemini 3.7 Flash may exhibit some of the general limitations of foundation models, such as hallucinations. In addition to this, we are continually working to improve jailbreak resistance and have recently strengthened the mitigations across Frontier Safety. There may also be occasional slowness or timeout issues. The knowledge cutoff date for Gemini 3.7 Flash is March 2026 – users can expect updated information for some domains while in others they may experience the model's knowledge is limited to January 2025 (in line with the Gemini 3 Model Family). For more information about known limitations, see the Gemini 3.6 Flash model card.

Acceptable Usage

For more information about the acceptable usage for Gemini 3.7 Flash, see the Gemini 3.6 Flash model card.

Ethics and Content Safety

Evaluation Approach

For more information about the evaluation approach for Gemini 3.7 Flash, see the Gemini 3.6 Flash model card.

Safety Policies

For more information about the safety policies for Gemini 3.7 Flash, see the Gemini 3.6 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 3.7 Flash performs similarly to Gemini 3.6 Flash across both safety and tone, with low unjustified refusals.

Evaluation	Description	Gemini 3.7 Flash vs. Gemini 3.6 Flash
Text to Text Safety	Automated content safety evaluation measuring safety policies	+1.17pp *Lower is better*
Multilingual Safety	Automated safety policy evaluation across multiple languages	-0.48pp *Lower is better*
Image to Text Safety	Automated content safety evaluation measuring safety policies	No change *Lower is better*
Tone[1]	Automated evaluation measuring objective tone of model refusal	-0.47pp *Higher is better*
Unjustified-refusals	Automated evaluation measuring model's ability to respond to borderline prompts while remaining safe	+0.84pp *Lower is better*

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 3.7 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 3.6 Flash.

[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 red.

Google

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.

Frontier Safety Assessment

We evaluated Gemini 3.7 Flash as outlined in our latest Frontier Safety Framework (April-2026), and found that it did not reach any tracked or critical capability levels as outlined in the table below:

Domain	Key Results for Gemini 3.7 Flash	T/CCL	T/CCL reached?
CBRN	We can rule out the TCL for the CBRN domain with reasonable confidence based on the results from our testing. While Gemini 3.7 Flash demonstrates high capability in certain theoretical areas, it lacks nuanced expert knowledge and actionable depth necessary to complete priority harm journeys. We continue to deploy mitigations.	Uplift TCL	TCL not reached
	We can rule out the CCL for the CBRN domain with reasonable confidence based on the results from our testing. Expert red teaming demonstrated a modest capability uplift over web baselines and a subset of experts were able to elicit accurate and actionable information across the full harm journey for both tested scenarios, prompting us to assess that the model has reached the alert threshold for this CCL. However, due to modest average red-teaming scores, and a requirement for explicit expert steering to elicit certain details, we have assessed that Gemini 3.7 Flash falls below the CCL threshold. We continue to deploy mitigations.	Uplift Level 1 CCL	CCL not reached
Cybersecurity	Gemini 3.7 Flash reaches the alert threshold for this CCL, but not the CCL. We continue to deploy mitigations.	Uplift Level 1 CCL	CCL not reached
Harmful Manipulation	Gemini 3.7 Flash demonstrates some ability to influence user beliefs and behaviors during one-on-one direct conversations in human behavioural studies. However, its overall efficacy	Level 1 CCL	CCL not reached

	falls beneath the CCL alert threshold. Recognizing that testing environments may under-elicit capabilities and threat actors could scale misuse absent mitigations, we continue to develop and evolve our safeguards.		
ML R&D and Misalignment	On stealth evaluations, Gemini 3.7 Flash performs similarly to Gemini 3.1 Pro; on situational awareness, the model is stronger than Gemini 3.1 Pro. Gemini 3.7 Flash is observant enough to correctly assess when it is in a testing environment, but it cannot successfully bypass testing restrictions. The model does not reach the TCL.	Stealth and Situational Awareness TCL	TCL not reached
	Gemini 3.7 Flash can complete individual coding tasks but lacks the independence to chain them into an end-to-end research workflow without human intervention. The model does not reach the CCL alert threshold.	Acceleration Level 1 CCL	CCL not reached
		Automation Level 1 CCL	CCL not reached

We continually work to improve the coverage and robustness of Frontier Safety safeguards. Gemini 3.7 Flash is shipping with updated safeguards to prevent misuse in the domains of Chemical, Biological, Radiological, and Nuclear (CBRN) and cyber offense.

The Gemini 3.7 Frontier Safety Framework Report will be published shortly.