

# How to use Gemini 1.5 Flash for free

At I/O 2024, Google announced several new AI models, upcoming projects, and countless AI features coming to its products. However, what is worth noting is the Gemini 1.5 Flash model.

This is an impressively fast and efficient model, offering multimodal capabilities and a context span of up to 1 million tokens (2 million via waitlist).

Although Gemini 1.5 Flash is small - Google hasn't revealed its size specifications - it scores high across all measures - text, image and audio. In the Gemini 1.5 technical report, Google revealed that the Gemini 1.5 Flash is far superior to larger models like the 1.0 Ultra and 1.0 Pro in many ways. Only in speech recognition and translation does it lag behind larger models.

Core Capability	1.5 Pro Relative to			1.5 Flash Relative to		
	1.5 Pro (Feb)	1.0 Pro	1.0 Ultra	1.0 Pro	1.0 Ultra	
Text	Math, Science & Reasoning	+5.9%	+49.6%	+18.1%	+30.8%	+4.1%
	Multilinguality	-0.7%	+21.4%	+5.9%	+16.7%	+2.1%
	Coding	+11.6%	+21.5%	+11.7%	+10.3%	+1.5%
	Instruction following	—	+9.9%	-0.2%	+8.7%	-1.2%
	Function calling	—	+72.8%	—	+54.6%	—
Vision	Multimodal reasoning	+15.5%	+31.5%	+14.8%	+15.6%	+1.0%
	Charts & Documents	+8.8%	+63.9%	+39.6%	+35.9%	+17.9%
	Natural images	+8.3%	+21.7%	+8.1%	+18.9%	+5.6%
	Video understanding	-0.3%	+18.7%	+2.1%	+7.5%	-8.1%
Audio	Speech recognition	+1.0%	+2.2%	-3.8%	-17.9%	-25.5%
	Speech translation	-1.7%	-1.5%	-3.9%	-9.8%	-11.9%

Table 10 | Detailed breakdown of the results presented in Table 1.

Unlike the Gemini 1.5 Pro which is a MoE (Mixture of Experts) model, the Gemini 1.5 Flash is a dense model, distilled online from the larger 1.5 Pro model for improved quality. In terms of speed, the Flash model outperforms all the smaller models available, including the Claude 3 Haiku, which runs on Google's custom TPU.

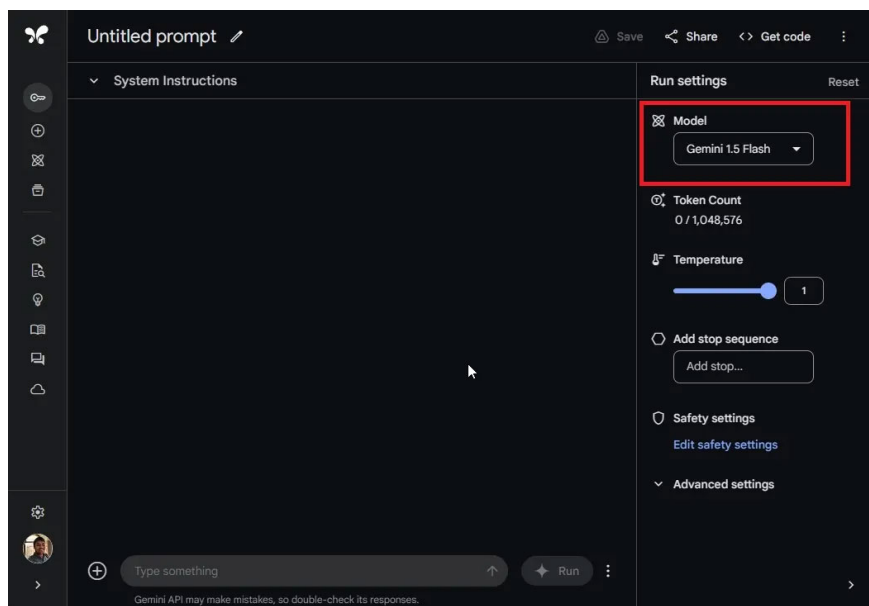
Language	Gemini 1.5 Flash	Gemini 1.5 Pro	GPT-3.5 Turbo	GPT-4 Turbo	Claude 3 Haiku	Claude 3 Sonnet	Claude 3 Opus
English	1.5	4.3	2.6	6.8	2.2	6.2	10.5
Japanese	4.3	10.9	12.9	35.4	10.7	23.9	46.6
Chinese	5.2	14.1	18.4	53.4	12.8	30	55.3
French	1.9	4.7	4.1	10.7	3.0	6.8	13.4

And its price is unbelievably low. Gemini 1.5 Flash costs \$0.35 for input and \$0.53 for output to process 128K tokens, \$0.70 and \$1.05 for 1 million tokens. It is much cheaper than the Llama 3 70B, Mistral Medium, GPT-3.5 Turbo and of course the larger models.

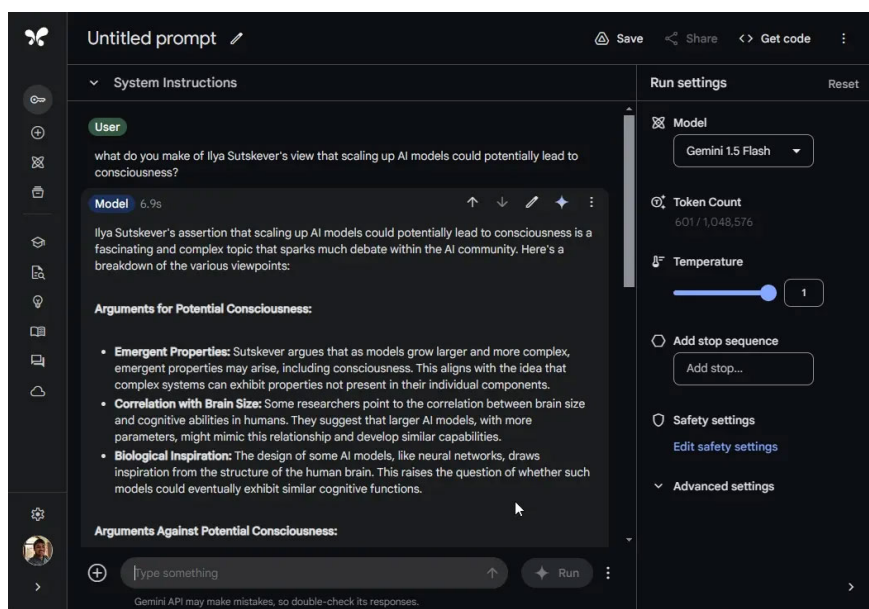
If you're a developer and need multimodal inference with a larger context window on the cheap, you should definitely take a look at the Flash model. Here's how you can try Gemini 1.5 Flash for free.

## How to use Flash Gemini 1.5 for free

1. Go to [aistudio.google.com](https://aistudio.google.com) and sign in with your Google account. There is no waiting list to use the Flash model.
2. Next, select the '**Gemini 1.5 Flash**' model in the drop-down menu.

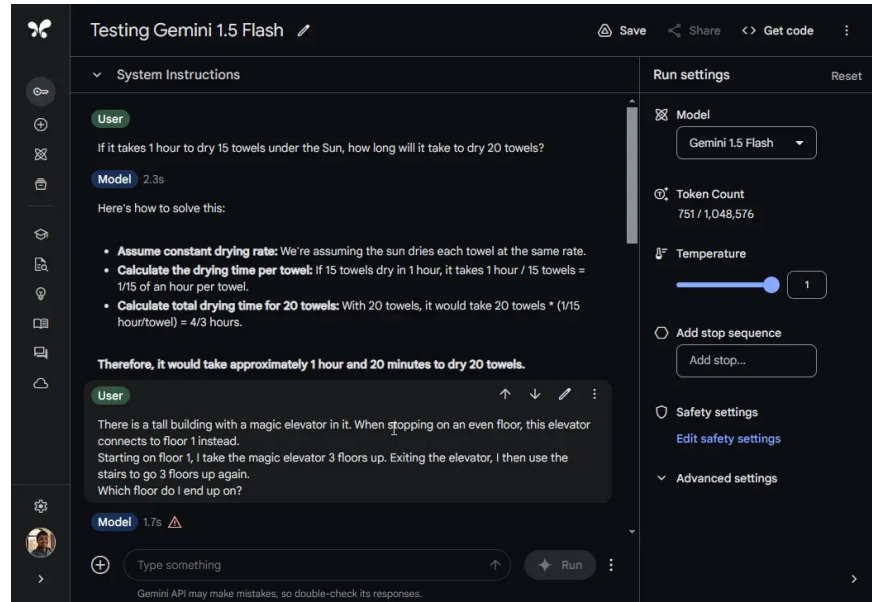


3. Now you can start chatting with the Flash model. You can also upload images, videos, audio clips, files and folders.



# First impressions of Gemini 1.5 Flash

Although Gemini 1.5 Flash is not a modern model, its advantages are breakneck speed, efficiency and low cost. It ranks below Gemini 1.5 Pro in capabilities and other larger models from OpenAI and Anthropic. However, the author has tried some of the theoretical suggestions used to compare ChatGPT 4o and Gemini 1.5 Pro.



It can only generate one correct answer out of 5 questions. It may not be very smart in terms of common sense, but for other applications that require multimodal capabilities and large context windows, it may be suitable for your use case. Additionally, the Gemini model is good at creative tasks that can bring value to developers and users.

Simply put, there is no AI model that is fast, efficient, multimodal, and has a large context window with near-perfect capabilities. Best of all, it's incredibly cheap.

What is your opinion on Google's latest Flash model? Share your opinions in the comments section below!

You finished reading the article "**How to use Gemini 1.5 Flash for free**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.