

A Deep Dive into Gemini

Christina Warren
DevRel at Google DeepMind
filmgirl@google.com
[@film_girl](https://twitter.com/film_girl)



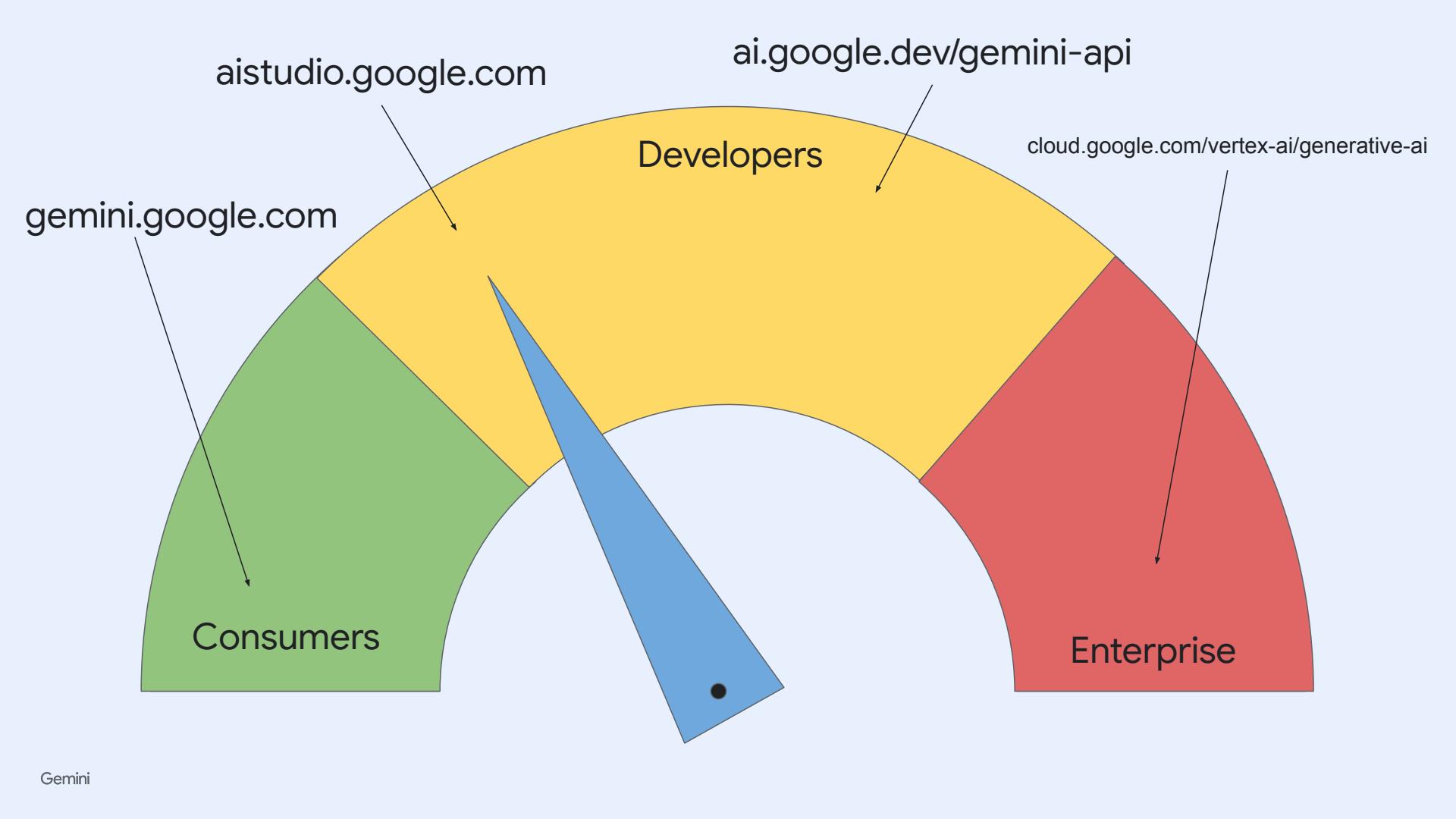
We have release notes!!!

goo.gle/aistudio-changelog

goo.gle/gemini-api-changelog

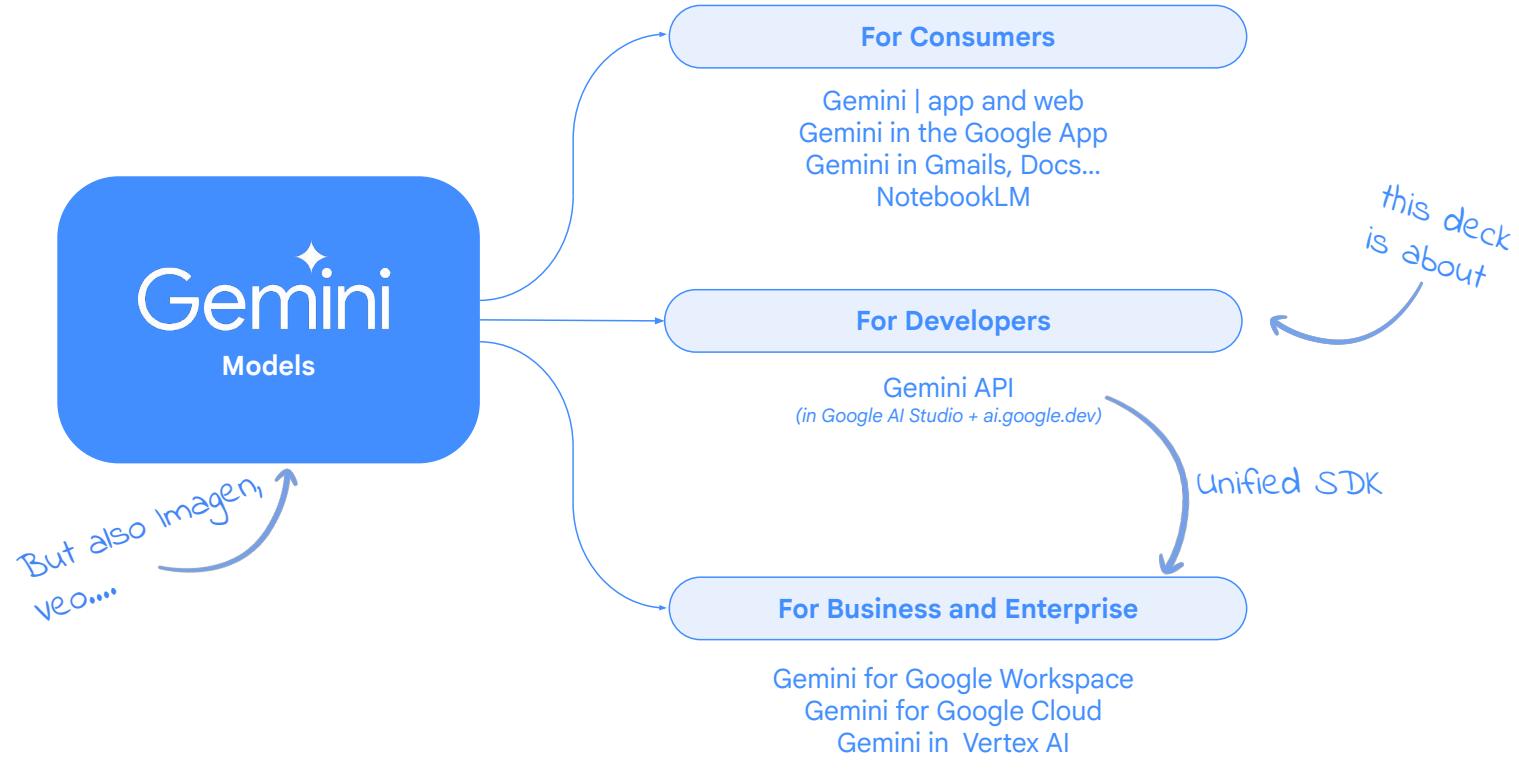
What is the Gemini API & Google AI Studio?





The Gemini Ecosystem

The most advanced AI from Google

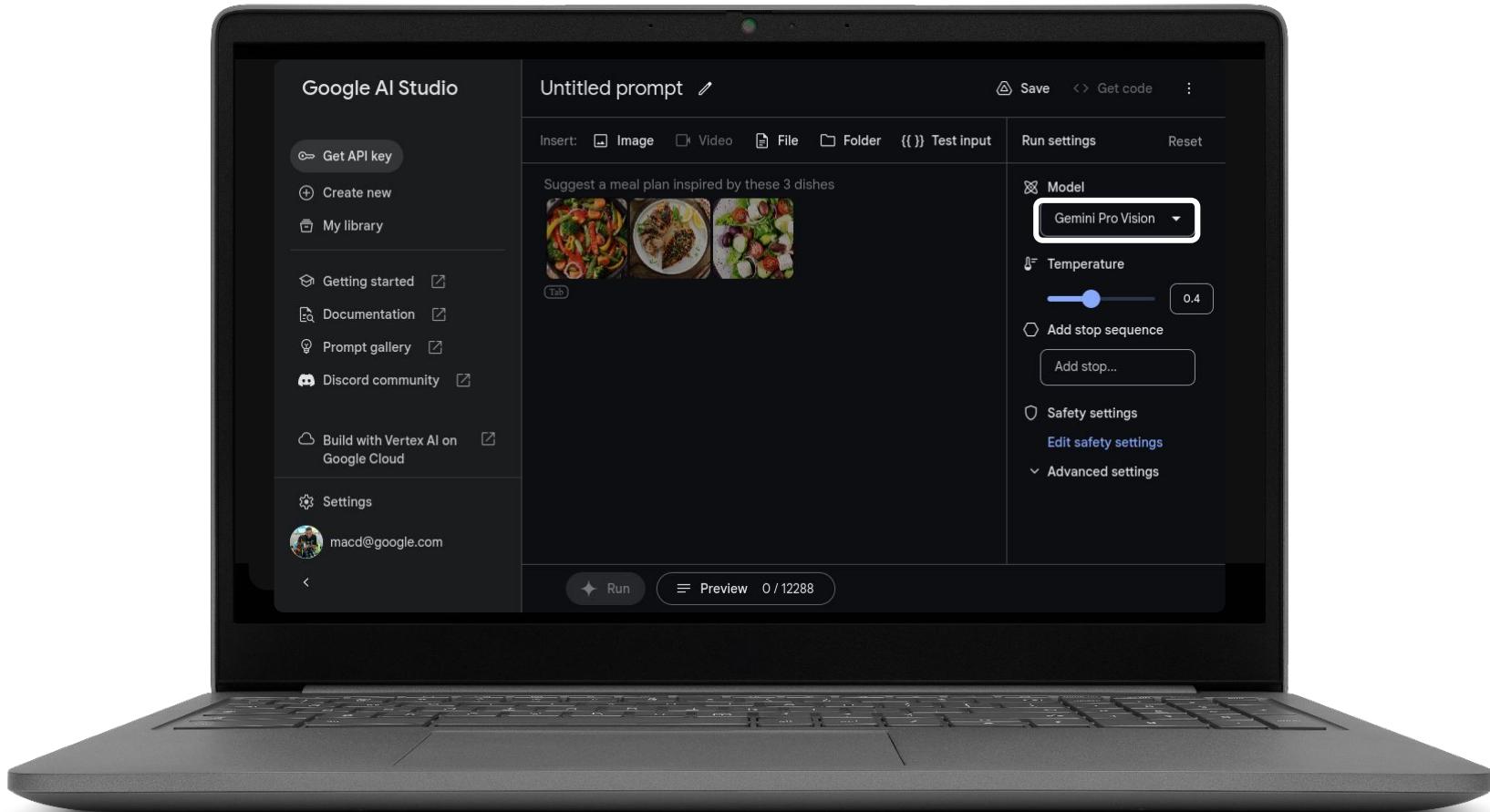


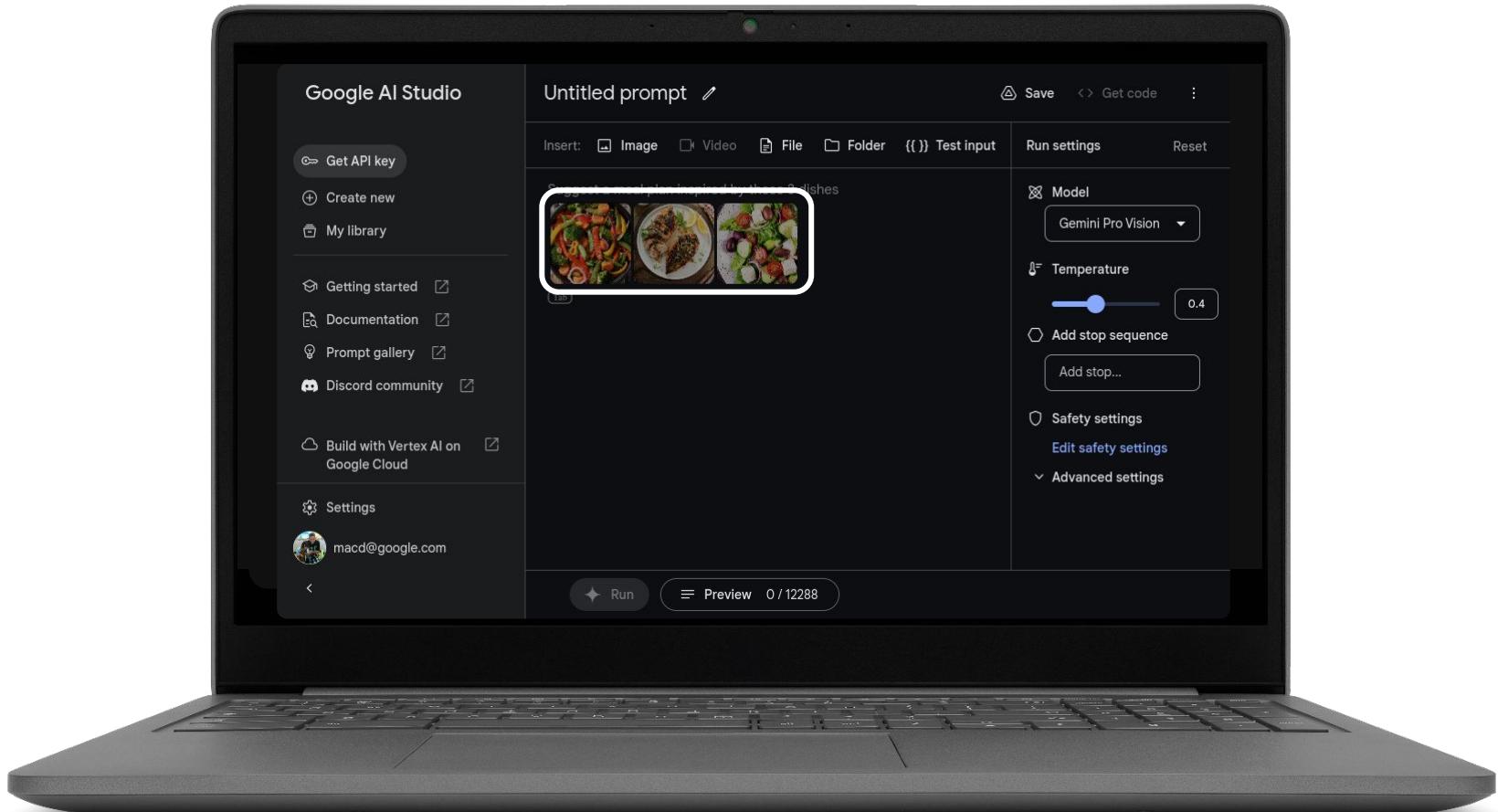
Google AI Studio

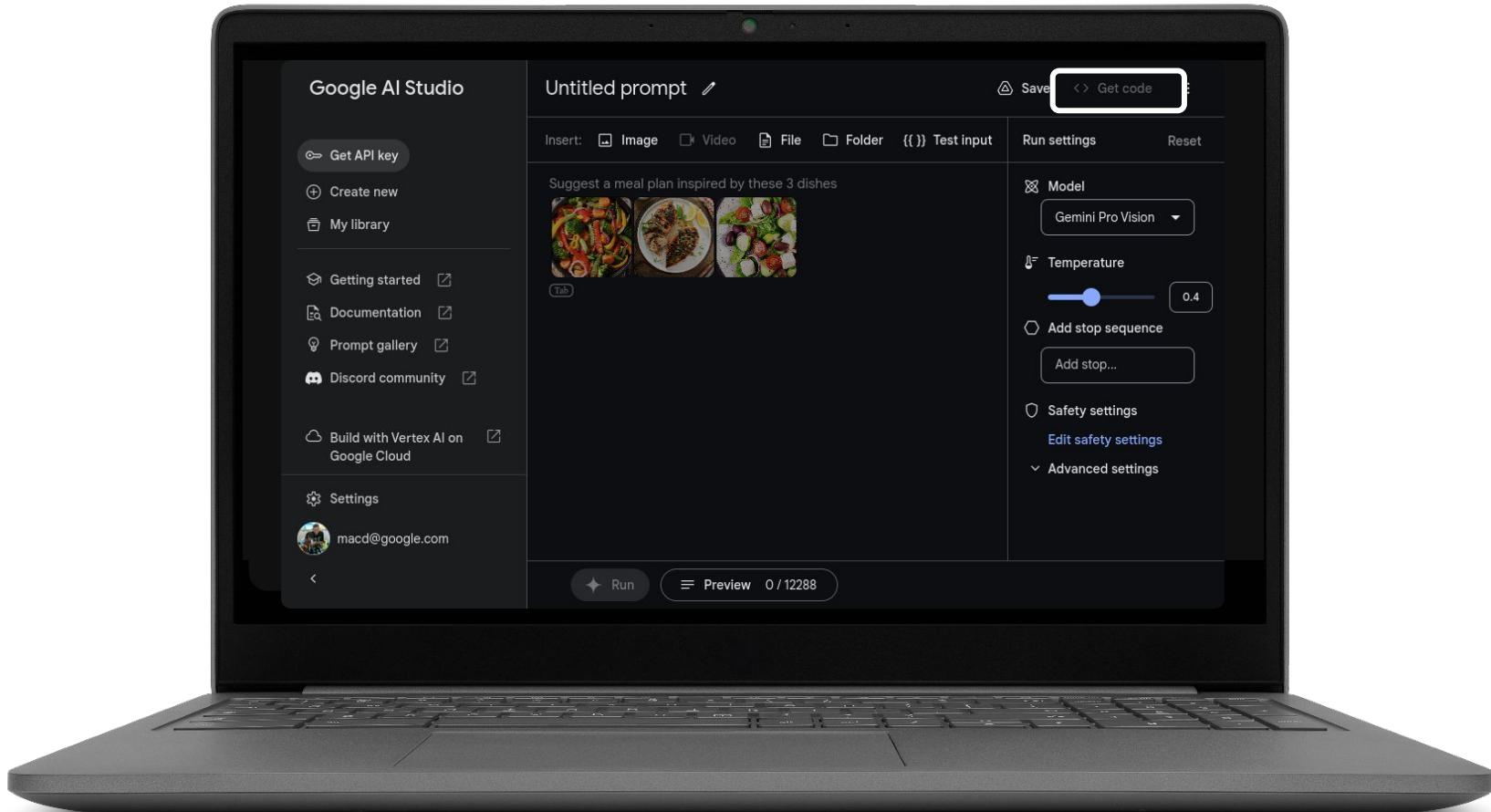
aistudio.google.com

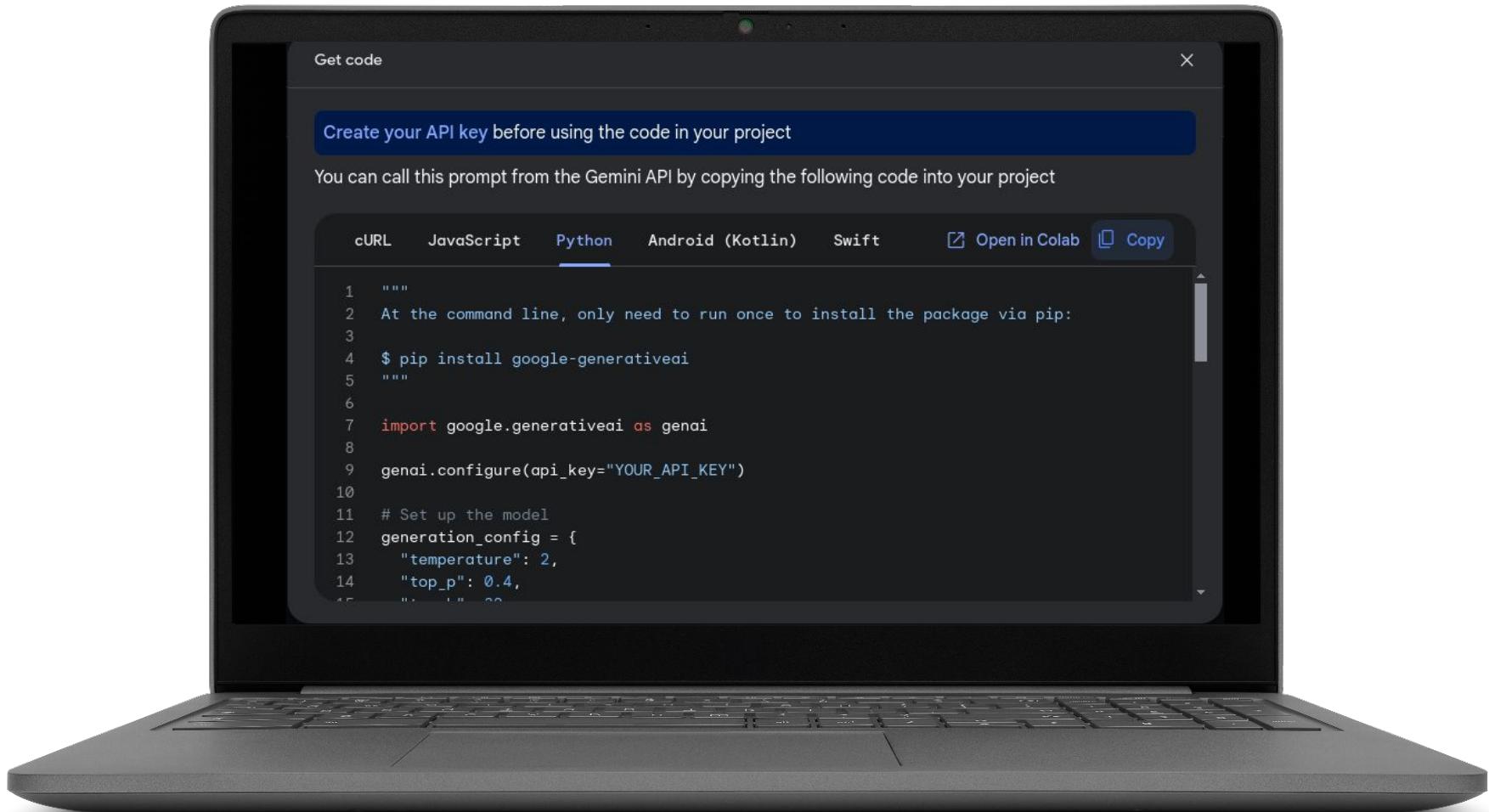
- ◆ Simplest way to test the latest Gemini models
- ◆ No Google Cloud knowledge or set-up required (all automatic)
- ◆ **Totally free** (including fine-tuning)
- ◆ Create, test, and save prompts
- ◆ Generate starter code
- ◆ Multimodal Live API demo with camera and screen sharing
- ◆ Starter apps: spatial/video understanding, maps explorer

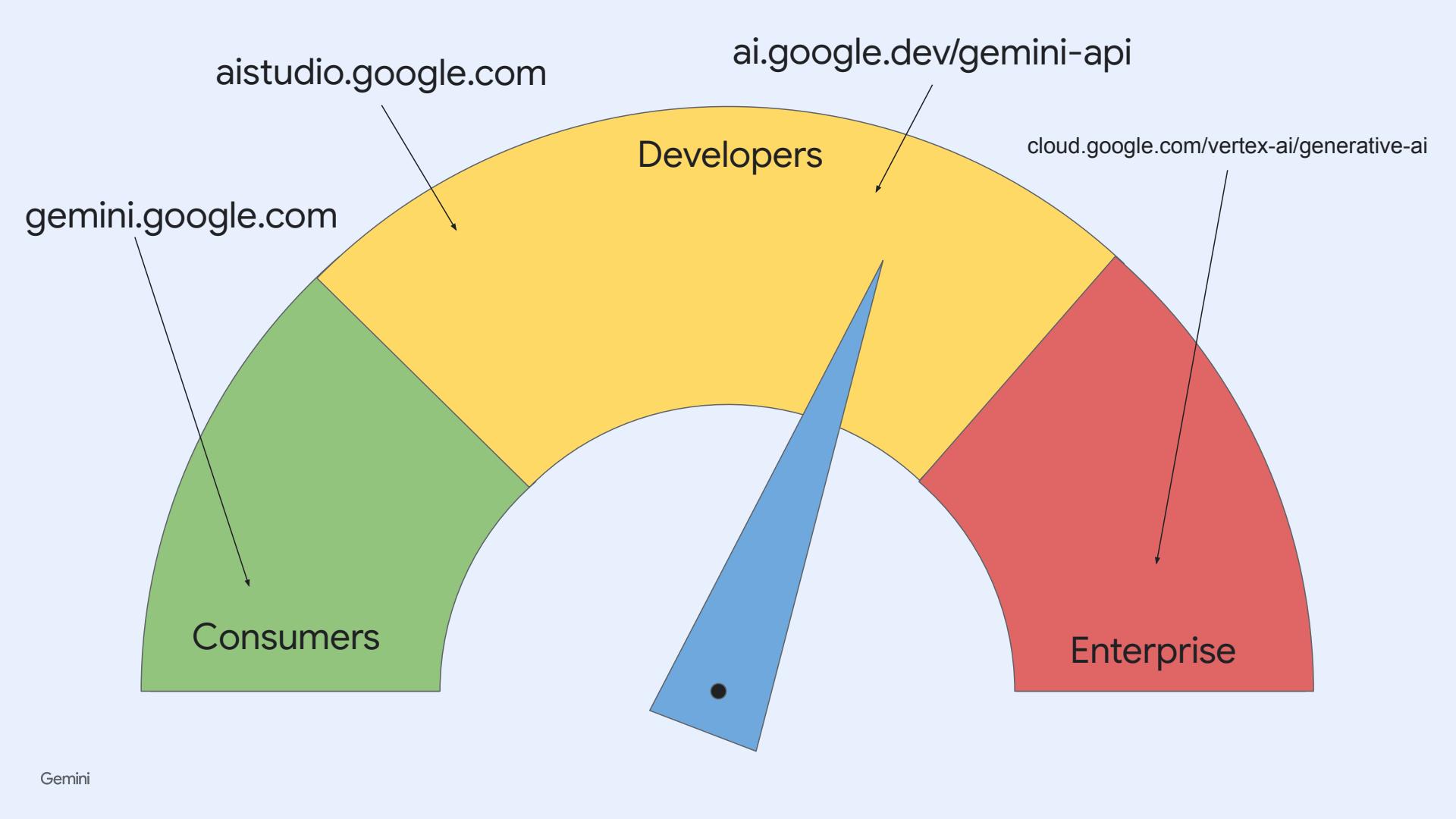












Gemini APIs

ai.google.dev/gemini-api

- ❖ Simplest way to use the Gemini models
- ❖ Easy auth with API keys
- ❖ No GCP set-up required
- ❖ Generous free tier

New unified SDK

Use the same SDK for Vertex AI and Gemini API



ai.google.dev/gemini-api/docs/sdks

66

It takes a minute to
create an API key, no
credit card nor GCP
knowledge required

99

aistudio.google.com/app/apikey



Get API key

API keys

You can create a new project if you don't have one already. By creating a project, you agree to the Google Cloud Platform Terms of Service, which you agree to when creating a project. You also agree to the Gemini API Terms of Service.

Use your API keys securely. Do not share them or embed them in your code. See the [Gemini API Terms of Service](#) for more information.

If you use Gemini API from a project that has billing enabled, your use will be subject to [pay-as-you-go pricing](#).

[Create API key](#)

Now
automatically
sent to Colab!

subject to the Google Cloud
Platform Terms of Service. Gemini Studio is subject to the
Gemini API Terms of Service.

Your API keys are listed below. You can also view and manage your project and API keys in Google Cloud.

Project number	Project ID	API key	Created	Plan	
...1234	your GCP project [edit]	...2qZ9	Sep 6, 2024	Paid	Go to billing
		...p3cd	Sep 4, 2024		View usage data
		...q5sd	Aug 21, 2024		
		...yU4c	Aug 20, 2024		
		...DsqE	Aug 5, 2024		
...9876	Another GCP project [edit]	...UmOU	Sep 6, 2024	Set up Billing	View usage data
		...zE1A	Sep 6, 2024		
		...vFO6	Sep 6, 2024		
		...vAkQ	Sep 6, 2024		
		...h3CU	Sep 4, 2024		
		...8Img	Sep 4, 2024		
		...DyVk	Sep 4, 2024		
		...cl0	Sep 3, 2024	Proprietary + Confidential	View usage data
		...L3Ns	Sep 3, 2024		

Welcome to the Gemini API Cookbook

This cookbook provides a structured learning path for using the Gemini API, focusing on hands-on tutorials and practical examples.

For comprehensive API documentation, visit ai.google.dev.

Navigating the Cookbook

This cookbook is organized into two main categories:

1. Quick Starts: Step-by-step guides covering both introductory topics ("Get Started") and specific API features.
2. Examples: Practical use cases demonstrating how to combine multiple features.

We also showcase **Demos** in separate repositories, illustrating end-to-end applications of the Gemini API.

What's New?

Here are the recent additions and updates to the Gemini API and the Cookbook:

- **Gemini 2.0 Flash experimental**: Explore the capabilities of the latest Gemini 2.0 Flash experimental model! See the [Get Started Guide](#).
- **Recently Added Guides**:
 - [Thinking model](#): Discover the thinking model capabilities.
 - [Gradio and Live API](#): Start working with Gradio to set-up your own live instance.

1. Quick Starts

The [quickstarts section](#) contains step-by-step tutorials to get you started with Gemini and learn about its specific features.

To begin, you'll need:

1. A Google account.
2. An API key (create one in [Google AI Studio](#)).

We recommend starting with the following:

- Authentication: Set up your API key for access.
- Get started: Get started with Gemini models and the Gemini API, covering basic prompting and multimodal input.

66

Play with the Gemini API
in 2 clicks.

99

goo.gle/cookbook



What's new?



What's new?

- ❖ [Dec 2024] Search Grounding
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Grounding with Google Search

goo.gle/gemini-api-cookbook-grounding

- ◆ Models have knowledge cut-offs
- ◆ Grounding with Google Search provides up-to-date information
- ◆ Gives Gemini the ability to browse the web and cite sources
- ◆ Prevents the model from hallucinating by searching for the relevant info
- ◆ 500 free RPD (requests-per-day) with Flash 2.0

```
response = client.models.generate_content(  
    model="gemini-2.0-flash",  
    contents="Who is Christina Warren?",  
    config = {  
        'tools': [{  
            'google_search': {}  
        }]  
    }  
)  
  
print(response.text)  
  
# Christina Warren is an American developer advocate, podcaster,  
and writer. She has worked for USA Today, The Unofficial Apple  
Weblog, DownloadSquad, Mashable, Gizmodo, Microsoft, and GitHub.  
She is also a co-host of the podcast "Rocket" on the Relay FM  
podcast network.
```

What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Gemini

New Models | Improved Capabilities | Efficient Inference

Gemini 2.0 Flash

Natively Multimodal
Decreased Latency
Highly Capable
Audio-Out Streaming
(public preview)
Built for Agents

Generally Available

Gemini 2.0 Flash-Lite

Built for high volume tasks where cost and latency matter
Eg. chat, video analysis, extracting content and data from long-form documents, and more
1M context window & multimodal support as 1.5 Pro (which has 2M+ tokens context)

Generally Available

Gemini 2.5 Thinking & Pro

Our best models for scaling across a wide range of tasks
Orchestrate complex tasks involving higher-order reasoning, more nuance, depth, and understanding than Flash
Top-tier quality across a range of complex tasks.

up to 2M context window

Experimental

What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Multimodal Live API

- ◆ Talk directly with Gemini in **real time**

- ◆ Multimodal input: text, voice, camera, screen sharing...

- ◆ Limited to **10mn** at the moment

- ◆ Compatible with tools

- ◆ Websockets instead of REST API

- ◆ Limited experience in colab,
Use [py code](#) for full experience

AI Studio demo

goo.gle/gemini-api-cookbook-liveapi

Talk with Gemini live

Interact with Gemini using text, voice, video, or screen sharing.



Talk to Gemini

Start a real-time conversation using your microphone.



Show Gemini

Use your webcam to share what you're looking at and get real-time feedback.



Share your screen

Share your screen to show Gemini what you're working on.

Gemini 2.0 for games



Games

Clash of Clans



What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Unified SDK

ai.google.dev/gemini-api/docs/sdks

- ◆ Same SDK for both Gemini API and Vertex AI
- ◆ Just change the authentication
- ◆ Python is now stable
- ◆ GO and Java, recently released in preview, Typescript coming up
- ◆ Still, some minor differences

What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Spatial understanding

- ❖ Huge improvement compared to 1.5
- ❖ Can find 2D bounding boxes, 3D ones, point
- ❖ Currently not compatible with structured output, use default output (y,x)
- ❖ Mix it with reasoning (eg. list allergens)

goo.gle/gemini-2-spatial

AI Studio example



What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Multi-tools

goo.gle/gemini-api-cookbook-live-tools

- ❖ Welcome to the agentic world!
- ❖ Mix **multiple tools** (function calling, code execution, grounding...) together
- ❖ Automatic function calling not available yet with Live API, you have to handle the calls yourself
- ❖ Which other tools are you interested in?

What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Thinking model

Try in AI Studio

[goo.gle/gemini-2-thinking-examples](https://www.google.com/gemini-2-thinking-examples)

- ◆ Analyze the question and strategize before answering
- ◆ You can see the model “**thoughts**”
- ◆ But only in **AI Studio** since it’s mostly useful when deep-diving in a specific prompt

Deep-research

[Get started with thinking colab](#)

- ❖ Gemini agent for Gemini chat built using grounding and function calling
- ❖ Plan first
- ❖ Then search, reason and report

What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Code execution

goo.gle/gemini-api-cookbook-code-execution

- ◆ Generates and runs Python code to answer your queries.
 - No back and forth like with function calling.
- ◆ Available in both **AI Studio** and the **Gemini API**:
- ◆ Once activated, the model decides when/if to use it.
- ◆ Now able to **output graphs!**

```
response = client.models.generate_content(  
    model = "gemini-2.0-flash",  
    contents = "Can you draw a graph of the first 50 values of the  
fibonacci serie?",  
    config = types.GenerateContentConfig(  
        tools=[types.Tool(  
            code_execution=types.ToolCodeExecution  
        )]  
    )  
)
```

```
response = client.models.generate_content(
    model = "gemini-2.0-flash",
    contents = "Can you draw a graph of the first 50 values of the
fibonacci serie?",
    config = types.GenerateContentConfig(
        tools=[types.Tool(
            code_execution=types.ToolCodeExecution
        )]
    )
)

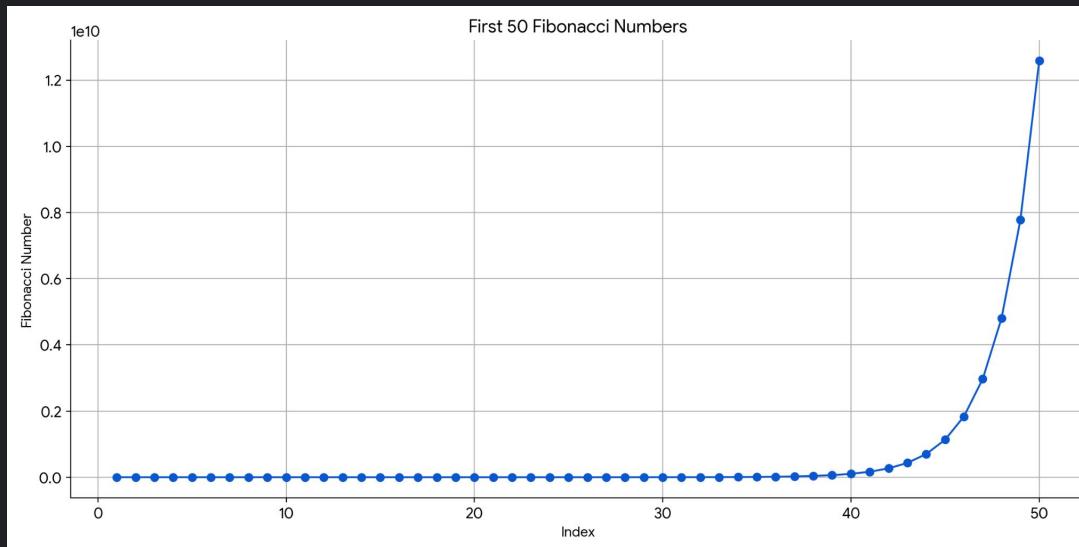
# Okay, I can help you with that. I'll generate the first 50 Fibonacci
numbers and then create a graph plotting the index (from 1 to 50)
against the Fibonacci number. I will use python to generate the numbers
and then display a graph.
```

```
import matplotlib.pyplot as plt
```

```
def fibonacci(n):
    fib_list = []
    a, b = 0, 1
    for _ in range(n):
        fib_list.append(b)
        a, b = b, a + b
    return fib_list
```

```
fib_numbers = fibonacci(50)
```

```
plt.figure(figsize=(12, 6))
plt.plot(range(1, 51), fib_numbers, marker='o', linestyle='--')
plt.title('First 50 Fibonacci Numbers')
plt.xlabel('Index')
plt.ylabel('Fibonacci Number')
plt.grid(True)
plt.yscale('linear')
plt.show()
```



What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Imagen

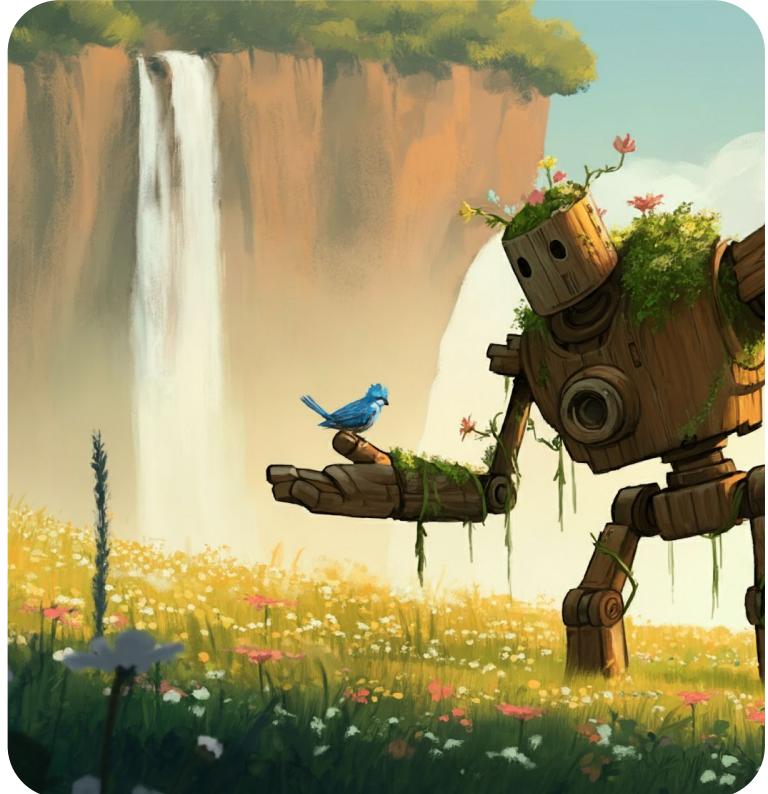
[goo.gle/gemini-api-cookbook-imagen](https://google.github.io/gemini-api-cookbook-imagen)

Prompting guide

- ◆ Our highest quality [text-to-image](#) model
- ◆ Generates stunning, detailed, artifact-free images.
- ◆ Understands complex prompts and diverse styles.
- ◆ Superior text rendering for creative uses.
- ◆ Built with robust safety and watermarking ([SynthID](#)).
- ◆ Paid tier only (0.03\$/image)

Imagen 3

A weathered, wooden mech robot covered in flowering vines stands peacefully in a field of tall wildflowers, with a small blue bird resting on its outstretched hand. Digital cartoon, with warm colors and soft lines. A large cliff with waterfall looms behind.





What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Gemini-embedding

goo.gle/gemini-api-cookbook-embeddings

Announcement

- ◆ New embedding model
- ◆ Multilingual embeddings
- ◆ Trained across diverse domains, including finance, science, legal, search, and more

What's new?

- ❖ [Dec 2024] First 2.0 model: Flash 2.0 experimental
- ❖ [Dec 2024] Multimodal Live API
- ❖ [Dec 2024] Unified SDK
- ❖ [Dec 2024] Spatial understanding
- ❖ [Dec 2024] Multi-tools
- ❖ [Dec 2024] Thinking model
- ❖ [Feb 2025] Code Execution
- ❖ [Feb 2025] Imagen
- ❖ [Mar 2025] New experimental embeddings model
- ❖ [Mar 2025] Gemini 2.5 Pro

Gemini 2.5 Pro

Announcement

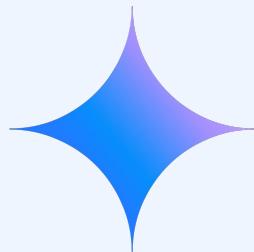
Benchmark		Gemini 2.5 Pro Experimental (03-25)	OpenAI o3-mini High	OpenAI GPT-4.5	Claude 3.7 Sonnet 64k Extended Thinking	Grok 3 Beta Extended Thinking	DeepSeek R1
Reasoning & knowledge							
Humanity's Last Exam (no tools)		18.8%	14.0%*	6.4%	8.9%	—	8.6%*
Science							
GPQA diamond	single attempt (pass@1)	84.0%	79.7%	71.4%	78.2%	80.2%	71.5%
	multiple attempts	—	—	—	84.8%	84.6%	—
Mathematics							
AIME 2025	single attempt (pass@1)	86.7%	86.5%	—	49.5%	77.3%	70.0%
	multiple attempts	—	—	—	—	93.3%	—
Mathematics							
AIME 2024	single attempt (pass@1)	92.0%	87.3%	36.7%	61.3%	83.9%	79.8%
	multiple attempts	—	—	—	80.0%	93.3%	—
Code generation							
LiveCodeBench v5	single attempt (pass@1)	70.4%	74.1%	—	—	70.6%	64.3%
	multiple attempts	—	—	—	—	79.4%	—
Code editing							
Aider Polyglot		74.0% / 68.6% whole / diff	60.4%	44.9%	64.9%	—	56.9%
			diff	diff	diff	—	diff

Prompt:

Make me a captivating endless runner game.
Key instructions on the screen. p5js scene,
no HTML. I like pixelated dinosaurs and
interesting backgrounds.]

What's new?

Veo



Audio-out

YouTube integration

Image-out





A cute black dog sits in a bathtub with its head covered in soap foam.
Symmetrical frame, blue tile background

Learn about AI from
the people building it.
[goo.gle/ai-podcast](https://goo.gl/ai-podcast)



aistudio.google.com

goo.gle/cookbook

goo.gle/gemini-api-eap-form

Thank you for listening.
Any questions?

Thank you for Joining Us

Tell us how we did!

Scan the code to take a 1 min
survey



[Vertex AI Free Trial](#)

