

What's new in the Gemini universe?

3rd generation, genMedia, vibe-coding and agents



Guillaume Vernade

Gemini Developer Advocate - Google Deepmind



Guillaume Vernade

Gemini Developer Advocate
Google Deepmind



+ Laurent Picard dans le chat

Models & Research

- LearnLM
- AlphaFold 3
- Gemini 1.5 Flash
- Imagen 3
- Gemma 2
- Veo
- Project Astra

2024

May — June — July — Aug — Sept — Oct — Nov — Dec

Gemini Live

NotebookLM
Audio Overviews

AI Overviews
VideoFX

- Genie 2
- Gencast
- Gemini 2.0 Flash
- Veo 2
- Gemini 2.0 Flash Thinking
- Project Mariner

2025

Jan — Feb — March — April — May

Gemini 2.0
Flash-Lite

Gemini 2.0 Pro

Weather
Next

Career
Dreamer

Veo 2 in
Dream Screen

AI Mode in
Search Labs

Live Video &
Screenshare
in Gemini App

Gemini Canvas

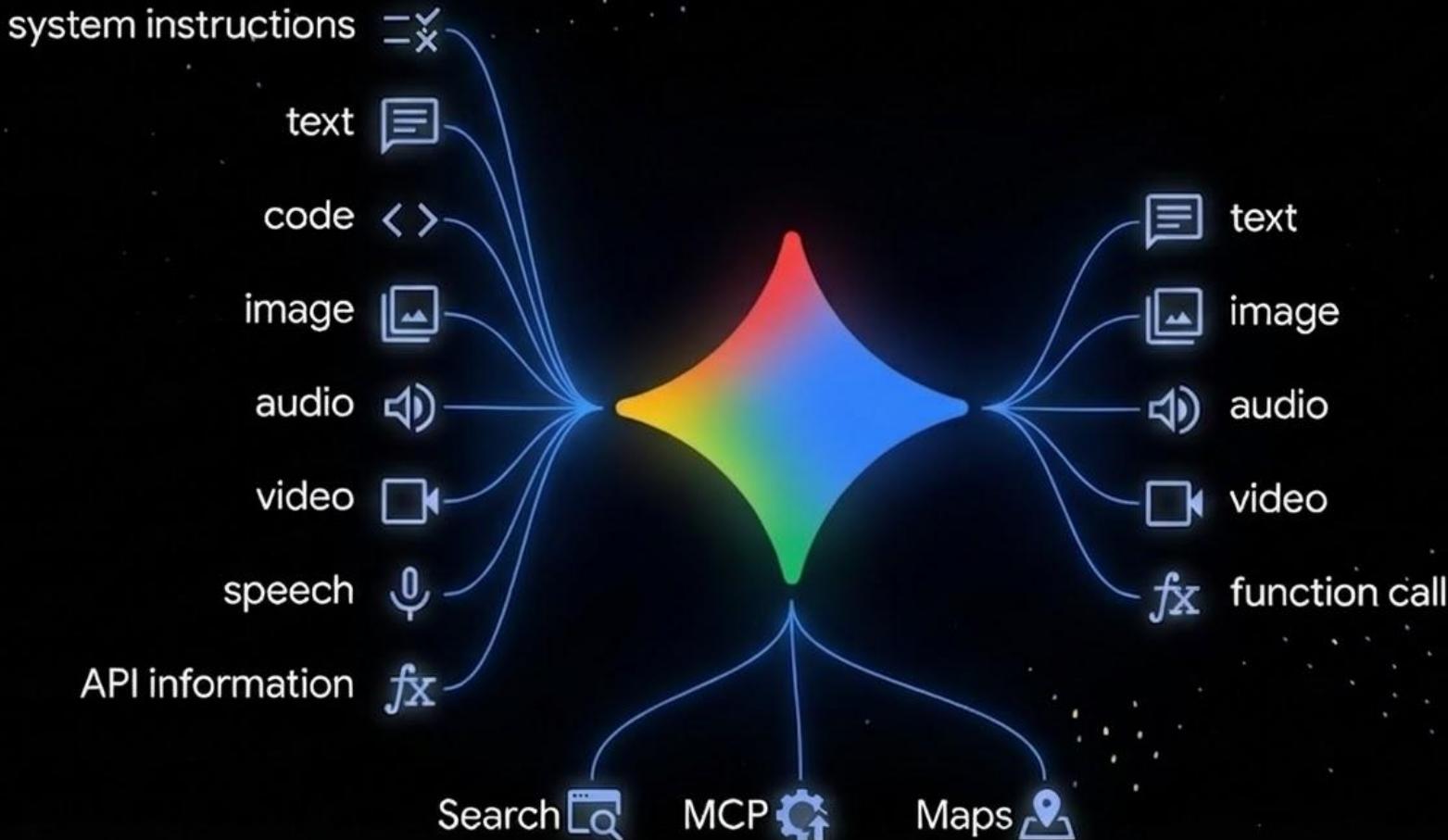
Gemini with
Personalization

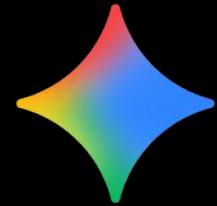
Products & Features

Shipping at Relentless Pace

- Gemma 3
- Gemini Robotics
- Chirp 3
- Gemini 2.5 Pro (03-25)
- Gemini 2.5 Flash
- AlphaEvolve
- Gemini 2.5 Pro (I/O Edition)

Gemini inputs and outputs





Gemini 3

SOTA Reasoning & Multimodality

Agentic Coding

Advanced Tool Use & Planning

Shifting to an era of action

UNDERSTANDING

Gemini 1

THINKING

Gemini 2

ACTION

Benchmarks

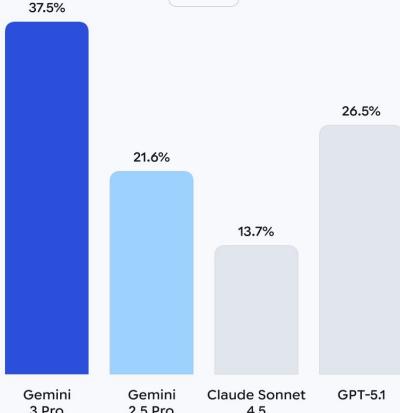
Gemini 3 Pro



Humanity's Last Exam

Reasoning & knowledge

Tools off



Methodology:

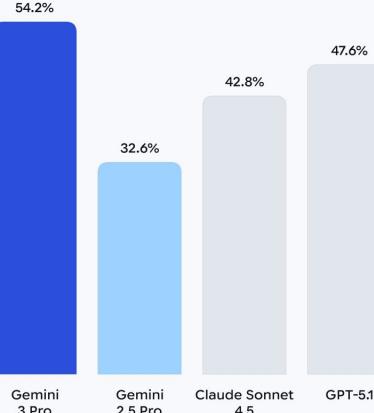
Gemini results: All Gemini 3 scores are pass @1 and were obtained by running with the Gemini API for the model-id gemini-3-pro with default sampling settings.

Humanity's Last Exam results for Gemini 2.5 Pro and Claude Sonnet 4.5 are from Scale AI leaderboard and GPT-5.1 from Artificial Analysis.



Terminal-Bench 2.0

Agentic coding



Methodology:

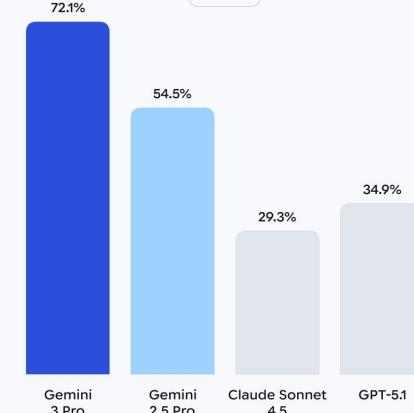
Gemini results were obtained by running with the Gemini API for the model-id gemini-3-pro and gemini-2.5-pro with default sampling settings. All of the results use the same Terminus+2 agent. Results come from <https://www.tberner.ai/leaderboard/terminal-bench/2.0?agents=Terminus+2>



SimpleQA Verified

Parametric knowledge

Tools off



Methodology:

All Gemini, Anthropic, and OpenAI results are pass @1. Gemini results were obtained by running with the Gemini API for the model-id gemini-3-pro and gemini-2.5-pro with default sampling settings. Results come from <https://www.kaggle.com/benchmarks/simpleqa-verified>.

Benchmarks

| Benchmark | Description | Gemini 3 Pro | Gemini 2.5 Pro | Claude Sonnet 4.5 | GPT-5.1 |
|------------------------------|--|--|----------------|------------------------|------------------------|
| Humanity's Last Exam | Academic reasoning | No tools 37.5% | 21.6% | 13.7% | 26.5% |
| ARC-AGI-2 | Visual reasoning puzzles | ARC Prize Verified 31.1% | 4.9% | 13.6% | 17.6% |
| GPQA Diamond | Scientific knowledge | No tools 91.9% | 86.4% | 83.4% | 88.1% |
| AIME 2025 | Mathematics | No tools With code execution 95.0% 100% | 88.0% — | 87.0% 100% | 94.0% — |
| MathArena Apex | Challenging Math Contest problems | 23.4% | 0.5% | 1.6% | 1.0% |
| MMMU-Pro | Multimodal understanding and reasoning | 81.0% | 68.0% | 68.0% | TBC |
| ScreenSpot-Pro | Screen understanding | 72.7% | 11.4% | 36.2% | 3.5% |
| CharXiv Reasoning | Information synthesis from complex charts | 81.4% | 69.6% | 68.5% | 69.5% |
| OmniDocBench 1.5 | OCR | Overall Edit Distance, lower is better 0.115 | 0.145 | 0.145 | 0.147 |
| Video-MMMU | Knowledge acquisition from videos | 87.6% | 83.6% | 77.8% | 80.4% |
| LiveCodeBench Pro | Knowledge acquisition from videos | Elo Rating, higher is better 2,439 | 1,775 | 1,418 | 2,243 |
| Terminal-bench 2.0 | Agentic terminal coding | Terminus-2 agent 54.2% | 32.6% | 42.8% | 47.6% |
| SWE-Bench Verified | Agentic coding | Single attempt 76.2% | 59.6% | 77.2% | 76.3% |
| t2-bench | Agentic tool use | 85.4% | 54.9% | 84.7% | 80.2% |
| Vending-Bench 2 | Agentic long term planning | Net worth (mean), higher is better \$5,478.16 | \$573.64 | \$3,838.74 | \$1,473.43 |
| FACTS Benchmark Suite | Held out internal grounding, parametric, MM, and search retrieval benchmarks | 70.5% | 63.4% | 50.4% | 50.8% |
| SimpleQA Verified | Parametric knowledge | 72.1% | 54.5% | 29.3% | 34.9% |
| MMMLU | Multilingual Q&A | 91.8% | 89.5% | 89.1% | 91.0% |
| Global PIQA | Commonsense reasoning across 100 Languages and Cultures | 93.4% | 91.5% | 90.1% | 90.9% |
| MRCR v2 (8-needle) | Long context performance | 128k (average) 1M (pointwise) 77.0% 26.3% | 58.0% 16.4% | 47.1% not supported | 61.6% not supported |

Benchmarks

LMArena



1501

#1 on LMArena

| Rank ↕ | Rank Spread ⓘ (Upper-Lower) | Model ↗ | Score ↓ |
|--------|--------------------------------|---|---------|
| 1 | 1 ↔ 2 | gemini-3-pro | 1501 |
| 2 | 1 ↔ 3 | grok-4.1-thinking | 1484 |
| 3 | 2 ↔ 5 | grok-4.1 | 1465 |
| 4 | 3 ↔ 8 | gemini-2.5-pro | 1451 |
| 5 | 3 ↔ 9 | claude-sonnet-4-5-20250929-thinking-32k | 1449 |
| 6 | 4 ↔ 8 | claude-opus-4-1-20250805-thinking-16k | 1449 |
| 7 | 4 ↔ 12 | claude-sonnet-4-5-20250929 | 1445 |

18 Nov 2025

Results are from voting pre-release,
and tagged as preliminary

Benchmarks

WebDev Arena

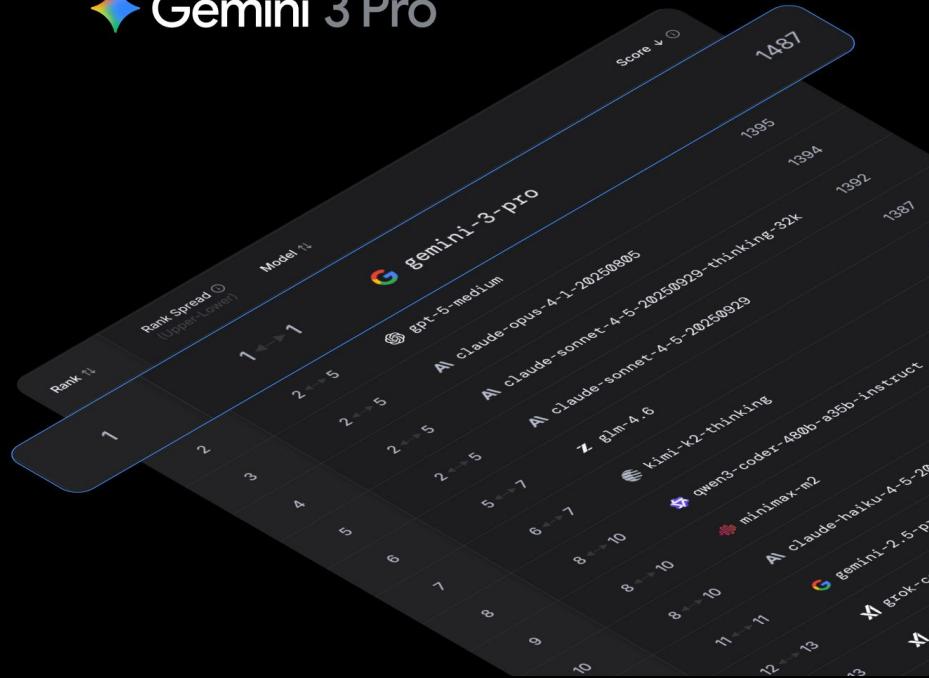
1487

#1 on WebDev Arena

18 Nov 2025

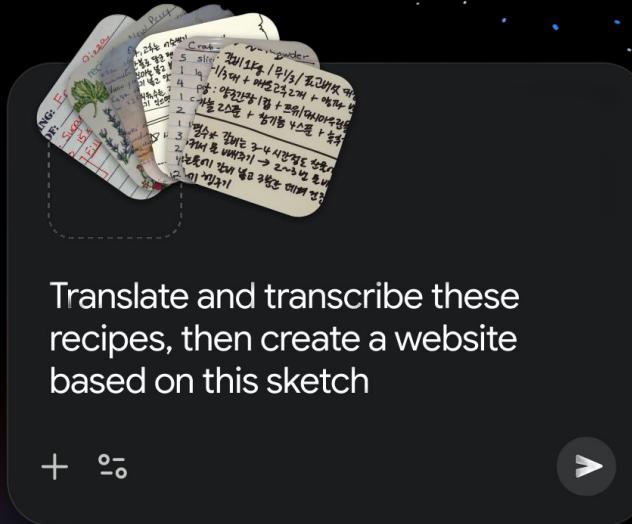
Results are from voting pre-release,
and tagged as preliminary

◆ Gemini 3 Pro



SOTA Reasoning & Multimodality

- Analyzes text, audio, video, and code within a single, industry-leading 1M context window
- Delivers highly factual reasoning for complex inputs (e.g., medical scans, logs)
- Handles difficult data like overlapping speakers or blurry text for maximum insight extraction



Translate and transcribe these recipes, then create a website based on this sketch

+ ⌂ ➤

Agentic Coding

- **Core Skill:** From code completion to system refactoring
- **Safety:** Comprehensive safety evaluations for code

★ Gemini 3



Advanced Tool Use & Planning

- Precise interaction with external APIs and systems with minimal prompt tuning
- Orchestrates complex, multi-step workflows, across disparate enterprise data
- Untangles ambiguous business problems to connect high-level strategy with autonomous execution

★ Gemini 3 Pro



Application



Gemini

New API features

- `thinking_level` parameter: `high`, `medium`, `low`
Control model's internal reasoning process
- `media_resolution` parameter
Granular control over multimodal vision processing
- `thoughtSignatures` to maintain reasoning context across API calls
Stricter validation and updated handling
- `Interactions API`: A unified, stateful interface designed for complex agentic workflows and server-side context management.

Responsible deployment

- Built from the ground up to be safe and responsible
- Our most secure model yet
- Comprehensively evaluated for safety

Nano-Banana in Gemini 3 Pro: From fun to functional

Nano-Banana in 2.5 Flash (Nano Banana)

Best for fast, conversational editing

- Character consistency
- Combining photos (up to 3)
- Local edits
- Design applications

Key audience: Casual consumers

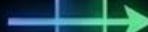
Latency: 10-15s

Nano-Banana Pro in 3 Pro (Nano Banana Pro)

Best for complex, sophisticated outputs

- All Nano-Banana strengths
- Precision controls
- SoTA Text rendering
- Reference images
- Enhanced world knowledge

Key audience: Professional creatives, students, designers, marketers



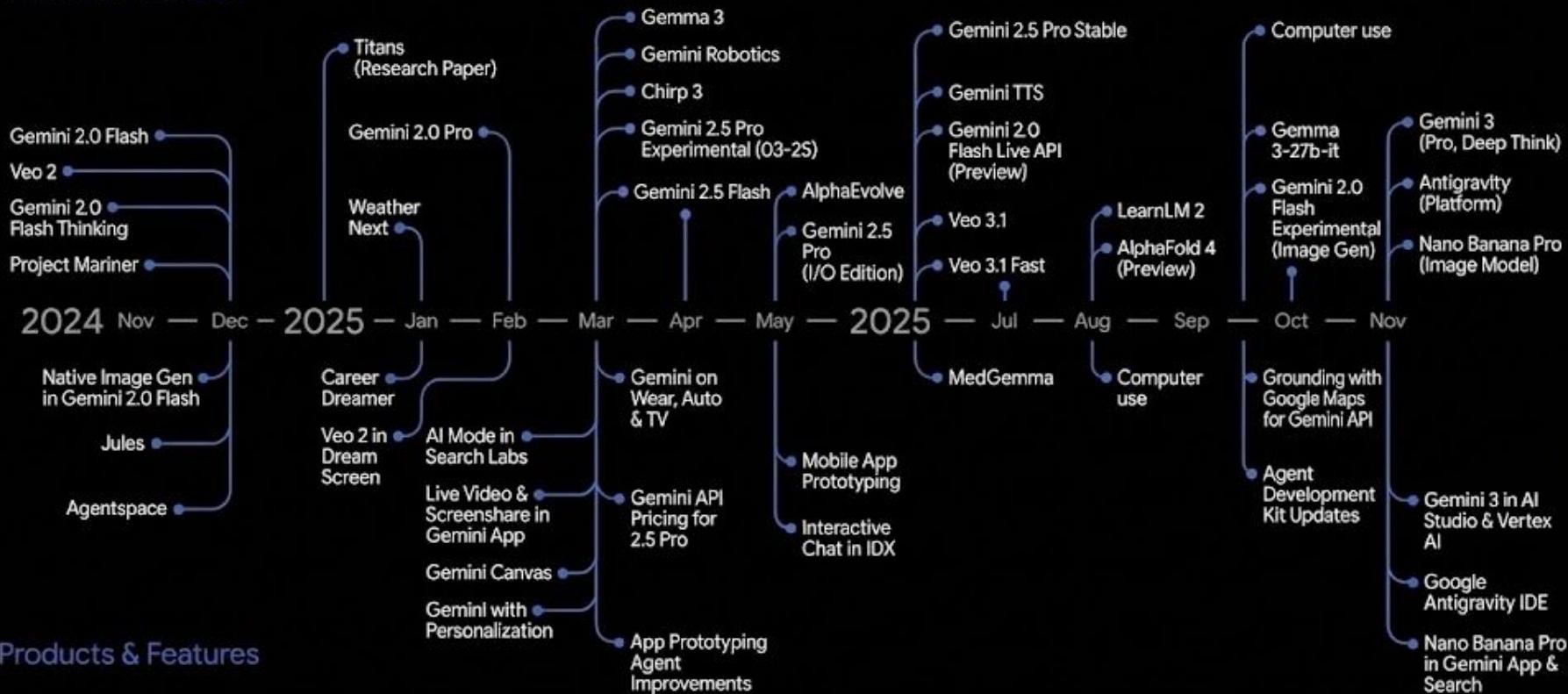
Search Grounding

Search the web then generate an image of isometric perspective, detailed pixel art that shows the career of Guillaume Vernade



Shipping at Relentless Pace

Models & Research

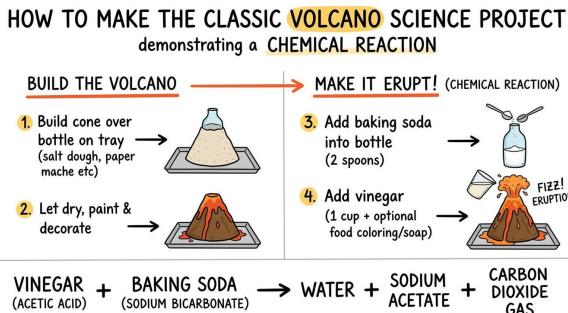


Products & Features

Use Case Exploration

Advanced Text rendering

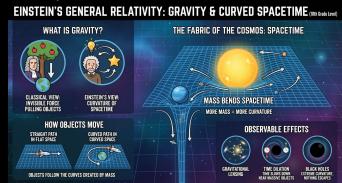
Create anything from comic panels, infographics, menus, invites and more with improved text and spelling accuracy.



Use Case Exploration

Works for i18n

Create anything from comic panels, infographics, menus, invites and more with improved text and spelling accuracy.



now make this infographic in spanish, keeping everything else the same

LA RELATIVIDAD GENERAL DE EINSTEIN: GRAVEDAD Y ESPACIOTIEMPO CURVADO (Nivel de 10º Grado)

¿QUÉ ES LA GRAVEDAD?

VISIÓN CLÁSICA: FUERZA INVISIBLE QUE ATRAÍE OBJETOS

VISIÓN DE EINSTEIN: CURVATURA DEL ESPACIOTIEMPO

CÓMO SE MUEVEN LOS OBJETOS

TRAYECTORIA RECTA EN ESPACIO PLANO

TRAYECTORIA CURVA EN ESPACIO CURVADO

LOS OBJETOS SIGUEN LAS CURVAS CREADAS POR LA MASA

EL TEJIDO DEL COSMOS: ESPACIOTIEMPO

LA MASA CURVA EL ESPACIOTIEMPO

MÁS MASA = MÁS CURVATURA

EFEKTOS OBSERVABLES

LENTE GRAVITACIONAL

DILATACIÓN DEL TIEMPO. EL TIEMPO SE RALENTEZA CERCA DE OBJETOS MASIVOS

AGUEROS NEGROS. CURVATURA EXTREMA, NADA ESCAPA

Use Case Exploration

Your photos explained

Create infographics and explainers with real world knowledge.



PALACE OF FINE ARTS

SAN FRANCISCO, CALIFORNIA

Colonnades: Covered walkways with repeating columns and arches

Main Rotunda: Roman-inspired architecture with Corinthian columns and ornate carvings

Originally constructed for the 1915 Panama-Pacific International Exposition. Rebuilt in the 1960s.

Lagoon Reflection: Creates a mirrored view, enhancing the tranquil atmosphere.

Morning Sun / Clear Sky

An annotated infographic about the Palace of Fine Arts. It highlights the Main Rotunda, Colonnades, and Lagoon Reflection. It also provides information about its construction for the 1915 Panama-Pacific International Exposition and its reconstruction in the 1960s. The background shows the building reflected in the water under a clear sky.

Turn this into an informative visual explainer about this place



One World Trade Center
Tallest building in Western Hemisphere (1,776 ft!)

30 Hudson Yards
Features "The Edge," an outdoor sky deck with a glass floor!

40 Wall Street
Was the world's tallest building for a very short time in 1930!

20 Exchange Place
Was the world's tallest bank building when built in 1931!

55 Water Street
Once NYC's largest office building by floor area!

An annotated infographic about the New York City skyline. It points to the One World Trade Center, 30 Hudson Yards, 40 Wall Street, 20 Exchange Place, and 55 Water Street, providing fun facts about each. The background shows the city skyline across the Hudson River.

Annotate the image with a fun fact for each of the top 5 famous buildings in the image

Use Case Exploration

Combine up to 14 photos

Upload up to 14 photos (6 with high fidelity) to combine elements, blend scenes, and transfer designs to create something entirely new.



Create a product photoshoot using this product, model and logo. Respect the overall aesthetic and color palette.



Use Case Exploration

Aspect ratios and 2K and 4K outputs for any project

Make your photos fit to any aspect ratio



Beyond Gemini models

GenMedia (Imagen4 & nano-banana)



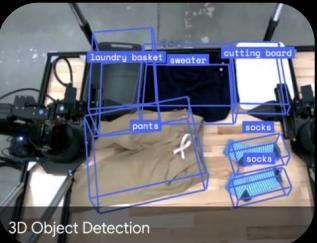
GenMedia (Veo3)



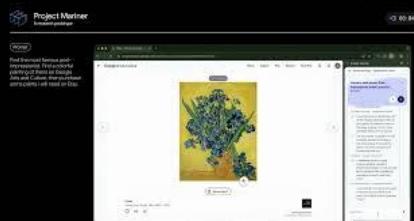
GenMedia (Lyria)



Gemini Robotics



Agents (Mariner, Jules)



Open models

Gemma 3

And more...



DOW JONES  -0.19% NASDAQ  +0.01% S&P 500  -0.09% AAPL  -0.31% NVDA  +1.36% MSFT  -0.55% AMZN  +0.57% META  -0.11%

TECH

Google's AI playground was its best-kept secret. Now everyone is finding out.

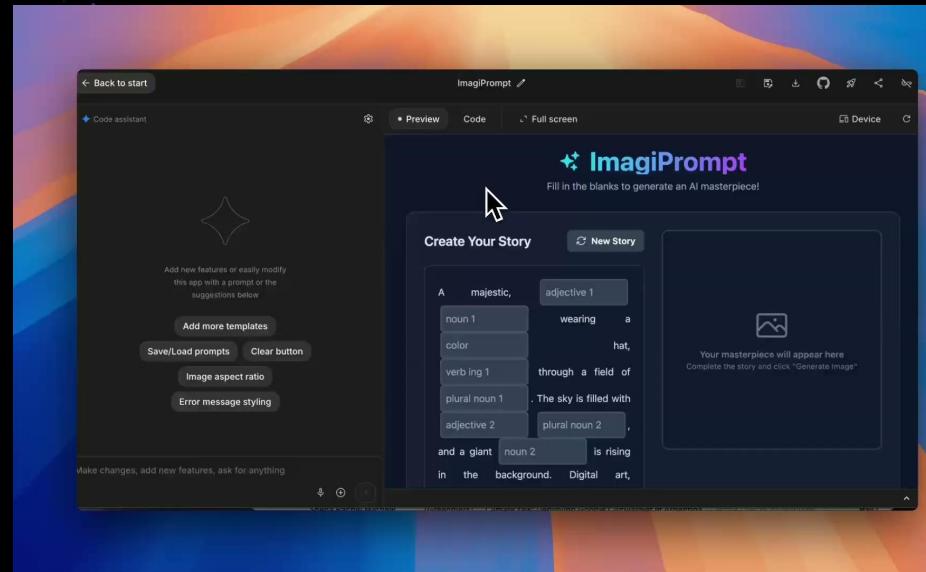
By [Hugh Langley](#) 



ai.studio

Gemini 2.5 Computer Use

Power agents that
can interact with user
interfaces

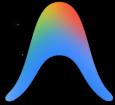


Try it on
gemini.browserbase.com

Jules

Your experimental
AI-powered code
agent

Introducing



Google Antigravity

A new **Agentic Development Platform** from Google DeepMind.

Experience Liftoff

Increased autonomy across workstreams

Alongside the synchronous agent embedded in your IDE, asynchronous agents and browser control further increase autonomy and enhance your work.

Trust, verification, and self-improvement

Intuitive, asynchronous user feedback across all surfaces is crucial for increasing trust. Verification is supported by artifacts and summaries. A central knowledge bank to store and retrieve information, driving continuous improvement.

An agent-first paradigm

Manage your asynchronous agents from a centralized homebase. Monitor their progress while you focus on your own tasks.

Beginners Guide

▲ Google Antigravity



Beginners Guide

▲ Google Antigravity



Beginners Guide

▲ Google Antigravity



Start new conversation in ▾ CollabWhiteboard

 View Inbox

|Ask anything, @ for context

+ ▾ Fast ▾ Gemini 3 Pro Preview



< > Open editor Use Playground





Gemini CLI

The Future of AI in Development

From Code Assistant to Team Member

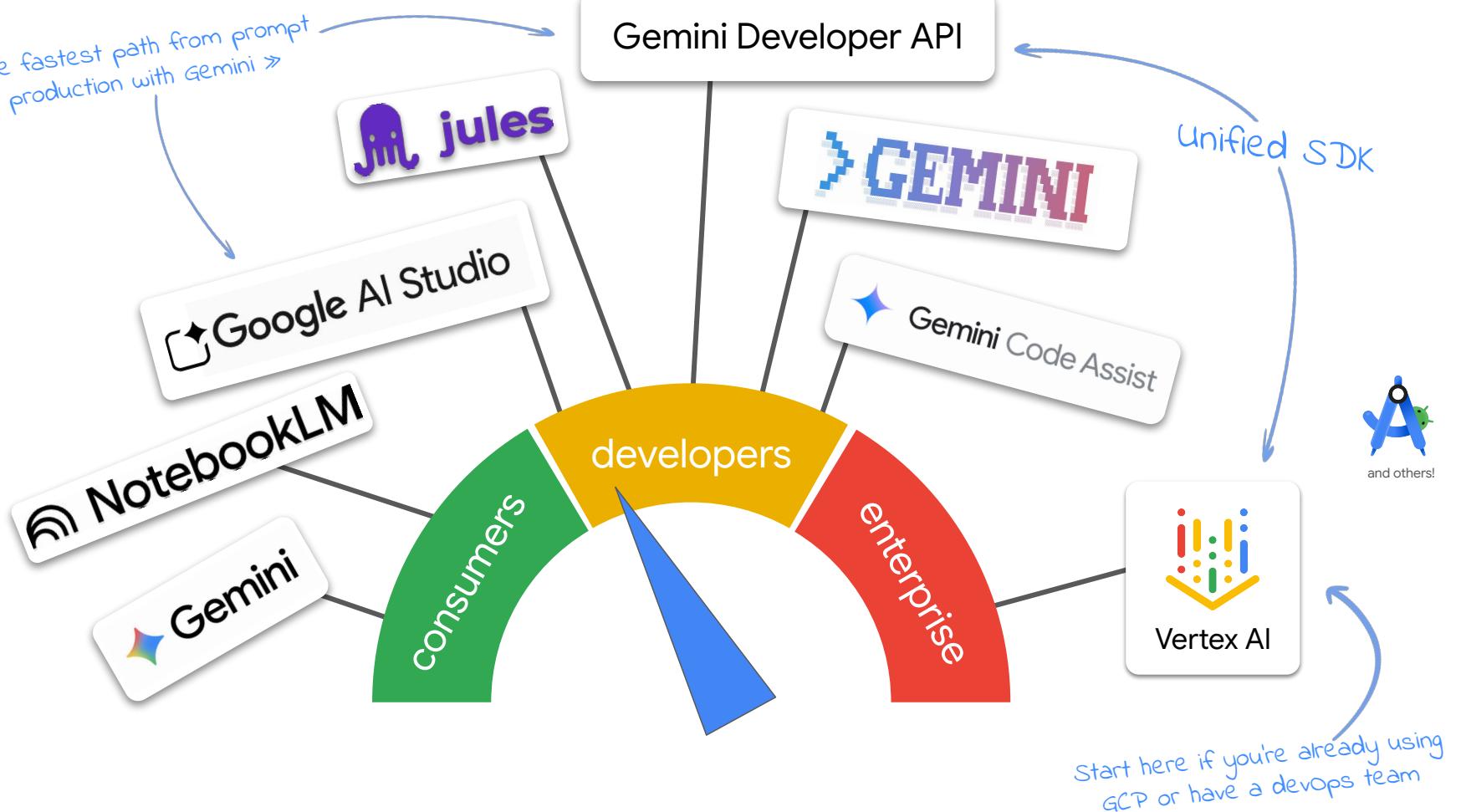
- From simple code completion to AI agents that can work semi-autonomously
- Next breakthrough: AI QA/QC and AI mock-ups

From coding to directing & architecturing

- Directing the AI agents
- Organizing projects for AI
- System instruction engineering

The Rise of No-Code and Low-Code

- Everybody can (no) code!
- Empowering people who are not developers to build applications.



Live-ish demo

Magic mirror

About a week of work using Jules with the live models, Veo and Lyria to go from the idea to the working POC.

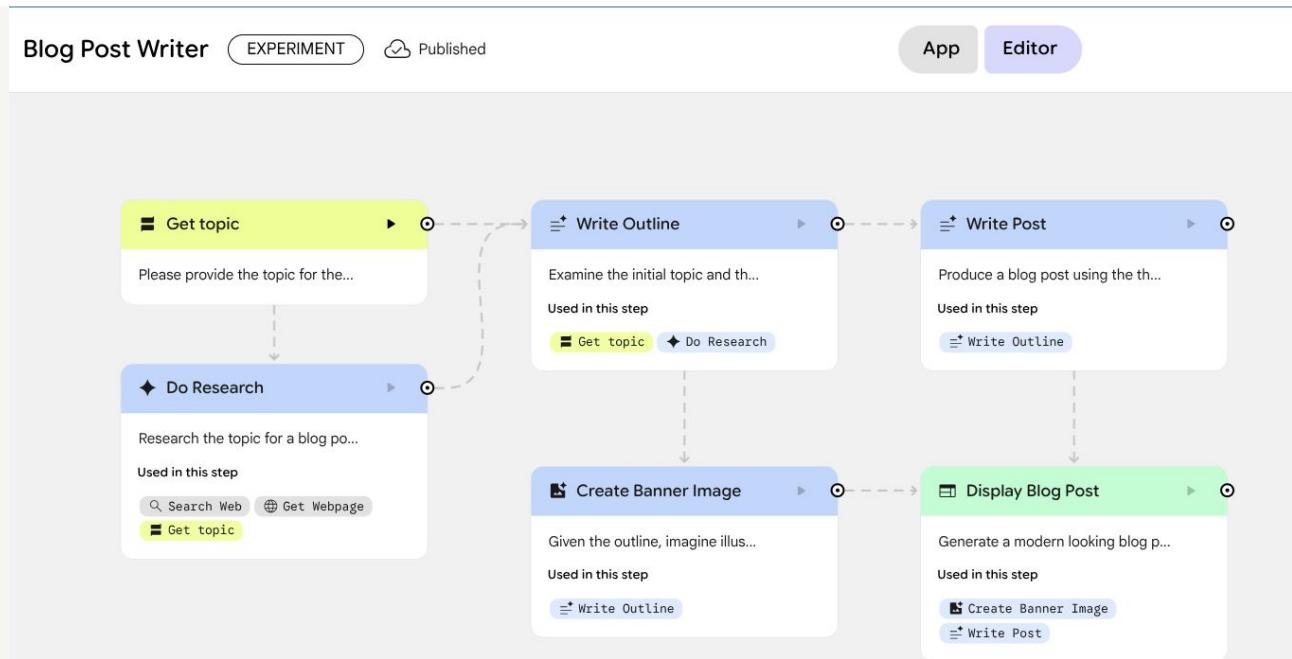
Try it yourselves:

goo.gl/devfest-magic-mirror



Opal

No-code app and agents builder



Very generous free (of charge) mana Tier

Standard API rate limits

The following table lists the rate limits for all standard Gemini API calls.

 Note: Any values that show * have no published rate limits.

| Free Tier | Tier 1 | Tier 2 | Tier 3 | Requests per minute | RPM | Tokens per minute | TPM | RPD | Requests per day |
|-------------------------------|--------|--------|--------|---------------------|-----|-------------------|-----|-------|------------------|
| Model | | | | | | | | | |
| Text-out models | | | | | | | | | |
| Gemini 2.5 Pro | | | | 5 | | 250,000 | | 100 | |
| Gemini 2.5 Flash | | | | 10 | | 250,000 | | 250 | |
| Gemini 2.5 Flash Preview | | | | 10 | | 250,000 | | 250 | |
| Gemini 2.5 Flash-Lite | | | | 15 | | 250,000 | | 1,000 | |
| Gemini 2.5 Flash-Lite Preview | | | | 15 | | 250,000 | | 1,000 | |

Very generous free (of charge) mana Tier

How Google Uses Your Data

When you use Unpaid Services, including, for example, Google AI Studio and the unpaid quota on Gemini API, Google uses the content you submit to the Services and any generated responses to provide, improve, and develop Google products and services and machine learning technologies, including Google's enterprise features, products, and services, consistent with our [Privacy Policy](#).

To help with quality and improve our products, human reviewers may read, annotate, and process your API input and output. Google takes steps to protect your privacy as part of this process. This includes disconnecting this data from your Google Account, API key, and Cloud project before reviewers see or annotate it. **Do not submit sensitive, confidential, or personal information to the Unpaid Services.**

The license you grant to Google under the "[Submission of Content](#)" section in the API Terms also extends, to the extent required under applicable law for our use, to any content (e.g., prompts, including associated system instructions, cached content, and files such as images, videos, or documents) you submit to the Services and to any generated responses.

Google only uses content that you import or upload to our model tuning feature for that express purpose. Tuning content may be retained in connection with your tuned models for purposes of re-tuning when supported models change. When you delete a tuned model, the related tuning content is also deleted.

If you're in the European Economic Area, Switzerland, or the United Kingdom, the terms under "[How Google uses Your Data](#)" in "[Paid Services](#)" apply to all Services, including Google AI Studio and unpaid quota in the Gemini API, even though they are offered free of charge.



This means we can't use your data in Europe!

(Thank you GDPR)



Questions?

ai.studio

goo.gle/cookbook

goo.gle/gemini-docs