

Building the Unimaginable: The Gemini 3.1 Pro Era.



A Definitive
Product
Showcase &
Strategic
Guide.

TL;DR: The era of prompt-to-text is over.
The era of **prompt-to-product** is here.



The Text Chatbot



The Agentic Team Member

We have officially shifted from text generation to full-stack application building. AI is no longer a tool you micromanage; it is an agentic partner that plans, executes, and iterates.

TL;DR: A generational performance upgrade masquerading as a point release.

+148% Reasoning Leap

**Native
Multimodality**

Zero Price Hike

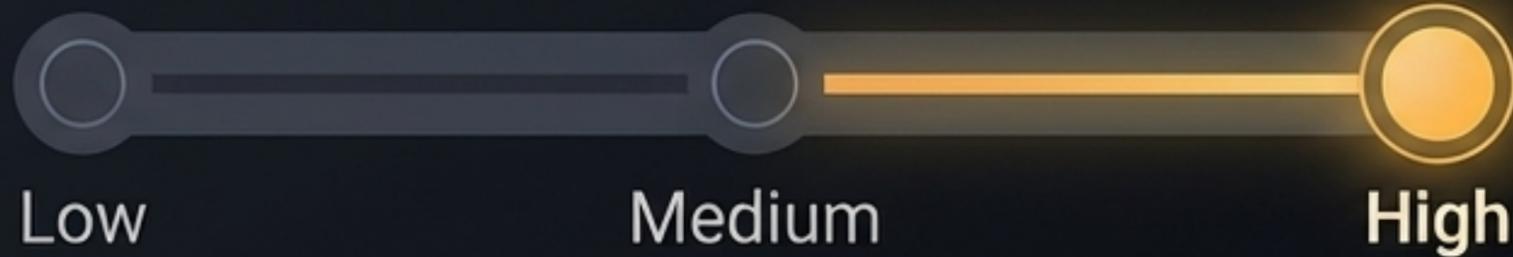
Google breaks tradition. The 0.1 increment usually implies a minor patch, but Gemini 3.1 Pro delivers a massive architectural leap in reasoning and output efficiency—at the exact same \$2 per 1M token price as Gemini 3 Pro.

TL;DR: Outputs up to 65,000 tokens in a single generation without truncation.



The **1M Context Window** native in production. End the era of copy-pasting chunks. Feed the model your entire repository, full legal contract sets, or years of customer feedback in a single prompt.

TL;DR: Pay for deep reasoning only when your codebase actually requires it.



77.1%
ARC-AGI-2
(up from 31.1%)

Doubled abstract reasoning. Gemini 3.1 Pro introduces configurable Thinking Levels, allowing you to scale compute exactly to your workload. Use 'Low' for rapid autocomplete, and 'High' for multi-step algorithmic debugging.

TL;DR: One unified model. Infinite input combinations.



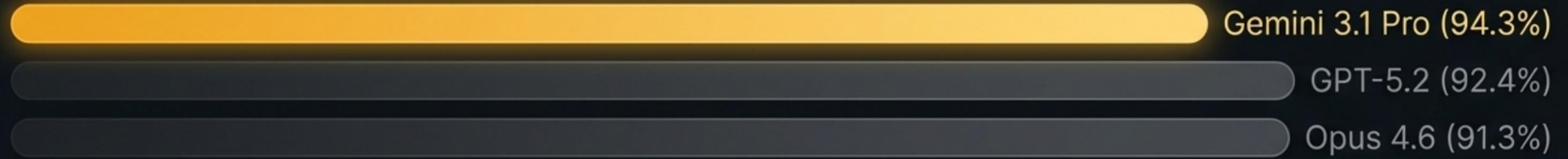
The **only frontier model** handling text, imagery, audio, and video natively at the API level. No translation layers. No lost fidelity.

TL;DR: The 2026 Frontier Landscape is defined by three giants.

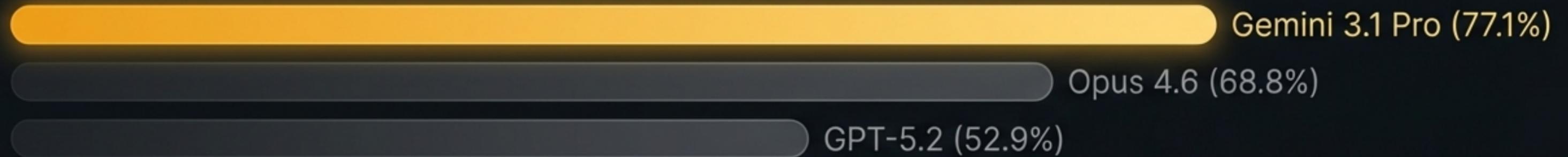
	Gemini 3.1 Pro	Claude Opus 4.6	GPT-5.4
Context	1M	200k standard	1M
Output	64k	128k	128k
Input Cost per 1M	\$2.00	\$5.00	\$2.50

TL;DR: 94.3% on GPQA Diamond makes it the smartest science model available.

GPQA Diamond - PhD Science



ARC-AGI-2 - Abstract Puzzles



Unmatched in deep, multi-step problem solving. Gemini 3.1 Pro leads the field in PhD-level scientific knowledge and abstract reasoning puzzles.

TL;DR: Production-grade coding performance capable of full-stack generation.

SWE-Bench Verified

Gemini: **80.6%** | Opus 4.6: **80.8%**

LiveCodeBench Pro

Gemini 3.1 Pro: **2887 Elo**

Terminal-Bench 2.0

Codex: **77.3%** | Gemini: **68.5%**

A **top-tier coding engine**. While specialized models slightly edge out terminal tasks, Gemini delivers **80.6%** on **SWE-Bench Verified**, resolving real-world GitHub issues end-to-end.

TL;DR: Frontier intelligence is no longer cost-prohibitive for high-volume SaaS.

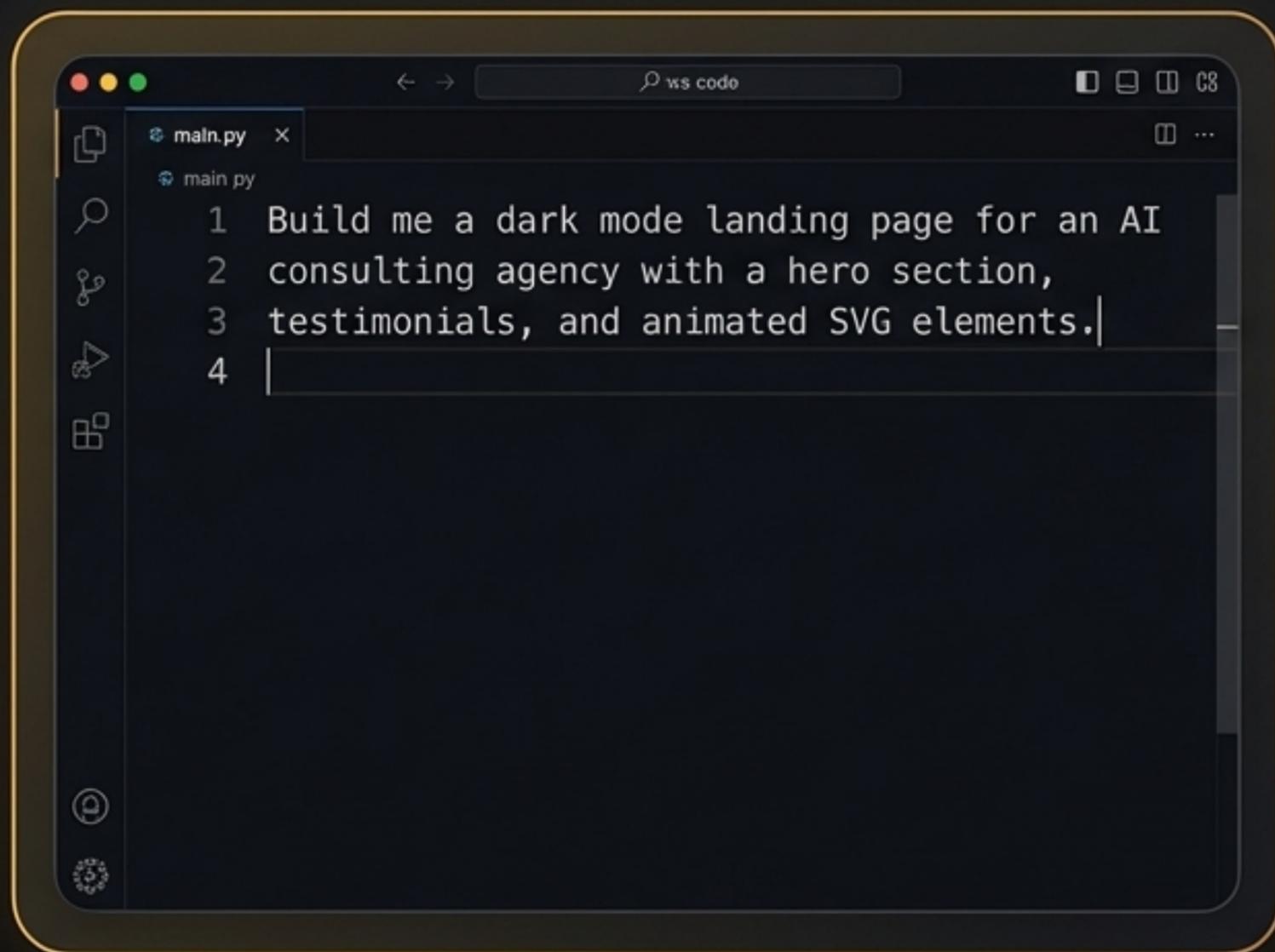
\$0.50 Context Cache Hit

\$2.00

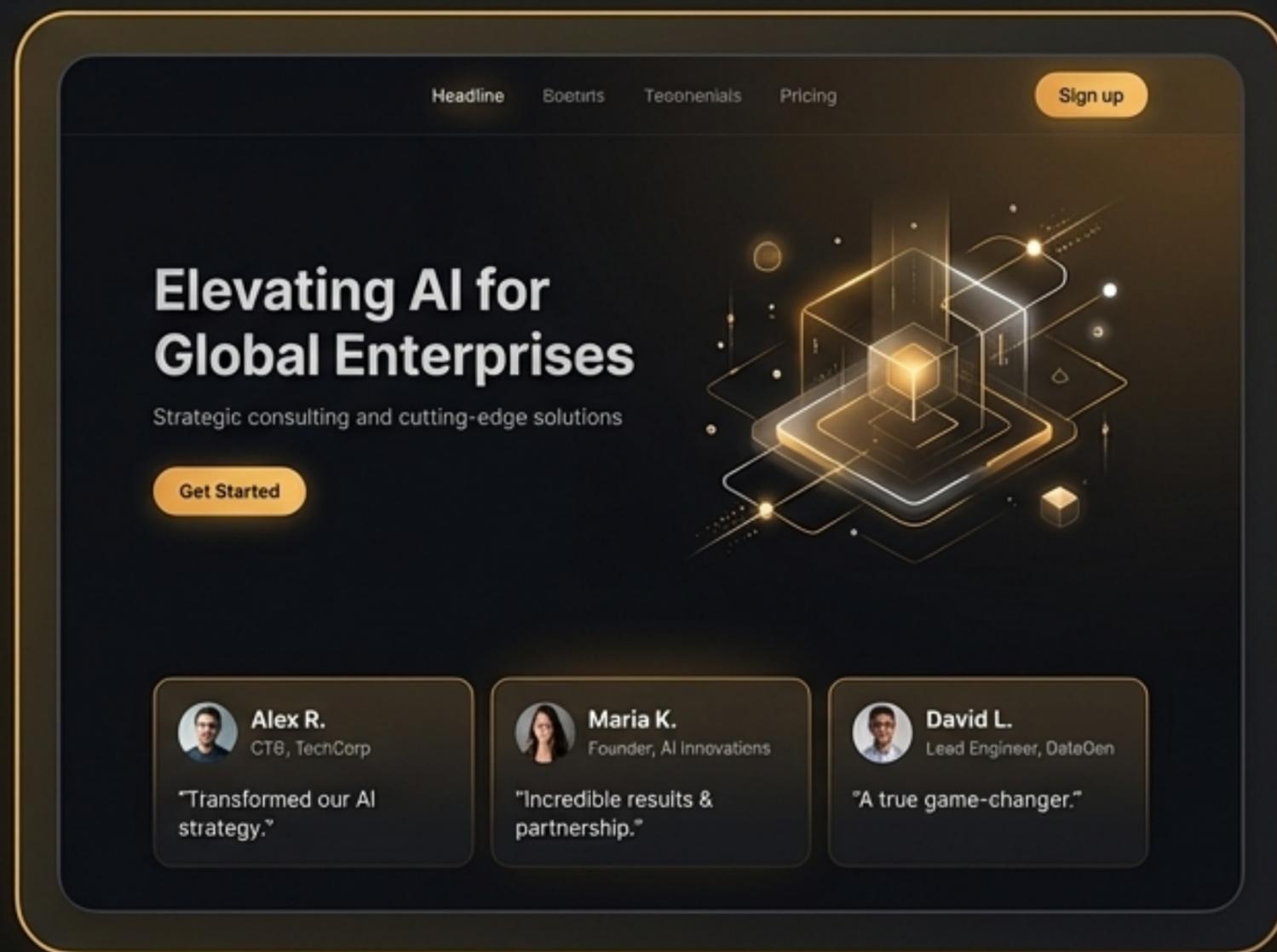
Opus 4.6 Input Costs: \$5.00 to \$15.00

The **Price-Performance King**. Gemini 3.1 Pro delivers Opus 4.6-level performance at up to **7.5x cheaper** input costs. Combine this with **\$0.50 Context Caching**, and repetitive codebase analysis becomes incredibly cheap.

TL;DR: Zero designers needed for interactive prototypes.



```
main.py x  
main.py  
1 Build me a dark mode landing page for an AI  
2 consulting agency with a hero section,  
3 testimonials, and animated SVG elements.  
4 |
```



Vibe Coding in action. Describe the intent, the style, and the functionality. Gemini 3.1 Pro handles the HTML, CSS, JavaScript, and pure-code animated SVGs in a single output.

TL;DR: Combine the strengths of both models in Google Antigravity.

**Claude Opus 4.6
(Thinking Mode)**

Generates deep system architecture and file structures.

**Gemini 3.1 Pro
(Fast Execution Mode)**

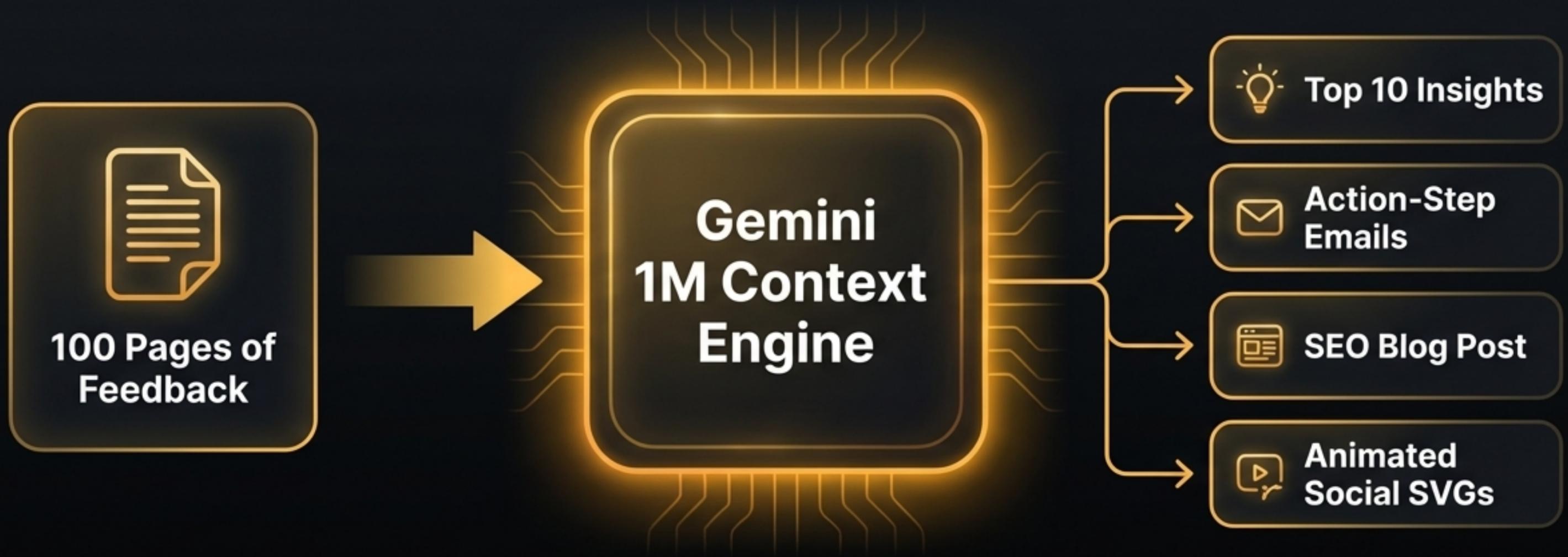
Codes all components, soft-body physics, and visual aesthetics.

**Google Antigravity
IDE**

The live 3D Minecraft clone.

The Ultimate Coding Workflow. Route complex architectural planning to Opus 4.6, and hand the implementation plan to Gemini 3.1 Pro to flawlessly execute front-end UI and game mechanics.

TL;DR: What used to take a marketing team a week now takes an afternoon.



One prompt, complete workflows. Dump massive datasets into the context window and have Gemini autonomously group, analyze, write, and design all necessary outputs without losing a single thread.

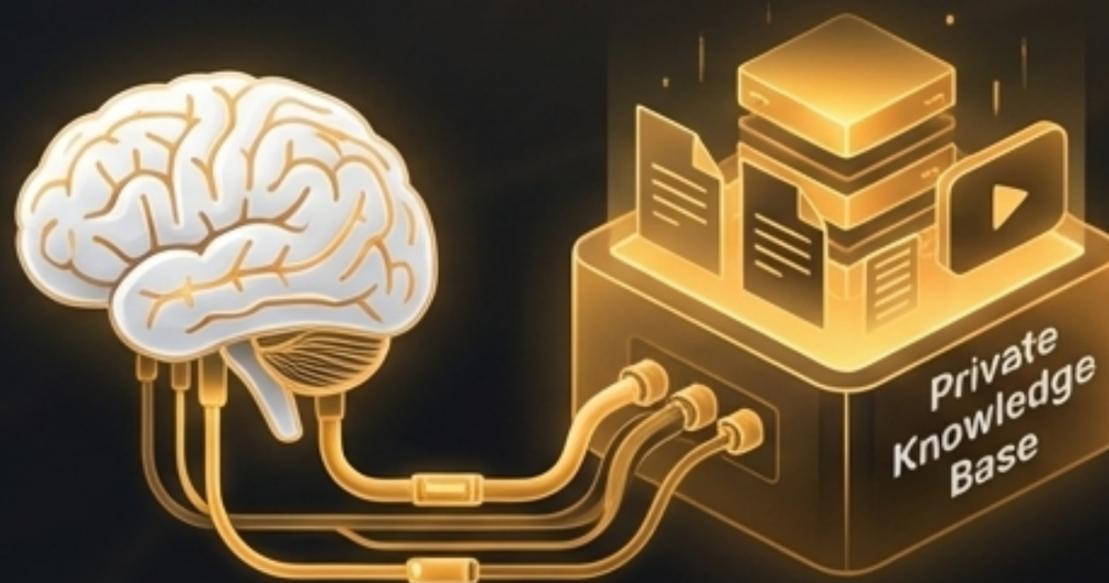
TL;DR: Eliminate hallucinations by grounding Gemini in your private data.

Standard LLM



Hallucinations

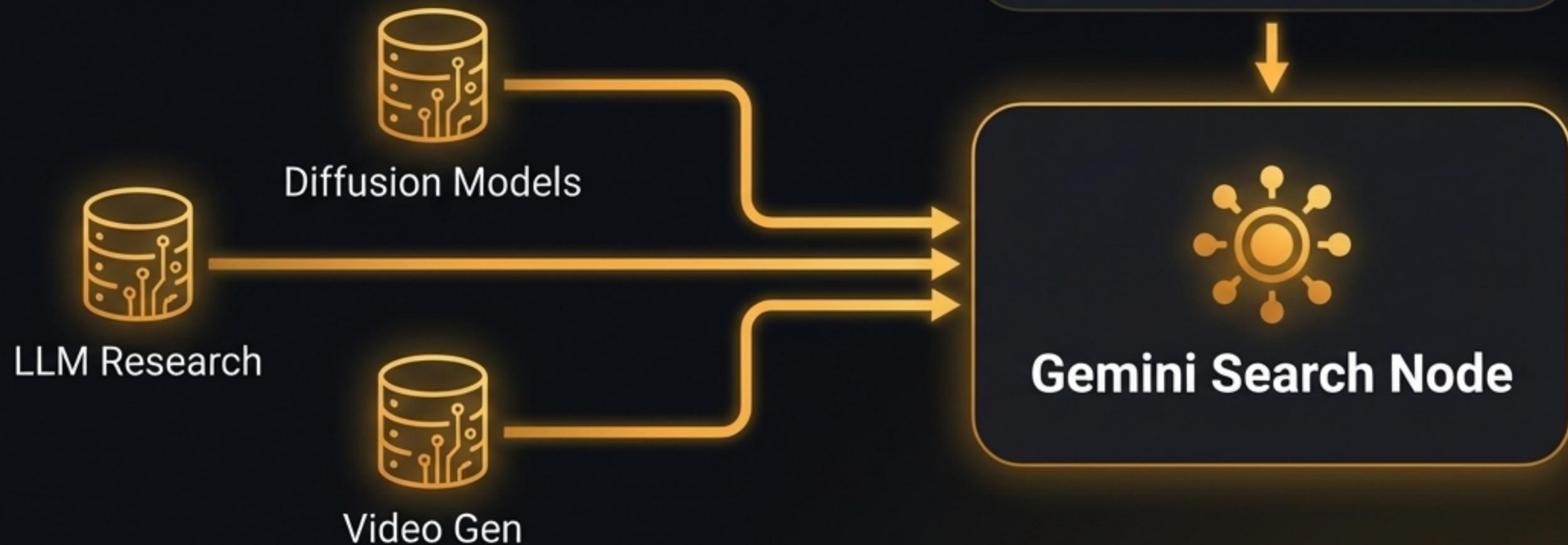
Gemini + NotebookLM



Grounded Truth

The Context Problem: Chatbots forget. The Solution: NotebookLM. By directly integrating Gemini with NotebookLM, every answer is grounded in your uploaded PDFs, spreadsheets, and YouTube transcripts with inline citations.

TL;DR: Compress 10 hours of research into 20 minutes.



Break the silos. While NotebookLM restricts you to one notebook, Gemini allows you to attach multiple notebooks to a single session. Synthesize insights across massive, disparate datasets while simultaneously searching the live web.

TL;DR: Build a **custom brain** for **every department** of your **business**.

Gem Configuration

Custom Instructions:

Act as my YouTube Strategist. |



Knowledge Base:

NotebookLM Analytics Data [🔗](#)

● Auto-sync Enabled

Gems are **permanent**, specialized AI assistants. Connect a **NotebookLM source** to a Gem, and it **auto-syncs**. When you add new documents to the notebook, your Gem instantly gets smarter without needing new prompts.

TL;DR: The fastest model
in the Gemini ecosystem.



45%
Faster
Generation

2.5x
Faster to
First Token

Enter Gemini 3.1 Flash-Lite.

Engineered specifically for high-volume, low-reasoning tasks
where speed is the absolute priority.

TL;DR: Millisecond performance for high-throughput enterprise pipelines.

\$0.25
per 1M Input

\$1.50
per 1M Output



Real-time Weather Dashboards



Bulk E-commerce Translations

Built for scale. At a fraction of the cost, Flash-Lite is the perfect API engine for real-time dashboards, massive e-commerce localization, and instant moderation of user-generated content.

TL;DR: Deploy agentic AI with complete confidence.

- ✓ CBRN Capability (Passed)
- ✓ Cyber Uplift Level 1 (Passed)
- ✓ Misalignment/Instrumental Reasoning (Passed)

Evaluated under the Frontier Safety Framework. Gemini 3.1 Pro has undergone rigorous testing against critical capability levels, ensuring enterprise-grade security even when operating with Deep Think enabled.

TL;DR: Zero friction to deployment.



Access the API wherever you work. From fast prototyping in AI Studio, to enterprise SLA environments in Vertex AI, to native IDE integrations in Copilot and Antigravity.

Stop Chatting. Start Building.

[Access the API](#)

The era of the Agentic Builder is officially live.