Introducing Agentic Vision in Gemini 3 Flash

Jan 27, 2026 · 4 min read

Read AI-generated summary

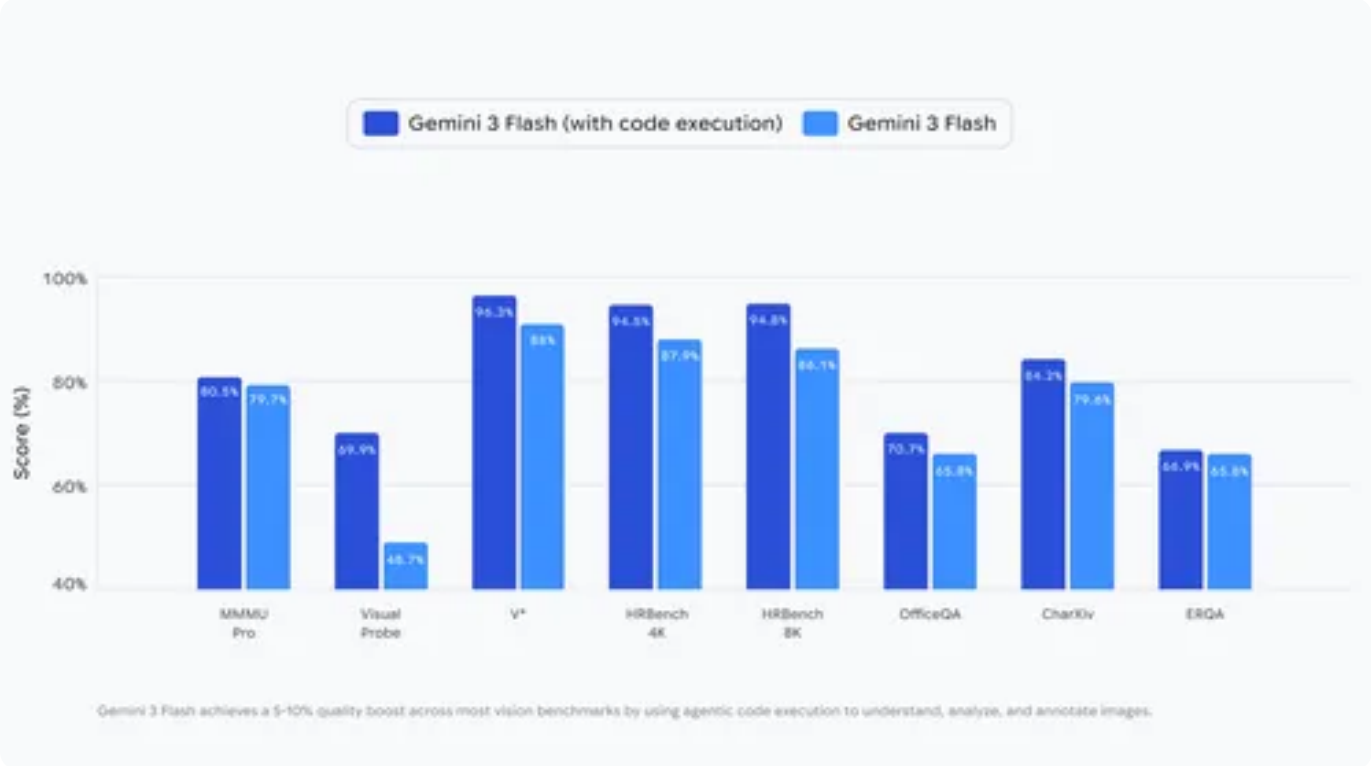
Agentic Vision, a new capability in Gemini 3 Flash, combines visual reasoning with code execution to ground answers in visual evidence.

Rohan Doshi
Product Manager, Google DeepMind

Frontier AI models like Gemini typically process the world in a single, static glance. If they miss a fine-grained detail — like a serial number on a microchip or a distant street sign — they are forced to guess.

Agentic Vision in Gemini 3 Flash converts image understanding from a static act into an agentic process. It treats vision as an active investigation. By combining visual reasoning with code execution, one of the first tools supported by Agentic Vision, the model formulates plans to zoom in, inspect and manipulate images step-by-step, grounding answers in visual evidence.

Enabling code execution with Gemini 3 Flash delivers a consistent 5-10% quality boost across most vision benchmarks.



Gemini 3 Flash achieves a 5-10% quality boost across most vision benchmarks by using agentic code execution to understand, analyze, and annotate images.

Agentic Vision: a new frontier AI capability

Agentic Vision introduces an agentic Think, Act, Observe loop into image understanding tasks:

1. **Think**: the model analyzes the user query and the initial image, formulating a multi-step plan.

2. **Act**: The model generates and executes Python code to actively manipulate images (e.g. cropping, rotating, annotating) or analyze them (e.g. running calculations, counting bounding boxes, etc).

3. **Observe**: The transformed image is appended to the model's context window. This allows the model to inspect the new data with better context before generating a final response.

Agentic Vision in action

By enabling code execution in the API, you can unlock many new behaviors, many of which are highlighted in our demo app in Google AI Studio. From big products like the Gemini app to smaller startups, developers have already started integrating the capability to unlock many use cases, including:

1. Zooming and inspecting

Gemini 3 Flash is trained to implicitly zoom when detecting fine-grained details.

PlanCheckSolver.com, an AI-powered building plan validation platform, improved accuracy by 5% by enabling code execution with Gemini 3 Flash to iteratively inspect high-resolution inputs. The video of the backend logs demonstrate this agentic process: Gemini 3 Flash generates Python code to crop and analyze specific patches (e.g., roof edges or building sections) as new images. By appending these crops back into its context window, the model visually grounds its reasoning to confirm compliance with complex building codes.

2. Image annotation

Agentic Vision allows the model to interact with its environment by annotating images. Instead of just describing what it sees, Gemini 3 Flash can execute code to draw directly on the canvas to ground its reasoning.

In the example below, the model is asked to count the digits on a hand in the Gemini app. To avoid counting errors, it uses Python to draw bounding boxes and numeric labels over each finger it identifies. This "visual scratchpad" ensures that its final answer is based on pixel-perfect understanding.

3. Visual math and plotting

Agentic Vision can parse high-density tables and execute Python code to visualize the findings.

Standard LLMs often hallucinate during multi-step visual arithmetic. Gemini 3 Flash bypasses this by

offloading computation to a deterministic Python environment. In the example below from our demo app in Google AI Studio, the model identifies the raw data, writes code to normalize prior SOTA to 1.0 and generates a professional Matplotlib bar chart. This replaces probabilistic guessing with verifiable execution.

What's next

We are just getting started with Agentic Vision.

- **More Implicit Code-Driven Behaviors:** Today, Gemini 3 Flash excels at implicitly deciding when to zoom in on small details. While other capabilities, such as rotating images or performing visual math, currently require an explicit prompt nudge to trigger, we are working to make these behaviors fully implicit in future updates.

- **More Tools:** We are also exploring how to equip Gemini models with even more tools,

including web and reverse image search, to ground its understanding of the world even further.

- **More Model Sizes**: Additionally, we also plan to expand this capability to our other model sizes beyond just Flash.

How to get started

Agentic Vision is available today via the Gemini API in Google AI Studio and Vertex AI. It is also starting to roll out in the Gemini app (access by selecting Thinking from the model drop-down). Developers can try the demo in Google AI Studio, or experiment with the feature in the AI Studio Playground by turning on "Code Execution" under Tools. Read the developer docs to learn more for (Vertex AI dev docs).

```
from google import genai
from google.genai import types

client = genai.Client()

image = types.Part.from_uri(
    file_uri="https://goo.gle/instrument-img",
    mime_type="image/jpeg",
)

response = client.models.generate_content(
    model="gemini-3-flash-preview",
    contents=[image, "Zoom into the expression pedals and tell me
    config=types.GenerateContentConfig(
        tools=[types.Tool(code_execution=types.ToolCodeExecution
    ),
)
```

```
print(response.text)
```

POSTED IN:

Developer tools Gemini models