

Coding Beauty · [Follow publication](#)

★ Member-only story

Yet Another Claude Model Just Shocked The World — Faster Than Sonnet 4.5 😮

3 min read · 5 days ago



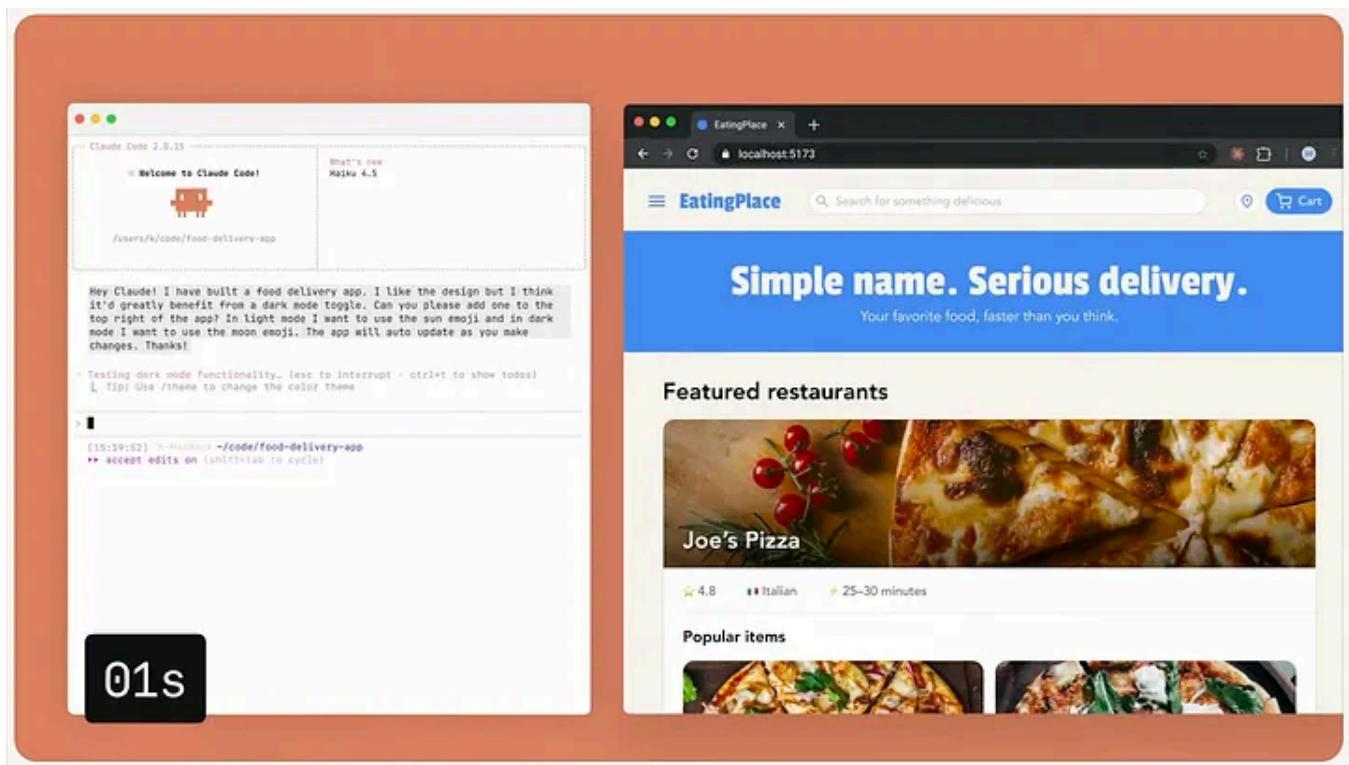
Tari Ibaba

Follow

Listen

Share

More

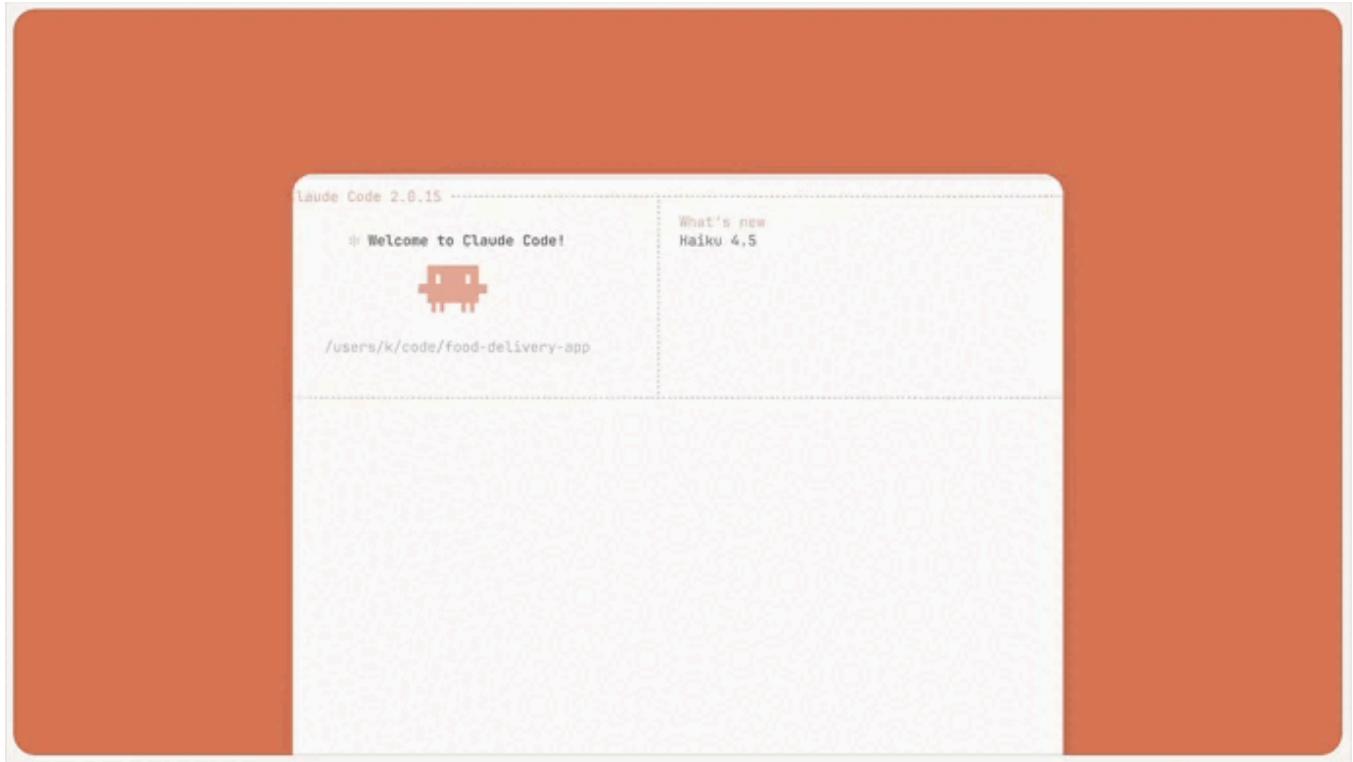


Well that was FAST.

Barely 2 weeks after Sonnet 4.5 and we have yet another incredible Claude model from Anthropic.

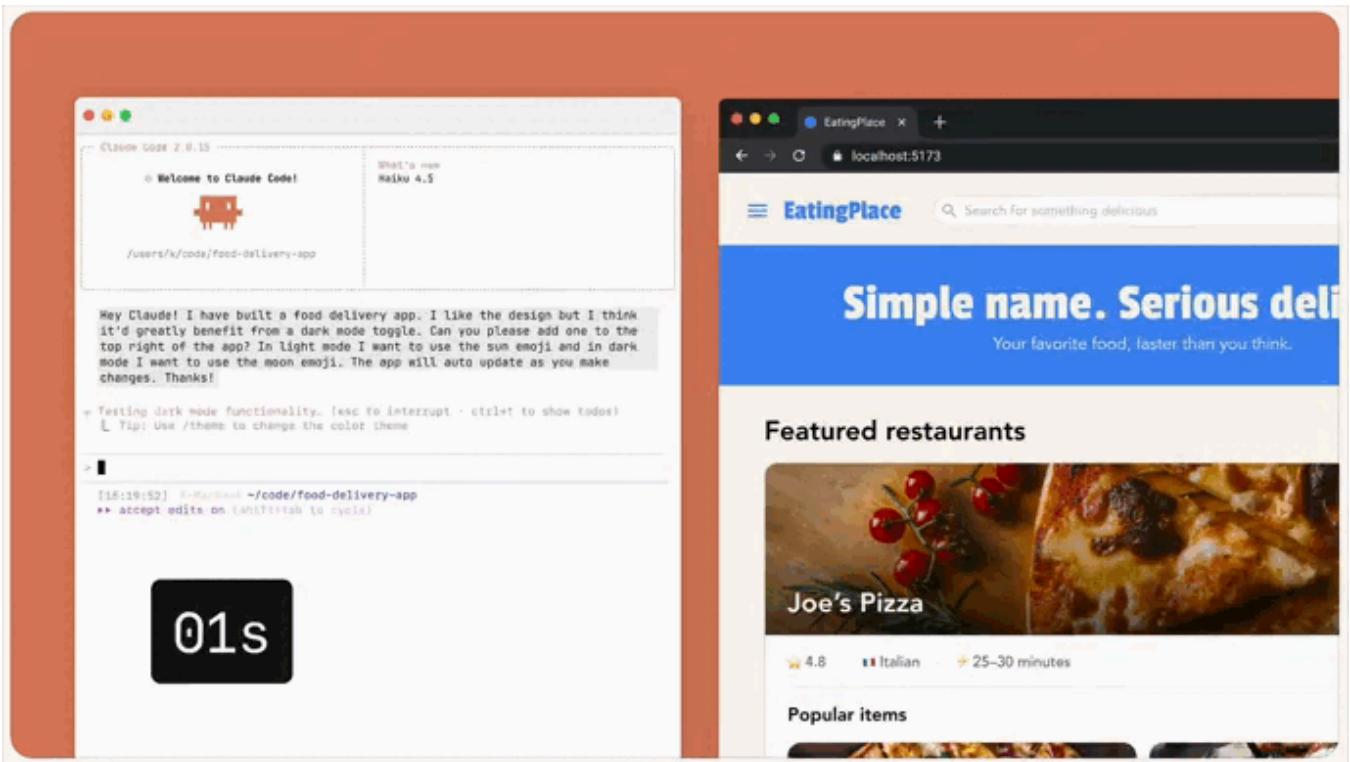
So many devs are getting blown away by the sheer speed. The speed is really something else.

The new Claude Haiku 4.5 gives you all the intelligence of Sonnet 4.5 at only a fraction of the cost and an unbelievable pace.



It's leaner, it's faster, it's cheaper — and it even comes with a brand new feature never seen before in this line of models.

Haiku 4.5 is built for pure speed. Speed is what this is all about.



This model is perfect for anything needing a real-time response — **chat assistants, customer service bots, or coding copilots**.

IDEs like Cursor have already added support for it — to make your development even smoother with quicker responses.

Being significantly faster than Sonnet 4.5 definitely makes for a much snappier user experience. Near-premium performance for much less.

This is a model that can be deployed at massive scale — making powerful AI more accessible for anyone looking to manage costs — even powering free-tier user experiences.

And the new Extended Thinking ability makes it even better.

With Extended Thinking you can boost complex reasoning and get even better accuracy with your coding — thinking harder to handle multi-step problems that would have previously required a slower, larger model like Sonnet.

Haiku 4.5 even supports Computer Use — which means it can autonomously use your PC to carry out complex tasks.

Manage files, fill out forms, handle spreadsheets and email... the possibilities are endless.

Haiku 4.5 processes both text and images and supports large context windows — up to ~200k tokens with generous output — making it suitable for managing full codebases and image-assisted tasks.

It also scores very well in industry benchmarks like SWE-bench.

Its speed also makes it ideal for **multi-agent systems**, where a larger model like Sonnet 4.5 could handle the overall plan, and multiple Haiku agents execute the parallel subtasks — like code refactoring or document analysis.

Claude Haiku 4.5 is a major step forward in making high-performance AI more affordable and deployable.

It's not just a smaller sibling to Sonnet — it's a purpose-built engine for real-time intelligence.

With new features like Extended Thinking and Computer Use it bridges the gap between speed and sophistication — giving you high-power state-of-the-art intelligence at scale.

• • •

MCP Server Cheatsheet

Supercharge your AI workflow with this powerful MCP Server Cheatsheet — 30+ MCP servers containing several dozens of tools, curated across memory, accuracy, cloud, and data, and much more to help you develop astronomically smarter and faster.

[Get a free copy here today.](#)

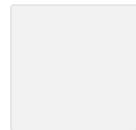
Programming

Software Development

Artificial Intelligence

Software Engineering

Technology



Follow

Published in Coding Beauty

21K followers · Last published 23 hours ago

Latest news, growing trends, and unique insights in the world of tech, AI, and computing.



Follow



Written by Tari Ibaba

93K followers · 13 following

Latest news, growing trends, and unique insights in the world of tech, AI, and computing.

Responses (3)





Bgerby

What are your thoughts?



Andrew Walsh MD, PhD.

11 hours ago

...

The WORST Model Anthropic has ever made!

<https://www.youtube.com/watch?v=VgaypFe2C7Q>



2

[Reply](#)



mohamad shahkhajeh

5 days ago

...

Haiku 4.5's integration with Cursor for faster responses seems particularly useful for real-time coding workflows.



1

[Reply](#)



Thomas Møller Jørgensen

3 days ago

...

What is the increased cost in terms of carbondioxide per question, on average?



1 reply

[Reply](#)

More from Tari Ibaba and Coding Beauty

 In Coding Beauty by Tari Ibaba

This Secret New Coding Model Just Shocked The World

The crazy thing is nobody knows exactly who's behind it.

 Sep 22  732  26



...

 In Coding Beauty by Tari Ibaba

Microsoft's new layoffs just confirmed every programmer's worst nightmare

The news hit hard.

Jun 6 2K 143



...

In Coding Beauty by Tari Ibaba

This new AI tool from Google just destroyed web & UI designers

Wow this is absolutely massive.

Sep 18 108 3



...

In Coding Beauty by Tari Ibaba

The secret code Google uses to monitor everything you do online 😲

This is so sneaky!

◆ Sep 13 702 17



...

[See all from Tari Ibaba](#)

[See all from Coding Beauty](#)

Recommended from Medium

 aakash

Perplexity Just Unleashed 10 FREE AI Agents That Do Your Entire Job (The “Comet” Shortcut)

Stop what you are doing. The era of the simple AI search engine is officially over.

◆ Oct 7 663 19



...



Alex Suzuki

Cretaceous Software Engineering—Why I don't use AI to write code.

👋 Hi! I'm a Cretaceous Software Engineer.

6d ago 641 28



...



In MeetCyber by Jules May

Vibe coding: the antidote

LLMs have shown themselves to be a terrible way to write programs. But the problem they address is real, and we aren't solving it any other...

 5d ago

 438

 29



...

 Ignacio de Gregorio

Can AI Solve Problems They Haven't Seen Before?

The Eternal Question, Answered.

 6d ago

 972

 20



...

 The CS Engineer

Forget JSON—These 4 Data Formats Made My APIs 5x Faster

A 1.2 KB JSON payload transformed a smooth request into a 120 ms wait. Switching formats cut that to under 20 ms for some endpoints.

◆ Sep 29 ⌘ 1.1K 🗣 32



...

 Akshay khale

Open source projects that got me excited

Discover Deno, Tauri, Ollama, Zod, Zen, AstroJS & Requestly—and why they matter

◆ Oct 11 ⌘ 84 🗣 1



...

[See more recommendations](#)