

Distill Launches \$4/Month Proxy for Cheaper Claude Code and Codex Workflows in Public Beta

One-line proxy from MyDream Labs helps reduce AI coding costs, wait time, and generated code volume while keeping existing provider credentials.

SHERIDAN, WY, UNITED STATES, July 1,

2026 /EINPresswire.com/ -- AI coding

agents and agentic workflows are

moving more developer work onto

usage-based model APIs, making

output-token waste a direct cost issue

for individuals and teams. MyDream

Labs today launched Distill, a drop-in

optimization proxy for Claude Code

and OpenAI Codex. In a 16-task internal

[benchmark](#), Distill measured reductions of up to 50.6%

in output-token use, up to 55.3% in wait time, and up to 30.7% in generated code volume.

Average reductions and full methodology are published on the Distill benchmark page.

“

Waste from vibe coding compounds fast — every unnecessary token is money now, every unnecessary line is maintenance cost later.”

*Viacheslav Bogdanov,
Founder, MyDream Labs*

As more AI coding tools move toward usage-based billing, independent developers and small teams have less room for inefficient agent output. For developers running agents on personal projects or client work, the cost lands directly on a personal or small-business card.

The economics of LLM APIs compound the issue: output tokens cost developers 5x more per token than input. The models behind today's coding agents were originally built for conversational AI — where thorough, multi-angle

answers are exactly what users want. That training is the right behavior for a chat interface and the wrong behavior for a coding agent, which needs one effective solution executed cleanly, not a decision tree. Distill optimizes around both: the verbosity that comes from chat-model training and the cost of output tokens.

Distill.codes



**Cheaper. Faster. Cleaner
AI coding**

for Codex & Claude Code

Distill.codes routes Claude Code and Codex through an optimization proxy for cleaner, cheaper AI coding.

Distill sits between the coding agent and the model provider, supporting both the Anthropic and OpenAI API formats. Unlike model-switching tools or agent replacements, Distill works transparently inside the existing Claude Code or Codex workflow with no changes to how the agent or model is used. Developers connect using their existing Anthropic or OpenAI account — requests are billed exactly as if the agent called the provider directly, without request logging or vendor lock-in.

The effect compounds across sessions: each task that generates less code leaves the next session with a leaner codebase, a faster agent, and a smaller bill.

[Public beta](#) highlights:

- Drop-in setup: one configuration line for Claude Code or Codex
- Works through your existing Anthropic or OpenAI account — no separate credentials to manage
- No request logging — code and prompts are not retained by Distill
- \$4/month for individuals, \$40/month for teams
- 3-day free trial

"Waste from vibe coding compounds fast — every unnecessary token is money now, every unnecessary line is maintenance cost later. Distill was how I fixed my own AI coding spend without changing how I work," said Viacheslav Bogdanov, founder of MyDream Labs.

"Before Distill, I was constantly asking the agent to simplify — cut this function, don't over-engineer, stay focused. With Distill the code comes out clean. I barely have to ask anymore," said one early beta tester.

Full internal results and benchmark methodology are published on the Distill benchmark page. Until a reproducible public benchmark is available, the most accurate test is the developer's own codebase and tasks. Results vary by task complexity; longer and more complex tasks tend to show larger gains.

The public beta is available today at <https://distill.codes>, with a 3-day trial built for side-by-side comparison on the developer's own work.

About Distill.codes / MyDream Labs LLC

Distill.codes is operated by MyDream Labs LLC. Its first product, Distill, is a drop-in proxy and developer tool for Claude Code and OpenAI Codex built for independent developers, freelancers, and small engineering teams who want agent speed without unpredictable token bills. The product was created by Viacheslav Bogdanov, a full-stack engineer with 25+ years in software development across web, mobile, and artificial intelligence products, including mobile apps with

10M+ downloads and LocalHelper.AI, a local voice assistant built on Whisper.cpp and Llama.cpp.

Viacheslav Bogdanov

MyDream Labs LLC

press@distill.codes

This press release can be viewed online at: <https://www.einpresswire.com/article/923594278>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2026 Newsmatics Inc. All Right Reserved.