

Kimi K3 Setup Guide

Set up Kimi K3 from Moonshot AI as the default model in Claude Code, Codex CLI, Pi, Aider, Cursor, or any OpenAI-compatible harness — in under 5 minutes.

Codex CLI

Pi Agent

Aider

Cursor

OpenRouter

3x

CHEAPER THAN CLAUDE

26.7%

LEGAL BENCHMARK (VS CLAUDE 14.2%)

9-10x

CHEAPER WITH PLANNER/WORKER SETUP

CREATED BY

Vikash Kumar

[linkedin.com/in/vikashyavansh](https://www.linkedin.com/in/vikashyavansh)

WHY THIS MATTERS

The first open source model to beat closed source

Kimi K3 from Moonshot AI just went open source. The weights are public. 20+ inference providers will be running it within days. Here is why you should be using it in your stack right now.

#1

Front-end code — beats every closed source model

26.7%

Legal tasks vs Claude at 14.2% — nearly double

3x

Cheaper than Claude per token

2.8T

Parameters — almost 2x DeepSeek V4 Pro

THE REAL METRIC

Cost per token is not what matters. Cost per task is. Kimi K3 completes the same tasks as Claude in fewer turns for 3x less per token — making the real cost difference even larger in practice.

STEP 1

Get your API access

Option A — OpenRouter (Recommended for all harnesses)

OpenRouter is an OpenAI-compatible API router that gives you access to Kimi K3 and 200+ other models through one key. Every harness in this guide supports it.

CREDENTIALS

```
# Sign up at openrouter.ai → Keys → Create Key
Base URL : https://openrouter.ai/api/v1
Model ID : moonshotai/kimi-k3
API Key  : your-openrouter-key
```

Option B — Moonshot Platform (Direct)

Use this if you want to build directly on the Moonshot platform or access Kimi-specific features.

CREDENTIALS

```
# Sign up at platform.moonshot.ai → API Keys → Create Key
Base URL : https://api.moonshot.cn/v1
Model ID : kimi-latest
API Key  : your-moonshot-key
```

Setup by harness

Codex CLI

TERMINAL — ADD TO ~/.ZSHRC OR ~/.BASHRC FOR PERMANENT SETUP

```
export OPENAI_BASE_URL="https://openrouter.ai/api/v1"
export OPENAI_API_KEY="your-openrouter-key"
```

Run with Kimi K3

```
codex --model "moonshotai/kimi-k3"
```

Pi Agent

~/.PI/CONFIG.YAML

```
model: moonshotai/kimi-k3
base_url: https://openrouter.ai/api/v1
api_key: your-openrouter-key
effort: high
```

Aider

COMMAND LINE

```
aider \
  --model openrouter/moonshotai/kimi-k3 \
  --openai-api-key your-openrouter-key \
  --openai-api-base https://openrouter.ai/api/v1
```

~/.AIDER.CONF.YML — PERMANENT SETUP

```
model: openrouter/moonshotai/kimi-k3
openai-api-key: your-openrouter-key
openai-api-base: https://openrouter.ai/api/v1
```

Cursor

Settings → Cursor Settings → Models → Add Custom Model

VALUES TO ENTER

```
Model name : moonshotai/kimi-k3
Base URL   : https://openrouter.ai/api/v1
API Key    : your-openrouter-key
```

Any OpenAI-Compatible Harness

These three variables work for any harness that accepts OpenAI-compatible configuration.

```
UNIVERSAL CONFIG

BASE_URL : https://openrouter.ai/api/v1
API_KEY  : your-openrouter-key
MODEL    : moonshotai/kimi-k3
```

STEP 3 — THE OPTIMAL SETUP

Planner / Worker configuration

Do not use Kimi K3 for everything. Use it as the planner. This is how you get the same output quality at 9 to 10x lower cost, based on Cursor's experiment rebuilding SQLite from an 835-page manual.



Result: Cursor's test showed this combination rebuilds complex projects (like SQLite from scratch) at 9 to 10x lower cost than using a single powerful model for everything. Same output. Fraction of the spend.

REFERENCE

When to use Kimi K3

USE KIMI K3

- Front-end development (best in class)
- 3D design and modeling tasks
- Legal document review (26.7%)
- Complex multi-step planning
- Architecture decisions
- Any task needing deep reasoning

USE A SMALLER MODEL

- Simple file edits
- Repetitive execution tasks
- Quick code snippets
- Tasks not requiring planning
- High-volume, low-complexity work
- Anything the worker handles well

Quick Reference

Harness	Base URL	Model ID
Codex CLI	openrouter.ai/api/v1	moonshotai/kimi-k3
Pi Agent	openrouter.ai/api/v1	moonshotai/kimi-k3
Aider	openrouter.ai/api/v1	moonshotai/kimi-k3
Cursor	openrouter.ai/api/v1	moonshotai/kimi-k3

Verify it is working

TEST PROMPT — RUN THIS IN YOUR HARNESS

"Who are you and what model are you running on?"

Expected: Kimi will identify itself as Kimi, made by Moonshot AI.
If it says Claude or GPT, check your BASE_URL and MODEL settings.

Ready to build with Kimi K3?

Follow these steps, set up in under 5 minutes, and start using the best open source model available today in your existing workflow.

- 1 Get your OpenRouter key at openrouter.ai
- 2 Configure your harness using the code in this guide
- 3 Run the verification prompt to confirm it works
- 4 Set up the planner/worker split to cut costs 9-10x

Vikash Kumar

Founder, BULDRR AI · Claude Code and n8n Automation Expert · \$1.44M saved for clients

[Connect on LinkedIn](#)

\$1.44M

SAVED FOR CLIENTS

20K+

LINKEDIN FOLLOWERS

28mo

BUILDING DAILY