

# Cursor AI Pricing Explained

7 min read · May 21, 2025



Laurent Kubaski

Follow



Listen



Share



More

July 2025 update: the Cursor pricing model changes a lot and as I was expecting, what I wrote back in this article back in May 2025 is not 100% accurate anymore. However it seems like Cursor messed up big time with one of their recent June 2025 pricing changes -> for more details, see this official statement from Cursor: [Clarifying Our Pricing](#) and also [What Happened To Cursor?](#)

OK, back to what I wrote back in May 2025:

## 0. Introduction

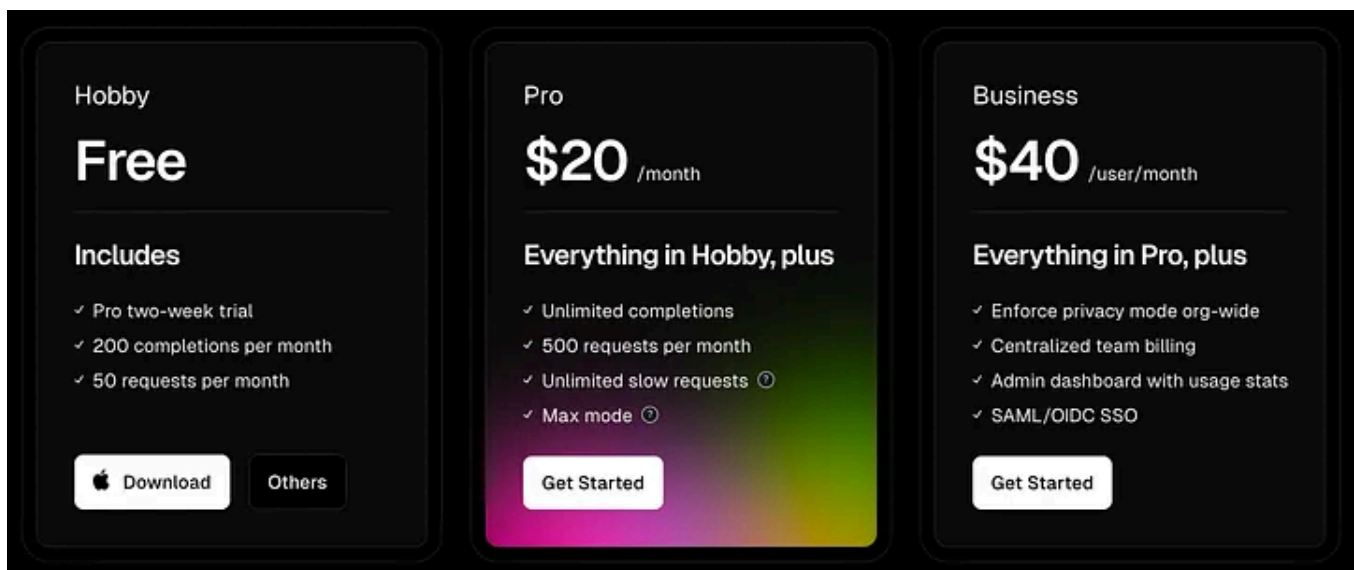
The [Cursor AI Code Editor](#) is an impressive piece of technology to say the least... but what's not so impressive is the way the Cursor pricing model is explained.

So if you're confused with what completions, requests, slow requests, free requests, premium models, free models mean and how much they cost, you're in the right place: I was initially confused too, which is why I've created this article to share my findings... you're welcome!

**Note:** the Cursor pricing model seems to change a lot so if you're reading this in the future, the monthly amounts I'm describing here may not be valid anymore... however the Cursor features I'm about to describe are unlikely to change.

## 1. Cursor Business Plans

Let's start with the [3 plans](#): Free, Pro & Business.



Cursor Pricing Page — 30th May 2025 ([source](#))

Those come with a certain number of *completions*, *requests* and *slow requests*... what are those?

### 1.1. Completions

This is what gets consumed when you use Cursor tab autocomplete feature.

For the Pro and Business plans, you get unlimited completions, however for the Free plan, Cursor contradicts itself:

- The pricing page states that you get 200 completions/month (see screenshot above)
- The tab completion documentation states that you get 2000 suggestions/month (see screenshot below)

So: is it 200/month or is it 2000/month? ([source](#))

### 1.2. Requests

This is what gets consumed when you use Cursor chat feature in combination with a *Premium* model. That last part is important because Cursor also gives access to *Free* models that don't consume requests (more on this later on).

There are 3 types of requests:

### **Type #1: Fast Requests for Premium models (see documentation)**

- The Free plan includes 50/month but only when using the GPT 4.1 Premium model.
- The Pro & Business plans include 500/month (the Pro **trial** plan only includes 150 though).
- When you've consumed all of them, you start using *Slow* requests: you can enable usage-based pricing to continue using Fast requests.

The Free plan includes 50 GPT 4.1 Fast request (source)

### **Type #2: Fast Requests for Free Models**

- The Free plan includes 500/month.
- The Pro & Business plans include an unlimited number.
- See one of the following chapter to understand what a "Free Model" is.

### **Type #3: Slow Requests (see documentation)**

- The Free plan doesn't mention how many are included.
- The Pro & Business plans include an unlimited number.

- You start using Slow requests when you've consumed all the Fast requests that are included by default in your plan.
- Slow requests are free... but they're slower... hence their name... get it? :)
- The Cursor documentation also calls those free requests: this is confusing because you also have "Fast Requests for Free Models"

### **1.3. Usage-based pricing**

When you've used all the Fast requests that are included in your plan, you can switch to usage-based pricing. This means that you can now use as many Fast requests as you want... but you pay for each one of those!

As I'm writing this, each additional Fast request will cost you \$0.04: that's exactly the price of the Fast requests that are included by default in the Pro plan (\$20 for 500 requests = \$0.04 per request).

### **1.4. Use your own API key**

An alternative to paying a fixed price of \$0.04 for every Fast request, is to use your own API keys: this way you pay the same price as if you were directly using the APIs of the various LLM providers like OpenAI or Anthropic.

If like me you like to use OpenRouter (this allows you to use the Open AI SDK with any model of your choice), here is how to configure the Cursor IDE:

The main problem with this approach is that **this will prevent you from using a lot of the Cursor IDE features**. For example, the agent mode doesn't work if you use your own keys:

#### **Agent mode doesn't work if you use your own API key**

Anyways: after entering my own OpenRouter API key, I've enabled the "Ask mode" and then asked Cursor to explain what a basic function in my project was doing. As expected, I **did** not consume any Fast request, and I ended up paying \$0.003 to OpenRouter for that.

#### **OpenRouter dashboard showing the Cursor request**

Yes, \$0.003 is way cheaper than the fixed Cursor Fast request price of \$0.04 BUT in this case I only consumed a very limited number of tokens: for larger requests, a fixed price of \$0.04 no matter how many tokens are used is certainly cheaper.

#### **1.5. Premium models vs Free models**

The number of Fast requests that get consumed depend on which model you use. For example on the screenshot below, you can see that Claude 3 Opus consumes 2.5 Fast requests for each interaction:

#### Number of Fast requests consumed by each model

But wait: as you can see in the screenshot above, some models — like “Cursor Small” — do not consume any Fast request: these are **Free models**.

As I’m writing this, Cursor supports the following Free models: Cursor Small, Deepseek v3, Gemini 2.5 Flash, GPT-4o mini (limited to 500 requests/day with the Free plan), Grok 3 Mini Beta,

## 2. How To Use A Free Model

On the [Cursor model page](#), find a free model, then hover your mouse on the copy-paste icon and click on it: this copies the internal name of the model (on the screenshot below it’s “cursor-small”) to the clipboard.

#### Finding the name for a free model

Then in the Cursor IDE, open the Cursor Settings page and if you don't see your Free model in the list, click "Add model", then paste the value you copied and hit "enter": the Free model should now be in the list (make sure that the checkbox is selected in front of it).

#### **Adding a model to the Cursor IDE**

Finally, in the Cursor IDE Chat window, open the model switcher, disable the "Auto" option and select your Free model in the list:

Additional notes:

- Obviously some of the Free models will not perform as well as the Premium models.
- AFAIK, Cursor doesn't document if there is a limit on how many times you can use a Free model.

### **3. Tracking Usage**

#### **3.1. Settings page**

The *first* location where you can track usage is the settings page:



- This page is **not** available when using the Free plan (it redirects to the dashboard page described in the next chapter)
- This page displays:
  - how many Fast requests have been consumed by Premium models and by Free models.
- This page does **not** display:
  - how many Fast requests have been consumed by usage-based pricing.
  - how many tab completions have been consumed.

Usage tracking on settings page when using a Pro **Trial** plan ([link](#))

### **3.2. Dashboard page**

The *second* location where you can track usage is the dashboard:

- This page displays:
  - how many Fast requests have been consumed by Premium models.
  - when the number of Fast requests is reset.
  - how many Fast requests have been consumed by usage-based pricing.
- This page does **not** display:
  - how many Fast requests have been consumed by Free models.
  - how many completions have been consumed.

Usage tracking on account dashboard when using a **Free** plan ([link](#))

## 4. To Recap

Note: all numbers are “per month” and are as of May 31th, 2025.

Feature	Free plan	Pro/Business plan
Tab completion	200 or 2000 ?	unlimited
Fast request (Premium model)	50 (with GPT 4.1)	500
Fast request (Free model)	500	unlimited.
Slow request	<not documented>	unlimited

Artificial Intelligence

Cursor



Follow

**Written by Laurent Kubaski**

108 followers · 3 following

I'm a product manager at Salesforce, more info here: <https://kubaski.com/>

## Responses (2)



Bgerby

What are your thoughts?



The Effective Project Manager

Jul 5



I tried a model and it told me I can only use GPT4.1 on the free tier. So I used that. After about 5 requests it told me I had reached my limit. So I went back to the Auto setting. It feels like the Auto setting is not smart. It can write code but... [more](#)



1

[Reply](#)



p0l

Jun 1




really interesting!



[Reply](#)

### More from Laurent Kubaski


 Laurent Kubaski

## MCP Resources explained (and how they differ from MCP Tools)

This article is part of my MCP Explained series and a follow up to my MCP Prompts Explained article.


Aug 1  71  3



 Laurent Kubaski

## Ollama prompt templates

This article is part my “Ollama Explained” series.


 Laurent Kubaski

## Connecting your Python MCP Server to Copilot Chat in Visual Studio Code

This article is part of my MCP Explained series.

Jun 20 🖱️ 5

🔖+ ⋮

 Laurent Kubaski

## Ollama endpoints options parameter

This article is part my “Ollama Explained” series.

Mar 16  37  1



See all from Laurent Kubaski

## Recommended from Medium




Edwin Lisowski

### **Claude, Cursor, Aider, Cline, Copilot: Which Is the Best One?**

1. Claude (Anthropic)

Jul 10  170  5



 Vikas Ranjan

## Using Cursor IDE Like a Pro: My Personal Guide to Building, Debugging, and Staying Sane

“Cursor is like having a junior engineer at your side—the better you train it, the better it performs.”

May 26  38  2



 Sale Skela

## Augment Code AI: A Long-Term Review (and Why I’m Keeping It in My Toolbox)

Alright, let's talk more about Augment Code AI Coding Agent — the coding assistant that's slowly turning me into a lazy (but efficient)...

May 24  131  2



BillfordX

## Stop Blaming Vibe Coding For Your Problems

Here's a cheat sheet if you want to dive into something vibe-coded and properly done.



Oct 15  4  1







In Software, AI and Marketing by Madhukar Kumar

## How to Build a Full-Stack App with an AI Coding Agent

A hands-on guide to creating a modern web application using agentic AI development

Jun 24  147  6



Arihant Tripathy

## Cursor vs Copilot: Let the Results Speak

I have been a long-time user of Github Copilot, ever since it became available through the Github Student Developer Pack, where it felt...

Jul 17  3



See more recommendations