

# Scrape any website with AI

Pull clean, structured data off a web page by describing what you want in plain English. No coding background needed to follow this.

## WHAT IT IS

### A free tool that reads web pages for you

ScrapeGraphAI visits a web page and pulls out exactly the information you ask for. You write the request the way you would say it out loud — "get me the product name, the price, and every review" — and it hands back a tidy list you can drop straight into a spreadsheet.

## WHAT IT DOES

### Turns messy pages into organised data

- Reads a page and extracts the fields you name
- Works across many pages at once, or straight from search results
- Handles product pages, reviews, directories, listings, contact pages
- Returns JSON — the format spreadsheets and automations expect
- Runs on your own computer, so your data stays with you

## WHAT IT REPLACES

### The copy-paste afternoon, or a paid scraping service

The usual options are paying a scraping service per page, or opening forty tabs and copying fields by hand. The free route used to mean writing code aimed at one exact page layout, which broke the moment that site redesigned. This skips all of it. You pay only for the AI reading each page — roughly a cent or two per page.

## BEFORE YOU START

### You need Python 3.12 or newer

Python is the free programming language this tool is built on. The version matters: ScrapeGraphAI refuses to install on anything older than 3.12, and most computers ship with an older one.

- **To check:** open Terminal (Step 1 below), type `python3 --version` and press Enter. If the number is below 3.12, or you get an error, download the latest version from **python.org** → **Downloads**, run the installer, then close and reopen Terminal.

## STEP 1

### Open Terminal

Terminal is the app for typing instructions to your computer directly. Everything below happens inside it.

#### 1 On Mac: press **⌘ + Space**, type **"Terminal"**, press Enter

On Windows: click Start, type "PowerShell", press Enter. A plain black-and-white window opens with a blinking cursor and nothing else. That is correct — it is meant to look empty.

## STEP 2

### Install the tool

#### 1 Paste this in and press Enter

```
pip install scrapegraphai
```

Dozens of lines scroll past for a minute or two. Normal — it is pulling in the parts it depends on. Wait until the cursor comes back.

#### 2 Then paste this one and press Enter

```
playwright install
```

This adds the invisible browser the tool uses to open pages. Skipping it is the most common reason people hit an error later.

## STEP 3

### Get your AI key

The tool needs an AI model to do the reading. The simplest option is OpenAI.

#### 1 Go to **platform.openai.com** → **sign up** → **Settings** → **API keys** → **Create new secret key**

Copy it immediately — the site will not show it again. You also need a payment card under Billing before it works.

#### 2 Tell your computer the key — replace the text inside the quotes with your own

```
export OPENAI_API_KEY="sk-your-key-goes-here"
```

This key is yours alone and works like a password attached to your card. Never paste it into a post, a screenshot, or a message to anyone.

## Run your first scrape

### 1 Paste this entire block into Terminal and press Enter

```
cat > scrape.py << 'EOF'
import json, os
from scrapegraphai.graphs import SmartScraperGraph

graph = SmartScraperGraph(
    prompt="Extract the business name, phone, and email.",
    source="https://example.com",
    config={"llm": {
        "api_key": os.environ["OPENAI_API_KEY"],
        "model": "openai/gpt-4o-mini",
    }, "headless": True},
)
print(json.dumps(graph.run(), indent=2))
EOF
```

This saves a small file called `scrape.py`. Nothing appears on screen, which means it worked.

### 2 Now run it

```
python3 scrape.py
```

Give it ten to twenty seconds. It opens the page, reads it, and prints your data as a neat list.

### 3 Change the two lines that matter

Swap the address after `source=` for the page you want, and rewrite the sentence after `prompt=` to name whatever fields you need. Those two lines are the whole tool. That's it.

#### GOOD TO KNOW

#### Three things that trip people up

| SITUATION                                                | WHAT TO DO                                                                                                                                                                                                                                         |
|----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| You want it free, with no card                           | Install Ollama from <a href="https://ollama.com">ollama.com</a> , then change the model line to <code>"model": "ollama/llama3.2"</code> and delete the <code>api_key</code> line. Slower and less accurate, but it costs nothing and runs offline. |
| Fields come back named differently each run              | Add a schema so every run returns identical keys. Without one the wording drifts between runs, which quietly breaks any spreadsheet you feed it into.                                                                                              |
| You point it at a file you saved rather than a live page | Pass the page's actual contents, not the filename. Handing it a filename fails silently and returns empty fields rather than an error.                                                                                                             |

#### BEFORE YOU SCRAPE AT SCALE

Each page costs a fraction of a cent and takes ten to twenty seconds, so a thousand pages is real money and several hours — test on five first. Check a site's terms before bulk-scraping. And if you are collecting contact details for outreach, personal email addresses are regulated in the EU and UK: you need a lawful basis and a working opt-out.