Codex CLI

- Usage with Ollama
  - Quick setup
  - Manual setup
  - Profile-based setup
- Web search through local Ollama

Install the Codex CLI. For the ChatGPT desktop app, see ChatGPT Desktop.

```
npm install -g @openai/codex
```

Usage with Ollama

Use a context window of at least 64k tokens for Codex.

Quick setup

```
ollama launch codex
```

When launched through `ollama launch codex`, Ollama refreshes the model catalog and uses a dedicated Codex profile for that session.

Type `/model` inside Codex to switch between your installed, tool-capable Ollama models without restarting. The starting model remains selected, and each model keeps its own context window, vision, and thinking settings. Pull additional models with `ollama pull <model>` before launching to include them in the picker.

To configure without launching:

```
ollama launch codex --config
```

To remove the Ollama launch profile and generated model catalog:

```
ollama launch codex --restore
```

Manual setup

To use `codex` with Ollama, use the `--oss` flag:

```
codex --oss
```

To use a specific model, pass the `-m` flag:

```
codex --oss -m gpt-oss:120b
```

To use a cloud model:

```
codex --oss -m gpt-oss:120b-cloud
```

Profile-based setup

For a persistent Codex CLI configuration, create `~/.codex/ollama-launch.config.toml`:

```
model = "gpt-oss:120b"
model_provider = "ollama-launch"
model_catalog_json = "/Users/you/.codex/model.json"

[model_providers.ollama-launch]
name = "Ollama"
base_url = "http://localhost:11434/v1/"
wire_api = "responses"
```

Then run:

```
codex --profile ollama-launch
```

Web search through local Ollama

Codex web-search requests sent through the Ollama profile are executed by Ollama for both local and cloud models. Sign in with `ollama signin` to use the web-search service.

To disable web search for a Codex session:

```
codex --profile ollama-launch -c 'web_search="disabled"'
```

□□□ #2
□ □□□ □□□ 2026-10-31 13:17:47 UTC
□ □□□ □□□ 2026-10-31 13:17:47 UTC