Practical AI Tools and Resources for Research

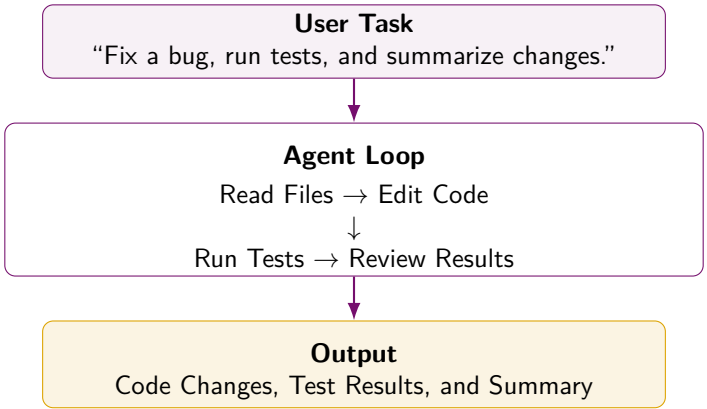
Codex • Cloud GPUs • LLM API

Siqi Rao

The Chinese University of Hong Kong, Shenzhen

2026-10-30

Codex: How Does a Coding Agent Work?



How Is It Different from a Chatbot?

Codex can read and edit project files, run tools and tests, and iterate on results within granted permissions.

Codex: Useful Resources and Tips

AGENTS.md

Project-level instructions:

- Repository structure
- Coding conventions
- Test commands
- Project constraints

Official Website

Skills

Task-specific instructions and resources for repeatable workflows.

Skills Documentation

Official Codex Documentation

Setup, core concepts, common workflows, and best practices.

Codex Documentation

Git Diff + Tests

Review code changes and validate outputs after the agent completes a task.

```
git diff
pytest
```

Tip: Define project rules in AGENTS.md, reusable workflows in Skills, and always review changes.

Cloud GPUs: Platforms and Selection Tips

Platforms and Resources

AutoDL	Website
Alibaba Cloud PAI	Website
Virtai Cloud	Website
AutoDL Guide	Quick Start

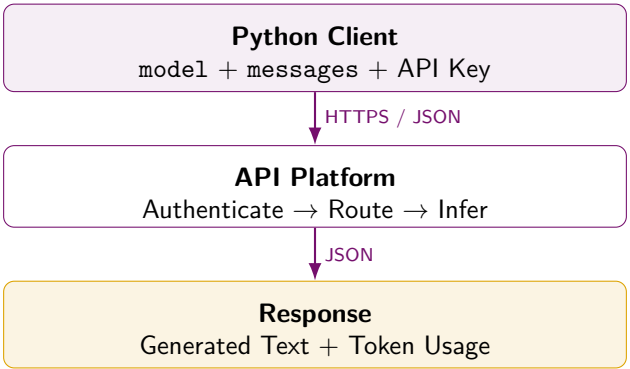
What Should You Check First?

- **GPU Memory**
  Model and batch size
- **Budget**
  On-demand vs. long runs
- **Environment**
  CUDA and frameworks
- **Storage**
  Data persistence

Reminder

Calling a hosted LLM API normally requires no rented GPU.
Local inference, fine-tuning, and training may require one.

LLM APIs: From Python to a Model Response



API Key / Base URL	Model / Messages
Authenticate your request and choose the provider.	Select the model and supply prompts and messages.

Popular LLM API Platforms: China and Beyond

Platform	Highlights	Links
DeepSeek	First-party model API	Documentation
Alibaba Cloud Bailian	Cloud model API services	Documentation
Volcano Engine Ark	Model serving platform	Website
SiliconFlow	Multi-model API access	Documentation
Zhipu AI	GLM model APIs	Documentation
OpenRouter (Global)	Unified multi-provider API	Quick Start

How to Choose a Provider

Start with your existing provider.
Compare model availability, API compatibility, costs, rate limits, and data policies.

Live Demo: Call an LLM with Python

```python
import os
from openai import OpenAI

client = OpenAI(
    api_key=os.environ["DEEPSEEK_API_KEY"],
    base_url="https://api.deepseek.com",
)

response = client.chat.completions.create(
    model=os.getenv(
        "DEEPSEEK_MODEL", "deepseek-chat"
    ),
    messages=[
        {
            "role": "user",
            "content": (
                "Explain overfitting in two sentences."
            ),
        }
    ],
)

print(response.choices[0].message.content)
```

Demo Steps

1. Install SDK
   `pip install openai`
2. Set the API key
3. Run the script
4. Change the prompt

Key Takeaway:

Switching providers often involves changing the base URL, API key, and model.
Check API compatibility.

Note: Verify model IDs against current provider documentation. Never commit or display API keys.

Resources and Takeaways

Documentation and Examples

- Codex
  Official Docs
- AGENTS.md
  Official Website
- AutoDL
  Quick Start
- DeepSeek API
  Official Docs
- OpenRouter
  Quick Start

Three Key Takeaways

1. **Codex**
   A coding agent that can use tools.
2. **Cloud GPUs**
   Choose platforms based on GPU memory, budget, and software environment.
3. **LLM APIs**
   Integrate models with a few lines of Python.

Questions?