



武汉大学
WUHAN UNIVERSITY

地球与空间科学技术学院
School of Earth and Space Science and Technology

AI Agent is All Your Need

胡扬

春季组会·分享

2026年3月12日

- 1. Introduction**
- 2. From LLMs to AI agents**
- 3. GitHub Copilot (学生认证·3年)**
- 4. What kind of us does the future need?**

1. Introduction

现场调研

问题：在大家的日常里，AI 已经被用在了哪些地方？

- AI for science：数学推导、vibe coding、实验设计、学习、论文...
- 日常生活：机器人、看病、产品推荐、炒股、心理治疗、植物辨别、小说
- **趋势：**① 对话框LLMs -> API工具调用自动化 -> 完整自主决策迭代的全流程 Agentic AI
- ② 大语言模型 -> 多模态大模型（文字、语音、图像、视频等更多信息感知能力）

AI Agent ≠ 全自动机器人，而是人机协同的智能合伙人

2026年1月以前的大模型排行

(最近两个月有更多模型发布)

司南OpenCompass 大语言模型官方自建榜单

官方闭源评测集

26 年 01 月榜 ▾

[能力维度说明](#)

[申请榜单数据](#)

[榜单规则](#)

综合评分

语言

知识

推理

数学

代码

工具调用

	模型 ▾	发布日期	类型 ▾	参数量 ▾	均分 ▴ ▾	语言 ▴ ▾	知识 ▴ ▾	推理 ▴ ▾	数学 ▴ ▾
	Gemini-3-Pro-Preview 闭源 · Google	2025/11/19 更新: 2026/2/11	对话	N/A	76.5	82.8	95.5	61.5	71.8
2	Qwen3-Max-2026-01-23 闭源 · Alibaba	2026/1/23 更新: 2026/2/11	对话	N/A	74.4	89.4	90.9	52.5	73.2
3	Claude Opus 4.5 (Thinking) 闭源 · Anthropic	2025/11/1 更新: 2026/2/11	对话	N/A	74.2	85.3	92.2	56.8	69.2
4	Kimi-K2.5 开源 · Moonshot	2026/1/27 更新: 2026/2/11	对话	1000B	73.3	87.3	92.8	54.7	72.9
5	GPT-5.2-2025-12-11 (high) 闭源 · OpenAI	2025/12/11 更新: 2026/2/11	对话	N/A	73.1	82.3	92	60.1	71.4
6	GLM-4.7 开源 · Zhipu AI	2025/12/23 更新: 2026/2/11	对话	355B	72.8	81.8	91.1	50.4	69.7
7	Kimi-K2-Thinking 开源 · Moonshot	2025/11/6 更新: 2026/2/11	对话	1000B	71.1	87.4	88.2	44.1	70.3
8	DeepSeek-V3.2 开源 · DeepSeek	2025/12/1 更新: 2026/2/11	对话	671B	70.8	81	92.5	51.2	68.7
9	Claude Sonnet 4.5 (Thinki... 闭源 · Anthropic	2025/9/29 更新: 2026/2/11	对话	N/A	69.4	85.1	91.2	50.7	68

3月6日发布的GPT-5.4:

	OpenAI		Run with maximum available reasoning effort.		Anthropic	Google
	GPT-5.4 Thinking	GPT-5.4 Pro	GPT-5.3 Codex	GPT-5.2 Thinking	Claude Opus 4.6	Gemini 3.1 Pro
OSWorld-Verified Computer use	• 75.0%	—	74.0%	47.3%	72.7%	—
WebArena-Verified Web browsing	• 67.3%	—	—	65.4%	66.4% Sonnet 4.6	—
GDPval Knowledge work tasks (wins or ties)	• 83.0%	82.0%	70.9%	70.9%	78.0%	—
BrowseComp Agentic browsing	82.7%	• 89.3%	77.3%	65.8%	84.0%	85.9%
SWE-Bench Pro Software engineering	• 57.7%	—	56.8%	55.6%	—	54.2%
GPQA Diamond Expert scientific reasoning (no tools)	92.8%	• 94.4%	92.6%	92.4%	91.3%	94.3%
FrontierMath Advanced mathematics (Tier 1 – 3 & Tier 4)	47.6%	• 50.0%	—	40.7%	40.7%	36.9%
	27.1%	• 38.0%	—	18.8%	22.9%	16.7%
Toolathlon Agentic tool use	• 54.6%	—	51.9%	45.7%	44.8%* Sonnet 4.6, no data reported for Opus 4.6	—

2. From LLMs to AI agents

➡ 从 LLMs 到 AI Agents，到底发生了什么本质变化？

绝大多数人现在接触最多的还是
对话式的大语言模型



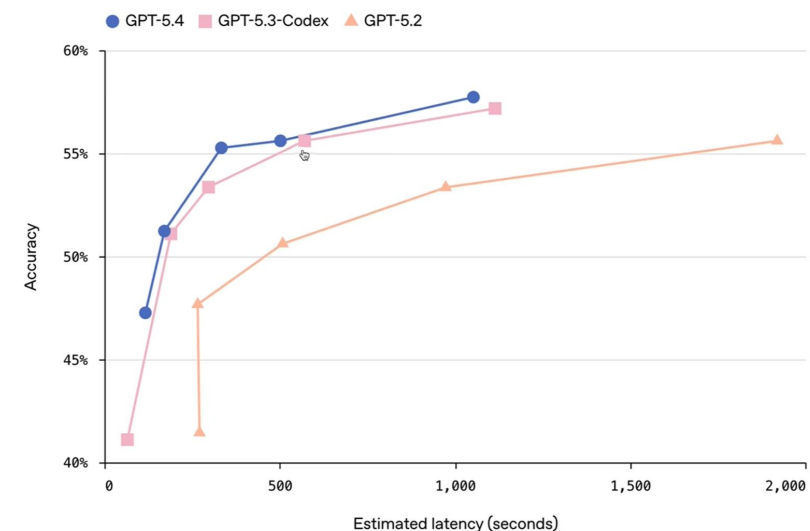
多模态方面：

视觉感知能力

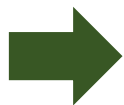
GPT-5.4 改进的计算机使用能力建立在更强的通用视觉感知能力之上：

- **MMMU-Pro** (视觉理解与推理)：GPT-5.4 达到 **81.2%** (GPT-5.2 为 79.5%)
- **OmniDocBench** (文档解析)：GPT-5.4 平均误差 **0.109** (GPT-5.2 为 0.140)

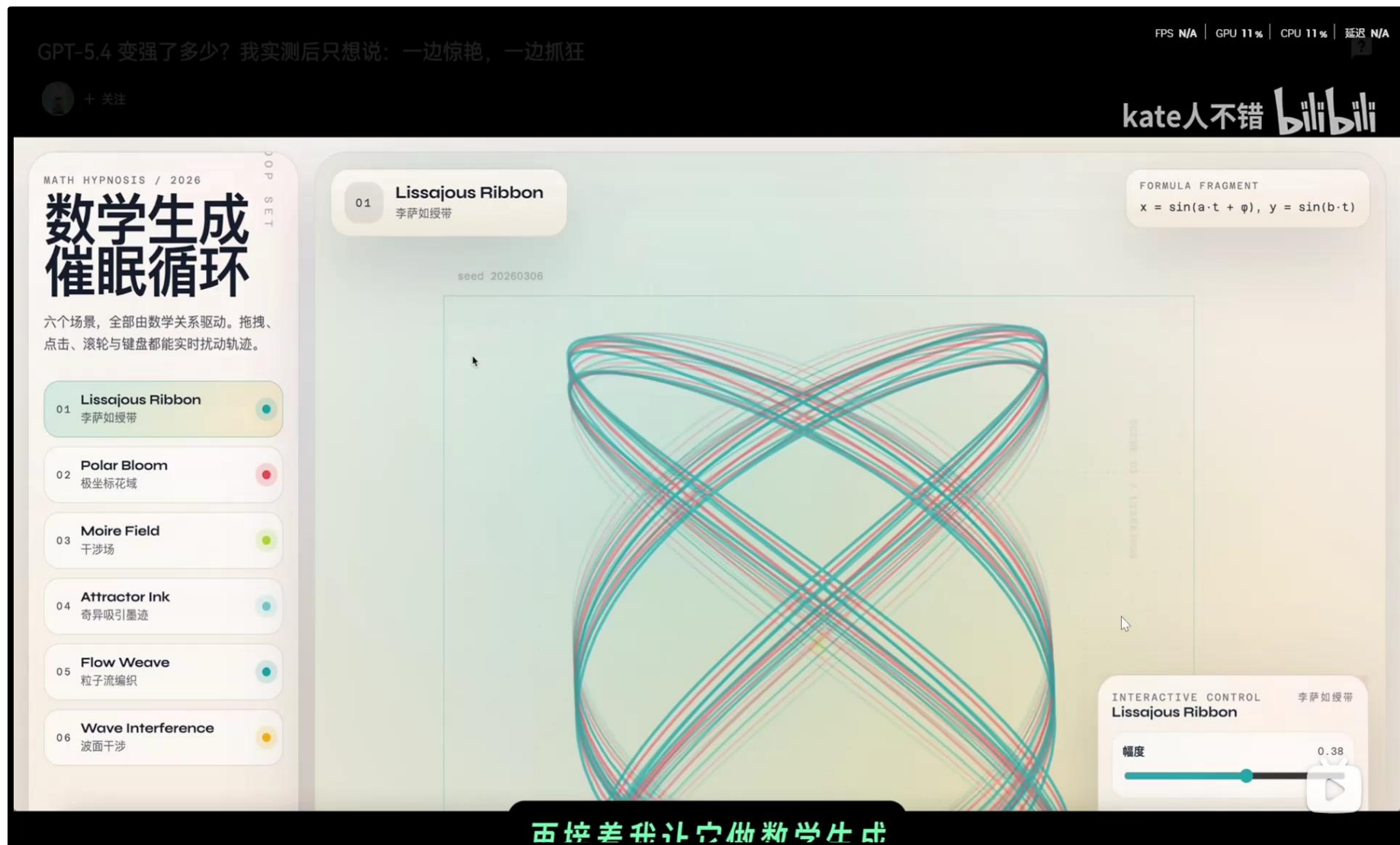
SWE-Bench Pro (public)



使用 GPT 5.4
编写各类项目



远远超过一个对话框
的使用范式!!!



LLM 解决单点问题

vs

AI Agent 实现完整目标

对话式的大语言模型：

让它写一段面波频散反演的代码，它给你一段代码；你让它推导一个波动方程的解析解，它给你一套推导过程。

但执行的主体永远是我们自己——代码报错了要自己排查；数据格式不对要自己调整。

AI Agent：

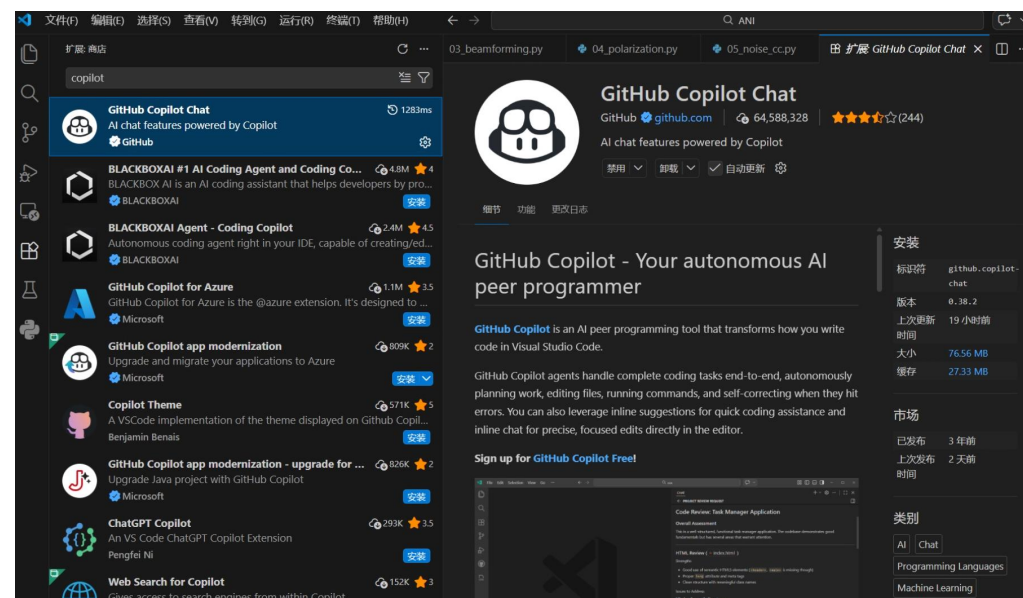
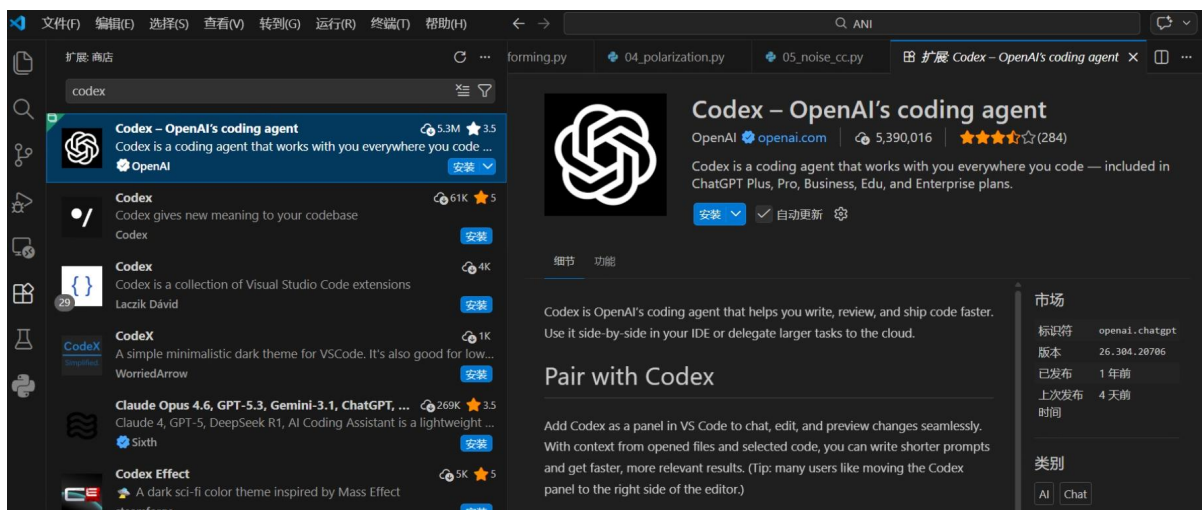
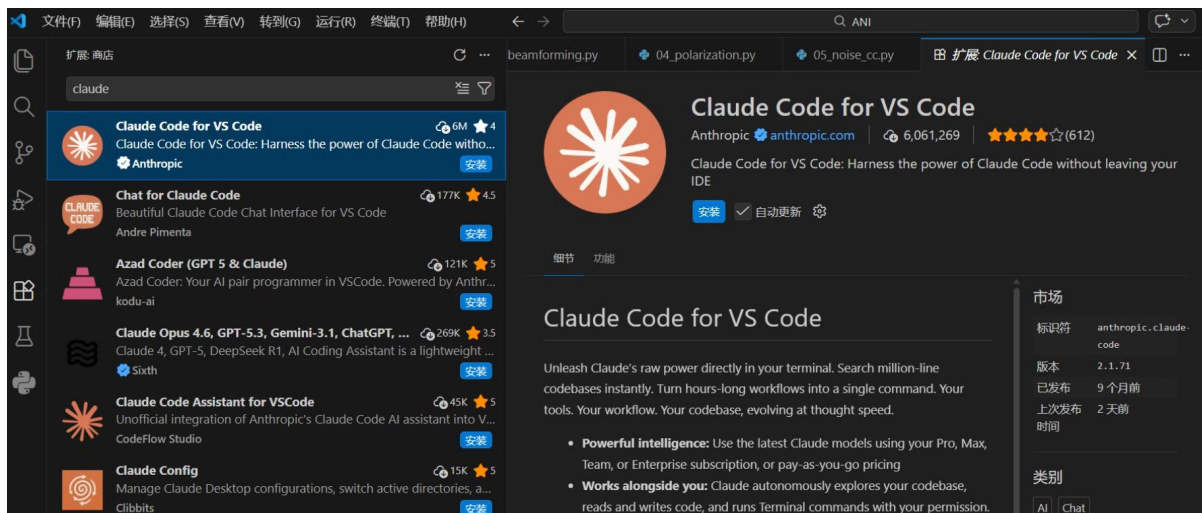
本质上是一个具备了“人类思维闭环”的智能合伙人。

它的核心能力是基于一个终极目标，自主完成任务拆解、工具调用、执行落地、报错排查、迭代优化的全流程，**最终直接交付你想要的结果。**

3. GitHub Copilot (学生认证·3年)

当前最火的几个AI-IDE (Integrated Development Environment, 集成开发环境)
VSCode、Cloud Code、Cursor、Codex

VScode 扩展兼容了绝大多数 vibe coding





TS runner-service.ts × runner-service.test.ts × TS runner-service.ts ×

```
1 import { type Runner } from '@prisma/client';
2 import prisma from '$lib/data/prisma';
3
4 export interface RunnerService {
5   getById(id: number): Promise<Runner | null>;
6 }
7
8 const runnerService: RunnerService = {
9   getById: async (id: number) => {
10     return await prisma.runner.findUnique({
11       where: { id },
12     });
13   }
14 };
15
16 export default runnerService;
```

CHAT

 **Edit with Copilot**

Agent mode

Start your editing session by defining a set of files that you want to work with. Then ask Copilot for the changes you want to make.

Copilot is powered by AI, so mistakes are possible. Review output carefully before use.

 or type # to attach context

 Add Context...  service.md ×

Edit files in your workspace in agent mode

 Agent ▾ Claude 3.5 Sonnet ▾ 

<https://github.com/features/copilot>

本次组会主要推荐的IDE：VScode 和 轻量 AI Agent：GitHub Copilot

一些概念：Spec、Skills 属于提示词工程的核心部分，MCP 是底层协议但落地依赖提示词工程

- **Spec** (Specification, 规格说明)：给 AI 的精准任务说明书。
- **Skills** (技能框架)：AI 完成编程任务所需的具体能力模块（如 Python/ObsPy 代码编写、数值模拟算法实现），也包括作为使用者“引导 AI 调用这些技能”的能力。
- **MCP** (Model Context Protocol, 模型上下文协议)：规范 AI 专属工具 / 数据交互的规则。



OBS 31.0.3 - 配置文件: 未命名 - 场景: 未命名

文件(F) 编辑(E) 视图(V) 停靠窗口(D) 配置文件(P) 场景集合(S) 工具(T) 帮助(H)



46% 缩放至窗口

未选择源

设置

滤镜

场景

场景

源

窗口采集

显示器采集

混音器

麦克风/Aux

桌面音频

转场动画

淡入淡出

时长 300 ms

控制按钮

开始直播

开始录制

启动虚拟摄像机

工作室模式

设置

退出

第二阶段规划

Visual

WPS Office

weixin

QQ

腾讯会议

机械革命电竞

Google

Microsoft

https://www.bilibili.com/video/BV1Abp1z6Eqs/?share_source=copy_web&vd_source=6dd124dde1ddc4f080683a9cca4d6657

Github Copilot 学生认证教程二维码如下:



https://www.bilibili.com/video/BV1WxcuznECw/?share_source=copy_web&vd_source=6dd124dde1ddc4f080683a9cca4d6657

vscode agent 快速使用教程二维码:



学生认证需要用到的材料格式



Ministry of Education Student Status Online Verification

Name: Yang Hu

Institution: Wuhan University

Level: Graduate student

Class: 2501

Major: Solid Geophysics

Student ID: 2025202140004

Duration: 3 years

Type: Regular Higher Education

Mode: Full-time

Student Status: Registered (Expected Graduation Date: July 2028)

4. What kind of us does the future need?

AI 不会取代那些有思想、有创造力、有好奇心的人

Water can flow or it can crash. Be water, my friend. —— Bruce Lee

地学



A global dataset for physics-ML seismic wavefield modeling and FWI

Shiqian Li^{1,3,4}, Zhi Li^{1,2}, Zhancun Mu¹, Shiji Xin¹, Zhixiang Dai⁵, Ruihua Zhang⁵, Kuangdai Leng⁶, Xiaodong Song^{2,1,5}, Yixin Zhu^{3,1,4,5}

1 Institute for Artificial Intelligence, Peking University

2 State Key Laboratory of Deep Earth and Mineral Exploration, School of Earth and Space Sciences, Peking University

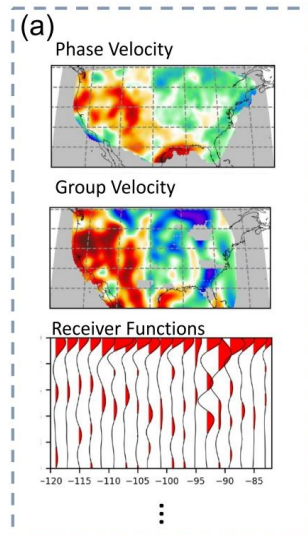
3 School of Psychological and Cognitive Sciences, Peking University

4 Beijing Key Laboratory of Behavior and Mental Health, Peking University

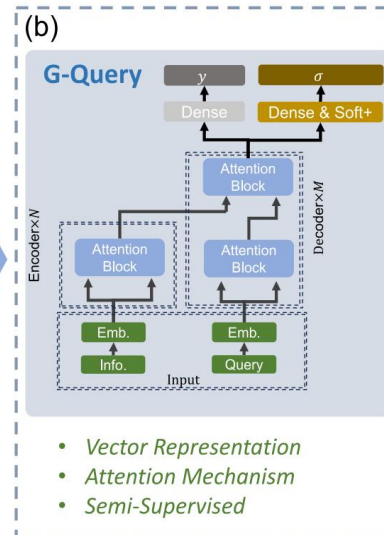
5 NVIDIA

6 Rutherford Appleton Laboratory

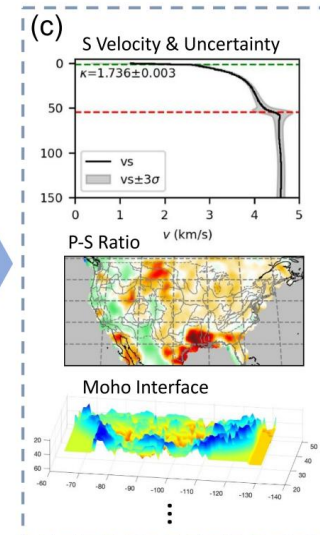
Multimodal Data



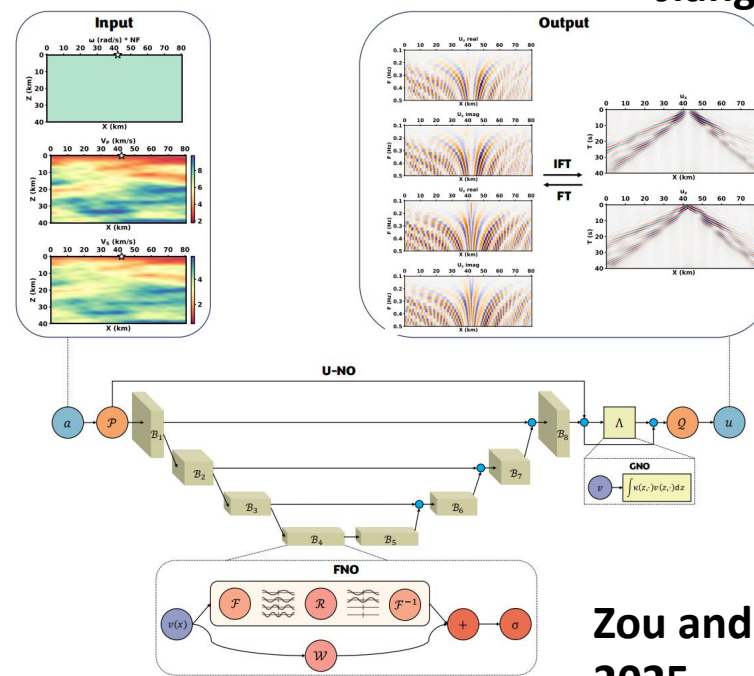
One-fit-all Framework



Multiple Targets



Jiang et al. 2025 JGR



Zou and Ross et al.
2025

8 地震学AI技术交流(1534) ?

群主 于子叶

我们也在做一个地震检测方向的agent, 欢迎大家提出意见, 目前刚写一个原型。🤖
<https://github.com/cangyeone/seismicx-agent>

群主 于子叶

9da8c304-ca90-48eb-8477-31b...
268.20KB

群文件

群主 于子叶

Auto-Model Builder

群主 于子叶

想的是能全流程参数自动优化, 并且使用自然语言来构建模型。

群主 于子叶

AI-Provider Configuration

群主 于子叶

Algorithmic Preferences

常见风险：

- ① AI 生成的地学代码语法正确，但地学参数设置错误
- ② 规避地学数据泄露和代码版权问题：AI 生成的代码，若基于自己的地学专业指令 / 现有开源代码（如 ObsPy 开源库），版权归自己所有；若 AI 直接复制他人受版权保护的代码，需承担侵权责任，因此需对 AI 生成的代码做版权溯源
- ③ AI 幻觉

附录：主流AI的付费情况以及对应权限

Cursor

个人方案

兴趣爱好版
免费版

包括：

- ✓ 无需信用卡
- ✓ Agent 调用次数有限
- ✓ Tab 自动补全次数有限

下载

Pro 专业版
\$20/mo.

包含 兴趣爱好版的全部功能，另外还包括：

- ✓ 更高的 Agent 调用额度
- ✓ 无限制 Tab 自动补全
- ✓ 云端 智能代理
- ✓ 超大上下文窗口

升级到 Pro 专业版

Pro+ 增强版 推荐
\$60/mo.

包含 Pro 专业版的全部功能，另外还包括：

- ✓ 所有 OpenAI、Claude 和 Gemini 模型的使用额度提升至 3 倍

升级到 Pro+ 增强版

Ultra 旗舰版
\$200/mo.

包含 Pro 专业版的全部功能，另外还包括：

- ✓ 所有 OpenAI、Claude 和 Gemini 模型的使用额度提升至 20 倍
- ✓ 新功能抢先体验

升级到 Ultra 旗舰版

按月计费

按年计费

个人方案

兴趣爱好版
免费版

包括：

- ✓ 无需信用卡
- ✓ Agent 调用次数有限
- ✓ Tab 自动补全次数有限

下载

Pro 专业版
\$16/mo.

包含 兴趣爱好版的全部功能，另外还包括：

- ✓ 更高的 Agent 调用额度
- ✓ 无限制 Tab 自动补全
- ✓ 云端 智能代理
- ✓ 超大上下文窗口

升级到 Pro 专业版

Pro+ 增强版 推荐
\$48/mo.

包含 Pro 专业版的全部功能，另外还包括：

- ✓ 所有 OpenAI、Claude 和 Gemini 模型的使用额度提升至 3 倍

升级到 Pro+ 增强版

Ultra 旗舰版
\$160/mo.

包含 Pro 专业版的全部功能，另外还包括：

- ✓ 所有 OpenAI、Claude 和 Gemini 模型的使用额度提升至 20 倍
- ✓ 新功能抢先体验

升级到 Ultra 旗舰版

Claude

Meet Claude Platform Solutions Pricing Resources Login Contact sales Try Claude

Pricing

Explore here

Free

Try Claude

\$0

Free for everyone

Try Claude

- ✓ Chat on web, iOS, Android, and on your desktop
- ✓ Generate code and visualize data
- ✓ Write, edit, and create content
- ✓ Analyze text and images
- ✓ Ability to search the web
- ✓ Memory across conversations
- ✓ Create files and execute code

Pro

For everyday productivity

\$17

Per month with annual subscription discount (\$200 billed up front), \$20 if billed monthly.

Try Claude

Everything in Free, plus:

- ✓ More usage*
- ✓ Includes Claude Code and Cowork
- ✓ Access to unlimited projects to organize chats and documents
- ✓ Access to Research
- ✓ Ability to use more Claude models
- ✓ Claude in Excel

Max

Get the most out of Claude

From \$100

Per person billed monthly

Try Claude

Everything in Pro, plus:

- ✓ Choose 5x or 20x more usage than Pro*
- ✓ Higher output limits for all tasks
- ✓ Early access to advanced Claude features
- ✓ Priority access at high traffic times
- ✓ Claude in PowerPoint (research preview)



Education Benefits

You have a current student coupon applied. Find more information on your benefits [here!](#)

Start an application



Coupon applied

Expires in over 1 year

Verified (benefits available) on October 28, 2025

Application Type: Student

Congratulations!

Your academic benefits, including Partner offers, are now available.

You can access Student Developer Pack offers [here](#) and redeem Copilot Pro via this [link](#).

We can't wait to see what you build with GitHub!

Your benefits will expire on October 28, 2027.

Comparing Copilot plans

The tables below show the features available in each Copilot plan.

	Copilot Free	Copilot Pro	Copilot Pro+	Copilot Business	Copilot Enterprise
Pricing	Not applicable	\$10 USD per month, or \$100 USD per year (free for some users)	\$39 USD per month, or \$390 USD per year	\$19 USD per granted seat per month	\$39 USD per granted seat per month
Premium requests	50 per month	300 per month	1500 per month	300 per user per month	1000 per user per month
Purchase additional premium requests at \$0.04/request	×	✓	✓	✓	✓

Codex与Copilot 相对便宜~10 美元/month