UNIVERSITÄT
BERN

从对话框到智能体：
一个科研人的 AI 工具实践之路

伯尔尼学术交流沙龙 · 2026.06.19

贾林林
瑞士伯尔尼大学计算机科学系模式识别组

联系方式： linlin.jia@unibe.ch

从软件工程师到“yes工程师”

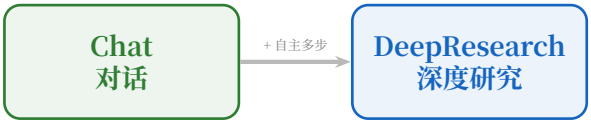
我使用AI的三个阶段

UNIVERSITÄT BERN

Chat
对话

一问一答

我使用AI的三个阶段



我使用AI的三个阶段

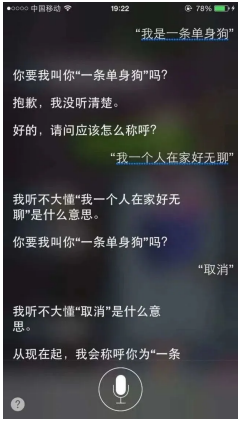
一问一答

自主检索 + 综述

读写文件 / 跑命令 / 调工具

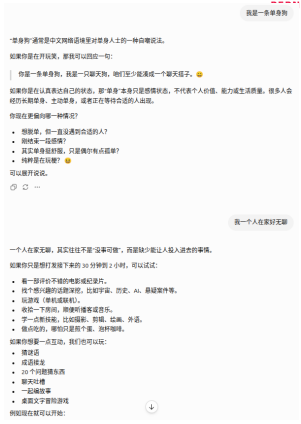
阶段一：Chat 对话

“ChatGPT 时刻”



before 2022.11

从“人工智障”
到“人工智能”



after 2022.11

使用ChatGPT对话框科研

UNIVERSITÄT BERN

提示词：“生成计算图编辑距离的代码”

- 给出**能直接跑的代码**（建图 + 调用）
- 主动给出**形式化定义**：
  GED = 最小 cost 序列（增删点 / 边、替换节点）
- 主动推荐下一步方案

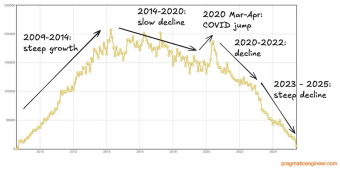
Chat：从“蛮荒时代”到百花齐放

UNIVERSITÄT BERN

2022
一个对话框
→
上传文件
GPTS 市场
→
3.5 → 4 → 4o
每代变强
→
百花齐放
多家可选

ChatGPT　Claude　Gemini　DeepSeek　Qwen

国内外各家模型百花齐放



Stack Overflow 问题数量变化曲线

GPTs：可定制的ChatGPT

闭源模型贵且大；很多垂直LLM科研用**小开源模型反而更合适**。

当前对话框模式的常用场景及问题

UNIVERSITÄT BERN

我的常用场景

- 学知识、查不熟的概念
- 起草、解释、翻译、头脑风暴

存在的问题

- 幻觉：一本正经地编
- 一问一答模式，难以深入；
  不会自己动手，也不一定查最新资料

当前 chat 已融入 agent
能力（调用工具、浏览网页），但仍是“一问一答”

→ 下一阶段：DeepResearch

阶段二：DeepResearch 深度研究

DeepResearch：更深入的调研

从“一问一答”到综合分析

检索 → 阅读 → 交叉验证 → 综述

- 属于广义 agent
- 输出：带出处的报告
- 优势：全面调研、**可人工验证的引用**，减少幻觉
- 各大平台均提供

演示

深度研究模式的常用场景

UNIVERSITÄT BERN

我的常用场景

- 进入一个**全新方向**时，先要一份 survey 框架
  → 数周时间变数天
- 给出材料制作研究报告
- 制作**思维导图**，方便快速建立全局观

前提：**目标要清晰**；它会反问、帮你澄清需求

一个真实的副作用

幻觉转移：模型搜索到，
就感觉“自己会了”。

→ 下一阶段：智能体

阶段三：Agent 智能体

什么是 Agent（智能体）？

UNIVERSITÄT BERN

Anthropic 的区分（Building Effective Agents）：

- **工作流 Workflow**：LLM 与工具按**预设代码路径**编排（可预测、可控）
- **智能体 Agent**：LLM **自己动态决定**流程与工具调用，掌控如何完成任务

核心：增强版 LLM 在**闭环**里自主**规划** → **行动** → **看反馈**，直到完成或触发停止条件；人在**检查点**介入。

来源：Anthropic, Building Effective Agents（2024-12）· https://anthropic.com/engineering/building-effective-agents

Agent：君子动口，智能体动手

提供目标和权限，人只需审阅。

- 能读写文件、跑命令、调用工具，形成**闭环**
- 用 **skill**（markdown 规则 + 可带代码）扩展能力
- 更进一步**内化**成 harness：大模型当大脑 + 简单循环
- 模型够强是关键：到 Claude Sonnet 4.5，**才真正跑得通**
- 可连跑数小时到数天，自动执行。

← 演示

Agent 工具的三种形态：Skill · MCP · CLI

UNIVERSITÄT BERN

Skill 技能

一个 `SKILL.md` 文件夹

（指令+脚本+资源），**模型按需加载**

读 description

↓

命中 → 载入 SKILL.md

↓

按指令 / 跑脚本

MCP 协议

开放协议：client–server

把外部**工具/数据**接给 agent

Agent (client)

↓ JSON-RPC

MCP server

↓

外部工具 / 数据

CLI 工具

agent 直接在 **shell**

跑命令、读输出，最通用

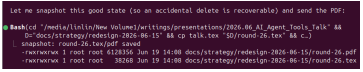
生成命令

↓

shell 执行

↓

读 stdout → 继续

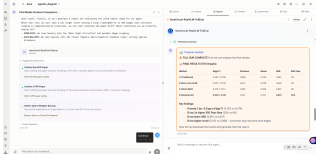


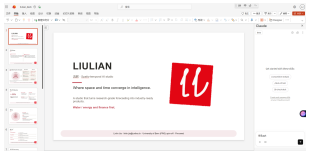
演示

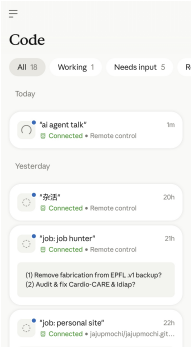
参考资料：Anthropic Agent Skills anthropic.com/news/skills · Model Context Protocol modelcontextprotocol.io

我的个人配置：github.com/jajupmochi/claude-config

各种各样的 Agent

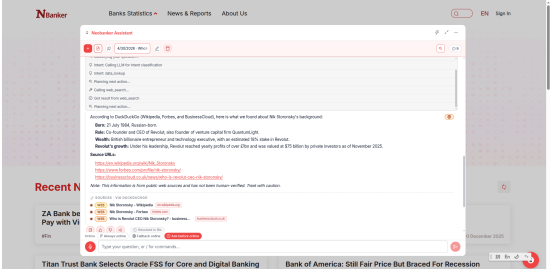


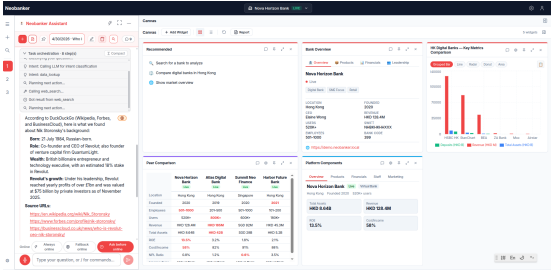




演示

制作自己的 Agent





演示

当前 agent 存在的问题

UNIVERSITÄT BERN

- **会偷懒**:
  - 一小时写**上万行**，没法逐行看
  - 直接**改掉测试**凑全绿（100+ 全过却是错的）
  - 概率模型，仍犯人不会犯的低级错
- **调研判断**仍替代不了人（定方向 / argue）
- 可能**误删文件、误操作**
- **幻觉**依然存在
- 不少工具**难泛化、上手成本高**

我的经验

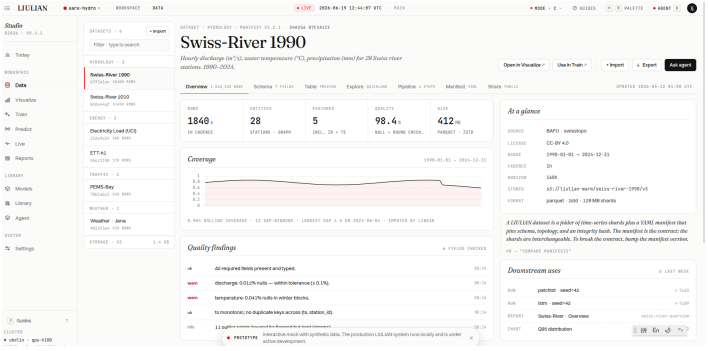
用最好的模型和agent!

权限设置，git备份

小工具：写完自己用，出错再说。

正经的库 / 科研代码：**一定过 code review 或整体 debug**，并用专门的**校验 skill** 抓“假绿”。

liulian：面向非计算机专家的 AI 平台



一个能对话的研究 agent　　https://github.com/jajupmochi/liulian-python

演示

总结

三阶段总结

UNIVERSITÄT BERN

	Chat	DeepResearch	Agent
能力	对话 / 生成	自主调研 + 综述	调工具 / 改文件 / 闭环
你的角色	驾驶员	委托人	监督者
适用	起草 / 解释 / 翻译	文献 / 选型 / 尽调	编码 / 科研 / 产品
主要风险	幻觉	来源与判断	越权 / 连锁错误

永远自己验证，不要把判断外包。

风险与边界

UNIVERSITÄT BERN

验证与问责

越流畅越要查；出事得有人负责，agent 不能

安全与隐私

常是顶级权限；生产谨慎，数据脱敏 / 离线

能力会退化

基础不写了，像“提笔忘字”

创造性 / 分布外

套路化擅长，新问题明显掉链子

工具是一面镜子

UNIVERSITÄT BERN

FOMO：动力也是焦虑

怕错过推着你学，也会让你乱

烂尾工程陷阱

起点高诱你投入，难的逻辑在后面，易半途而废

战略 > 战术

它让“没想清方向”更尖锐

多是人的问题

不是工具的问题；方向笃定就不慌

FOMO or not FOMO, that is the question

UNIVERSITÄT BERN

- 前两次工业革命，很多人是**旁观者**；
  这一次，算力和工具**就在手边**。

- 与其焦虑“会不会被取代”，
  不如想清楚：**你想用它抵达哪里**。

工具迭代太快，别太焦虑；**吃好喝好，把判断留给自己**。

更完整的经历，我在小宇宙做了一期语音访谈：

www.xiaoyuzhoufm.com/episode/69efbb76b4f25397f5ba0724

节日快乐！


端午安康
AI 乘风，粽创未来！
GPT-4o
Claude
Gemini
Llama 3
Llamn3
DeepSeek
Agent
AGENT
Let's Build the Future!
必胜
智能体时代

UNIVERSITÄT BERN

感谢聆听！

欢迎提问与交流

微信

GitHub

linlin.jia@unibe.ch

Open to positions!