

大模型的探索与实践

Introduction to Large Language Models

§ 1.1 大模型初探

Introduction

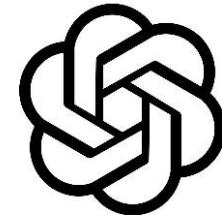
滕佳烨
上海财经大学
www.tengjiaye.com

大模型 Large Models

常见大模型：



Qwen



ChatGPT



deepseek



豆包大模型

KIMI

Gemini

<https://arxiv.org/pdf/2307.06435>

大语言模型 Large Language Models, LLM

越大的模型，效果一般越好

[1] Scaling Laws for Neural Language Models.

大语言模型 Large Language Models, LLM

越大的模型，效果一般越好

- 海量的数据

GPT-3 training data [1]:9

Dataset	# tokens	Proportion within training
Common Crawl	410 billion	60%
WebText2	19 billion	22%
Books1	12 billion	8%
Books2	55 billion	8%
Wikipedia	3 billion	3%

大语言模型 Large Language Models, LLM

越大的模型，效果一般越好

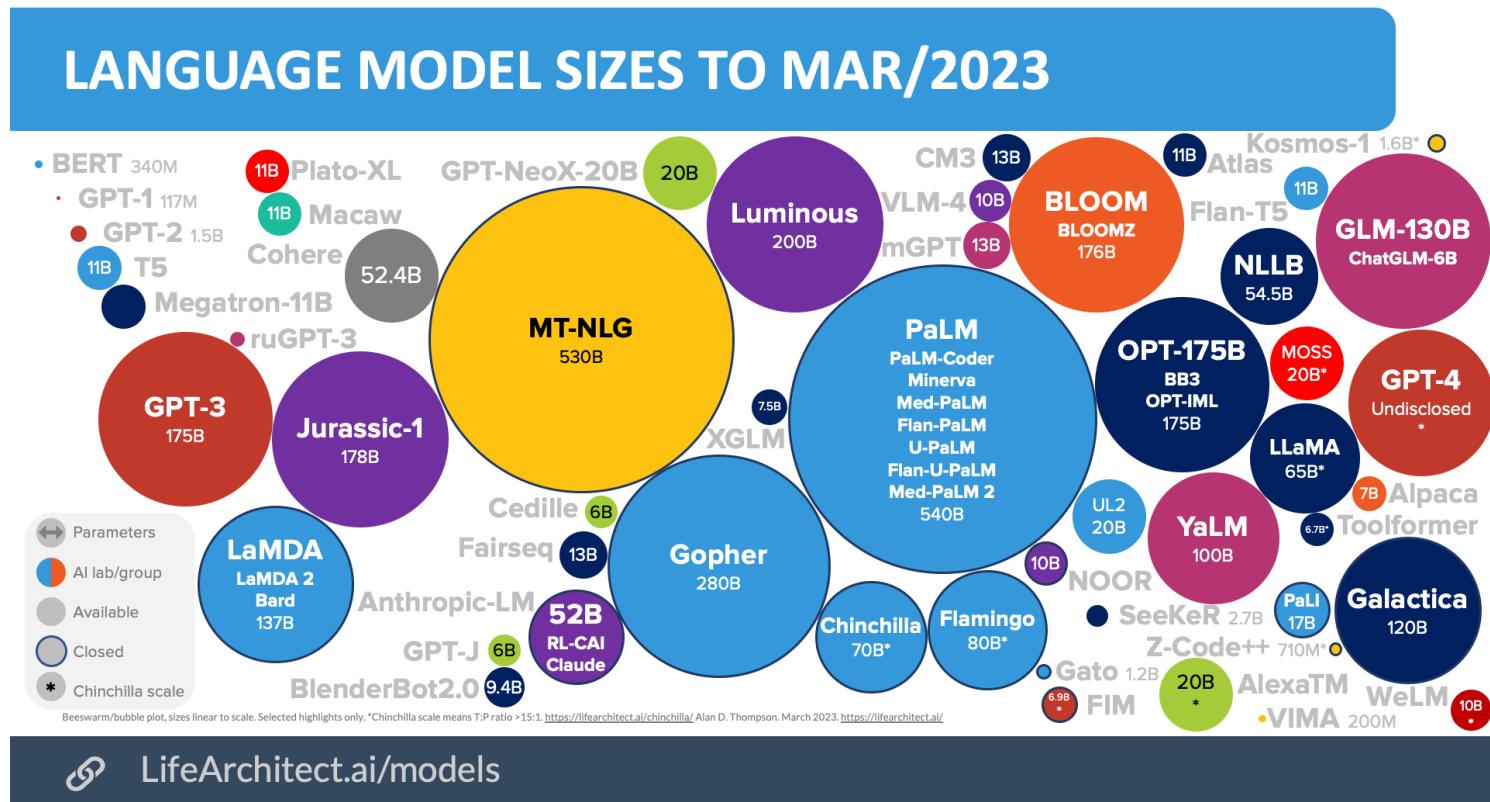
- 海量的数据
- 巨大的参数规模

Benchmark (Metric)	DeepSeek V2-0506	DeepSeek V2.5-0905	Qwen2.5 72B-Inst.	LLaMA-3.1 405B-Inst.	Claude-3.5- Sonnet-1022	GPT-4o 0513	DeepSeek V3
Architecture	MoE	MoE	Dense	Dense	-	-	MoE
# Activated Params	21B	21B	72B	405B	-	-	37B
# Total Params	236B	236B	72B	405B	-	-	671B

大语言模型 Large Language Models, LLM

越大的模型，效果一般越好

- 海量的数据
 - 巨大的参数规模



大语言模型 Large Language Models, LLM

越大的模型，效果一般越好

- 海量的数据
- 巨大的参数规模
- 庞大的计算资源

DeepSeek-V3 训练使用了2048块NVIDIA H800 GPU！

Training Costs	Pre-Training	Context Extension	Post-Training	Total
in H800 GPU Hours	2664K	119K	5K	2788K
in USD	\$5.328M	\$0.238M	\$0.01M	\$5.576M

Table 1 | Training costs of DeepSeek-V3, assuming the rental price of H800 is \$2 per GPU hour.

大语言模型 Large Language Models, LLM

越大的模型，效果一般越好

但耗费的资源也就更多更贵

- 训练时所需资源多
- 推理时所需时间长

有限的资源应该怎么利用？模型该多大？数据该多少？

[1] Scaling Laws for Neural Language Models.

大语言模型 Large Language Models, LLM

对话能解决很多问题

大语言模型 Large Language Models, LLM

原始问题

机器翻译



完形填空



机械控制



股票投资



...

对话形式

我有一个句子XX, 请帮我翻译成英文形式。

我吃了一个苹[]，很甜。请填写空格部分。

现在外界环境35度，我应该如何调控我的制冷环境？

昨天的大盘情况是XXX，我应该怎么调配我的资产？

...

大语言模型 Large Language Models, LLM

利用对话进行资料查询

线形可分数据



线性可分数据是指可以通过一个线性超平面将不同类别的数据点完全分开的数据集。具体来说，在二维空间中，线性可分意味着可以用一条直线将两类数据点分开；在三维空间中，可以用一个平面将两类数据点分开；在更高维的空间中，则可以用一个超平面将两类数据点分开。

数学定义

给定一个数据集 $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$ ，其中 \mathbf{x}_i 是特征向量， $y_i \in \{-1, 1\}$ 是类别标签。如果存在一个权重向量 \mathbf{w} 和一个偏置 b ，使得对于所有 i 满足：

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 0$$

则称该数据集是线性可分的。

例子

假设有一个二维数据集，包含两类点：

- 类别1: $(1, 2), (2, 3), (3, 3)$

类别2: $(2, 1), (4, 2), (5, 1)$

给 DeepSeek 发送消息

深度思考 (R1)

联网搜索



给 DeepSeek 发送消息

深度思考 (R1)

联网搜索



<https://chat.deepseek.com/>

大语言模型 Large Language Models, LLM

利用对话进行图像生成

创建一张漫画风格的烹饪意大利面教程的图片

图片已创建



Figure from GPT-o3

大语言模型 Large Language Models, LLM

利用对话进行代码生成

The screenshot shows a Jupyter Notebook environment with a sidebar and a main workspace. The workspace contains an untitled notebook file named 'Untitled-1.ipynb'. The code in the notebook is as follows:

```
model = MLP()
n_params = count_parameters(model)

# 计算理论泛化误差边界
def theoretical_bound(n, p):
    return 100 * np.sqrt(np.log(n) * p / n) # 乘以100转换为百分比

theoretical_errors = [theoretical_bound(n, n_params) for n in sample_sizes]

# 设置中文字体
plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签
plt.rcParams['axes.unicode_minus'] = False # 用来正常显示负号

# 绘制泛化误差曲线和理论边界
plt.figure(figsize=(10, 6))
plt.plot(sample_sizes, generalization_errors, 'bo-', label='实际泛化误差')
plt.plot(sample_sizes, theoretical_errors, 'r--', label=r'理论边界 $\sqrt{\log(n) * p / n}')
plt.xscale('log')
plt.xlabel('训练集大小')
plt.ylabel('泛化误差 (%)')
plt.title('训练集大小与泛化误差的关系')
plt.legend()
plt.grid(True)
plt.show()
```

To the right of the notebook, there is a 'Chat' interface with a message from 'Trae AI': "我来帮你编写一个使用MLP进行MNIST数据集分类的代码，并通过不同数据量来观察泛化误差。我们将使用PyTorch来实现这个实验。". Below the message, there is a code snippet labeled 'Untitled-1.ipynb' which is identical to the one in the notebook. At the bottom of the interface, there is a note: "#上下文" and "Claude-3.5-Sonnet".

Trae from ByteDance

大语言模型 Large Language Models, LLM

纸上得来终觉浅, 绝知此事要躬行

为什么不打开Trae玩一玩呢?

用对话的方式写代码 Vibe Coding

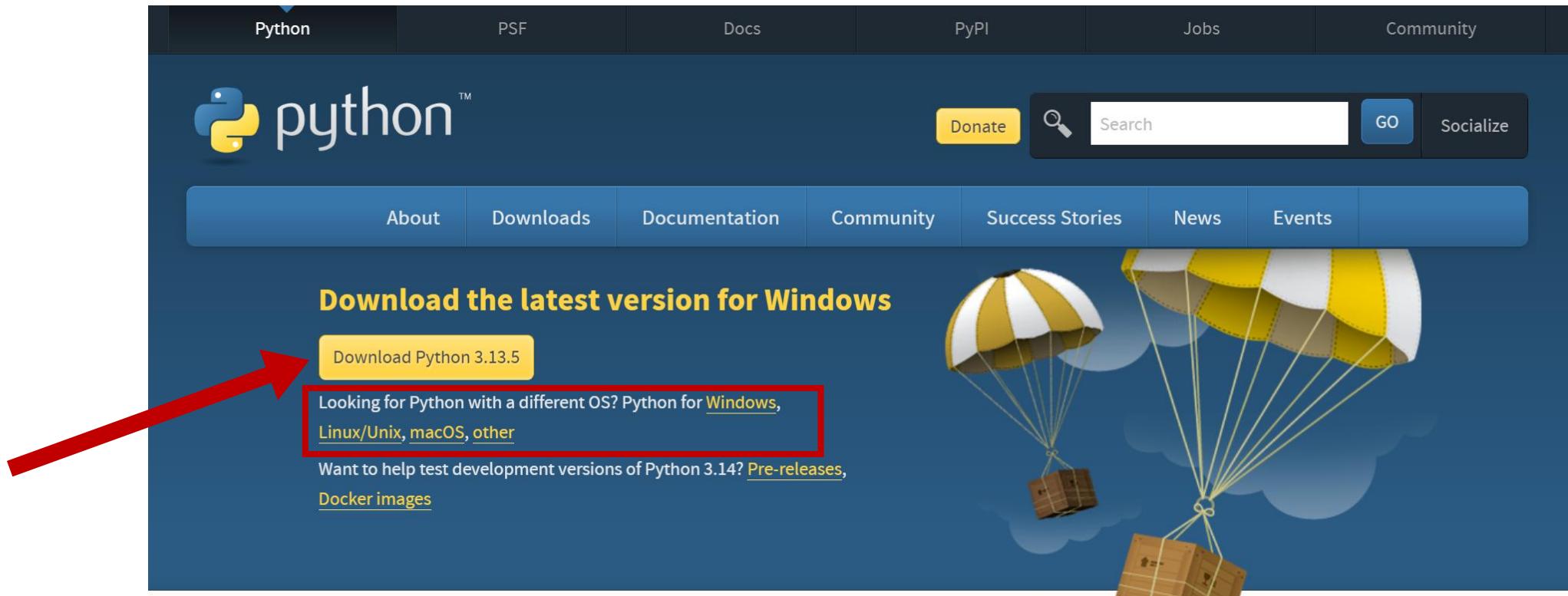
下载Trae: <https://www.trae.cn/>

The screenshot shows the homepage of the Trae website. At the top, there is a navigation bar with links for 'IDE', '插件' (Plugins), '文档' (Documentation), '社区' (Community), '活动 NEW', '登录' (Login), and '下载 IDE' (Download IDE). The main headline is '智能无限，协作无间'. Below it is a brief description: 'Trae, 致力于成为真正的 AI 工程师 (The Real AI Engineer)。Trae 旗下的 AI IDE 产品, 以智能生产力为核心, 无缝融入你的开发流程, 与你默契配合, 更高质量、高效率完成每一个任务。' Two download buttons are visible: a purple button labeled '立即获取 Trae IDE' and a red button labeled '查看所有下载选项'. A large red arrow points from the text above to the red download button.

- 有条件的也可以下载cursor, 界面内容差别不大
- 注意一下自己是windows系统还是mac系统
- 安装相信大家都会...

用对话的方式写代码 Vibe Coding

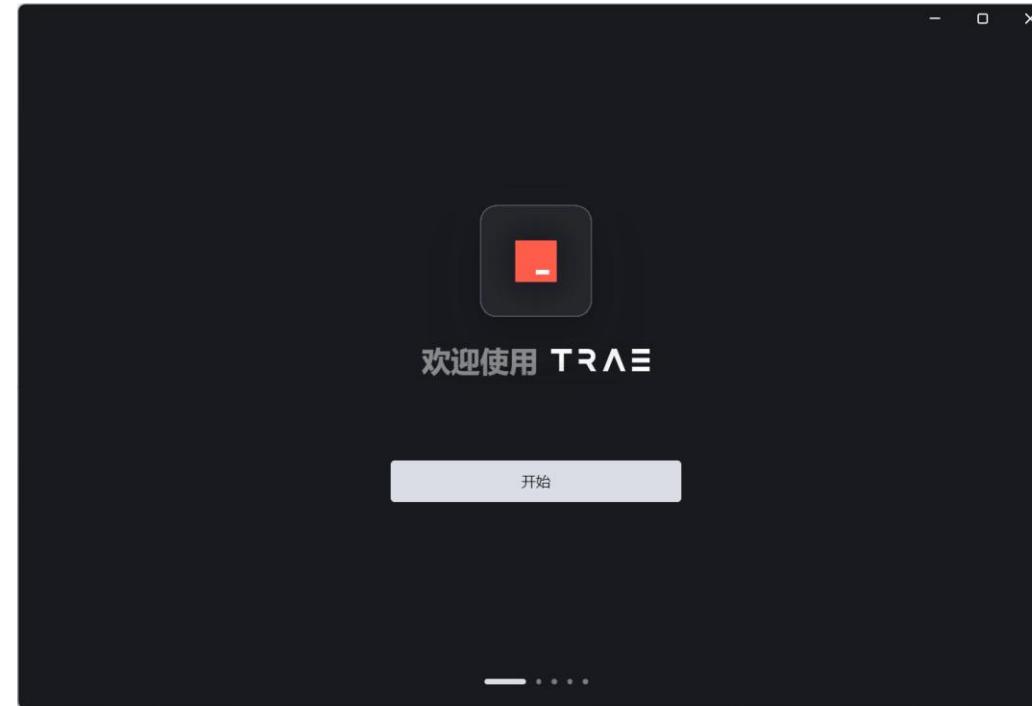
下载python: <https://www.python.org/downloads/>



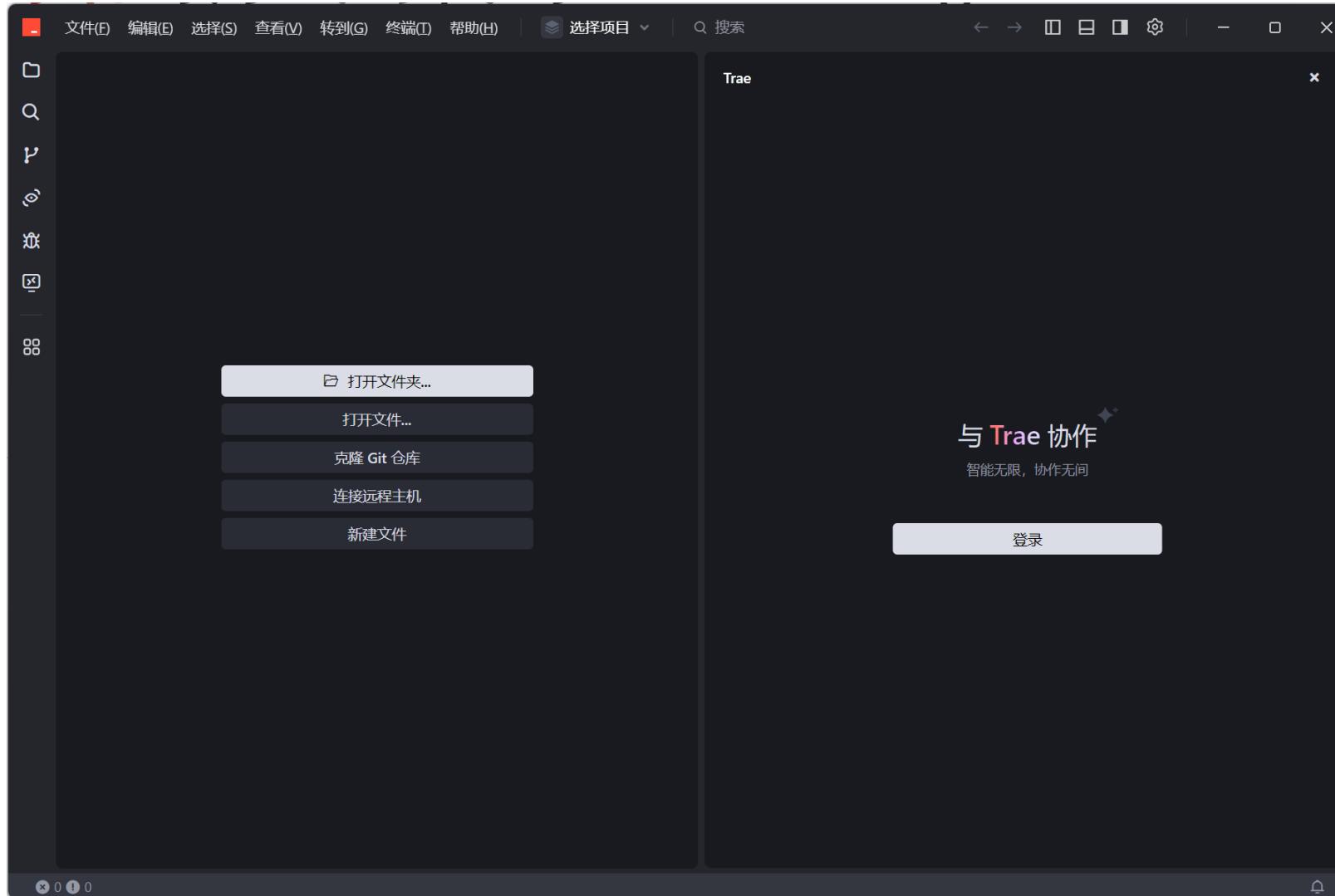
- 注意一下自己是windows系统还是mac系统
- 安装相信大家都会...

用对话的方式写代码 Vibe Coding

打开 Trae



用对话的方式写代码 Vibe Coding

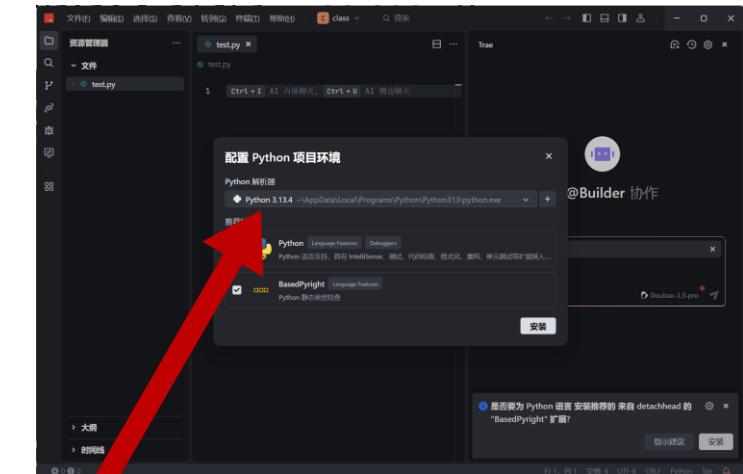
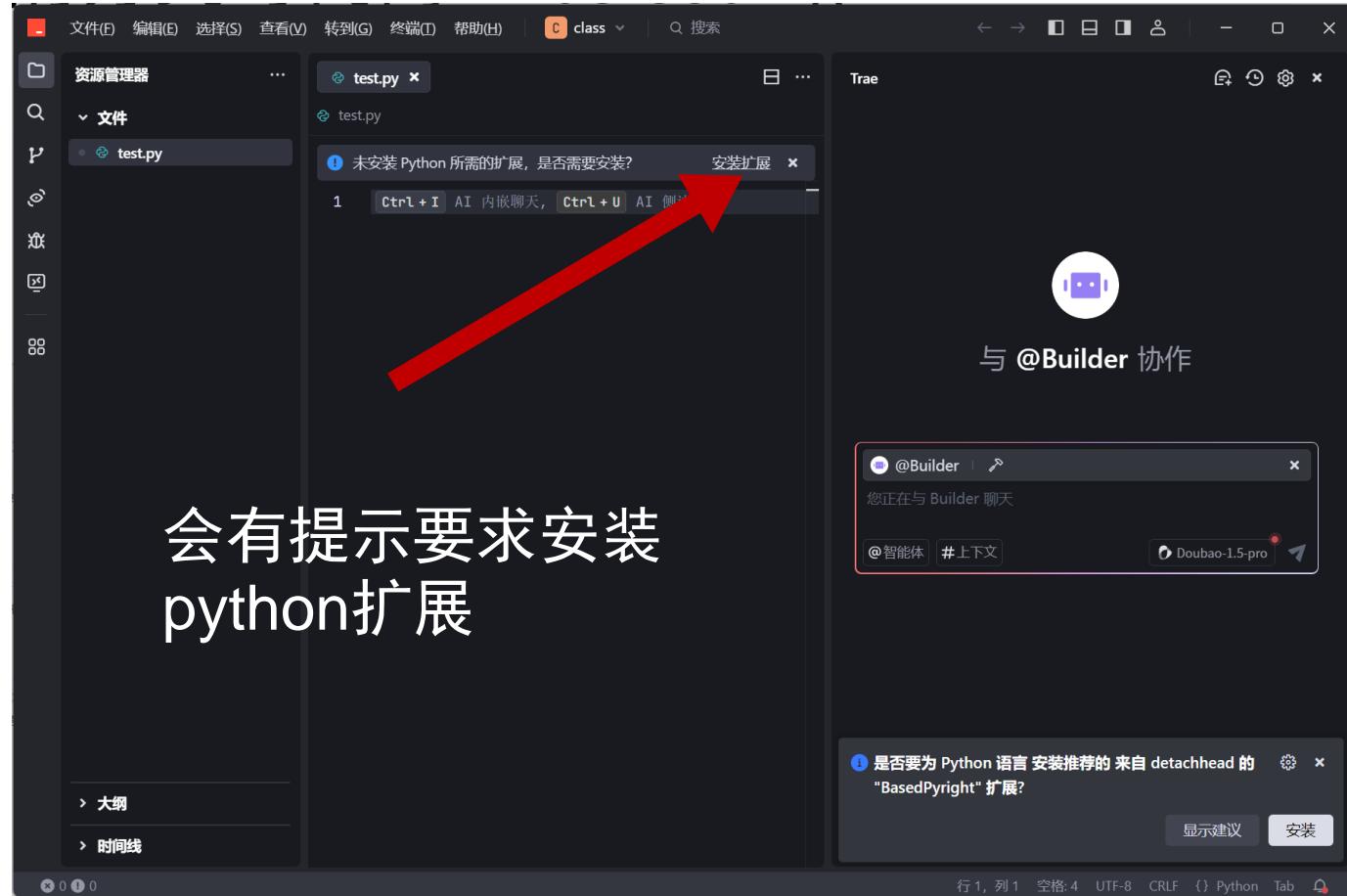


- 可以用手机号直接登录
- 可创建一个新的文件夹后打开

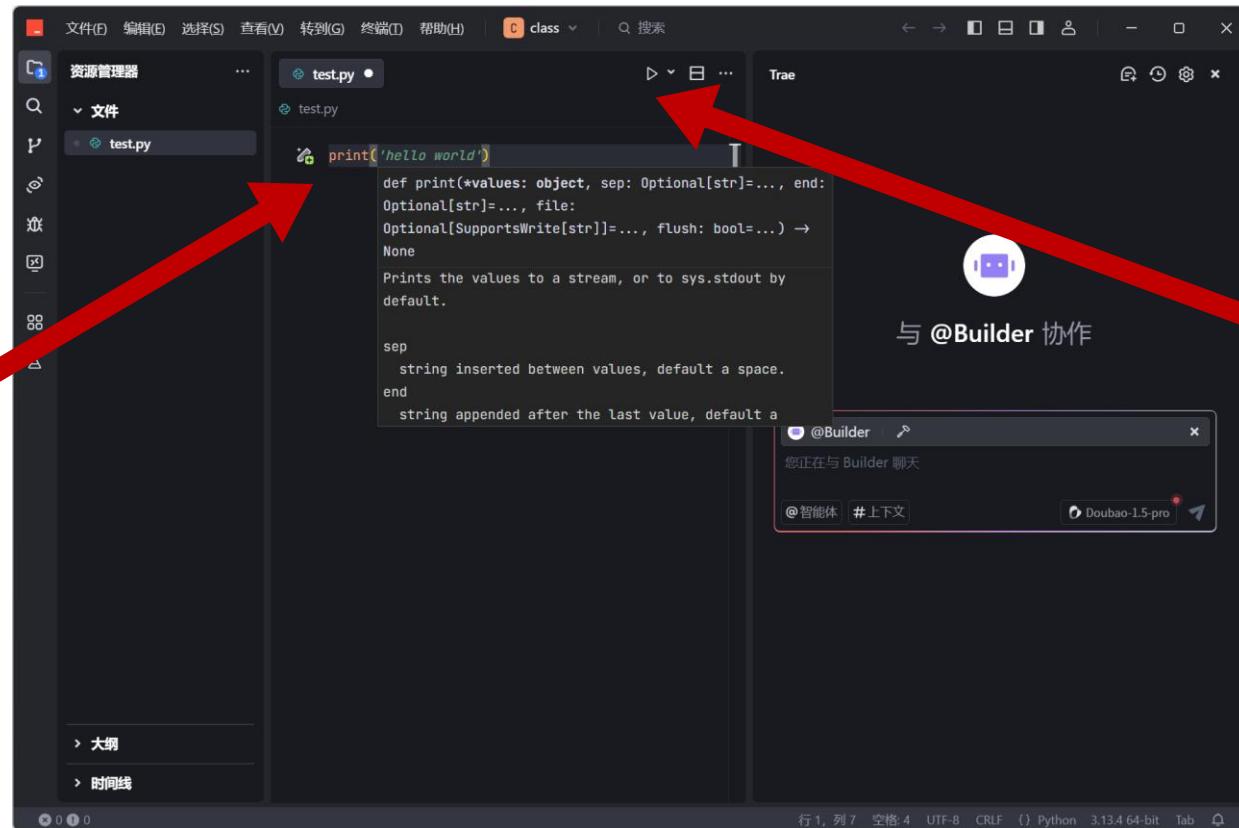
用对话的方式写代码 Vibe Coding



用对话的方式写代码 Vibe Coding



用对话的方式写代码 Vibe Coding



自动补全功能

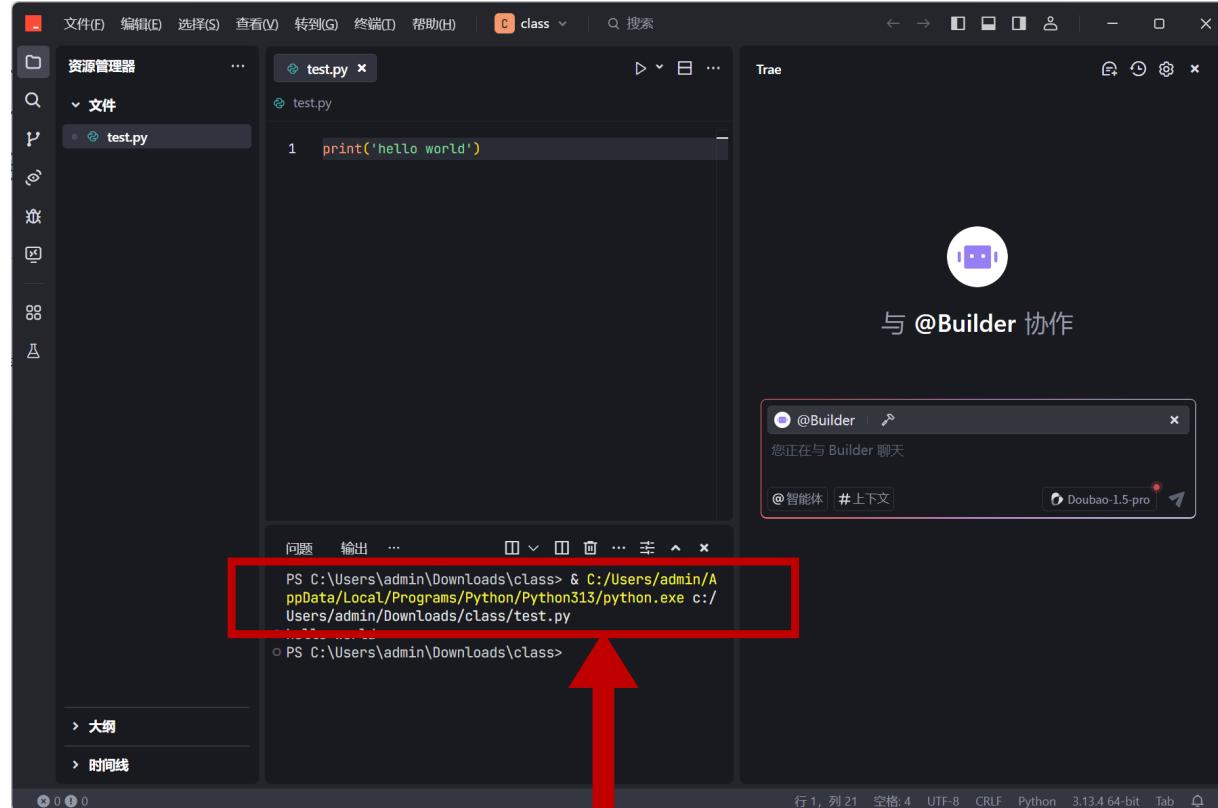
点击这里进行编译

用对话的方式写代码 Vibe Coding

输出结果

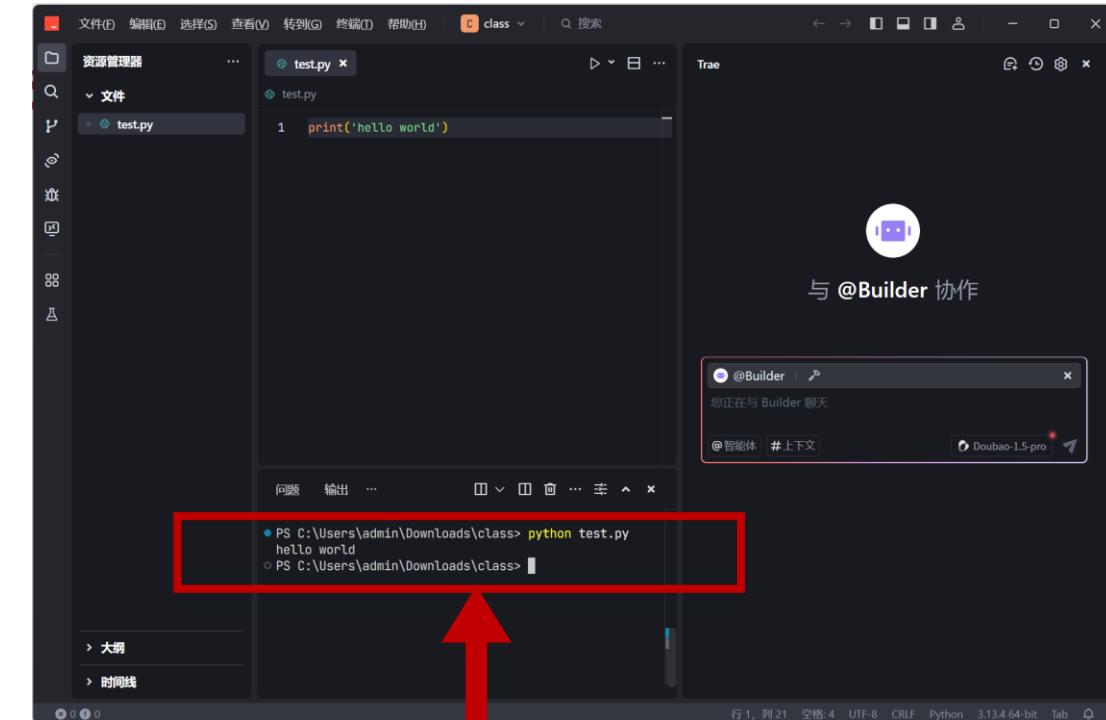
```
PS C:\Users\admin\Downloads\class> & C:/Users/admin/AppData/Local/Programs/Python/Python313/python.exe c:/Users/admin/Downloads/class/test.py
hello world
PS C:\Users\admin\Downloads\class>
```

用对话的方式写代码 Vibe Coding



```
PS C:\Users\admin\Downloads\class> & C:/Users/admin/AppData/Local/Programs/Python/Python313/python.exe c:/Users/admin/Downloads/class/test.py
hello world
PS C:\Users\admin\Downloads\class>
```

很长很丑



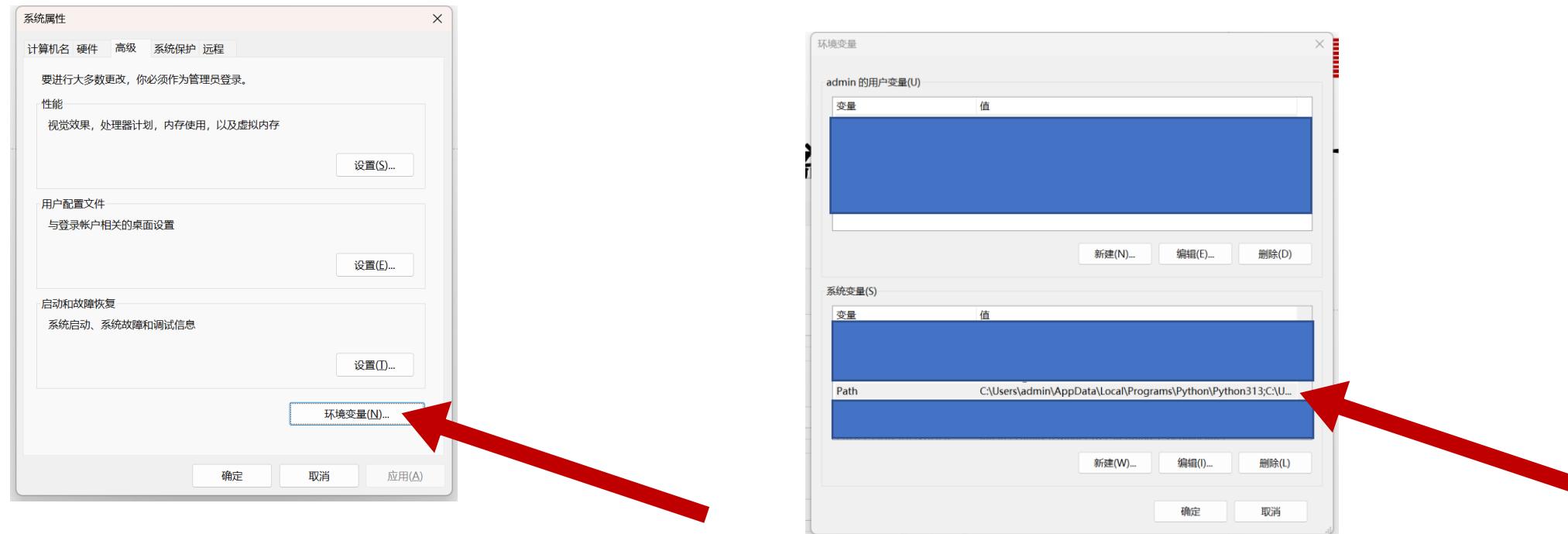
```
PS C:\Users\admin\Downloads\class> python test.py
hello world
PS C:\Users\admin\Downloads\class>
```

试试你们的电脑是否能运行?

用对话的方式写代码 Vibe Coding

添加环境变量

直接在系统中搜索环境变量，即可打开属性界面



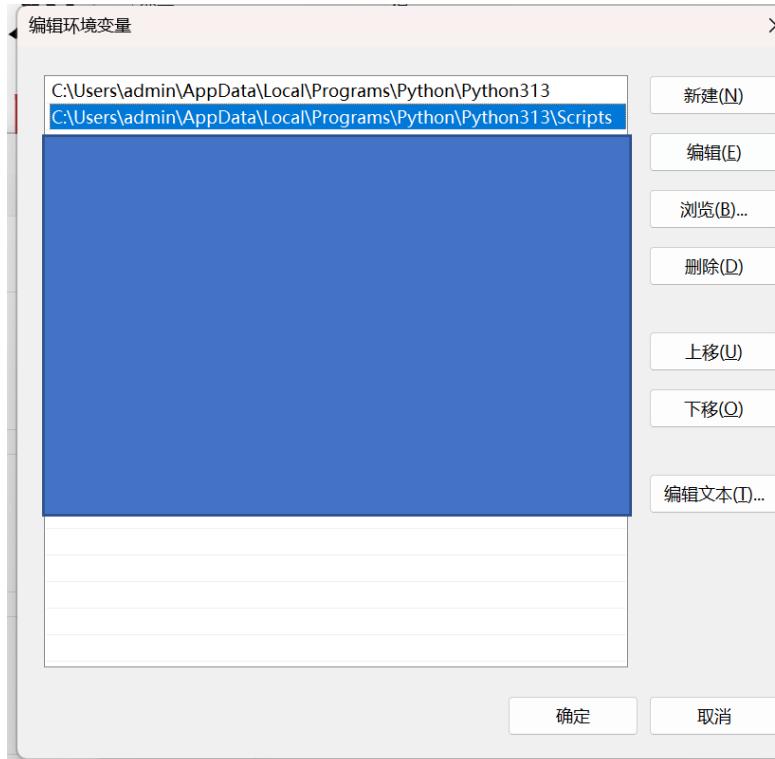
用对话的方式写代码 Vibe Coding

添加环境变量

加入python对应文件地址

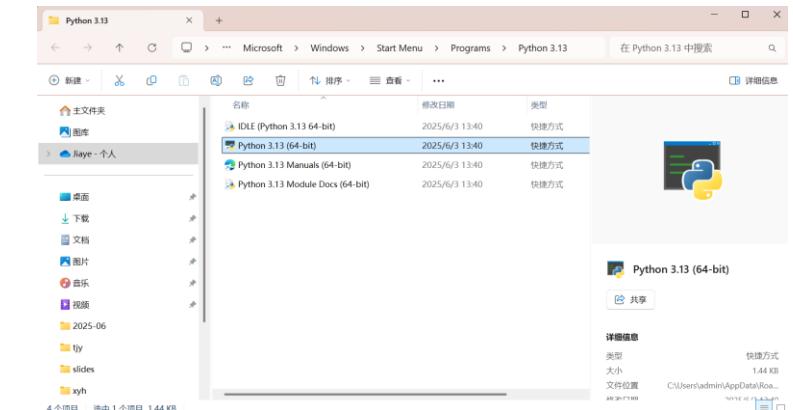
C:\Users\admin\AppData\Local\Programs\Python\Python313

C:\Users\admin\AppData\Local\Programs\Python\Python313\Scripts



注意：

1. 我们的地址可能不一样。你可以去看自己的 python 安装到了哪个地址。但 python 的快捷启动地址（从搜索界面直接进去的地址）与真实地址往往不一样



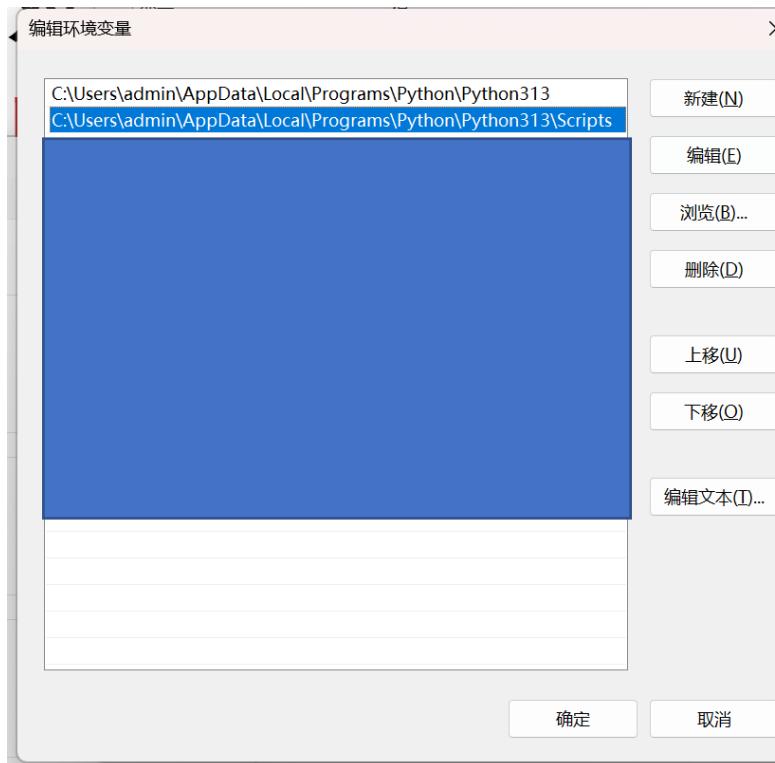
用对话的方式写代码 Vibe Coding

添加环境变量

加入python对应文件地址

C:\Users\admin\AppData\Local\Programs\Python\Python313

C:\Users\admin\AppData\Local\Programs\Python\Python313\Scripts

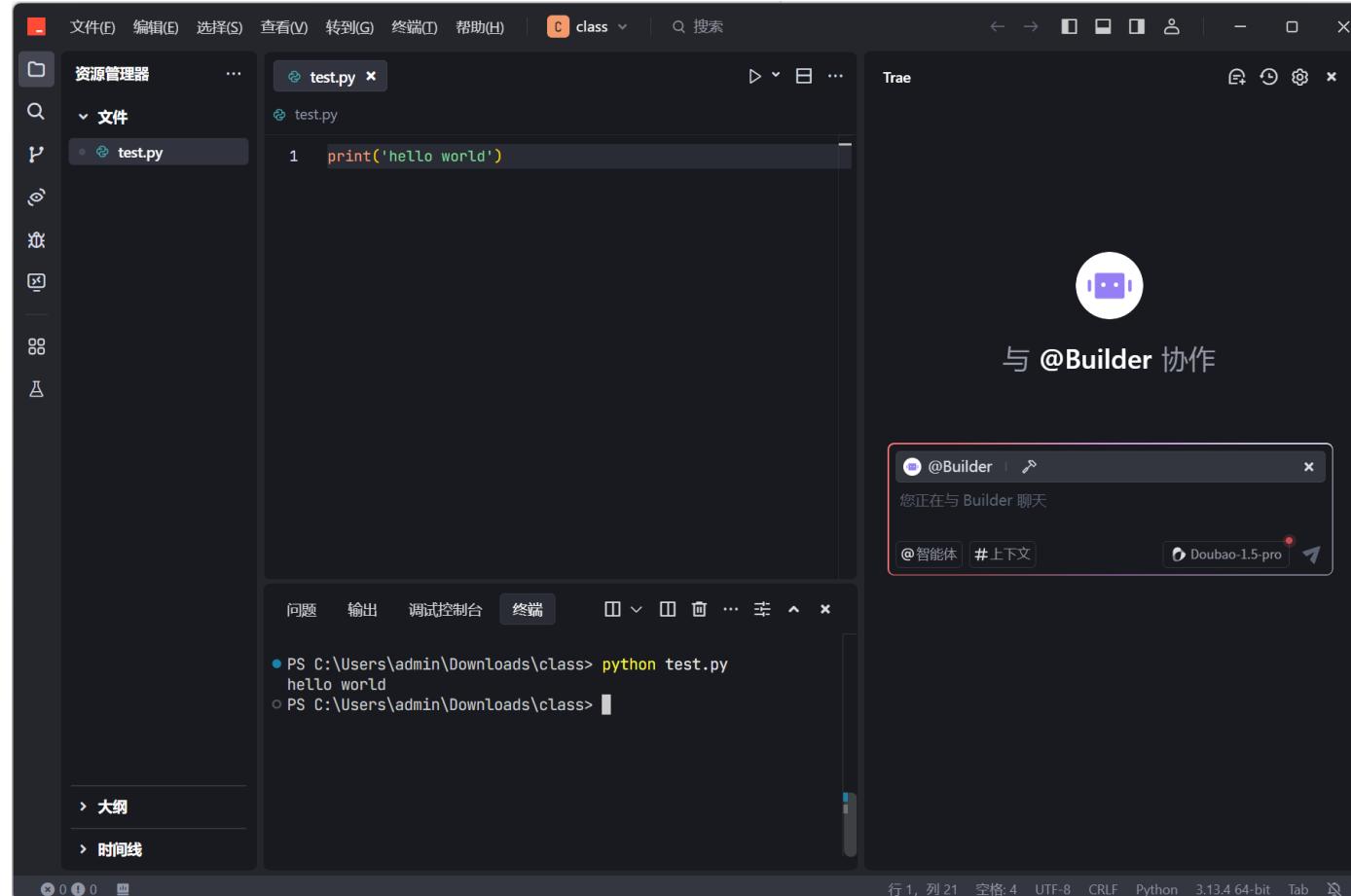


注意：

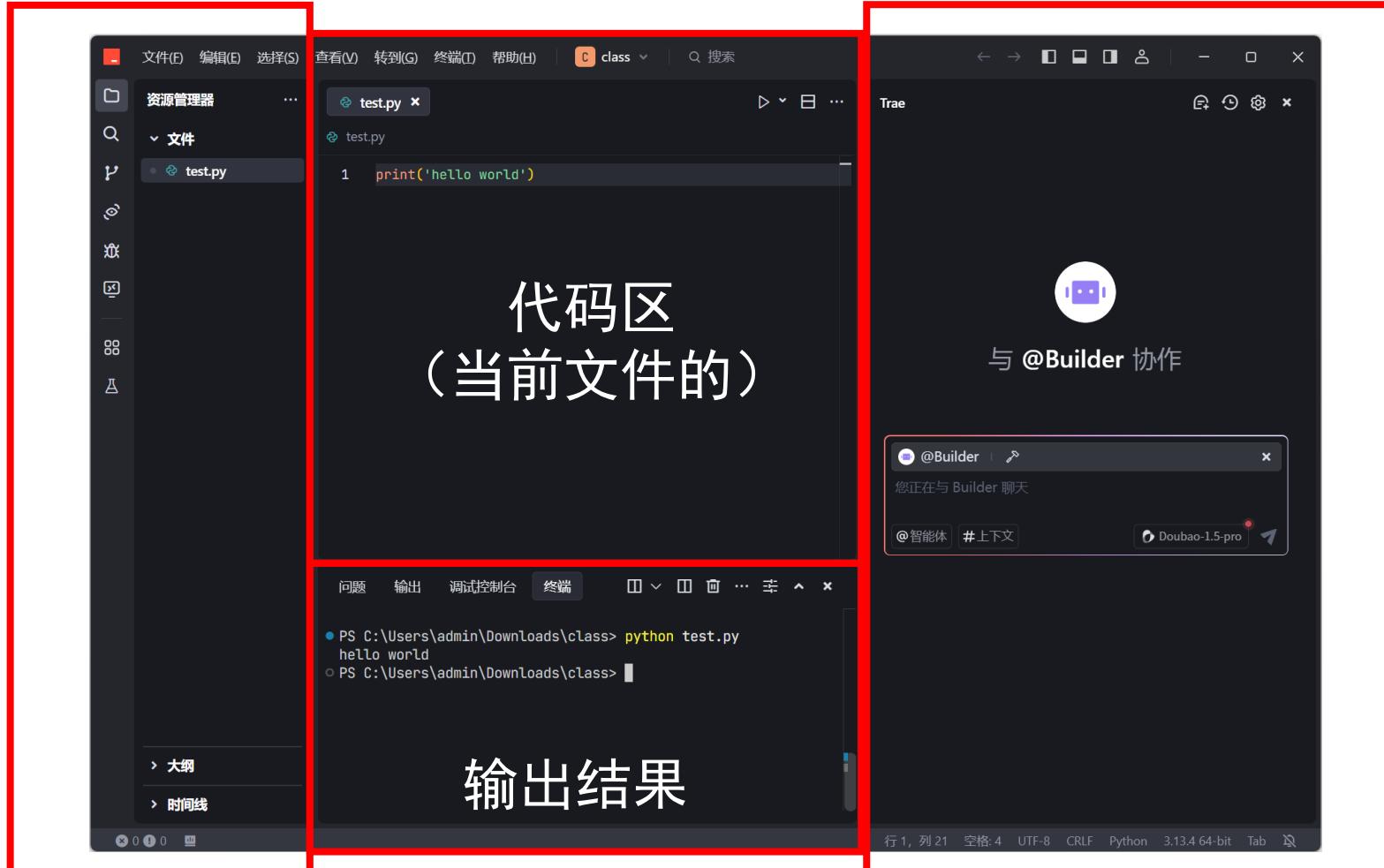
1. 我们的地址可能不一样。
2. 要同时加上有/无scripts 的
3. 变量顺序非常重要。如果新建的环境变量在下面，可能会优先运行system的某个指令，以至于无法在terminal内直接成功运行python

用对话的方式写代码 Vibe Coding

使用terminal



用对话的方式写代码 Vibe Coding



包含电脑
中的文件

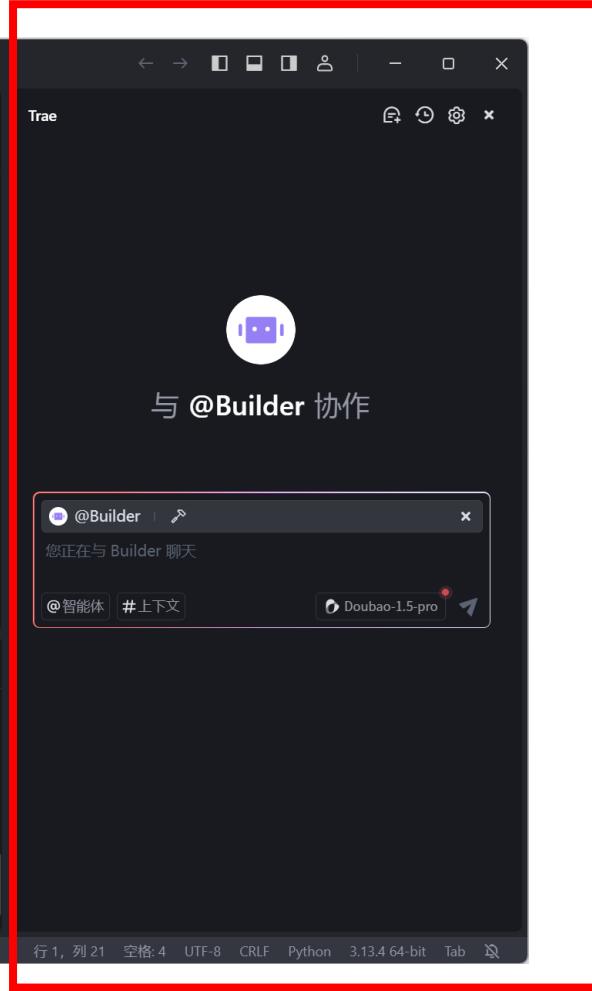
AI 交互

用对话的方式写代码 Vibe Coding

代码区
(当前文件的)

```
PS C:\Users\admin\Downloads\class> python test.py
hello world
PS C:\Users\admin\Downloads\class>
```

输出结果

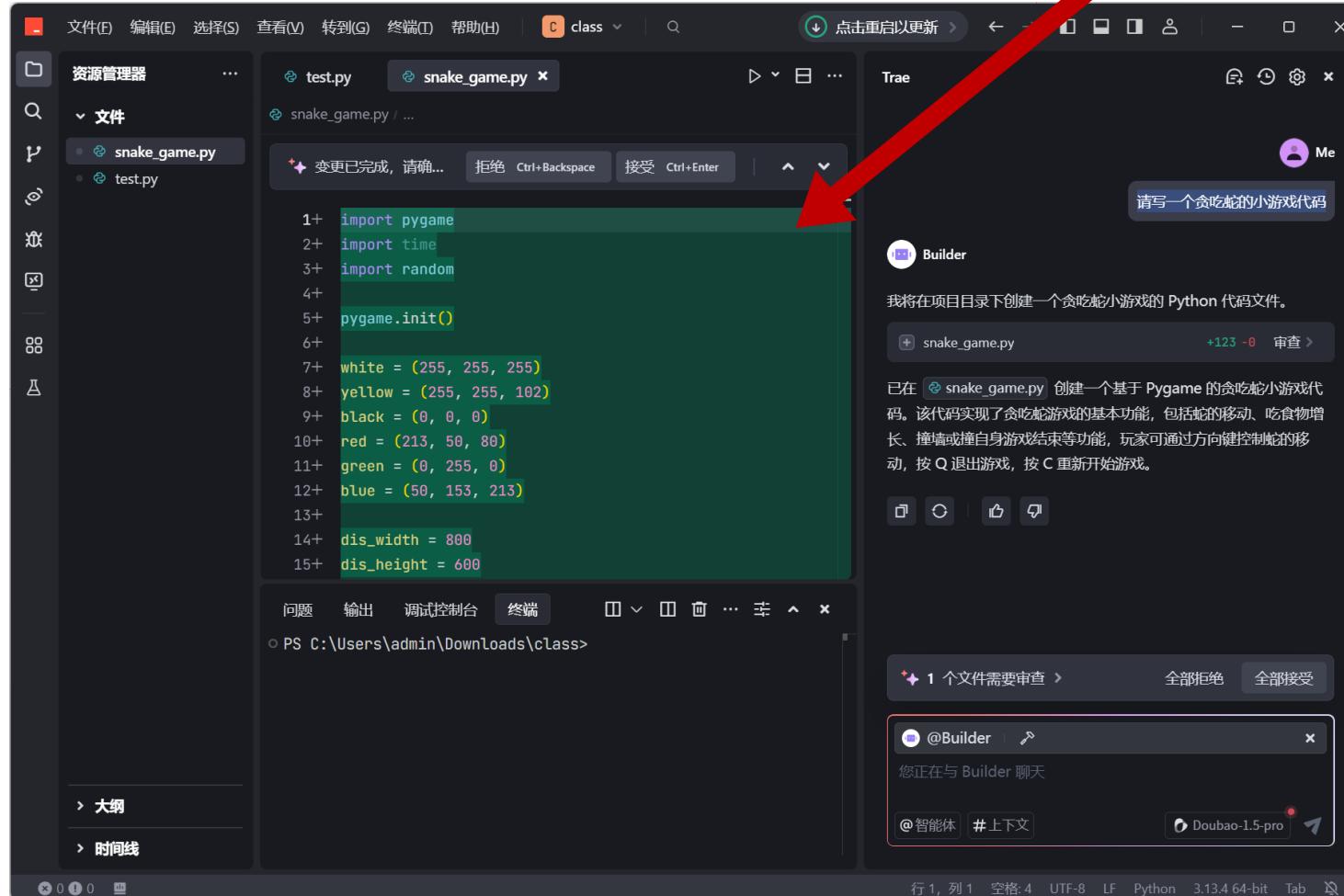


在这里输入如下prompt:

请写一个贪吃蛇的小游戏
代码

用对话的方式写代码 Vibe Coding

AI 写的部分，建议运行成功后再accept



在这里输入如下prompt:

请写一个贪吃蛇的小游戏
代码

用对话的方式写代码 Vibe Coding

The screenshot shows a code editor interface with a dark theme. On the left is a file explorer with files 'test.py' and 'snake_game.py'. The main area displays the content of 'snake_game.py':

```
1+ import pygame
2+ import time
3+ import random
4+
5+ pygame.init()
6+
7+ white = (255, 255, 255)
8+ yellow = (255, 255, 102)
9+ black = (0, 0, 0)
10+ red = (213, 50, 80)
11+ green = (0, 255, 0)
12+ blue = (50, 153, 213)
13+
14+ dis_width = 800
15+ dis_height = 600
```

Below the code editor is a terminal window with a red border around its output area. The output shows:

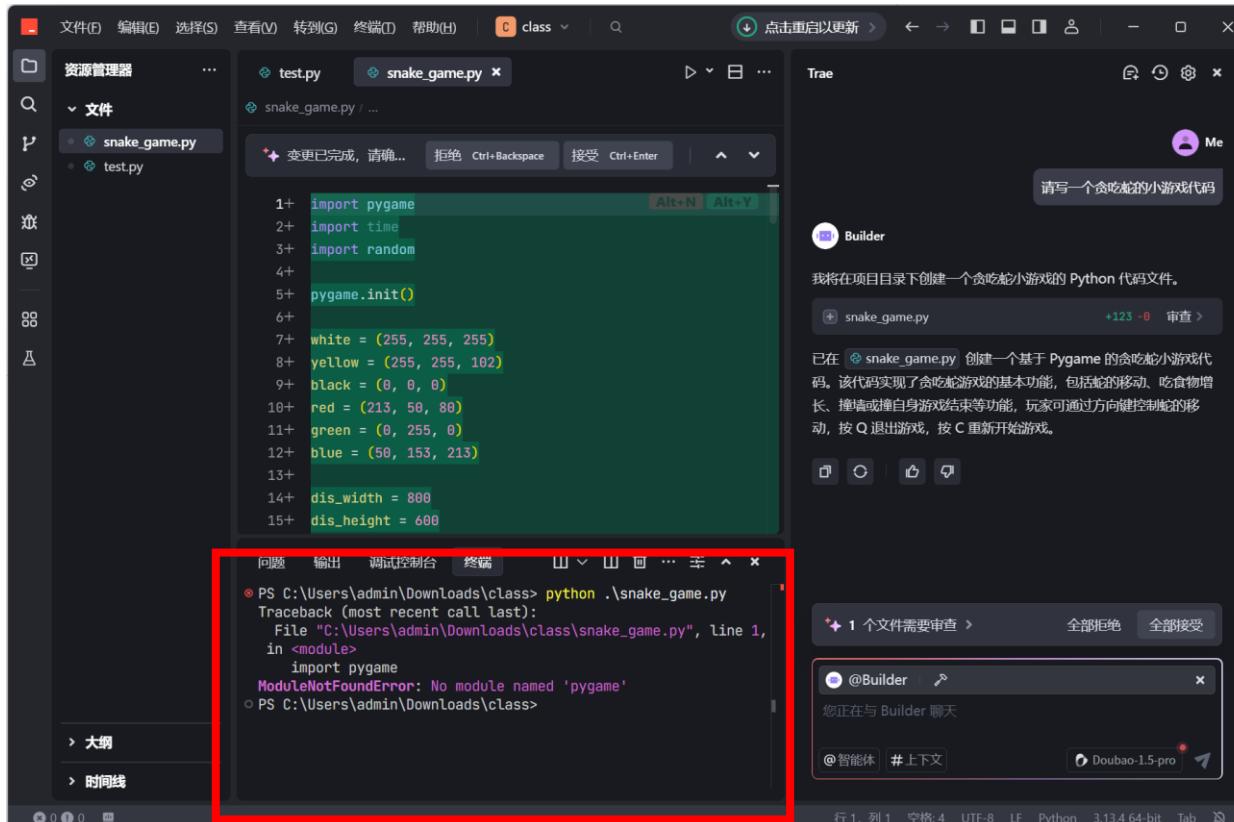
```
PS C:\Users\admin\Downloads\class> python .\snake_game.py
Traceback (most recent call last):
  File "C:/Users/admin/Downloads/class/snake_game.py", line 1, in <module>
    import pygame
ModuleNotFoundError: No module named 'pygame'
PS C:\Users\admin\Downloads\class>
```

在这里输入如下prompt:

请写一个贪吃蛇的小游戏
代码

报错了，怎么办？

用对话的方式写代码 Vibe Coding

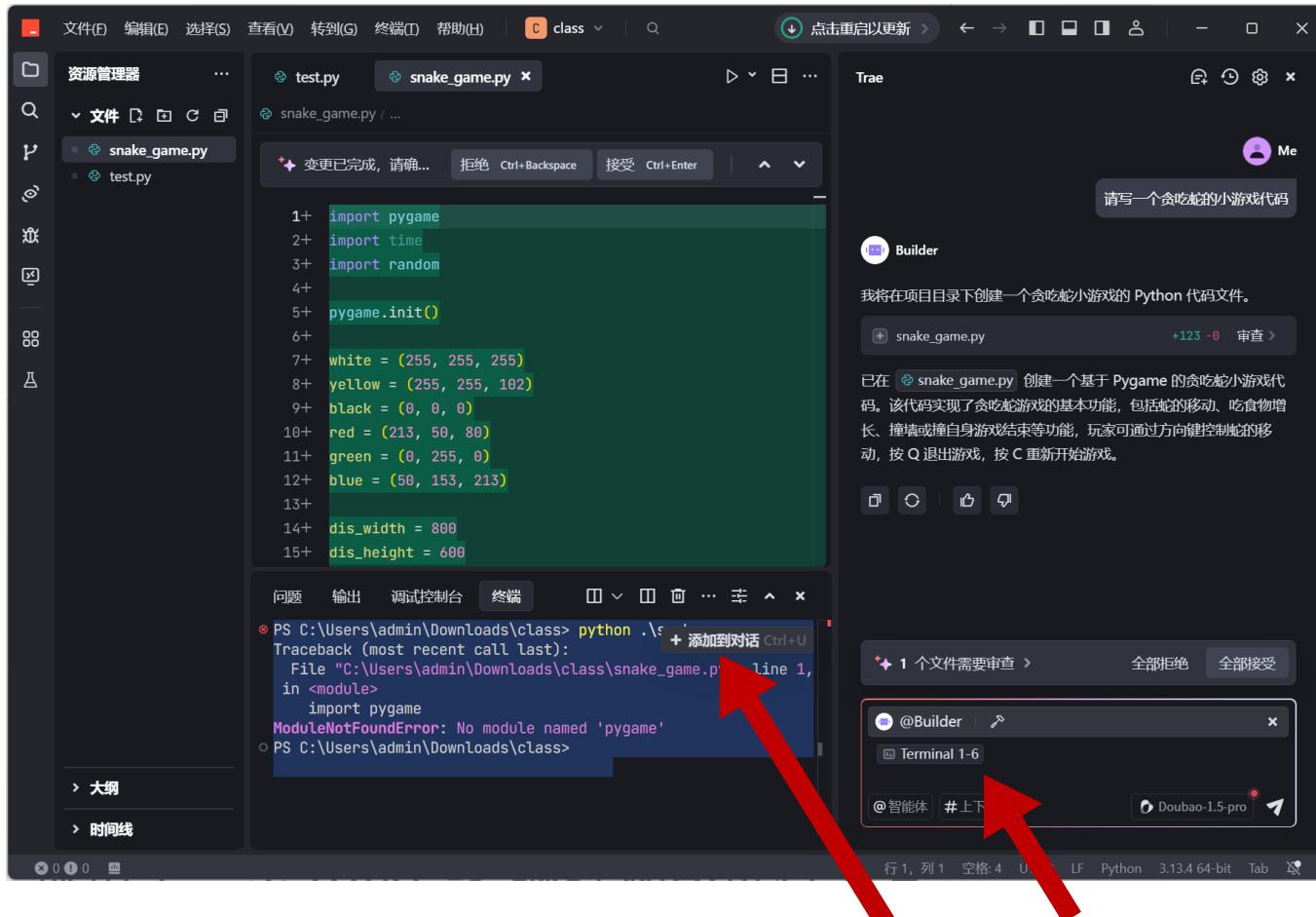


首先复制报错的命令行，并粘贴至AI交互界面

- > 告诉AI你要问哪
- > 选中后点击添加到对话

报错了，怎么办? -> 继续和AI进行交互

用对话的方式写代码 Vibe Coding



首先复制报错的命令行，并粘贴至AI交互界面

- > 告诉AI你要问哪
- > 选中后点击添加到对话

报错了，怎么办? -> 继续和AI进行交互

用对话的方式写代码 Vibe Coding

A screenshot of a code editor showing a Python script named `snake_game.py`. The script contains code to initialize Pygame and set up colors. A terminal window shows a `ModuleNotFoundError` for the `pygame` module. An AI interface window titled "Builder" suggests installing `pygame` using `pip install pygame`.

```
1+ import pygame
2+ import time
3+ import random
4+
5+ pygame.init()
6+
7+ white = (255, 255, 255)
8+ yellow = (255, 255, 182)
9+ black = (0, 0, 0)
10+ red = (213, 50, 80)
11+ green = (0, 255, 0)
12+ blue = (50, 153, 213)
13+
14+ dis_width = 800
15+ dis_height = 600
```

PS C:\Users\admin\Downloads\class> python .\snake_game.py
Traceback (most recent call last):
File "C:\Users\admin\Downloads\class\snake_game.py", line 1,
in <module>
import pygame
ModuleNotFoundError: No module named 'pygame'
PS C:\Users\admin\Downloads\class>

正在等待你的操作 ·

从下次开始自动运行命令和 MCP 工具 ·

从报错信息可知，系统缺少 pygame 模块，我将运行命令安装 pygame 库。

pip install pygame

首先复制报错的命令行，并粘贴至AI交互界面

简单描述一下你的问题，例如：这里报错了，是为什么。AI会给你答案 -> 直接运行一般就可以

报错了，怎么办? -> 继续和AI进行交互

用对话的方式写代码 Vibe Coding

The screenshot shows a code editor with two tabs: `test.py` and `snake_game.py`. The `snake_game.py` tab contains the following code:

```
1+ import pygame
2+ import time
3+ import random
4+
5+ pygame.init()
6+
7+ white = (255, 255, 255)
8+ yellow = (255, 255, 102)
9+ black = (0, 0, 0)
10+ red = (213, 50, 86)
11+ green = (0, 255, 0)
12+ blue = (50, 153, 213)
```

The terminal output shows an error message:

```
ModuleNotFoundError: No module named 'pygame'
PS C:\Users\admin\Downloads\cClass> pip3 install pygame
pip3 : 无法将“pip3”项识别为 cmdlet、函数、脚本文件或可运行程序的名称。请检查名称的拼写，如果包括路径，请确保路径正确，然后再试一次。
所在位置 行:1 字符: 1
+ pip3 install pygame
+ ~~~~~
    + CategoryInfo          : ObjectNotFound: (pip3:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException
PS C:\Users\admin\Downloads\cClass>
```

A floating window titled "Builder" provides assistance:

- 参考了 1 个上下文
- 从报错信息可知，系统缺少 pygame 模块，我将运行命令安装 pygame 库。
- 终端 > pip install pygame
- 执行 pip 命令报错，提示无法识别 pip 命令，推测 Python 环境未正确配置，我将尝试运行 python -m pip install pygame 命令来安装 pygame 库。
- 终端 > python -m pip install pygame
- 手动终止输出
- 1 个文件需要审查 > 全部拒绝 全部接受
- @Builder 您正在与 Builder 聊天

首先复制报错的命令行，并粘贴至AI交互界面

简单描述一下你的问题，例如：这里报错了，是为什么。AI会给你答案 -> 直接运行一般就可以

仍然报错 -> 继续找原因

报错了，怎么办? -> 继续和AI进行交互

用对话的方式写代码 Vibe Coding

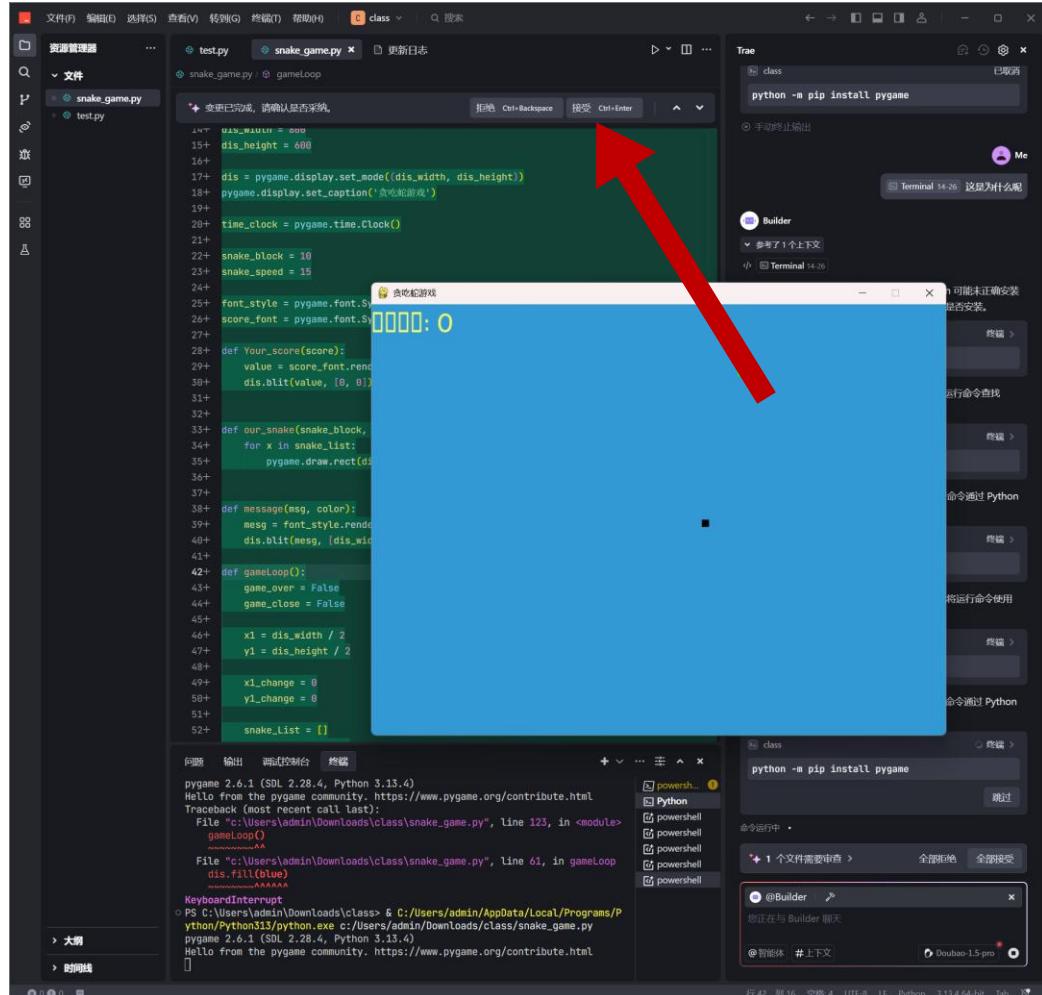
The screenshot shows a code editor with a Python file named `snake_game.py` containing code for a snake game. Below it is a terminal window showing the command `python -m pip install pygame` being run, which fails because the pip command is not found. The AI interface on the right is used to troubleshoot this issue.

首先复制报错的命令行，并粘贴至AI交互界面

简单描述一下你的问题，例如：这里报错了，是什么。AI会给你答案 -> 直接运行一般就可以

仍然报错 -> 继续找原因
AI可以帮你一步步地进行解决

用对话的方式写代码 Vibe Coding



最终，调试成功！

当然，这个程序还可以进一步优化

- 看起来没有引入中文字体
- 像素点可能有点太小了
- ...

别怕，继续和AI进行交互
(但建议这个能跑通的程序可以
accept一次修改~)

用对话的方式写代码 Vibe Coding

```
file 编辑 选择 查看 转到 终端 帮助 class 搜索
资源管理器 ... test.py snake_game.py x 更新日志
文件 ...
snake_game.py
test.py
+ 变更已完成, 请确认是否采纳。
拒绝 Ctrl+Backspace 接受 Ctrl+Enter
Trac Python 安装路径下的 pip 可执行文件。
class where python -m pip
前序命令未读取到 pip 路径输出, 接下来我们将运行命令通过 Python 模块方式使用 pip 安装 pygame 库。
class python -m pip install pygame
从输出可知 Python 没有安装 pip 模块, 接下来我们将运行命令使用 ensurepip 来安装 pip。
class python -m ensurepip --upgrade
pip 已成功安装, 版本为 25.1.1, 接下来我们将运行命令通过 Python 模块方式使用 pip 安装 pygame 库。
class 运行被中断 python -m pip install pygame
○ 手动终止输出
Me 现在已经可以运行了, 但所有显示中文的地方都无法正常显示。请帮我解决。
Builder 用不支持的字体在游戏中中文显示异常, 我将把 bahnscrift 和 comicsans 替换为支持中文的 SimHei (黑体), 以此解决游戏中中文显示异常的问题。
snake_game.py:1-200
代码可知, 中文显示依赖字体样式, 原代码使用的字体可能不支持中文, 我将修改字设置以支持中文显示。
snake_game.py
问题 输出 调试控制台 终端
pygame 2.6.1 (SDL 2.28.4, Python 3.13.4)
Hello from the pygame community. https://www.pygame.org/contribute.html
Traceback (most recent call last):
File "C:\Users\admin\Downloads\class\snake_game.py", line 123, in <module>
    gameLoop()
  File "C:\Users\admin\Downloads\class\snake_game.py", line 64, in gameLoop
    pygame.display.update()
KeyboardInterrupt
PS C:\Users\admin\Downloads\class> & C:/Users/admin/AppData/Local/Programs/Python/Python313/python.exe c:/users/admin/Downloads/class/snake_game.py
pygame 2.6.1 (SDL 2.28.4, Python 3.13.4)
Hello from the pygame community. https://www.pygame.org/contribute.html
○ PS C:\Users\admin\Downloads\class>
大明 时间线
@Builder 您正在与 Builder 聊天
@智能体 #上下文 Doubao-1.5-pro
行 42, 列 16 空格 4 UTF-8 () Python 3.13.4 64-bit Tab Esc
```

```
file 编辑 选择 查看 转到 终端 帮助 class 搜索
资源管理器 ... test.py snake_game.py x 更新日志
文件 ...
snake_game.py
test.py
+ 变更已完成, 请确认是否采纳。
拒绝 Ctrl+Backspace 接受 Ctrl+Enter
Trac Python 安装路径下的 pip 可执行文件。
class where python -m pip
前序命令未读取到 pip 路径输出, 接下来我们将运行命令通过 Python 模块方式使用 pip 安装 pygame 库。
class python -m pip install pygame
从输出可知 Python 没有安装 pip 模块, 接下来我们将运行命令使用 ensurepip 来安装 pip。
class python -m ensurepip --upgrade
pip 已成功安装, 版本为 25.1.1, 接下来我们将运行命令通过 Python 模块方式使用 pip 安装 pygame 库。
class 运行被中断 python -m pip install pygame
○ 手动终止输出
Me 你的分数: 0
Builder 已修改 @snake_game.py 中的字体设置, 将 bahnscrift 和 comicsans 替换为支持中文的 SimHei (黑体), 以此解决游戏中中文显示异常的问题。
snake_game.py
问题 输出 调试控制台 终端
gameLoop()
File "C:\Users\admin\Downloads\class\snake_game.py", line 64, in gameLoop
    pygame.display.update()
KeyboardInterrupt
PS C:\Users\admin\Downloads\class> & C:/Users/admin/AppData/Local/Programs/Python/Python313/python.exe c:/users/admin/Downloads/class/snake_game.py
pygame 2.6.1 (SDL 2.28.4, Python 3.13.4)
Hello from the pygame community. https://www.pygame.org/contribute.html
○ PS C:\Users\admin\Downloads\class> & C:/Users/admin/AppData/Local/Programs/Python/Python313/python.exe c:/users/admin/Downloads/class/snake_game.py
pygame 2.6.1 (SDL 2.28.4, Python 3.13.4)
Hello from the pygame community. https://www.pygame.org/contribute.html
○ PS C:\Users\admin\Downloads\class>
大明 时间线
@Builder 您正在与 Builder 聊天
@智能体 #上下文 Doubao-1.5-pro
行 42, 列 16 空格 4 UTF-8 () Python 3.13.4 64-bit Tab Esc
```

用对话的方式写代码 Vibe Coding

现在，去完成属于你的代码吧！

课程与考核 Logistics

主讲教师：滕佳烨 www.tengjiaye.com

线上联系：tengjiaye@sufe.edu.cn

线下联系：统计与数据科学学院主楼1307 (需提前约时间)

2016-2020 上海财经大学统计与管理学院，统计学实验班 本科

2020-2024 清华大学交叉信息研究院，计算机科学与技术 博士

2024- 上海财经大学统计与数据科学学院 助理教授

主要研究方向：机器学习理论，大模型理论

主要荣誉：2025 CCF理论计算机科学博士学位论文激励计划提名（全国每年五个），2024 清华大学优秀毕业生（全院1人）、2023 清华大学王大中奖学金（全校20人）、2022 清华大学国家奖学金、2020 上海市优秀毕业生、2019 全国大学生数学竞赛（非数组）国家一等奖、2018/2017全国大学生数学建模竞赛国家二等奖。

课程与考核 Logistics

主要内容

Week 1-2: 大模型初探

- 大模型的基本工具
- 大模型的历史发展

Week 3-5: 大模型应用基础

- 大模型的基本应用与特点

Week 6: 开题展示

Week 7-12: 大模型应用前沿

- CoT, ICL, RAG...

Week 13-15: 大模型理论

- Transformer ...

Week 16: 期末展示

课程与考核 Logistics

平时成绩(30')

- 考勤成绩(10'):
 - 请假需要【提前】联系助教报备
 - 允许至多一次无故缺勤，第二次开始缺勤每次扣总评 2 分
- 课程作业(20'):
 - 共 12 次作业，可以选择其中 5 次完成
 - 作业大多为开放性题目，提交【非空白】的【合理】作业即为满分
 - 无故迟交扣总评 1 分，不交扣总评 4 分

开题报告(20')

- 第【五】周周末前提交开题报告
- 第【六】周进行开题展示

期末展示(50')

- 第【十五】周周末前提交课程设计论文
- 第【十六】周课堂进行期末展示

课程与考核 Logistics

平时成绩(30')

开题报告(20')

- 第【五】周周末前提交开题报告
 - 由于班级人数较多，可以自行 1-3 人进行组队，一旦选定无法更改
 - 开题报告要求在 **【一页纸以内】**
 - 开题报告应包含项目背景、相关文献、项目预期、大概计划时间点
- 第【六】周进行开题展示
 - Slides 最长两页，一页项目背景（为什么做），一页项目预期（怎么做）
 - 展示时长另行通知
- 开题报告评分标准：趣味性或有意义(10')，可行性(5')，格式规范(5')

期末展示(50')

- 第【十五】周周末前提交课程设计论文
- 第【十六】周课堂进行期末展示

课程与考核 Logistics

平时成绩(30')

开题报告(20')

期末展示(50')

- 第【十五】周周末前提交课程设计论文 (30')
 - 课程设计论文要求在**【三页纸以内】**
 - 论文应包含项目技术架构图、用户手册、技术报告等，可根据项目类型进行调整
 - 论文评分标准：趣味性或有意义(10')，工作量与深度(10')，表达清晰格式规范(10')
- 第【十六】周课堂进行期末展示 (20')
 - 以海报形式进行期末展示
 - 以组为单位，每个组至多可以投三票
 - 可以投给自己小组，但三票必须投给三个不同的组
 - 最终展示每个组总票数，以此为评分依据

总结 Take-away Messages

大语言模型

- 大 -> 训练数据大、模型大
- 语言 -> 对话能解决很多问题

大模型的工具

- 尝试使用Trae
- 很多任务可以由代码完成，而大模型能够参与写代码工作 (Vibe Coding)

第一次作业：利用大模型生成一个小游戏，要求

- 有和大模型的交互过程（不可以一句话直接生成）
- 玩法可以没有创新，但得有趣