

# 周梓洋

电话: +1 604 603 6794

国内: 157 5781 7693

邮箱: zhouziyang0630@163.com

生日: 2005年6月

个人主页: zhouziyang.me



## 教育背景

2023.09-2027.6

英属哥伦比亚大学(University of British Columbia)

统计学 (理学学士)

- 主修课程: 微积分、多元积分、线性代数、应用线性代数、概率论、统计学、似然估计、统计模型、软件工程原理、数据分析(R, Python)、数据结构与算法(C, C++)、机器学习(Python)、NLP、线性回归、聚类分析、应用程序开发(Java)、随机过程分析、算法设计

## 项目特长

- AI 与 Agent:** 熟悉 Vibe Coding、Multi-Agent 协作、本地模型调用、Agent 记忆机制、提示词(prompt)工程
- 机器学习:** 熟悉 NLP 文本处理、tokenization、词向量、主题建模、随机森林、逻辑回归与模型评估
- 数据分析:** 熟悉 Pandas 数据清洗、特征工程、可视化 Dashboard、实验复盘与结果解释
- 工程实践:** 具备从需求拆解、原型实现、部署调试到人工校验 AI 输出的完整项目迭代经验

## 实践经历

2025年7月-8月

国泰海通证券 IPO 项目组实习

IPO 实习生

- 参与某企业北交所 IPO 项目, 协助完成行业研究、财务数据整理与上市申报材料分析。
- 使用 Excel / Python 对企业经营数据、财务指标及行业可比公司进行分析, 提高数据处理效率与准确性。
- 协助项目组进行招股书相关数据核验与信息披露整理, 支持上市材料准备流程。
- 参与市场与行业数据调研, 对企业收入结构、成长性及风险因素进行分析与总结。

2025年1月-5月

SpeedUp 教育机构

兼职数学教师

- 提供一对一辅导与学术支持, 帮助学生成绩平均超出班级 10%以上。
- 因教学条理清晰、擅长简化抽象数学概念而获得学生好评。
- 向本科生讲授高级微积分课程, 重点讲解多元积分与复杂问题解决技巧。

## 项目经历

2026年6月

腾讯 AI-HR 新人 90 天成长系统 (腾讯 AI-HR Project)

独立工作者

项目重点: 基于腾讯 AI-HR 场景, 使用 Multi-Agent 协作、AI 辅助网页部署与数据库接入, 将产品想法推进为可演示项目。

- 设计新人 90 天成长路径原型, 将新人画像、成长任务、导师反馈和 HR 视角串联成完整 AI-HR 使用场景。
- 实践 Codex + CodeBuddy 的 Multi-Agent 协作方法, 让不同 AI 工具互相调用, 承担实现、部署调试、逻辑审查和表达优化。
- 使用 AI 辅助完成网页部署、数据库接入和问题排查, 并人工检查流程合理性、数据保存、异常回退和演示稳定性。

2026年5月 - 现在

本地 Qwen 中文短篇生成 Pipeline

独立工作者

项目重点: 使用本地 Qwen 模型和 Python 流程控制, 模拟 Agent 自驱式写作, 让模型自己提问、回答、记录并持续生成小说。

- 使用本地 Qwen 模型搭建中文短篇生成流程, 将小说创作拆成设定生成、大纲规划、分段续写、上下文回顾和成稿。
- 通过 Python 控制本地模型调用节奏, 让模型围绕角色、剧情和前文内容自己生成问题、回答问题, 并用于后续写作。
- 设计简化版 Agent 记忆机制, 使模型读取此前设定、片段和日志, 在后续内容中保持人物关系和剧情走向一致。

2026年5月 - 现在

ZiyangZhou.com 个人作品集网站

独立工作者

项目重点: 集中展示目前所有项目成果, 将作品集、GitHub 仓库、项目说明和简历内容整理成面向招聘场景的展示入口。

- 设计并搭建中英文个人作品集网站, 集中展示目前完成的项目、经历、简历、写作和联系方式。
- 使用 AI 辅助梳理信息架构, 把不同项目介绍、技术背景、仓库链接和中英文文案整理成统一展示结构。
- 将网站作为简历项目的证据中心, 使面试官可以从简历跳转到作品集, 再进入 GitHub 查看实际代码和项目成果。

2026年2月 - 2026年4月

ChatGPT API 浏览器工具

独立工作者

项目重点: 让本地模型和 Agent 通过 API/浏览器工具调用更强模型能力, 在长文本任务中减少约 3 成 token 浪费并提升生成质量。

- 设计本地 API 桥接工具, 使本地模型或 Agent 能够调用 ChatGPT 相关能力完成长文本生成、内容改写和多轮任务推进。
- 探索本地模型 + 外部 API 工具的协作方式, 让本地模型负责拆任务和流程控制, 复杂生成任务交给更强模型处理。
- 针对长文本任务中 token 消耗高、上下文容易变乱的问题, 将任务拆成多次接口调用, 减少重复输入和无效上下文占用。

2026年1月 - 2026年4月

NLP 文本分析与主题建模项目 | 机器学习项目

独立工作者

技术栈: Python, Jupyter Notebook, Scikit-learn, spaCy, Gensim, GloVe, LDA。

- 完成文本预处理、词向量探索和主题建模流程, 覆盖从数据清洗到结果解释的 NLP 实验链路。
- 使用 GloVe 词向量分析词相似度、最近邻、词表覆盖和表示偏差, 理解预训练 embedding 的适用边界。
- 使用 CountVectorizer 与 LDA 构建主题模型, 并检查 word-topic 与 document-topic 关系, 解释不同文本集中的主题分布。

2026年1月 - 2026年4月 **Airbnb 房源价格预测建模 | 机器学习项目** 独立工作者

技术栈: Python, Pandas, Scikit-learn, Random Forest, ColumnTransformer, GridSearchCV

- 基于纽约 Airbnb 房源数据构建价格预测模型, 完成任务定义、训练/测试集划分、探索性分析、特征工程和模型评估流程。
- 针对地理位置、房型、可用天数、评论数量等房源特征进行清洗和编码, 并使用 ColumnTransformer 统一处理数值与类别变量。
- 比较 Dummy Regressor 与 Random Forest 等模型表现, 通过交叉验证、 $R^2$ 、MSE 等指标评估模型效果。
- 使用 GridSearchCV 调整模型参数, 并在结果分析中讨论房源价格预测的误差来源和数据局限。

2026年1月 - 2026年4月 **文本分类与 Logistic Regression 优化 | 机器学习项目** 独立工作者

技术栈: Python, Scikit-learn, Logistic Regression, CountVectorizer, GridSearchCV, Cross-validation

- 基于推文数据构建二分类文本分类模型, 使用 CountVectorizer 将文本转化为 bag-of-words 特征。
- 从 Dummy Classifier 基线模型开始, 逐步训练 Logistic Regression, 并通过交叉验证评估模型泛化表现。
- 调整 max\_features、正则化强度等参数, 比较不同文本向量化和模型配置对验证集表现的影响。
- 分析模型预测概率和 learned coefficients, 解释哪些词语特征对分类结果影响较大。

2026年1月 - 2026年4月 **Recipe 文本聚类与无监督学习分析 | 机器学习项目** 独立工作者

技术栈: Python, Scikit-learn, K-means, DBSCAN, SentenceTransformer, UMAP

- 基于菜谱名称和文本信息进行无监督聚类实验, 比较 bag-of-words 与 sentence embeddings 对短文本聚类效果的影响。
- 使用 K-means、DBSCAN 等聚类方法分析菜谱文本相似性, 并观察不同算法在噪声文本和短文本场景下的表现差异。
- 使用 UMAP 对高维文本向量进行降维可视化, 辅助解释聚类结果和主题分布。
- 对聚类结果进行人工解释, 识别菜谱名称过长、语义信息不足导致的混合聚类和噪声问题。

2026年1月 - 2026年4月 **Adult Census 数据预处理与分类 Pipeline | 机器学习项目** 独立工作者

技术栈: Python, Scikit-learn, ColumnTransformer, Pipeline, Decision Tree, One-hot Encoding

- 基于 Adult Census 数据集构建监督学习分类流程, 完整处理训练/测试集划分、特征类型识别、预处理和模型评估。
- 将特征拆分为数值、类别、二元、序数和需删除字段, 并分别应用缺失值填补、标准化、one-hot encoding 等处理方式。
- 使用 Scikit-learn Pipeline 和 ColumnTransformer 封装预处理逻辑, 减少数据泄漏风险并提升实验可复现性。
- 比较 Dummy Classifier 与 Decision Tree 等模型的交叉验证结果, 用基线模型判断学习模型是否真正带来提升。

2026年1月 - 2026年4月 **Spotify 歌曲属性决策树分类 | 机器学习项目** 独立工作者

技术栈: Python, Scikit-learn, DecisionTreeClassifier, Cross-validation, Model Visualization

- 基于 Spotify 歌曲属性数据训练决策树分类模型, 分析音频特征与类别标签之间的关系。
- 从手写 decision stump 和 toy example 入手, 理解决策树的划分逻辑、树深度、叶节点和过拟合风险。
- 使用 10-fold cross-validation 比较 Dummy Classifier 与 Decision Tree 的表现, 并通过训练分数和验证分数差诊断过拟合。
- 调整 max\_depth 等超参数, 选择验证表现更稳定的模型, 并在 held-out test set 上进行最终评估。

2025年8月 - 2025年11月 **AI 股票分析与可视化工具** 独立工作者

- 项目重点: 使用 AI 辅助完成数据分析工具的需求拆解、模块整理和可视化迭代, 将个人投研观察转化为可交互 Dashboard。
- 设计面向个人投研复盘的 A 股分析 Dashboard, 将行情、逐笔成交、资金流、大单行为和日内价格变化整理成可视化界面。
- 使用 AI 辅助拆分数据获取、清洗、指标计算和图表展示模块, 把早期零散脚本整理成更容易维护的项目结构。
- 在迭代中区分观察指标和交易建议, 尝试异常检测、短期方向分类和行为聚类等实验模块, 用于探索数据特征。

2024年9月 - 现在 **中国玻璃期货量化交易策略** 独立工作者

技术栈: TradingView, Pine Script, 回测分析, Python, Pandas, Matplotlib.

- 围绕中国玻璃期货设计量化交易策略, 基于历史行情复盘交易信号、趋势阶段和风险控制逻辑。
- 使用回测结果评估策略在不同市场阶段的表现, 关注收益、回撤、交易频率和参数稳定性, 而不是只展示单一收益数字。
- 结合手续费、滑点和参数敏感性复盘策略风险, 使项目重点从收益展示转向策略逻辑是否可解释。

2024年1月 - 2024年5月 **睡眠追踪 App** 独立工作者

技术栈: Java, Swing, OOP, JSON Persistence, JUnit, UML.

- 基于用户故事实现 Java 睡眠记录桌面应用, 支持记录睡眠、查看统计、设置目标、分类汇总、保存和加载数据。
- 使用面向对象设计拆分 Sleep、SleepOverall、Goals、Date、EventLog 等模型类, 支撑睡眠记录、目标管理和事件日志功能。
- 实现 JSON 读写持久化, 并使用 JUnit 覆盖模型类与 persistence 行为, 保证核心功能可测试。

## 技术技能

- **编程语言:** Python, JavaScript, Java, C++, C, R, Vibe Coding, AI Agent
- **数据分析与可视化:** Pandas, NumPy, Matplotlib, ggplot, EDA (探索性数据分析)、数据清洗、统计分析、数据可视化
- **机器学习:** Scikit-learn、tidymodels、回归模型、分类模型、KNN、Random Forest、Gradient Boosting、XGBoost、特征工程、交叉验证、超参数调优、时间序列预测
- **无监督学习与 NLP:** K-Means、DBSCAN、Topic Modeling (LDA)、TF-IDF、Word Embedding、文本向量化、语义分析、文本聚类
- **开发工具:** Git、GitHub Actions、Streamlit、TradingView (Pine Script)
- **计算机基础:** 数据结构与算法、面向对象编程 (OOP)、多线程、数据库管理、软件开发基础