

# 钟元浩

(+86)189-2896-1614 | ZephyrZhong248@gmail.com | [github.com/zyh3699](https://github.com/zyh3699)<sup>1</sup>

## 教育背景

北京理工大学 | 计算机科学拔尖班, 计算机学院 | 工学学士 2023.09—2027.06

**GPA: 3.77/4.0 (专业前 2%)**, 荣获国家奖学金、校一等奖学金及“优秀学生”称号; 全国大学生数学建模竞赛一等奖、全国大学生数学竞赛二等奖、亚太地区数学建模竞赛二等奖; Kaggle 竞赛 *House Prices Prediction* 排名 Top 3%。

**专业课程:** 程序设计方法与实践 (100)、互联网应用开发基础训练 (100)、数据结构与算法 (95)、线性代数 (92)、数值分析 (92)、离散数学 (91)、概率论与数理统计、数据库原理与设计、知识工程、组合数学等。

**学术成果:** 《Doctopus: A System for Budget-aware Structural Data Extraction from Unstructured Documents》, *SIGMOD 2025*(CCF A), 一作; 《BestCore: Minimum Core Set Selection under Dirty Data Scenarios》, *VLDB 2025*(CCF A), 一作。

**专利与出版:** 《Deem: 一种面向深度学习模型的数据训练方法》, 《Doctopus: 一种基于预算约束的高效数据提取方法》。

## 技术能力<sup>2</sup>

- 编程语言:** 精通 Python (Sklearn, PyTorch)、SQL, 熟练使用 C++、C, 熟悉 Java Matlab、JavaScript。
- 开发工具与平台:** VSCode、PyCharm、Jupyter notebook、IntelliJ、Git、Gitee、GitHub、GitLab。
- 人工智能与机器学习:** 熟练掌握 Transformer 系列模型 (如 LLaMA、Qwen、GPT、Deepseek 等); 熟悉 SFT、RLHF 以及端到端部署流程; 了解 RAG、Agent 等 LLM 应用技术; 了解 DeepSpeed、Megatron 分布式训练框架。
- 数据分析与可视化:** 熟练使用  $\text{\LaTeX}$ , Excel, SPSS, 数据清洗与可视化经验丰富。

## 工作经历

广州灵犀互动娱乐有限公司 | 技术中台算法实习生 2024.07—2024.08

- 负责用户数据异动分析算法的设计、实现与优化。**运用 SQL 进行高效数据查询, 结合 Python 和 Excel 完成数据清洗、处理与可视化工作。构建了涵盖登录-注册比率、曝光量、转化率、消费金额等核心指标的用户活跃度评估体系, 精准定位系统及业务异常。
- 参与广告投放盈利预测模型的构建与优化。**全程参与从数据选择、特征工程到算法建模与报告撰写的流程, 通过对比优化后选择 XGBoost 模型, 成功提升验证集上 F1-score 至 **0.73**, 为广告投放策略提供有力支持。

## 项目经历

Clue 中国文学语言理解评测 | *CCL25-Eval* 2025.03—今

- 应用 BERT、RoBERTa、GPT 等 Transformer 模型进行情感分析任务, 利用 GPT、T5 深度生成模型进行文学语言风格转换与生成, 使用 CycleGAN 实现风格迁移, 结合知识图谱辅助理解古代文学文本。
- 从多维度全面评估模型在文学语言理解方面的能力, 最终在评测中成功进入前 10%。

BestCore 脏数据情景下的最小核心集选择 | 算法设计与实现 2025.03—今

- 基于深度学习训练策略理论, 设计并实现了针对脏数据场景的梯度近似算法, 结合多目标优化策略, 选择最小核心集。
- 针对多种机器学习与深度学习模型进行性能测试, 识别系统瓶颈, 通过贪心算法优化核心集选择过程, 成功缩短训练时间约 30%, 将核心集规模压缩至原数据集的 3%-5%。
- 该项目为实验室研究项目的一部分。

Doctopus 基于预算约束的非结构化数据提取系统 | 主要维护者 (开源项目) 2025.01—2025.03

- 设计并实现了一个高效的数据提取框架, 结合索引构建、质量评估与动态选择策略, 在预算约束下实现高效的非结构化数据提取。
- 通过动态选择属性提取方案, 如 OpenIE、PZ、Lotus 等大模型, 最大化预算下的数据提取准确率。
- 引入回归测试与 GitHub Actions 进行持续集成, 采用标准工作流进行开发与问题修复。
- 独立负责项目文档撰写、算法设计、前后端开发以及报告与视频制作, 项目成果被 *SIGMOD 2025 Conference* 接受。

## 个人总结

- 本人性格乐观开朗、成绩优异、具备较强的自驱力与团队协作能力, 善于解决问题与创新思考。
- 英语水平优秀 (六级成绩 604), 具备较强的英语沟通与学术写作能力, 能够流利进行工作交流。
- 熟悉大模型架构与算法, 具备扎实的机器学习、深度学习基础, 并拥有丰富的数据分析与开源项目经验, 擅长技术写作, 持续关注前沿技术发展。

<sup>1</sup> 下划线内容包含超链接。 <sup>2</sup> 与求职岗位无关的技能省略或用灰色表示。