

萧尧

本简历总体由 ChatGPT 翻译自英文版，可能存在语义错误或不顺，敬请谅解。

(+86) 186-2182-3612 | ✉ yaoxiao@g.harvard.edu | 🏠 charlie-xiao.github.io | 🌐 Charlie-XIAO | 📄 [yao-xiao-200073244](https://yao-xiao-200073244.github.io)

教育背景

哈佛大学 | 理学硕士 | 计算科学与工程

2024.09 – 至今

- GPA: NA/NA, 进行中...

上海纽约大学 | 理学学士 | 荣誉数学 | 计算机科学

2020.09 – 2024.05

- 荣誉数学 GPA: 4.00/4.00, 包括: 线性代数、数学建模、概率论、数值分析等。
- 计算机科学 GPA: 3.97/4.00, 包括: 算法、操作系统、计算机网络、软件工程等。

出版物

标有 [†] 的作者按形式顺序排列, 否则按贡献排序

- Yuejie Wang, Qitong Men, **Yao Xiao**, Yongting Chen, and Guyue Liu. 2024. ConfMask: Enabling Privacy-Preserving Configuration Sharing via Anonymization. In *Proceedings of the ACM SIGCOMM 2024 Conference (ACM SIGCOMM'24)*. Association for Computing Machinery, New York, NY, USA, 465–483. doi:10.1145/3651890.3672217
- Shengbin Yue, Shujun Liu, Yuxuan Zhou, Chenchen Shen, Siyuan Wang, **Yao Xiao**, Bingxuan Li, Yun Song, Xiaoyu Shen, Wei Chen, Xuanjing Huang, and Zhongyu Wei. 2024. LawLLM: Intelligent Legal System with Legal Reasoning and Verifiable Retrieval. In *29th International Conference on Database Systems for Advanced Applications (DASFAA'24)*. doi:10.48550/arXiv.2309.11325
- Xinyu Li[†], **Yao Xiao**[†], and Yuchen Zhou[†]. 2023. Efficiently Visualizing Large Graphs. doi:10.48550/arXiv.2310.11186

研究经历

通过匿名化实现网络配置共享时的隐私保护 | SIGCOMM'24 | GitHub

2022.10 – 2024.08

指导教师: 刘古月教授, guyue.liu@gmail.com

- 提出了 ConfMask 框架, 以系统性地匿名化网络配置中的拓扑和路由信息。
- 设计了适用于不同协议的匿名化算法, 减少了被反匿名化的风险, 同时确保匿名化后网络的实用性。
- 严格证明了网络匿名化后保留了路由的等效性和多个实用的网络性质。
- 负责了端到端网络配置匿名化系统的整体实现和代码及结果可复现性的评估。

大规模语言模型推理的高效分布式系统 | 本科毕业设计

2023.09 – 2024.01

指导教师: 刘古月教授, guyue.liu@gmail.com

- 根据只有自注意力层依赖 KV 缓存的特性, 实现了其余层超出 KV 缓存限制的大批量请求处理。
- 在自注意力层中动态批处理预填充和解码, 以减轻不同的 Transformer 输入长度造成的并行气泡。
- 将多个短的注意力计算与最长的计算进行打包, 同时交换 KV 缓存以最小化开销。

高效可视化大规模图 | 本科生研究基金 | ArXiv

2022.05 – 2022.08

指导教师: 薛杰教授, jiexue@nyu.edu

- 基于 t-SNE 算法设计了专门用于图的 t-SGNE 算法, 利用节点之间的邻接关系实现了 6.7 倍的计算效率提升。
- 提出了基于拉普拉斯特征映射和最短路径的图嵌入算法 SPLEE, 以适配 t-SGNE 算法。
- 将 SPLEE 和 t-SGNE 结合, 用于可视化包含 30 万个节点和 100 万条边的大规模图, 提升约 10% 的可视化效果。

工作经历

Scikit-learn | 开源项目 | 核心开发者 | [GitHub](#)

2023.04 – 至今

技能: Python, Cython, JavaScript, Sphinx, scikit-learn, numpy, scipy, pandas, polars

- 参与项目维护, 例如单元测试覆盖率、代码重构、入参验证改进、自动化 GitHub 工作流等。
- 主导了稀疏数组支持、模型 UI 显示、网站与文档等工作。
- 重新设计了 scikit-learn 主网站, 进行了各种文档与网站 UI / UX 改进。
- 贡献了 118 个已合并的拉取请求, 包括 bug 修复、性能改进、新特性和文档改进等。
- 作为项目维护者, 对问题进行分诊、审查拉取请求, 并指导新贡献者参与项目。

复旦大学 DISC 实验室 | 实验室助理 | [DASFAA'24](#) | [GitHub](#)

2023.05 – 2023.08

技能: Python, PyTorch, HuggingFace, 大语言模型, 指令调优

- 负责构建了 40.3 万条法律知识指令数据, 使用法律三段论提示以提高问答专业性。
- 基于 Baichuan 13B Chat, 微调了智能法律服务专用大语言模型 DISC-LawLLM。
- 设计了一个可验证的知识检索模块, 以注入外部知识并增强输出的真实性。
- 负责构建了包含主客观多个维度的智能法律系统评估方案。

教学经历

- 线性代数, MATH-SHU.0140, 上海纽约大学, 助教, 2024 秋季学期
- 微积分 II, MATH-SHU.0131, 上海纽约大学, 助教, 2021 秋季学期, 2023 秋季学期
- 操作系统, CSCI-UA.0202, 纽约大学, 助教, 2023 春季学期

项目经历

Deskulpt: 跨平台桌面组件定制工具 | [GitHub](#)

2024.03 – 至今

技能: Rust, TypeScript, Tauri, React, Vite, SWC

- 开发了一个跨平台高度可定制的桌面组件系统, 支持使用 React / TypeScript 编写桌面组件。
- 在 Deskulpt 中集成了丰富的桌面组件开发工具, 提供编辑器和类型提示等, 以达到更流畅的组件创建和开发体验。
- 使用 Tauri 构建 Deskulpt, 以确保系统安全性并兼容 Windows、macOS 和 Linux 平台。
- 在后端利用 Rust 的异步能力, 确保 UI 和系统资源之间的流畅响应与互动。
- 实现了多种安全保障, 例如 CSP 保护、文件系统访问限制、前端能力限制等。

CampusHelper: 微信 / 支付宝小程序

2023.12 – 至今

技能: TypeScript, MongoDB, 微信 / 支付宝云, 小程序框架

- 开发了微信 / 支付宝小程序 CampusHelper, 一个供校内同学们交易自己的闲置劳动力以创造价值的公开交易平台。
- 利用小程序云 (包括云函数、数据库和存储等功能), 提升了小程序数据管理和服务可靠性。
- 通过云技术、懒加载、列表虚拟化等手段优化了小程序性能。
- 构建了一个简洁、一致且用户友好的界面, 提升了整体用户体验。
- 在 2023 支付宝小程序开发者大赛中获得了二等奖 (第 4 名)。

YouTube 界面自定义插件 | 课程项目 | [GitHub](#)

2023.02

- 开发了一个火狐浏览器插件, 支持更改色彩主题、重新排列和自定义 YouTube 界面元素。
- 创建了功能文档和贡献指南, 并在 Mozilla 插件商城发布了 v1.0 自分发版本。

经济体中不平等交易过程模拟 | 课程项目 | [论文](#) | [GitHub](#)

2022.12

- 通过细致的随机交易函数模拟了经济体中的不平等交易过程, 以反映现实中经济体的交易过程。
- 发现现实中经济体内的最终财富分布符合 gamma 分布或 beta prime 分布的形状。

- 用弹性系统模拟将陀螺仪，使用微分方程公式化系统并通过欧拉方法求解。
- 模拟了多个陀螺仪的垂直堆叠，并发现在假设中间轴可弯曲的情况下，它们将遵循陀螺进动。

荣誉奖项

- [1] **荣誉学位**，上海纽约大学，2024
- [2] **一级认证**，[CRLA 国际辅导培训项目](#)，2024
- [3] **二等奖，第四名**, [支付宝小程序开发者大赛](#), 2023
- [4] **Meritorious Winner**，[美国数学建模大赛](#)，2023
- [5] **院长荣誉名单**，上海纽约大学，2020 – 2021，2021 – 2022，2022 – 2023

技术能力

- [1] **编程技能**: 精通 Python, Rust, C, TypeScript; 熟悉 Java, C++, Julia, MATLAB
- [2] **框架和软件包**: Tauri; React; Numpy, Scipy, Pandas, Polars, Scikit-learn, Matplotlib, Seaborn, PyTorch
- [3] **DevOps**: CircleCI, Github Actions; Docker; Git / GitHub; AWS, 微信 / 支付宝云; 计算机网络; Linux
- [4] **语言**: 英语流利