宋昕乐

ericS311@ucla.edu | 131-2227-3608 | github.com/EricSongXinLe

教育经历

加州大学洛杉矶分校 (UCLA) 2023 年 9 月 - 2027 年 6 月（预计）

- 计算机工程专业理学学士 GPA: 3.9/4.0 获院长荣誉提名 UCLA 电子工程和计算机荣誉协会(HKN)成员
- 2023 年获得校 John Richard Leffler 奖学金（唯一获奖的国际学生），2025 年获 John DeGroff Haller 奖学金
- 相关课程：软件开发，计算机体系结构，数据结构，算法，并行计算

专业技能

- 编程语言：C/C++，Python，CUDA，JavaScript（Node.js/React），Shell
- 开发工具：Git，Docker，vLLM，Linux，GDB，Perf，CMake，Apache，OpenMP
- 系统能力：操作系统，调度与资源管理，缓存优化，基础设施稳定性设计，并行计算

实习经历

腾讯科技（上海）有限公司 2025 年 8 月 - 2025 年 9 月

- 适配并部署腾讯自研智能体平台至中科海光 K100-AI/BW1000 及 NVIDIA H20/A10，完成神农 2 系列约 10 个大模型的精度验证与模型性能测试。
- 独立开发可复用的 Python 自动化测试脚本，实现数据准备-模型加载-输出采集-结果生成的全流程自动化，将完整测试流程耗时从 7 - 8 小时缩短至约 30 分钟，显著提升测试效率。
- 参与 vLLM-0.9.2-dcu 的模型权重解密与 license 鉴权模块移植，保证模型正确解密，密文/明文模型精度一致，同时确保无 license 时强制中断加载，提升模型交付安全性。

项目经历

社团搜索网站 (Find Your Clubs) | JavaScript, React, Node.js, MongoDB, Git 2024 年 4 月 - 2024 年 6 月

- 领导 6 人团队开发 UCLA **首个智能社团**搜索平台，部署于腾讯云 VPS，服务**上百名**用户寻找适配社团。
- 构建**用户认证，推荐系统，智能搜索**的 **REST API**，支持账号类型权限区分，提升数据安全与查询效率。
- 编写**通用组件**如 **ClubBlock**，优化 React 前端结构与页面响应性能，加载速度**提升 20%**，显著提升用户体验。
- 处理**密码加密**（MD5）、输入校验、数据库认证等安全问题，提升**系统稳定性**与隐私保护。
- 部署于**腾讯云** 2C2G 4Mbps 实例，在轻负载场景下**稳定支持多用户访问**与平台演示，**独立完成全过程 DevOps** 流程。

Brewin 解释器 | Python, 抽象语法树 (AST), 静态类型, Git 2024 年 10 月 - 2024 年 12 月

- **独立使用 Python** 开发 Brewin 语言**解释器**，实现**静态类型检查**、用户**自定义结构体**和**类型转换**，扩展语言功能。
- 设计**作用域管理**单元，实现 for/if 语句、函数调用、静态作用域等核心功能，变量搜索效率提升 **30%**。
- 优化解释器，减少 **AST 解析开销**。通过字典存储**变量作用域**，加速变量查找，提升解释器吞吐量 **55%**。
- 构建全面的**单元测试**框架，覆盖所有边界情况，严格符合 Brewin 语言规范，测试通过率达 100%。

带五级流水线的 RISC-V CPU 模拟器 | C++, 计算机体系结构, GNU Make, Git 2025 年 1 月 - 2025 年 3 月

- 运用 **C++**开发基于 **RISC-V** 指令集的**五级流水线** CPU 模拟器，支持指令解码、执行，并实现**乱序执行**机制。
- 设计并实现 **GSharePlus** 分支预测器，结合全局和局部历史信息，并采用动态选择器优化预测策略。
- 实现**寄存器重命名**(RAT)以消除数据冒险，优化**保留站**+**重排序缓冲区**+**公共数据总线**以解决数据冒险。

科研经历

UCLA 开源计算机体系结构与系统研究(ORCAS)实验室 2025 年 1 月 - 至今

- 研究 **TPU** 计算单元在 **Vortex 开源架构 GPGPU** 的实现来加速矩阵乘法计算，提升芯片 **AI** 训练/推理性能
- 在仿真环境中成功实现了自定义 **RISC-V** 矩阵乘法指令，验证了新指令在多数据类型下的准确性和性能。
- 使用 **C++**仿真平台搭建 intrinsic 指令进行性能分析，新指令在 **512×512 矩阵**乘法中加速约 **2.2 倍**。