

Emanuel20 (Yiding Sun)

☎ +86 15522420639 ✉ emanual20.sun@gmail.com 🏠 [Github: Emanuel20](#)

🎓 教育经历

中国人民大学	2022.09 – 2025.06
人工智能 硕士 高瓴人工智能学院 GPA: 3.85/4.0	北京
南开大学	2018.09 – 2022.06
计算机科学与技术 本科 计算机学院 GPA: 3.84/4.0 专业排名: 6/122	天津

💼 工作经历

美团外卖推荐算法组 - 推荐算法实习生 2022.02 – 2023.12

- 为了解决线上解释目前和推荐模型评分与物品特征对齐不显著的现象，提出了一种基于最大化解释和评分/物品特征互信息的强化学习框架改善了基石模型的性能。相关成果以第二作者在投于CCF-A类会议。
- 分析线上日志，开展用户实验，验证用户对线上推荐解释标签的感知能力，从多方面分析用户对解释标签的偏好，就改进标签线上展出策略提出建议。相关成果以第一作者在 **QUARE 2022 (CCF A workshop)**上报告。
- 设计了一种综合考虑有用性和真实性的线上推荐解释标签的定量评价框架，在线上日志数据上验证了评价框架的有效性。相关成果以第二学生作者身份在**QUARE 2022 (CCF A workshop)**上报告。

🔧 科研经历

Yulan大模型 2023.02 – 至今
基石模型数据组负责人 高瓴人工智能学院，中国人民大学

- 阅读LLaMA等基石模型论文，复现the pile数据集的代码和新闻部分。搭建中文数据通用清洗框架ZHEM，使用该框架处理得到Yulan中文预训练数据集，经1.3B GPT-NeoX检验预训练数据质量和the pile可比。
- 作为主要作者（数据组负责人）参与Yulan-Chat-2项目 [\[Code\]](#)，在LLaMA-2的基础上扩充中文词表与上下文长度，并使用高质量中英文双语数据进行增量预训练和指令微调，同时提升了模型的中英文基础语义和理解能力。模型与同期基于LLaMA-2的大模型相比，具有显著性能优势。

🏠 项目经历

nkucs.icu 经验交流平台 [\[Code\]](#) 2020.01 – 至今

- 为了打通学校信息交流壁垒，作为网站创始人，使用 Github Pages + Docsify 轻量解决方案搭建并长期维护平台。以南开大学学生为用户群体，累计访问量达 20k 人次。目前该项目在 Github 获得76 Stars，22 Forks。

NKU-TEDA ARM编译器 [\[Code\]](#) 2021.04 – 2021.08

- 在2021年全国大学生计算机系统能力大赛编译系统设计赛（华为毕昇杯）中作为主要参与者，在Raspberry Pi上使用 C++ 实现类 C 语言从词法分析、语法分析、中间代码生成及优化和目标代码生成的完整编译过程。
- 设计基于 llvm 的 SSA 中间代码形式思想进行代码优化，实现数据流分析、循环展开、自动向量化等优化Pass，在某些性能测试样例上达到可比gcc-O2的效率。该项目获得比赛全国二等奖。

编译系统原理、高级语言程序设计助教 [\[2021Code\]](#) [\[2022Code\]](#) 2021.03 – 2022.01

- 协助授课教师完成课程改革，重构实验课实验框架代码，编写实验指导书，讲授实验课。

⚙️ 获奖情况

中国人民大学研究生新生奖学金, 二等学业奖学金	2022, 2023
南开大学本科优秀毕业生	2022
南开大学公能奖学金、九五之芯奖学金、教育部-华为智能基座“未来之星”奖学金	2021
中国大学生程序设计竞赛 (CCPC-2019) 秦皇岛站优胜奖	2020
南开大学第15届校程序设计竞赛铜牌	2019

📌 求职意向

岗位：大模型基石模型训练算法，大模型微调算法，推荐算法，编译器开发 工作地：北京