

# Emanuel20 (Yiding Sun)

☎ +86 15522420639    ✉ [emanual20.sun@gmail.com](mailto: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: 3.81/4.0	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

北京三快科技有限公司	2022.02 – 至    今
推荐算法实习生	北京

- 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
基石模型数据组组长	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预训练数据集，重点负责清洗和处理代码、新闻类数据，经小规模（1.3B）模型检验预训练数据和the pile可比。
- 搭建中文数据通用清洗框架ZHEM，并使用该框架处理得到约200B token Yulan中文预训练数据集。

## 🏠 项目经历

nkucs.icu经验交流平台 <a href="#">[Code]</a>	2020.01 – 至    今
--	------------------

- 作为网站创始人，使用 Github Pages + Docsify 搭建并长期维护学习经验交流平台；
- 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62 Stars，23 Forks。

NKU-TEDA ARM编译器 <a href="#">[Code]</a>	2021.04 – 2021.08
--	-------------------

- 使用 C++ 不以任何框架为基础实现 Raspberry Pi 上编译器，可实现 SysY（类C）语言从预编译、词法分析、语法分析、中间代码生成及优化和汇编代码生成的完整编译过程。
- 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，在一定比例的性能测试样例上达到可比gcc-O2的效率。
- 该项目获得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）全国二等奖。

编译系统原理、高级语言程序设计助教 <a href="#">[2021Code]</a> <a href="#">[2022Code]</a>	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	2022.10
南开大学本科优秀毕业生	2022.05
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	2021.12
南开大学学业优秀奖学金、创新奖学金、公能奖学金	2020.12
南开大学第15届校程序设计竞赛铜牌	2019.04