

# 何逸宸 Yichen He

☎ (+86) 181-6516-2739 | ✉ hyc2026@yeah.net | 🏠 hyc2026.github.io | 📧 hyc2026 | 🎖 预备党员



## 教育经历

- 北京航空航天大学 (双一流; 985), 计算机学院, 智能信息处理研究所 (自然语言处理方向) 2021.9 - 2024.1  
硕士 - 计算机技术 (保研) 年级排名: 4/148 (前 3%); 加权平均分: 92.7; GPA: 3.86/4
- 北京航空航天大学 (双一流; 985), 计算机学院 2017.9 - 2021.6  
本科 - 计算机科学与技术 年级排名: 16/257 (前 6%); 加权平均分: 90.7; GPA: 3.81/4
- 相关课程成绩: 数据结构 (100), 离散数学 (100), 数理统计 (97), 人工智能原理 (94), 算法设计与分析 (93), 深度学习 (90)

## 实习经历

- 字节跳动 - AILab-NLP - 算法实习生 2022.5 至今
- 内容质量项目组, 审查头条和抖音发文的内容质量。本人主要工作如下:
  - 识别伪科学发文内容, 通过收集历史审核数据并进行特征增强, 训练文本二分类模型对写作风格进行检测, 判断文本是否为伪科学内容。完成模型从训练、评估到上线部署, 准确率提高 18%, 头条伪科学内容占比由 6.2% 降低至 3.8%。
  - 识别虚假新闻发文内容, 通过 ① 将权威媒体发布的文章录入信源库; ② 搭建临时信源库平台 (后端), 方便审核人员录入非官方发布但真实的内容作为临时信源; 等工作辅助进行虚假识别。准确率提高 5%, 信源覆盖率提升 10%。
  - 识别存在谣言的发文内容, 通过 “题问: [文章] 是否蕴含 [谣言点]” 和 “包含判断依据的回答” 数据对来微调百亿参数级别的大语言模型, 并针对难样例构造特殊的提示, 使模型理解具体审核规则, 准确率由 59% 提升至 85%。同时尝试了高效参数微调方法, 并完成了模型的量化和部署上线。  
🔗 <https://hyc2026.github.io/#/fakenews/README>
- 微软亚洲研究院 - System Research Group - 研究实习生 2021.11 - 2022.5
- 论文: Yanjie Gao, Yichen He, Hongyu Zhang, Haoxiang Lin, ... *An Empirical Study on Low GPU Utilization of Deep Learning Jobs*, Microsoft Systems Journal, 2022. (学生一作)
  - 主要贡献: 研究深度学习模型训练平台中 GPU 利用率低的问题, 对从 400 个 GPU 利用率低的案例中挖掘出的 14 种具体原因进行分析归类, 并针对每类问题提出改进方案。  
🔗 <https://hyc2026.github.io/#/GPUUtil/README>

## 项目经历

- 提交消息自动生成技术研究 - 项目负责人 2022.8 至今
- 论文: Yichen He, Liran Wang, Kaiyi Wang, ... *COME: commit Message Generation with modification embedding*, Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis, 2023. (CCF-A)
  - 主要贡献: 1. 使用最小编辑距离细粒度地表示代码更改; 2. 设计了一种自监督的生成式任务, 使词嵌入包含代码更改的上下文语义信息; 3. 通过决策算法融合了基于检索和基于翻译的方法。该方法在多个数据集上都达到了最好的效果。
  - 论文: Liran Wang, Yichen He, ... *Exploring the Correlation between Commit and Issue for Automatic Commit Message Generation*, Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering, ASE 2023. (在投)
  - 主要贡献: 1. 公开了一个大型的、包含更丰富的代码更改语义描述信息的数据集; 2. 提出生成提交消息的新范式 (提取, 对齐, 微调), 显著提高了提交消息的生成质量。  
🔗 <https://hyc2026.github.io/#/come/README>
- C 语言子集编译器 - 参赛选手 全国大学生计算机系统能力大赛编译系统设计赛全国总决赛一等奖 2020.6 - 2020.8
- 构建 C 语言子集编译器, 在实现基本功能的同时进行了常量传播, 死代码删除, 基本块合并, 循环展开, 循环不变量外提, 图着色, 函数内联, 公共子表达式删除, 内存变量传播, 运算强度削弱, 寄存器分配等优化。
  - 个人职责: 将编译器前端生成的中间代码转化成静态单赋值形式, 并参与了公共子表达式删除、死代码删除、内存变量传播和运算强度削弱等中间代码的优化。  
🔗 <https://hyc2026.github.io/#/compile/README>

## 专业能力

语言 英语 六级 532  
编程 Python, Golang, C/C++ (CCF-CSP 计算机软件能力认证排名前 6.54%)

## 获奖经历

- 2020 全国大学生计算机系统能力大赛编译系统设计赛全国总决赛 一等奖  
2020 智荟杯浦发百度高校极客挑战赛算法赛道 一等奖  
2020 全美大学生数学建模竞赛 Honorable 奖  
2019 十一届全国大学生数学竞赛 二等奖  
2019 全国大学生数学建模竞赛北京赛区 二等奖  
2019 北京市第三十届大学生数学竞赛 二等奖

其他奖项: 获得三好学生 2 次, 校级优秀生 1 次。  
多次获得学习优秀奖学金、学科竞赛奖学金、创新创业奖学金、社会工作奖学金, 其中一等奖 3 次、二等奖 7 次。