

龚启华

年龄: 28 | <https://gongqihua.github.io/> | gongqihua0115@gmail.com
电话: 13391026115 | 微信: gongqihua0115

教育背景

- | | |
|-----------------------|-----------------|
| • 约翰霍普金斯大学 | 08/2021-05/2023 |
| 硕士, 电气与计算机工程 | GPA: 3.5/4.0 |
| • 密苏里大学哥伦比亚分校 | 06/2018-05/2021 |
| 本科, 电气与计算机工程 (辅修: 数学) | GPA: 3.6/4.0 |

技能

- 软件: C++, C#, Python, Go, MATLAB, Unity/Unreal。
- 硬件: Arduino, Raspberry pi, Circuit design, AutoCAD, OrCAD Pspice。
- AI 相关: Machine learning, Deep learning, Audio signal process, Medical imaging process, Bioelectric signal process, Natural Language Processing。

工作经历

1. 华为技术有限公司 (OD) 05/2025-至今
职位: 算法工程师
 1. 基础业务开发与算法支持
 - a) 分词器 & 注音器: 自研 Python 分词器, 基于 Bigram 插值平滑与 Trie+DAG+DP, 在多用户集评测中相较于 jieba 精确度从 55% 提升至 71%; 自研注音器采用最大前缀匹配+笛卡尔积多音消歧, 相较第三方工具, 精度从 90% 提升至 95%, 支撑端侧词库拼音准确性。
 - b) 通讯录人名识别: 设计姓名模式识别算法, 融合百家姓及回流人名统计, 在通讯录导入场景中, 人名识别精度从 56.7% 提升至 90.7%, 覆盖度从 32% 提升至 87.5%。
 - c) 语音 ASR 个性化实体提词器: 构建用户词库 NER+分词后百科兜底的实体识别模式, 以 bert-chinese 为基模并用回流数据补训, 打通云侧语音业务, 实现千人千面语音输入; 正例从 55% 提升至 58%, 负例从 17% 降至 7.7%。
 - d) 分类词库体系: 自研分类词库由 143 类 62 万条大幅扩充至 8943 类 4883 万条, 并交付三批共 13 个专业词库, 新增 50.8 万百科条目; 维护每日回流数据与 PySpark 统计任务, 为词库提供持续的一元/二元词条支撑。
 2. 输入法云侧热词体系构建
 - a) 设计回流热词发现云侧流程, 构建云侧每日一元统计计算, 在每日一元的基础上, 开发基于 15-30 天时序加权的热词挖掘算法, 在随机单日测试集下平均覆盖率 78.6%, 远超讯飞热词 (24.4%) 与大搜热词 (6.5%)。
 - b) 补充第三方热词来源, 开发 AI 热点发现 Agent, 整合 Bing、搜狗、B 站、微博等多源数据, 由 AI 完成内容真实性判断、相关性分析与智能热词摘取, 通过 WebSocket 实时推送; 在随机单月标注测试集评估中额外补充 10% 热词来源, 补齐输入来源热词所欠缺的网络热梗和新词。
3. 大模型开发及训练
 - a) 接手基于 FSMT Encoder-Decoder 架构的云侧 max 模型, 定期执行业务 FST 补训, 基于端侧词库与回流数据制作新词语料, 持续更新模型出词能力。
 - b) 优化 max 模型性能和云侧部署, 完成原模型的量化与 ONNX 转换, 并用 C++ 重构推理逻辑, 在压测测试集下, 推理速度提升 15%, 精度仅下降 2%。
 - c) 基于 Qwen3B 模型的新云侧联想模型训练, 持续更新修改训练集, 进行探索式的预训练, 通用 SFT 和业务 SFT 微调, 截止目前联想首选率较原 max 模型提升 4-5 个百分点。

2. **WorldQuant Brain 量化研究顾问（兼职）** 08/2024-05/2025
- 参与 WorldQuant 平台的 Alpha 因子挖掘项目，通过分析海量金融数据开发预测性信号。
 - 在全球研究员社区协作讨论，学习因子标准化流程及组合优化方法（加权方式、因子降噪）
 - 通过平台构建因子库，涵盖技术指标、统计套利及情绪数据，优化因子组合的 IC 和夏普比率
3. **东富龙科技股份有限公司** 07/2023-08/2024
- 职位: 软件开发工程师
- 冻干系统 PID 温控算法研发。
 - Windows 平台的大型冻干系统上位机软件开发，WPF 和 QT 平台控制软件。
 - 基于 Raspberry pi 平台的小型实验机嵌入式控制软件的开发。
 - 数据库的维护和优化(Sql server, Mysql)。
 - 参与利用计算机视觉实现制药结晶生产精确控制的实验平台软件开发。

研究经历

1. **肌电假肢控制** 05/2022-09/2022
- 职位: 学生研究员 导师: Dr. Nitish V. Thakor, Professor of Biomedical Engineering
- 从植入动物体内的电极无线传输数据和搭建课题组实验的框架。
 - 解码动物模拟信号以控制虚拟假肢。
 - 收集从不同电极测量的动物模拟数据，并使用神经网络模型和机器学习进行数据分析和模型构建。
 - 使用 Python 和微控制器 (Arduino) 辅助实验设置所需的系统任务。
2. **基于深度学习的颈椎骨折图像检测** 09/2022-12/2022
- 导师: Dr. Najim Dehak, Associate Professor of Electrical and Computer Engineering
- 预处理每个椎骨的骨骼图像，进行语义分割并提取和分离切片。
 - 开发和测试与放射科医生性能相匹配的机器学习模型，用于检测和定位构成颈椎的七个椎骨的骨折。
 - 预测图像每个颈椎(C1-C7)骨折的概率。
 - 制作 2D 和 3D 动画直观显示骨折检测结果。
3. **蛋白质对象检测分析** 01/2020-05/2021
- 职位: 学生研究员 导师: Dr. Dong Xu, Director of Information Technology
- 通过对象检测实现精确的蛋白质细胞分析。
 - 设计基于 Fast-RCNN 网络的深度学习模型以检测细胞显微图像不同结构体的位置。
 - 实现基于 LSTM 深度学习模型的识别引擎，结合 OCR 工具和 Tesseract。
 - 利用少量图像数据扩展小型子项目，以减少现有模型机器学习中的识别和标注错误。
 - 完成蛋白质基因模型中 T-bar 和化合物的机器学习和预测标注。

研究兴趣方向

- 自然语言处理
- 信号处理和医学影像中的机器学习/深度学习
- 生物电信号处理和音频信号处理

荣誉/奖学金

- Dean's List 12/2018, 04/2019, 12/2019, 04/2020, 12/2020, 04/2021
- Dean's Scholarship 12/2018, 04/2019, 12/2019, 04/2020, 12/2020, 04/2021