

# 彭煜峰

pengyufenguni@mail.ustc.edu.cn | 187-1996-5521 | 安徽省合肥市金寨路 96 号 (230026)

## 教育背景

中国科学技术大学，人工智能与数据科学学院

2022.09-2026.07

GPA: 3.0/4.3

排名: 35/50

核心课程: 人工智能原理与技术 (4.3)、数据结构 (3.7)、计算系统概论 (3.7)、概率论与数理统计 (3.7) 等。

荣誉: 2022 年优秀新生奖学金银奖、2023 年度奋进奖学金 (15%)。

## 比赛经历

阿里天池大赛

大模型生成文本检测

2025.05-2025.06

- 问题概述: 目前大模型生成文本在很多领域可以做到以假乱真, 这造成了很多诸如学术不端的现象, 需要参赛者基于深度学习实现对文本的二元预测, 判断文段是否由大语言模型生成。
- 特征工程: 利用统计学习方法, 构造如 TF-IDF, 词源丰富度等特征, 结合 BERT 等基于注意力机制的模型实现复杂特征提取。
- 模型框架: 基于 SimCLR 对比学习框架, 构造正负样例对, 以对比损失和分类损失加权的形式构造损失函数, 并采用多重传统机器学习模型如支持向量机、随机森林, 最近邻查询并以 Stacking 方法融合实现高泛化能力的二元文本分类器。
- 比赛成绩: 在目前的官方 A 榜测试中的到 89.90 分, 排名前 25%。

Kaggle

Binary Prediction with a Rainfall Dataset

2025.03-2025.04

- 问题概述: 基于某地区过去几年的气压、温度等天气信息, 利用机器学习方法预测该地区未来一段时间的降雨情况。
- 模型设计: 利用卷积操作和残差连接构件深度神经网络, 以前馈神经网络作为分类头实现传统二元分类模型。
- 比赛成绩: 在官方公开榜单中进入前 10%。

## 项目经历

基于强化学习的五子棋智能体

2024.07

- 项目内容: 基于 Actor-Critic 网络和简单自我博弈实现在设计好的奖励函数、动作空间中对智能体的五子棋技能训练。
- 存在问题: 模型在基于卷积的深度残差网络中收敛性较差, 可以考虑采用更复杂的深度学习网络架构。

文本版面分析模型

2023.12

- 问题描述: 在大模型文本理解的任务中, 传统的定长文本分词往往无法达到很好的效果, 因此如果能够利用文档的结构化信息, 或能大大提高模型理解效果。
- 算法策略: 在 CDLA 数据集上对 YOLO 模型进行微调, 并结合排序树算法对模型输出的文本块内容按照阅读顺序进行排序, 从而得到完整的文本结构化信息。

## 科研经历

Fractional Spike Differential Equations Neural Network with Efficient Adjoint Parameters Training

2025.04-2025.06

- 研究内容: 利用神经微分方程优化脉冲神经网络的训练, 提出一个基于 PyTorch 的工具包以简化脉冲神经网络的搭建和训练。
- 参与贡献: 论文相关实验代码编写, 模型调优, 部分论文撰写 \ 共同一作。
- 接收状态: NeurIPS 2025 在投。

发明专利: 一种基于语义理解的专利检索方法

2023.11-2024.12

- 研究内容: 利用 BGE 模型实现文本嵌入, 基于文本向量的余弦相似度实现语义模糊检索, 利用 RAG 技术搭建智能知识库。
- 参与贡献: 基于 Milvus 向量数据库及 RAG 技术, 实现支持模糊检索的知识库系统; 利用 LLaMA-Factory 实现对 Qwen 等主流大语言模型的分布式微调训练;

## 专业技能

英语水平: 通过 CET4、CET6, CET4 口语 A 级; 随导师撰写全英文论文, 口语流利。

代码能力: 熟练使用 PyTorch, Numpy, Scikit-Learn 等深度学习框架; 熟悉 Milvus, Faiss 等向量数据库; 能利用 FastAPI, Flask 等框架进行后端开发; 熟悉 C 语言语法。

开发工具: 熟练使用 Git 实现代码的版本控制; 熟悉 Windows, MacOS 和 Linux 等多种操作系统; 熟练使用 LaTeX 撰写论文。