

姓名：邓宏

应聘岗位：算法工程师

年龄：22

学校：成都东软学院

籍贯：四川省阿坝州市

电话：180-1577-6247（微信同号）

电子邮箱：1500507371@qq.com



教育经历

2021-2025 年毕业

成都东软学院

网络工程(本科)

GPA: top15%

主修课程：线性代数、算法与数据结构、概率论与数理统计、网络编程、Python、c 语言、c++、计算机组成原理



实习经历

2024. 9-2025. 4

成都宏软科技实业有限公司

参与基于 Transformer 架构的多模态大模型研发，负责模型参数调优、数据预处理及训练流程优化

使用 Python/OpenCV 处理大规模图像数据集，完成数据清洗、标注与增强，为模型训练提供高质量数据支持。

参与大模型项目，配合团队成员完成模型选择与训练，开发 RAG 与 Agent。

在导师指导下完成技术文档撰写与代码规范优化，提升团队协作效率。



专业技能

基础技能：

熟练掌握 Python，能够编写高质量代码，调试和优化代码。

掌握 Git 版本控制，高效管理代码分支与合并。

熟练使用 PyCharm、VScode 等开发工具。熟悉 MySQL、Redis、MongoDB 及向量数据库，具备查询优化能力。

熟悉 Linux 系统。熟练 C++和 C 语言基础知识。

计算机视觉技能：

熟练掌握 YOLO 系列（v3/v5/v8）、R-CNN 等目标检测框架

熟悉使用 LabelImg 和 CVAT 等工具进行数据标注，优化训练数据集。

熟练使用 OpenCV、Numpy、Pillow 等视觉处理库

熟悉 SoftmaxLoss、Center-loss、ArcfaceLoss 等损失函数，优化图像分类与检测任务性能。

复现过 ResNet18 网络，经典 CNN 网络（例如 VGG 网络）的设计与应用。

熟悉 UNet++、Mask R-CNN 图像分割技术。

熟悉 PyTorch 中实现模型训练，并转换成 ONNX 和 TensorRT 格式加速推理部署。

大模型技能：

熟悉 Transformer 和 GPT-4 deepseek 等大模型架构，并在自然语言处理（NLP）任务中积累了一定的经验。

掌握 LoRA、QLoRA 高效微调技术。掌握 RAG，并且有调优经验。具备模型微调、量化、剪枝等优化经验。

熟悉 deepseek、qwen 等文本及多模态大模型。熟悉 LangChain、LangGraph、prompt、LoRA 等大模型相关技术。

熟悉 vLLM、xInference 推理框架，了解批处理和 KV 缓存优化。

了解分块技术。了解 difi 等技术。了解分布式训练，了解 Docker，Flask 等模型部署工具。



校园经历

2024. 1-2024. 5

体考姿态监测与评估系统

项目目标：在体育考试场景下，老师往往需要同时监考多个学生，存在工作量大、难以精准评估学生动作规范性的问题。本项目旨在开发一套体育考试姿态监测与评估系统，通过算法对学生在体育考试中的动作进行实时姿态分析，帮助老师高效、准确地监考，提供客观公正的评估依据。

项目方案

选用 YOLOv8-pose 模型作为姿态检测的核心算法，凭借其强大的关键点检测能力，精准捕捉学生在体育考试中的动作姿态。采用 RK3588 作为部署硬件，结合 OpenCV 图像处理技术，实现从视频流采集、关键点检测到姿态分析的全流程处理。通过数据增强技术扩充数据集规模，提升模型对不同场景和动作的适应能力。运用超参数优化与正则化方法，确保模型在实时监测过程中的稳定性和准确性。最终将模型转换为 RKNN 格式，部署在 RK3588 硬件上，实现低延迟、高效率的姿态监测与评估。

个人贡献

- 数据采集与预处理：**通过多种渠道收集公开数据集和自采集数据，涵盖引体向上等体育考试动作。运用数据增强技术，有效扩充训练集规模，丰富数据多样性。利用 OpenCV 对采集到的视频数据进行清洗，去除模糊帧并规范化帧率，显著提升图像质量，为模型训练提供高质量数据支持。
- 数据标注：**熟练使用 LabelImg 等专业工具，对 2000 帧图像进行关键点标注，涵盖肩、肘、膝等 17 个关键点，确保标注数据的准确性与可靠性，为模型训练奠定坚实基础。
- 模型训练与优化：**协助团队深入调研并选择 YOLOv8-pose 模型，参与模型训练全过程。通过优化超参数，将关键点检测精度 PCKh@0.5 提升至 91.5%，并将误检率成功控制在 5% 以内，显著提升模型检测性能。

项目成果：系统成功实现实时姿态检测，进行姿态分析，PCKh@0.5 91.5%，检测速度超 30 FPS。



项目经历

2024. 12-2025. 4

新媒体智能创作平台（多 Agent, AIGC, RAG）

项目目标：为新媒体公司开发一个智能创作平台，助力提升营销能力，降本增效，提供一站式文案撰写、品牌化设计素材生成和智能排版服务，实现高效优质内容创作。

项目方案

项目采用多 Agent 协作体系，涵盖需求分析、内容创作以及质量评估等多个子代理，形成高效的创作流程。精心设计通用提示词与特定提示词，优化模型生成效果，以实现精准且富有创意的内容创作。集成文案生成模型与文生图模型，通过精心设计的接口与提示词策略，使两者相辅相成，达成图文并茂的创作佳效。运用 API 调用方式，有效降低成本，同时确保系统的高效稳定运行。选用 Qwen2.5 作为文案生成的主模型，Stable Diffusion 3 担当文生图的主模型，以强大的模型性能为创作质量保驾护航。构建 RAG 机制，聚焦女装领域，实现对相关问题的精准回答与内容创作，提升内容的专业性与针对性。收集海量营销文案数据与开源图片库及品牌海报样本，进行深度清洗、精准标注风格与类别，并依据女装款式、风格进行细致分类，借助 Milvus 云数据库进行高效存储与管理。

个人职责

- 数据处理：**主导数据采集工作，对数据清洗，有效提升数据质量，为模型训练提供高质量的“养料”。
- Agent 流程与提示词开发：**精心设计 Agent 流程，巧妙优化各环节，显著提升创作效率与质量，确保平台能够快速响应用户需求并输出优质内容。专注于优化提示词，通过深入研究模型特性与创作需求，增强模型对不同创作任务的理解与生成能力，使创作内容更贴合用户期望，为平台的创意注入活力。
- RAG 开发：**使用 LangChain 构建 RAG 管道，检索服装信息并生成提示词。

项目成果：成功开发上线新媒体智能创作系统，文案生成准确率 85%，有效满足新媒体行业对高效优质内容创作的需求，用户反馈良好。

2024. 9-2025. 1

员工自主智能服务助手（RAG）

项目目标：为一家公司的员工自助服务平台开发一个智能助手，提供 24h 自动咨询服务，支持员工随时查询公司内部业务和员工福利等问题。

项目方案

项目基于 Qwen2.5-7B-Instruct 模型，充分利用其出色的中文任务处理能力和适中的参数量，使其完美适配本地部署需求。运用 LoRA 技术对该模型进行微调，借助精心准备的 800 条标注数据，将模型损失降低至 0.05，测试集评估显示准确率高达 91%，BLEU 分数达到 0.85，从而确保模型在对话和问答场景中的卓越表现。构建高效的 RAG 系统，利用 Milvus 实现文档向量索引的存储与高速检索，结合 Qwen2.5-7B-Instruct 的自然语言处理能力生成精准问答。引入 LangChain 框架，深度集成检索与生成模块，支持链式调用和上下文管理，使得检索模块在测试集上的精确率和召回率分别达到 95% 和 90%，F1 分数更是高达 0.92，为多轮对话的连贯性和复杂查询的精准回答提供坚实保障。采用 vLLM 对 Qwen2.5-7B-Instruct 进行优化，显著提升推理效率，将检索时间缩短 70%，同时确保准确率维持在 90%。

个人贡献

- 1. **需求分析**：与项目团队成员紧密合作，深度对接客户需求，全面把握文档类型、查询场景、服务器硬件配置及合规性要求等关键要素。精心设计 RAG 和微调方案，精准选择 Qwen2.5- 7B- Instruct 模型，确保本地部署的可行性与高效性。
- 2. **数据收集**：收集并整理客户提供的历史查询记录，制作出 1000 条高质量标注数据。
- 3. **RAG 开发**：熟练运用 LangChain 框架，深度集成 Qwen2.5- 7B- Instruct 模型，精心构建 RAG 流程，包括 Milvus 检索与模型生成等环节，实现高效精准的信息检索与问答生成，提升系统的智能化水平。

项目成果：

成功在客户服务器上部署基于 RAG 和微调的员工解答智能助手，与客户原有员工自助服务平台无缝集成，回答准确率达到 91%，项目按时交付，验收通过率 100%，获得用户高度评价。



自我评价

作为一名积极向上的大学生，我在学习过程中，注重理论知识的积累，同时也重视实践能力的培养。同时也并积极参与了多项不同的的实践活动和竞赛这些经历极大地拓宽了我的视野，并提升了我的实际操作能力。