



## 张泽铭

年龄：24

政治面貌：党员

籍贯：河北省邢台市 求职意向：计算机视觉 / LLM 应用开发工程师

电话：15811516045 电子邮箱：23021211861@stu.xidian.edu.cn

### 教育背景及成果

西安电子科技大学（本科），电子信息工程

2019.09 - 2023.06

- 奖项荣誉：星火杯校级一等奖、两次校级奖学金

西安电子科技大学（硕士），电子科学与技术

2023.09 - 2026.06

- 实验室：所在 VIPSL 实验室隶属于高新波教授（现西安电子科技大学校长、中国高等教育学会常务理事，原重庆邮电大学校长）团队，主要研究方向：图像超分辨率算法研究
- 荣誉：校级一等奖学金、校级二等奖学金、实验室一等奖学金、2024 学年‘优秀学生干部’
- 成果：Neurocomputing（一作论文在投）EfficientDynamicSelective-AttentionNetworkforImageSuper-Resolution

### 专业技能

- 深度学习&Transformer核心：熟练掌握pytorch及主流大模型（GPT、DeepSeek、llama等）原理与训练流程，能结合 LoRA/PEFT 等微调方法完成模型的开发与迭代。对 CV、NLP 领域深入了解，具备手动构建与改造深度神经网络的实战经验。
- LLM 工程落地：精通 Prompt Engineering、CoT 等关键技术，能针对复杂场景（RAG 问答、Chatbot、Agent 等）定制与优化大模型应用流程；深入理解向量检索（Faiss / Milvus）、BM25 等机制，具备 Embedding 微调、Rerank、RLHF 等能力，系统提升问答系统的相关性与响应精准度。
- 后端开发：熟悉 SpringBoot、MySQL、MongoDB、Redis、MinIO、RabbitMQ 等组件的使用与集成；具备基于 RESTful API 的服务开发经验，支持异步处理、权限管理与向量数据库集成。
- 其他技能：熟练使用Git、Maven、Docker、Latex、figma等工具，有丰富的Linux服务器项目部署经验。有一定英语阅读、写作、交流能力（CET-6 通过）。

### 实习经历

智能合同风险识别-华为质量与流程 IT 部门

2025.7 - 2025.9

项目简介：负责构建CRM系统中的AI合同风险识别产品。系统基于FlowChain编排，支持对复杂格式招标合同中的潜在风险条款进行自动识别，输出包括合同片段原文、识别出的风险点、对应风险条目以及判定依据。

核心优化与实现：

1. 文本解析与多源知识库构建：结合段落结构 prompt 与 OCR 图像识别进行语义切分与结构还原，提升表格、扫描件内容提取准确性；构建法规条文、风险样例、专家经验等多维知识库，支持上下文增强与多模态信息注入。
2. 检索流程优化：构建原文与摘要双向量库，融合 BM25 与 SimCSE/BGE 向量检索，结合权重匹配与 Small-to-Big 策略提升长尾问题召回率；引入 Rerank 模型并配置 TopK 去重与分数过滤（阈值 0.1），显著提升排序相关性。
3. 生成流程优化：引入核心实体注入、指代消解与答案出处标注机制，增强语义连贯性与可解释性；通过意图识别触发段落级二次召回，提升流程型问题的上下文覆盖与回答准确性。

- 4. 动态 Prompt 工程与 Query 调优：配置 FewShot/CoT 模板并按任务注入风险定义与合同背景，动态调整检索范围与生成方式；对领域术语、拼写变体等问题引入标准化处理与 Query 改写策略。
- 5. 多模态识别与生成增强：集成 Pangu-135B 多模态大模型解析图像内容并格式化注入文本索引，提升图文混合合同的信息可获取性。

项目成果：

- Top5 召回准确率由 65% 提升至 87%，P@3 提升约 30%；流程型问题上下文识别准确率提升 25%。
- 合同风险识别准确率超 90%，多轮问答保持一致性与可追溯性；生成内容支持风险溯源与合规质检。

项目经历

智能机票销售与客服助手 2025.2 - 2025.6

项目简介：构建融合本地知识库检索（RAG）与实时网络搜索的多模态智能机票助手系统，服务于航空公司在客服问答、售票培训与政策解答等场景。系统覆盖退改签规则、航线时刻、票价政策、行李额度等核心知识，支持自然语言多轮对话、卡片式推荐与上下文追问，有效提升客服与销售人员的知识获取效率与专业答复质量。

核心工作：

- 1.RAG 检索与知识构建: 使用 LlamaIndex + Milvus + BGE 构建航空知识库；采用 BM25 与向量召回混合策略，将 Recall@10 从 53% 提升至 89%；所有文档入库前进行“统一切块 + 标题拼接”处理，提升罕见问法下的召回覆盖能力。
2. 实时 WebSearch 强化应答: 集成 SerperSearch API 实时获取航班变更、天气、疫情、机场运营等最新动态；使用 Query Expansion 与关键词抽取技术精筛网页，构造 QA 对话对，供大模型推理判断；确保对航班延误、政策变更、特价信息等时效性强问题给出准确答复。
3. 多 Agent 工作流与上下文管理: 引入 ReAct + Reflection + MCP 机制，将用户请求智能拆分为：政策检索、实时搜索、订单查询、改签执行等任务；由不同 Agent 协同完成子任务。
4. 后端架构与异步调度机制: 后台使用 SpringBoot + MySQL 实现多级用户权限管理；支持上传航空条款、内部销售指南至 MinIO，并提供在线可编辑能力；所有问题向量化编码后入库 Milvus；使用 RabbitMQ 异步分发问答请求，解决高并发场景下响应阻塞问题，将首 Token 时延控制在 1.2 秒内。

8K超高清画质提升芯片，海信——西电联合实验室项目 2023.12-2024.8

项目简介：与海信信芯微联合研发国内首颗全自研8K画质AI芯片，负责核心画质增强IP模块算法设计与轻量化优化。项目聚焦超分辨率重建、动态插帧、感知增强与芯片部署技术，在 4K/8K@60fps 下实现了画质优化与实时推理的兼顾。

核心工作：

1. 设计多任务渐进式视频画质提升模型，构建“失真复原—结构增强—感知增强”三阶段子网络，提升视频的色彩、清晰度与纹理质量。
2. 主导自建数据集的清洗与样本筛选，提出场景标签驱动的数据预处理策略，解决 Repeat、Databroken 等场景问题。
3. 参与设计基于特征金字塔的运动视频插帧模块，优化大尺度运动下帧间对齐与融合精度。
4. 负责模型剪枝、特征级知识蒸馏与分级量化策略设计，压缩模型参数至 13K–25K，显著提升硬件部署效率。

个人介绍

- 兴趣爱好广泛，喜欢运动。多次参加校级、院级羽毛球比赛，并获得混合团体赛冠军、亚军。
- 作为学生干部，有较强的语言表达能力、组织协作能力、分析和解决实际问题的能力。
- 勤于学习，乐于思考，善于发现，做事情有较强的计划性，对问题有独到的领悟和理解。