

孙文涛

(+33) 077-488-4069 · wentao.sun@nokia.com · AI Agent 算法工程师

教育背景

法国国立应用科学学院 (INSA) & Nokia Bell Labs, 联合培养博士 (CIFRE)	2026.2 - 2029.2
研究方向: 语言与推理模型在电信网络管理中的应用 Nokia 导师: D. Verchere INSA 导师: M. Acher	
巴黎综合理工大学 (IPP École Polytechnique), 计算机科学与工程硕士	2023.9 - 2025.10
排名 1/14, 法国奖学金获得者	
华南理工大学, 信息工程, 工学学士	2020.9 - 2024.6
华南理工-巴黎综合理工“3+2”本硕连读项目, 校级奖学金	

实习经历

非凸科技 (正瀛资产), 量化 AI Agent 实习生	2025.10 - 至今
• 独立设计面向 A 股市场的实时金融新闻智能分析系统 DataDig, 构建 Kafka 消费 → LLM 多维特征提取 → 三层增量聚类 → 事件融合的全异步管线, 基于 Dagster 调度 8 路并行 snippet 抽取任务, 每 5 分钟增量更新 (Python 15000+ 行)	
• 设计三层实时聚类架构: Layer 1 基于 HAC ($\text{cosine} \geq 0.8$) 去重近似新闻; Layer 2 (≥ 0.6) 语义聚类并生成摘要; Layer 3 通过 LLM 结构化输出 (Pydantic Schema) 判断 merge/new/ignore, 将实时 cluster 融合至 365 天长期事件表	
• 设计 Cluster ID 跨周期稳定性算法: Jaccard 相似度映射 UUID + 两阶段后置修复 (anti-split/anti-merge) 保证增量 HAC 下旧 cluster 边界稳定; embedding 稳定化机制 ($\text{cosine} \geq 0.9$ 时复用旧向量) 抑制 LLM 措辞波动的级联影响	
Nokia Bell Labs, AI 研究实习生, 巴黎	2025.3 - 2025.10
• 开发 Graph RAG 系统, 通过结构化文本生成增强 LLM 推理, 在 Corr2Cause 基准上实现 F1 提升 47.5% ($32.71 \rightarrow 48.26$)	
• 实现三阶段架构: 正则约束 JSON 工具调用确保 100% 模式合规; 知识图谱构建实体精度 95%、消歧准确率 92%; NL-to-Cypher 查询语法有效性 100%	
• 核心创新: 混合 BM25-向量检索实现实体消歧、语义感知文档分块、基于 Qwen3-32B 的工具调用与结构化输出, 显著降低幻觉并提升多跳推理能力	
中国科学院自动化所, 机器人实习工程师 (乔红院士团队)	2024.3 - 2024.9
• 设计轮式人形机器人复杂感知导航系统, 实现全身级路径规划与厘米级重定位精度	
• 融合固态雷达 (fastlio2+octomap) 与深度相机 (Intel D435i), 采用 EKF 融合多源位姿数据, 重定位时间减少 80%	
• 开发基于 SLAM 的 RRT* 路径规划插件, 后重构迷你规划器代替 Nav2, 减少 30% 计算量	

项目经历

基于 ASR 开源模型的语音识别系统, 优秀毕业设计	2024.2 - 2024.5
• 基于 Whisper/FunASR 构建语音识别 Web 应用, 接入智谱 AI 进行文段纠错与翻译, 小语种识别率达 95%	
• 后续接入机器人系统, 实现语音指令识别并基于 LLM prompt 自动生成行为树	

技术能力

- 编程语言: Python (熟练)、C++
- 框架/工具: PyTorch, Transformer, LangChain, ROS2 Humble, Docker
- 大模型: Llama, Qwen, DeepSeek; RAG/Agent/工具调用/结构化输出

语言能力

- 英语: 雅思 6.5 法语: A2 普通话: 一级乙等