Wentao Sun (孙文涛)

wentao.sun@nokia.com | 07 74 88 40 69 | Paris, France | wentaopoly | wentao-sun

Education

INSA & Nokia Bell Labs, Language & Reasoning Models for Telecom Network Management — France, Feb 2026 – Feb 2029
- Nokia Supervisor: D. Verchere
- INSA Supervisor: M. Acher

Institut Polytechnique de Paris (IPP) | École Polytechnique, Computer Science & Engineering — Palaiseau, France, Sept 2023 – Oct 2025
- Ranked 1/14, French scholarship recipient

South China University of Technology (华南理工大学), Information Engineering — Guangzhou, China, Sept 2020 – June 2024
- SCUT–École Polytechnique '3+2' joint BSc-MSc program
- University scholarship recipient

Experience

Feiou Technology (正瀛资产), Quant AI Agent Intern — Remote, Oct 2025 – present, 5 months

Independently designed a real-time financial news intelligent analysis system (DataDig) for the global market.
- Built full async pipeline: Kafka consumer → LLM multi-dimensional feature extraction → 3-layer incremental clustering → event fusion, orchestrated by Dagster with 8-way parallel snippet extraction, incremental updates every 5 min (Python 15,000+ lines)
- Designed 3-layer real-time clustering: Layer 1 HAC (cosine ≥ 0.8) deduplication; Layer 2 (≥ 0.6) semantic clustering with summary generation; Layer 3 LLM structured output (Pydantic Schema) for merge/new/ignore decisions, fusing real-time clusters into a 365-day long-term event table
- Designed cross-cycle Cluster ID stability algorithm: Jaccard similarity UUID mapping + two-phase post-repair (anti-split/anti-merge) for stable cluster boundaries under incremental HAC; embedding stabilization (cosine ≥ 0.9 reuses old vectors) to suppress cascading effects from LLM phrasing variation

Nokia Bell Labs, AI Research Intern — Paris, France, Mar 2025 – Oct 2025, 8 months

Developed a Graph RAG system enhancing LLM reasoning through structured text generation.
- Achieved F1 improvement of 47.5% (32.71 → 48.26) on the Corr2Cause benchmark
- Implemented 3-stage architecture: regex-constrained JSON tool calling with 100% schema compliance; knowledge graph construction with 95% entity precision and 92% disambiguation accuracy; NL-to-Cypher query with 100% syntax validity
- Core innovation: hybrid BM25-vector retrieval for entity disambiguation, semantic-aware document chunking, Qwen3-32B-based tool calling and structured output, significantly reducing hallucination and improving multi-hop reasoning

Institute of Automation, Chinese Academy of Sciences (中科院自动化所), Robotics Intern Engineer (Prof. Qiao Hong's Team) — Beijing, China, Mar 2024 – Sept 2024, 7 months

Designed complex perception-navigation system for wheeled humanoid robots.
- Achieved whole-body path planning and centimeter-level relocalization accuracy
- Fused solid-state LiDAR (fastlio2 + octomap) with depth camera (Intel D435i), used EKF to fuse multi-source pose data, reducing relocalization time by 80%
- Developed SLAM-based RRT* path planning plugin, then refactored mini planner to replace Nav2, reducing computation by 30%

Projects

ASR-based Speech Recognition System Feb 2024 – May 2024

Outstanding Graduation Project — Built a speech recognition web app based on Whisper/FunASR, integrated Zhipu AI for text correction and translation, achieving 95% recognition rate for minority languages. Extended to robot system for voice command recognition with LLM-prompt-based automatic behavior tree generation.

Skills

Programming Languages

Frameworks & Tools

Large Language Models

Languages

Chinese (Mandarin)
Native speaker (一级乙等)

English
IELTS 6.5

French
A2