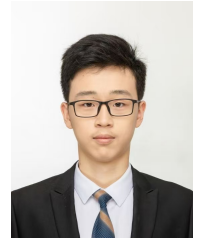


# Jiaheng Sun

B.Eng. Candidate in Computer Science  
Northeastern University, Shenyang, China

✉ [jiahengsun086@gmail.com](mailto:jiahengsun086@gmail.com) ☎ +86 150 6607 3956

🐙 [github.com/sjh04](https://github.com/sjh04)



## RESEARCH INTERESTS

My research focuses on **reliable and generalizable LLM agents**, with an emphasis on diagnosing and correcting failure modes. I develop controlled evaluations that expose failures in reasoning and tool use, and I study how structured feedback, external knowledge, and persistent memory can improve agent decisions. I am particularly interested in tool-using and multimodal agents operating on complex spatio-temporal tasks.

*Keywords:* LLM agents, reasoning and tool use, diagnostic evaluation, persistent memory, feedback-driven learning, multimodal agents, and spatio-temporal reasoning.

## EDUCATION

Northeastern University, China | B.Eng. in Computer Science  
School of Computer Science and Engineering

Sep. 2023 – Jun. 2027 (expected)  
Shenyang, China

- GPA: 4.159/5.000 | Rank: 17/214 (top 7.98%)
- Research advisors: [Prof. Tiancheng Zhang](#) and [Asst. Prof. Hengyu Liu](#) (Aalborg University)
- English: IELTS Academic 7.0 | College English Test Band 6 547/710

University of Oslo, Norway | Visiting Student Researcher  
Department of Informatics (IFI), hosted by [Prof. Tor Skeie](#)

Aug. 2026  
Oslo, Norway

- Conducting a full-time research visit under the **DSTrain** project on LLM-based trajectory planning for autonomous driving. Pursuing an independent first-author study and contributing to benchmark and software development.

## PUBLICATIONS AND PREPRINTS

<sup>†</sup> denotes equal contribution (co-first authorship).

1. [Jiaheng Sun<sup>†</sup>](#), [Hengyu Liu<sup>†</sup>](#), [Zhihong Cui](#), Haoyu Wang, Xinyuan Zhang, [Amir Taherkordi](#), [Tor Skeie](#), [Jianwei Zhang](#), [Kaize Shi](#), [Tiancheng Zhang](#). “VLM-Guided Diffusion Models for Vehicle Trajectory Prediction.” *World Wide Web Journal (WWWJ, CCF B)*, **under major revision**, 2026.

**Contribution:** implemented the RL prior generator and retrieval-augmented VLM reward module; designed and conducted experiments and ablations; drafted the manuscript.

2. [Jiaheng Sun](#), [Hengyu Liu](#), [Zhihong Cui](#), [Tiancheng Zhang](#), et al. “TripWire: A Capability-Diagnostic Benchmark for Spatio-Temporal Data Agents” (working title). **In preparation**, targeting ICLR 2027.

**Contribution:** benchmark and scoring design; implementation of the evaluation harness (task builders, deterministic replay, scoring package, LLM-as-judge); construction of the verified task bank.

## PATENTS

1. [Tiancheng Zhang](#), [Jiaheng Sun](#) (2nd inventor), [Hengyu Liu](#), Haoyu Wang, Huimin Zhu, Mengxin Wang. “A Trajectory Prediction Method for Autonomous Driving Based on VLM, PPO and Diffusion Models.” **Chinese invention patent**, application filed and under substantive examination.

## RESEARCH EXPERIENCE

Shenyang YaTrans Network Technology Co., Ltd.  
Research Intern, NLP Group; worked with CTO Dr. Quan Du

Jun. 2026 – Jul. 2026  
Shenyang, China

- \* **System and leadership.** Led a four-person team in developing a local Qwen3.5-4B agent framework for autonomous task planning, ReAct-based tool use, and automated task execution.
- \* **My role:** Designed and implemented the active-memory subsystem, combining hybrid retrieval and reranking with budget-aware compression, conflict and poisoned-memory detection, and lifecycle management. Built a labelled corpus and six reproducible evaluation suites; on the controlled testbed, improved Hit@3 from 0.20 to 1.00, intercepted 90% of poisoned memories, and raised memory-dependent task success from 0/10 to 10/10.

- \* **Problem.** Diffusion-based trajectory predictors are sensitive to the initial noise prior. However, a vision-language critic provides non-differentiable scalar feedback and lacks the domain knowledge needed to evaluate driving trajectories reliably.
- \* **Approach.** Froze the diffusion backbone and used reinforcement learning to translate VLM rewards into scene-conditioned refinements of the initial noise prior. A retrieval-augmented VLM supplies the rewards and retrieves prior evaluations to accumulate domain knowledge. It operates only during training, eliminating inference-time overhead.
- \* **My role:** Co-first author. Implemented the RL prior generator and RAG-VLM reward module in the NAVSIM and nuScenes pipeline, conducted experiments and ablations, and drafted the manuscript.

- \* **Problem.** Aggregate task accuracy cannot reveal which capabilities an LLM data agent lacks. For trajectory data (GPS and AIS), tabular shortcuts can also bypass spatial reasoning and conceal these limitations.
- \* **Approach.** Developed a task-first pipeline that converts public data requests into replayable tasks with gold-standard checks. Capability milestones replace a single aggregate score, shortcut probes identify tasks solvable by non-spatial baselines, and a difficulty gate excludes saturated, non-diagnostic test cases.
- \* **My role:** Designed the benchmark and scoring framework. Implemented task builders, deterministic replay, the scoring package, and LLM-as-judge evaluation for code and report artefacts.
- \* **Status.** The framework runs end to end. The verified task bank remains under construction, and no cross-agent comparisons are reported yet.

## SELECTED PROJECTS AND COMPETITIONS

- \* Developed **EDPHO**, a hybrid optimisation strategy coupling NSGA-II with dynamic programming for a high-dimensional, non-convex, multi-constrained collaborative deployment problem.
- \* **My role:** Led the formulation of the core optimisation models and wrote most of the final report.

## HONORS AND AWARDS

## SKILLS

- **Programming:** Python, C/C++; PyTorch, NumPy, pandas, scikit-learn, Matplotlib; Git, Linux,  $\LaTeX$ .
- **Research tools and datasets:** autonomous-driving benchmarks and simulators (nuScenes, NAVSIM, closed-loop planning metrics); trajectory and mobility datasets (GeoLife, T-Drive, Porto taxi, AIS); spatial data stacks (PostGIS, MobilityDB).
- **Models and methods:** diffusion models, reinforcement learning (PPO), vision-language models, retrieval-augmented generation, LLM-as-judge evaluation, and agent evaluation harnesses.