TUM International Graduate School of Science and Engineering

Where Do LLMs Get Lost? A Multi-Scale Diagnostic Benchmark for Spatial Navigation

Yuhan Jiang, Chair of Cartography and Visual Analysis, project funded by IGSSE

Project Outline/Mission

1. What spatial representations best support LLM navigation?
2. Where do LLMs fail when navigating across spatial scales?
3. What levels of spatial decision-making should be delegated to LLMs in hybrid navigation agents?

Key Figures

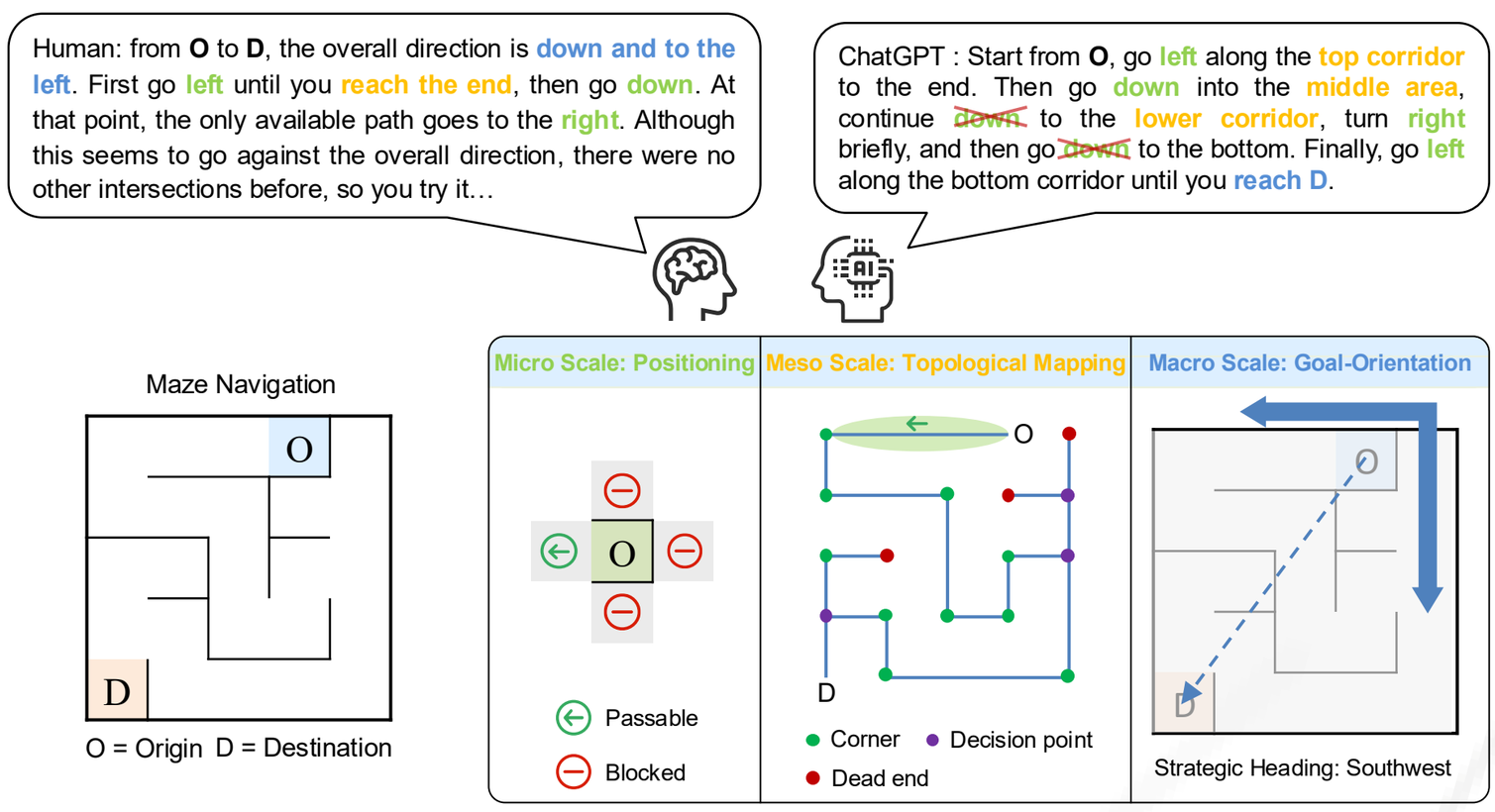

Fig 1. Decomposing maze navigation into micro, meso, and macro scale spatial reasoning

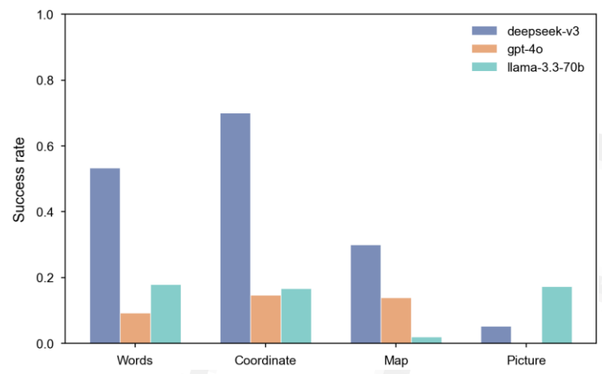

Fig 2. Success rate across four input representations on small mazes

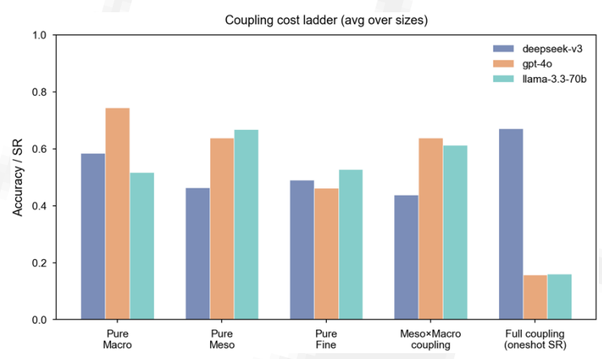

Fig 3. Performance under Cross-Scale Coupling

Conclusion

1. Coordinate-based representations achieved the most consistent performance in our maze navigation benchmark, outperforming maps, text, and images.
2. LLMs possess spatial abilities at individual scales, but fail when decisions must be coordinated across Micro, Meso, and Macro spatial scales.
3. Hybrid navigation agents should delegate high-level spatial decisions to LLMs, while low-level path execution is handled by deterministic algorithms.

Team

Principal Investigators:
TUM: Prof. Dr. Liqiu Meng
Chair of Cartography and Visual Analytics
School of Engineering and Design

International Partner: Prof. Dr. Peng Luo
MIT Senseable City Lab
Massachusetts Institute of Technology

Doctoral candidates:
Yuhan Jiang