Minghao Guo

(919) 306-3350 mhkwok2016@gmail.com LinkedIn Google Scholar Homepage

Profile

- Computer Science Ph.D. specializing in agentic retrieval, persistent memory, personalization, and evaluation. Builds systems that retrieve and reason over enterprise knowledge and long-term multimodal interactions, and evaluation frameworks that measure how personalized agents behave.

Education

- **Rutgers University** — **New Brunswick, NJ**
  Ph.D. in Computer Science; Advisor: Dimitris N. Metaxas — *Sep. 2023 – Dec. 2026*
- **North Carolina State University** — **Raleigh, NC**
  M.S. in Computer Science — *Aug. 2021 – May 2023*
- **Sichuan University** — **Chengdu, China**
  B.E. in Computer Science and Technology — *Sep. 2016 – Jun. 2020*

Experience

Zoom — **Bellevue, WA**
Agentic Retrieval R&D Intern — *May 2026 – Aug. 2026*

Evaluation Framework for Context Memory of Zoom Meeting

- Built an end-to-end evaluation framework for Zoom's meeting knowledge graph, generating synthetic meeting transcripts from event timelines with aligned ground-truth claims and QA pairs to measure retrieval quality and answer grounding.
- Designed a gated evaluation-and-improvement loop integrating extraction, clustering, retrieval, and claim-level scoring; improved claim recall from 0.625 to 0.750 on a representative evaluation set.
- Created a 260-question single-meeting benchmark covering action items, ownership, relationships, blockers, and temporal changes, enabling fine-grained diagnosis of retrieval and knowledge-capture failures.

Rutgers University — **New Brunswick, NJ**
Ph.D. Researcher — *Sep. 2023 – Present*

MemEye: A Visual-Centric Evaluation Framework for Multimodal Agent Memory

- Developed a visual-centric benchmark for evaluating how multimodal agent memory systems retain, retrieve, and reason over information from long-term visual interactions.
- Proposed a two-dimensional evaluation taxonomy covering memory granularity, from scene-level context to pixel-level details, and evidence usage, from single observations to temporal state evolution.
- Evaluated 13 memory methods across four vision-language model backbones, revealing substantial weaknesses in fine-grained visual recall, temporal change tracking, and multi-observation evidence synthesis.

Seeing, Maintaining, Learning: Evolution of Multimodal Agent Memory

- Led 80 authors and coordinated taxonomy design and research synthesis, organizing methods by experience representation and the evolution of memory-management policies, from static retrieval to self evolving.
- Analyzed 252 memory architectures and 76 benchmark records, identifying gaps in evidence preservation, temporal consistency, and bounded lifelong memory.

DeepSieve: Information Sieving via LLM-as-a-Knowledge-Router

- Built an agentic RAG framework that decomposes complex questions, routes subquestions to documents, structured data, and APIs, and iteratively filters retrieved evidence.
- Enabled reasoning across heterogeneous sources without merging their indexes; evaluated multi-hop QA on MuSiQue, 2WikiMultiHopQA, and HotpotQA.

Individual Turing Test: Personalized Agent via RAG, LoRA & Hybrid Memory

- Built a personalized agent by fine-tuning Qwen2.5-7B with LoRA on longitudinal personal WeChat chat data, and compared fine-tuning, RAG, and hybrid memory through blind judgments by the user's acquaintances.
- LoRA fine-tuning alone beat a RAG-only baseline in blind pairwise judgments (win rate 0.608, $p < 0.001$), and combining RAG with LoRA performed best, showing personalization benefits from both memory retrieval and model adaptation.
- Proposed an individual Turing test framework where acquaintances evaluate the authenticity of the AI digital twin against the real user.

Tsinghua University **Beijing, China**
Ph.D. Research Intern *May 2025 – Aug. 2025*

TwinVoice: A Multi-dimensional Benchmark Towards Digital Twins via LLM Persona Simulation

- Designing **TwinVoice**, the first **real-world** benchmark to evaluate Digital Twin agents (LLM-based systems that emulate a target human's linguistic style, preferences, and personality).
- Curating large-scale datasets across **three tracks**: Social Behavior, Personal Dialogue, and Narrative Persona, covering both **discriminative** and **generative** tasks.
- Leveraging **authentic conversational history** to assess style imitation fidelity, memory consistency, and persona alignment, avoiding synthetic or crowd-annotated biases.

Selected Publications

1. **MemEye: A Visual-Centric Evaluation Framework for Multimodal Agent Memory.**
   Minghao Guo*, Qingyue Jiao*, et al.
   *Under review at **ICLR 2027**.* [Paper]
2. **Seeing, Maintaining, and Learning: The Evolution of Multimodal Agent Memory.**
   Minghao Guo* (project organizer), Qingyue Jiao*, et al.
   *Under review at **TMLR**.* [Paper] [Project & Resources]
3. **DeepSieve: Information Sieving via LLM-as-a-Knowledge-Router.**
   Minghao Guo, Qingcheng Zeng, et al.
   ***Findings of EACL 2026**.* [Paper] [Code]
4. **Individual Turing Test: A Case Study of LLM-based Simulation Using Longitudinal Personal Data.**
   Minghao Guo, Ziyi Ye, et al.
   ***SIGIR 2026**, short paper.* [Paper]
5. **TwinVoice: A Multi-dimensional Benchmark Towards Digital Twins via LLM Persona Simulation.**
   Bangde Du*, **Minghao Guo***, et al.
   ***ACL 2026**.* [Paper]
6. **OpenPM: Auditable Point-in-Time Evaluation for LLM Portfolio-Management Agents.**
   Xinying Cai*, **Minghao Guo***, et al.
   *Forthcoming in **SIGKDD Explorations**, Dec. 2026.* [Paper]
7. **ReaGAN: Node-as-Agent-Reasoning Graph Agentic Network.**
   Minghao Guo, Xi Zhu, et al.
   ***HCRS @ TheWebConf 2026**.* [Paper]

** Equal contribution.*

Skills

Programming & Systems: Python, SQL, C++, Bash, Git, Docker, Linux

ML & Retrieval: PyTorch, Hugging Face, vLLM, RAG, vector search, knowledge graphs, LLM agents, post-training (SFT, LoRA)

Evaluation: benchmark design, retrieval and claim-level evaluation, multi-hop QA, personalized-agent evaluation