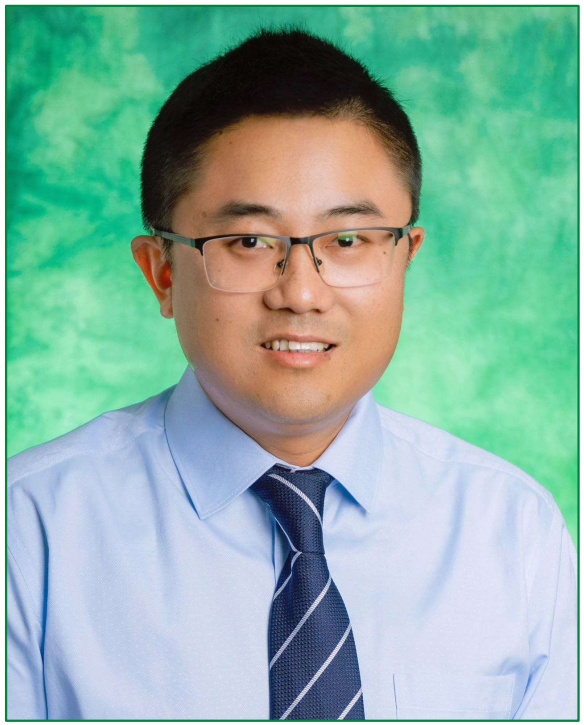
About Me



NAME

Haihua Chen, Ph.D.

TITLE

Assistant Professor, Department of Data Science;
Director, IDEA Lab; Affiliated Faculty, Health Informatics

YEARS AT UNT

5 years on faculty (since 2022);
UNT Ph.D. (08. 2017 – 01. 2022

RESEARCH FOCUS AREAS

- Large language models & multimodal AI
- Data quality & data-centric AI
- AI for digital libraries & cultural heritage
- Science of science & scholarly evaluation
- Health & legal informatics

60+ peer-reviewed papers · 2,200+ citations · Associate Editor, Information Processing & Management, *The Electronic Library*, and *Data Intelligence*

ORAL-AI: Multimodal LLMs for Oral History Collections

Making hours of recorded oral history searchable, answerable and summarizable with **multimodal LLMs**, without hallucination or privacy loss.

$490,396

IMLS NLG-L · PI · 2026–29
with UIUC (Co-PI J. Chen)

1 Benchmark

Test GPT, Gemini, Llama on 30 queries: grounding, hallucination, bias, privacy

2 Build

Align transcript, audio & video; evidence-cited RAG on local open models

3 Evaluate

80 users vs. keyword search; 30 interviews with librarians & public

4 Share

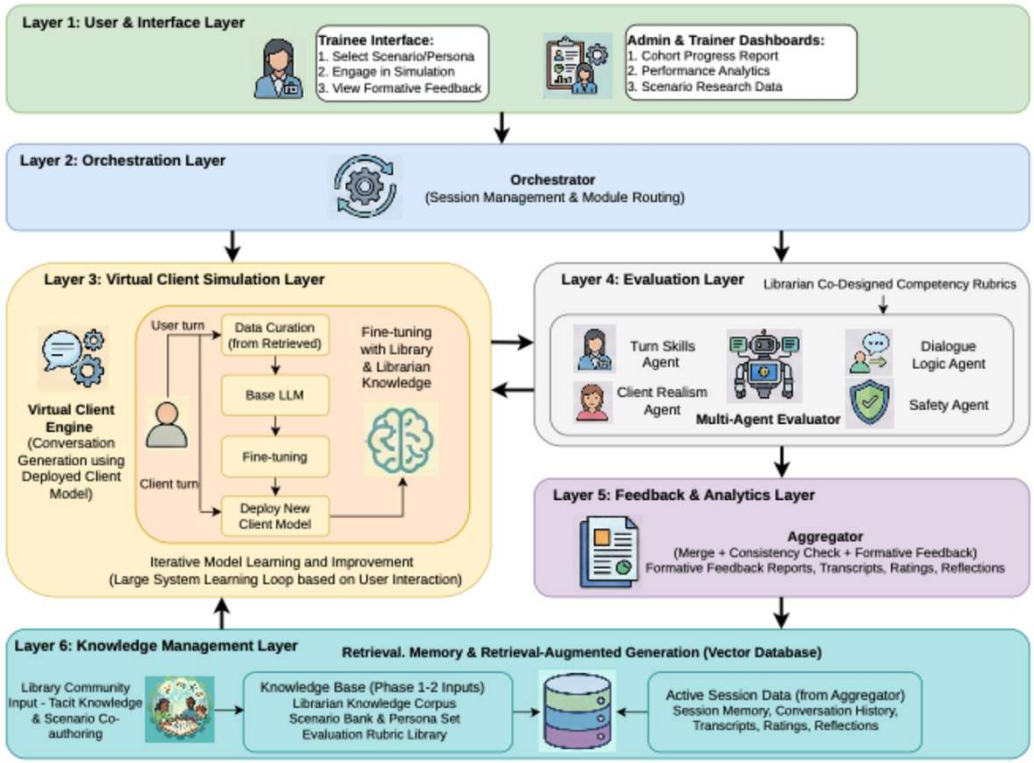
Open benchmark, national guidelines, training for libraries

Testbeds: Densho Digital Repository · LoC Civil Rights History Project · LoC Veterans History Project

Runs on: UNT Data Science GPU cluster (H100 / H200), so collection data never leaves campus

ALiT: AI Platform for Librarian Lifelong Training

Librarians practice hard patron encounters with **AI-simulated patrons** and get instant, rubric-based feedback: anytime, anywhere, at no risk to real patrons.



$382,268

IMLS LB21 · PI · 2026–29
with Amigos Library Services

Interpersonal

De-escalation, active listening, cultural responsiveness

Traditional library knowledge

Reference interviews, readers' advisory, search guidance

Digital literacy support

Walking patrons through devices, online tools, misinformation

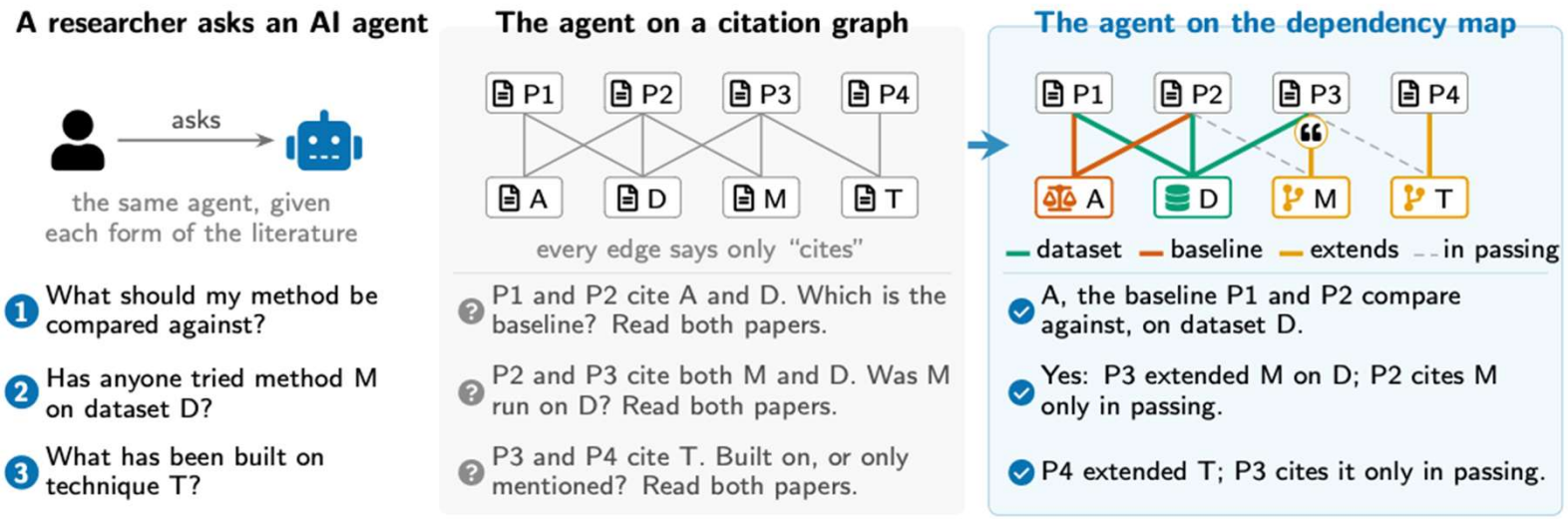
Management & advocacy

Outreach, policy communication, explaining library value

Enhancing the Open-Access AI/ML Literature as an AI-Ready Scientific Dataset

$223,523

NSF AI Datasets · UNT-PI · 2027–30 with NIU
Pending

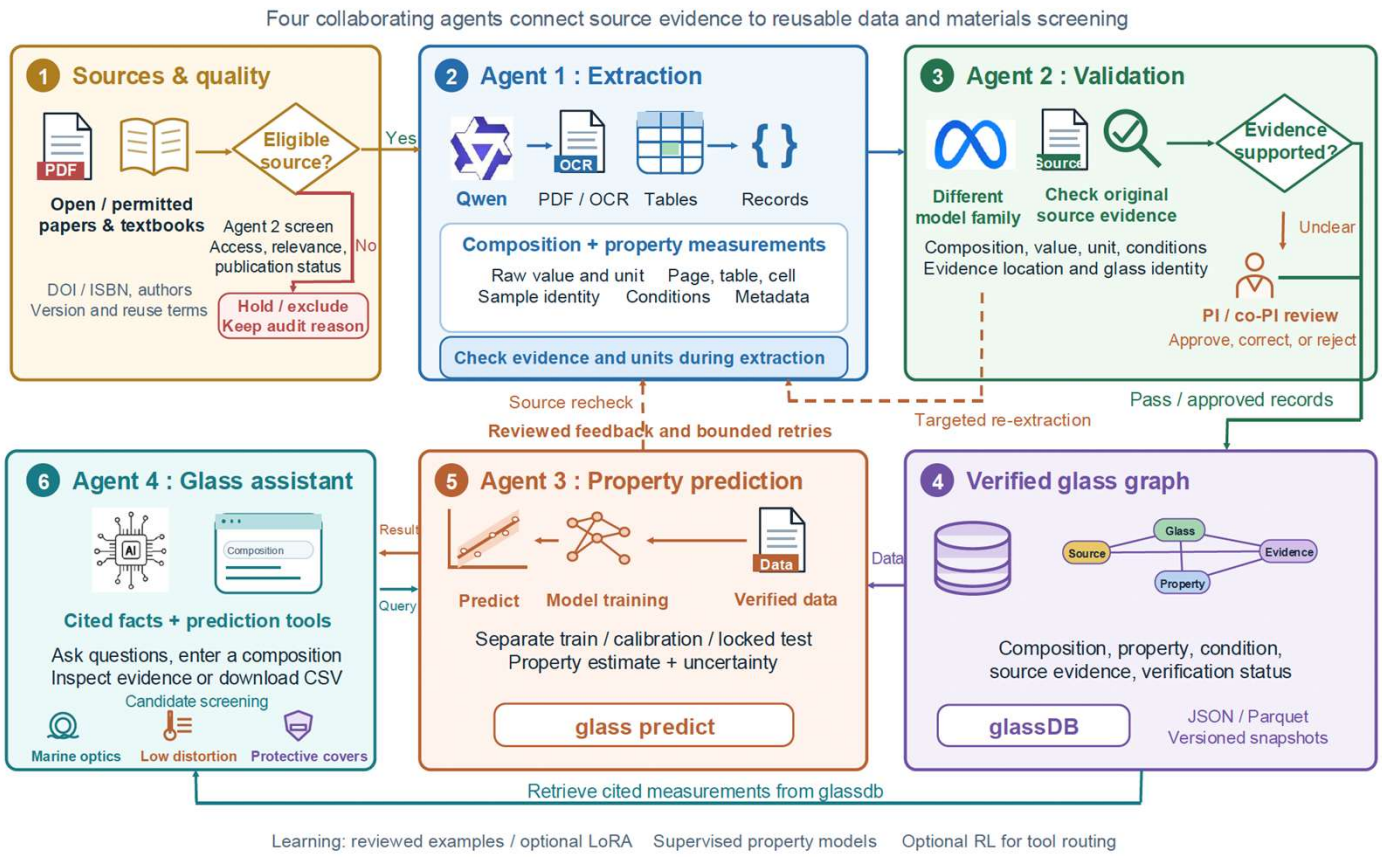


The objective is to build a validated, linked dependency map that people and AI systems can query.

GlassAI: A Multiagent System for Glass Knowledge Discovery and Property Prediction

$120,000

Amazon Research Award, PI
Pending



Let's Collaborate

Data Analytics & Statistics

- Study design & mixed-effects models for user studies (ORAL-AI: 80 participants)
- Measuring bias & reliability of LLM outputs
- Survey methods: faculty AI-in-teaching study

Health Informatics

- LLM data augmentation for clinical NLP (medical concept normalization)
- Trustworthy LLMs for patient-facing health information
- Joint NIH / NSF Smart Health proposals

Data Science

- Multimodal & agentic AI systems
- Benchmarks and evaluation for LLMs
- Joint NSF proposals (IIS, CISE, Smart Health)
- Co-advising Ph.D. & REU students

I bring: GPU cluster access (H100/H200) · LLM evaluation pipelines · IDEA Lab students (20+) · International or national collaboration · Publications on top journals and conferences

haihua.chen@unt.edu · haihua0913.github.io · IDEA Lab