

Education

University of Illinois at Urbana-Champaign

Illinois, United States

B.S. IN COMPUTER SCIENCE; INCOMING M.S. IN COMPUTER SCIENCE; BS GPA: 3.85/4.00

B.S. 2022–2026; M.S. 2026–2027

- Achieved A+/A grades in rigorous coursework including Advanced Natural Language Processing, Computational Photography, Systems Programming, Algorithms, Numerical Methods, Computer Architecture, Data Structures, Probability and Statistics, and Linear Algebra; earned a perfect A+ (100/100) in the graduate-level Advanced NLP course.

Research Profile

First author of an ICML 2026 main conference paper on LLM multi-agent communication protocols; co-first author of MultiAgentBench, an ACL 2025 main conference benchmark with 200+ citations.

Publications

ProtocolBench: Which LLM Multi-Agent Protocol to Choose?

HONGYI DU, JIAQI SU, JISEN LI, LIJIE DING, YINGXUAN YANG, PEIXUAN HAN, XIANGRU TANG, KUNLUN ZHU, JIAXUAN YOU

Sep 2025

- arXiv: arxiv.org/abs/2510.17149
- ICML 2026.

ModelingAgent: Bridging LLMs and Mathematical Modeling for Real-World Challenges

CHENG QIAN, HONGYI DU, HONGRU WANG, XIUSI CHEN, YUJI ZHANG, AVIRUP SIL, CHENGXIANG ZHAI, KATHLEEN MCKEOWN, HENG JI

May 2025

- arXiv: arxiv.org/abs/2505.15068
- EMNLP 2025 Findings

MultiAgentBench: Evaluating the Collaboration and Competition of LLM Agents

KUNLUN ZHU, HONGYI DU, ZHAOCHEN HONG, XIAOCHENG YANG, SHUYI GUO, ZHE WANG, ZHENHAILONG WANG, CHENG QIAN, XIANGRU TANG, HENG JI, JIAXUAN YOU

Mar. 2025

- arXiv: arxiv.org/abs/2503.01935
- ACL 2025 Main Conference. Widely referenced benchmark for LLM-based multi-agent collaboration (200+ citations).

Ψ-Bench: Evaluating Persona-Sensitive Influencing in Persuasive Dialogues

PEIXUAN HAN, HONGYI DU, JIAYU LIU, YIHANG SUN, JIAXUAN YOU

May 2026

- arXiv: <https://arxiv.org/abs/2606.02754>.
- Status: Under review at EMNLP 2026.

EscapeBench: Towards Advancing Creative Intelligence of Language Model Agents

CHENG QIAN, PEIXUAN HAN, QINYU LUO, BINGXIANG HE, XIUSI CHEN, YUJI ZHANG, HONGYI DU, JIARUI YAO, XIAOCHENG YANG, DENGHUI ZHANG, YUNZHU LI, HENG JI

Dec. 2024

- arXiv: arxiv.org/abs/2412.13549
- ACL 2025 Main Conference

Research Experience

SocioGenesis: Long-Horizon Organization Simulation for LLM Agents

Champaign, United States

LEAD RESEARCHER – INDEPENDENT RESEARCH PROJECT; ADVISED BY PROF. JIAXUAN YOU

Jun. 2026 – Present

- Content:** Develop SocioGenesis, a long-horizon LLM-agent organization testbed that studies how agent companies form persistent organizational capabilities under realistic constraints such as roles, skills, deadlines, meetings, tool access, fatigue, budget, repositories, and shared workspaces.
- Method:** Design professional profile graphs as structured decision priors over role, skill, memory, communication style, evidence threshold, risk preference, ownership drive, and decision preferences, moving beyond persona-style prompting toward graph-grounded organizational heterogeneity.
- Environment and Evaluation:** Build OrgEnv, a startup-style product-development environment with repositories, documents, task boards, customer events, and community-profile evaluators; compare homogeneous agents, role-only agents, aligned professional-profile agents, shuffled-profile controls, and institution-formation pipelines.
- My Role:** Lead problem formulation, benchmark design, profile-graph construction, simulation harness implementation, ablation design, evaluation protocol design, and product-market assessment for long-horizon agent organization research.

EvolvingAgents: Bio-Inspired Design and Evolution for LLM-Based Multi-Agent Systems

Champaign, United States

UNDERGRADUATE RESEARCH ASSISTANT AT BLENDER LAB & U LAB – SUPERVISED BY PROF. GENE ROBINSON, PROF. HENG JI,
AND PROF. JIAXUAN YOU; IN COLLABORATION WITH PROF. JAMES ZOU

Jun. 2025 – Jun. 2026

- **Content:** Developed **EvolvingAgents**, a two-pillar bio-inspired LLM multi-agent framework that couples within-task coordination mechanisms with across-task organizational adaptation.
- **Mechanism Design:** Co-designed **stigmergic memory**, **typed reflex signaling**, and an **LLM-driven adaptive immune pathway** to address memory pollution, routine coordination overhead, and non-persistent failure recovery in LLM-based multi-agent systems.
- **Meta-Evolution:** Designed a **Meta-Evolver** that clusters recurrent experience into auditable runtime components with lineage, usage evidence, ratings, promotion, and retirement, enabling persistent adaptation beyond static role or workflow prompts.
- **Nature Environment:** Led the design and implementation of the **Nature Environment**, a long-horizon ecological-civilization testbed with partial observability, resource scarcity, seasonal pressures, hidden recipe discovery, cultural transmission, and civilization-level metrics.
- **Evaluation:** Contributed to experiments on SWE-bench Verified, DSGym, and Nature pilots, showing that bio-inspired mechanisms behave as **conditional affordances**: they help when supplying a missing coordination capability and interfere when adding directive overhead to an already competent policy.
- **My Role:** Proposed the civilization-style benchmark vision, led Nature scenario and evaluation design, coordinated interfaces across simulation, observation, memory, evolution, and analysis modules, and contributed to the paper's technical narrative and June 2026 manuscript completion.

ATLAS: Adaptive Technologies for Learning and Reskilling

Champaign, United States

RESEARCH LEAD FOR GRAPH-GROUNDED PERSONALIZATION AND LEARNER MODELING AT UIUC ATLAS MOONSHOT PROJECT –

ADVISED BY PROF. CHENGXIANG ZHAI, PROF. JIAXUAN YOU, PROF. NANCY M. AMATO, PROF. DILEK HAKKANI-TUR, AND

May 2026 – Present

PROF. JIAWEI HAN

- **Content:** Lead the graph-grounded personalization and learner modeling direction in the UIUC ATLAS Moonshot project, an Upskill-as-You-Go AI education and reskilling ecosystem connecting learners, training providers, and employers through trustworthy AI assistance.
- **Method:** Design Open Reskill Graph structures that decompose curricula, job requirements, and learning materials into atomic skill nodes, dependency edges, prerequisite relations, evidence links, and adaptive microlearning units.
- **Learner Modeling:** Develop uncertainty-aware learner-state representations that track mastery, missing prerequisites, evidence quality, learning goals, constraints, and preference signals for graph-grounded recommendation and diagnosis.
- **Evaluation:** Lead evaluation design for graph-grounded human-AI learning interactions, including skill-gap diagnosis, learning-gain measurement, recommendation quality, and agentic pedagogy metrics for distinguishing productive AI-pilot behavior from passive autopilot usage.
- **My Role:** Own the personalization and learner-modeling track, including problem formulation, graph schema design, mastery-state modeling, benchmark construction, evaluation protocol design, and alignment with the broader ATLAS ontology and reskilling graph.

ProtocolBench: Which LLM Multi-Agent Protocol to Choose?

Champaign, United States

UNDERGRADUATE RESEARCH ASSISTANT AT U LAB (UIUC) – SUPERVISED BY PROF. JIAXUAN YOU

May 2025 – Oct. 2025

- **Content:** Developed ProtocolBench, a protocol-agnostic benchmark for LLM-based multi-agent communication (A2A, ACP, ANP, and Agora) and ProtocolRouter, a scenario-aware selector that chooses one protocol per module.
- **Method:** Built a three-layer evaluation stack (protocol adapters, scenario harness, metrics/logger) and designed representative scenarios (GAIA QA, safety tech, streaming queue, and fail-storm recovery) to jointly evaluate task utility, latency, resilience, and security.
- **My Role:** Acted as team lead and first author, responsible for benchmark design, scenario specification, protocol behavior analysis, and the end-to-end implementation and training of ProtocolRouter.
- **Outcome:** Showed that learned routing achieves 63.3% scenario / 81.7% module selection accuracy and matches or surpasses the best single-protocol baselines, including lower mean latency than ACP in streaming and faster recovery than A2A in fail-storm. Code | arXiv:2510.17149. Accepted to ICML 2026.

MultiAgentBench: Evaluating the Collaboration and Competition of LLM Agents

Champaign, United States

UNDERGRADUATE RESEARCH ASSISTANT AT BLENDER LAB – SUPERVISED BY PROF. JIAXUAN YOU AND PROF. HENG JI

Aug. 2024 – Feb. 2025

- **Content:** Developed MultiAgentBench, a benchmark for LLM-based multi-agent systems across coding, games (e.g., Minecraft), and research-style tasks, focusing on planning, memory, and collaborative abilities.
- **Method:** Designed the Werewolf Arena framework to simulate decentralized social interactions and evaluate theory-of-mind reasoning, leadership, cooperation, and emergent distrust among LLM agents.
- **My Role:** Co-first author; conceived and independently implemented the Werewolf scenario and analysis pipeline, and performed detailed behavioral analysis of multi-agent interactions. Discovered the emergent conservative “witch” behavior in the Werewolf environment.
- **Outcome:** Showed that stronger foundation models tend to develop mutual distrust and internal friction in social deduction games; benchmark has been recognized as a standard reference for LLM-based multi-agent collaboration and behavior evaluation. Code | arXiv:2503.01935.

Novelty Detection for RAG-Grounded LLM-Generated Research Content

Champaign, United States

RESEARCHER – SUPERVISED BY PROF. JIAWEI HAN

Aug. 2025 – Present

- **Content:** Study originality and novelty detection for LLM-generated long-form answers under RAG settings, distinguishing genuinely new contributions from paraphrased evidence, retrieved facts, and shallow recombinations.
- **Method:** Design a multi-dimensional novelty taxonomy covering restatement, semantic recombination, evidence-supported extension, and genuinely novel hypothesis generation, together with annotation guidelines robust to heavy paraphrasing.
- **System:** Build an originality-check pipeline that aligns generated claims with retrieved evidence, flags lexically transformed but semantically duplicated content, and down-weights pseudo-novel outputs in RAG-based generation.
- **Evaluation:** Apply the framework to long-form machine learning research proposals generated under different RAG configurations, jointly analyzing novelty, grounding, coherence, and proposal quality.
- **My Role:** Lead problem formulation, taxonomy design, pipeline implementation, and experimental analysis for an ongoing course-to-paper research project in CS598JH.