Muhua (黄牧华) Huang

muhua@stanford.edu · muhua-h.github.io · LinkedIn · Google Scholar

Computational social scientist and Stanford Organizational Behavior PhD researcher studying how AI systems represent human traits and reshape collective behavior. I build large-scale LLM-agent experiments and psychometric methods to investigate human–AI collaboration, culture, values, and emergent social dynamics.

Education

Stanford Graduate School of Business, PhD in Organizational Behavior — 2025 – 2030
University of Chicago, MA in Computational Social Science, GPA: 3.9/4.0 — 2023 – 2025
University of British Columbia, BA in Computer Science & Psychology (Honours), GPA: 4.2/4.3 — 2018 – 2023

Additional training: International AI Evaluation Programme Fellow; SICSS-Beijing (Outstanding Project, scholarship); Google Research Mentorship Program Scholar.

Industry Experience

Student Researcher | Google DeepMind — Jun – Sep 2026

Contributing Researcher and Writer | Stanford HAI & Google DeepMind — 2026

- Co-developed *Organizing Intelligence*, contributing to conceptualization, research, and writing of a 16-topic AI and organizations primer.

Research Intern | Microsoft Research Asia — Jul – Oct 2024

- Built a scalable multi-agent platform with cognitive modules, RAG, and parallel processing to study how LLM communities develop values and social structures.
- Designed experiments varying value diversity, population size, and time; analyzed network change and emergent language, culture, and institutions with social network analysis, GraphRAG, and NLP.

Publications

AI Simulation

- **Huang, M.**, & Evans, J. (2026). Institutions as cached computation for resource-rational negotiation. *Behavioral and Brain Sciences*, 49, e144.

AI Character & Value Alignment

- **Huang, M.**, Zhang, X., Soto, C., & Evans, J. (2026). Designing AI-Agents with Personalities: A Psychometric Approach. *Personality Science*, 7.
- Bai, Y., Duan, S., **Huang, M.**, et al. (2026). IROTE: Human-like Traits Elicitation of LLMs via In-Context Self-Reflective Optimization. *AAAI*, 40(36), 30040–30048.
- Yao, J., Yi, X., Duan, S., Wang, J., Bai, Y., **Huang, M.**, et al. (2025). Value compass benchmarks: A comprehensive, generative and self-evolving platform for LLMs' value evaluation. *ACL*.
- Kim, H., Yi, X., Bak, J., Yao, J., Lian, J., **Huang, M.**, et al. (2025). The Road to Artificial SuperIntelligence: A Comprehensive Survey of Superalignment. *SuperIntelligence - Robotics - Safety & Alignment*, 2(1).

Social Science

- Savalei, V., & **Huang, M.** (2025). Fit Indices Are Insensitive to Multiple Minor Violations of Perfect Simple Structure in Confirmatory Factor Analysis. *Psychological Methods*.
- Zhang, X., **Huang, M.**, Sun, J., & Savalei, V. (2025). Improving the Measurement of the Big Five via Alternative Formats for the BFI-2. *Journal of Personality Assessment*, 108(2), 250–271.
- Laurin, K., Engstrom, H., & **Huang, M.** (2024). What will my life be like when I'm 25? How social class predicts kids' answers to this question, and how their answers predict their futures. *Journal of Social Issues*, 80(4), 1433–1459.

Leadership & Service

Computational Social Science Workshop Coordinator | University of Chicago — 2024 – 2025
Organized the weekly MACSS workshop featuring speakers on advanced computational methods and interdisciplinary research.

Teaching Assistant | University of British Columbia — 2021 – 2022
Taught Human Computer Interaction Methods and Systematic Program Design; rated 4.7/5 and received three return offers.

Crisis Respondent | Kids Help Phone; 230+ hours and 550+ people supported — 2019 – 2020

Peer Reviewer | *Personality Science*; *Behaviour & Information Technology*

Methods: LLM agents/GABM, psychometrics/SEM, social network analysis, ML/NLP, behavioral experiments, RAG, Python, R, SQL.

Recognition: Stanford GSB & EDGE Fellowships; UChicago Thesis Award; MSRA Outstanding Intern; OpenAI Researcher Access.