Tiancheng Hu

tiancheng.hu.job@gmail.com | tiancheng.hu | Google Scholar

Cambridge Ph.D. candidate | 2025 Apple Scholar in AI/ML | Gates Cambridge Scholar

Research: LLM simulation; AI safety; personalization; evaluation science ; human–AI interaction; societal impact.

Education

Jan 2027 (expected) **University of Cambridge** — Ph.D. in Computation, Cognition and Language
Advisor: Nigel Collier.

2022 **ETH Zürich** — M.Sc. in Electrical Engineering and Information Technology
GPA: 5.65/6.00; thesis at EPFL Data Science Lab.
Advisors: Manoel Horta Ribeiro, Robert West, and Andreas Spitz.

2020 **The University of Texas at Dallas** — B.S. in Electrical Engineering
Advisor: Carlos Busso; GPA: 3.99/4.00; *summa cum laude*.

Research Experience

Apple | **Research Intern** Feb–Jul 2026
Researched **LLM post-training for complex instruction following**.
Cupertino, California. Mentors: Maartje ter Hoeve, Rin Metcalf, Andrew Silva.

International AI Safety Report 2026 | **Expert Section Writer** Aug–Dec 2025
Co-authored the *Current Capabilities* chapter on frontier LLMs and agentic systems.
Report chaired by Yoshua Bengio.

Bocconi University | **Visiting Ph.D. Researcher** Sep–Dec 2024
Researched LLM simulation of human behavior and individual differences.
Milan NLP Lab. Hosts: Dirk Hovy and Paul Röttger.

Selected Publications

* Equal contribution. Titles link to publications; full list: Google Scholar.

LLM Simulation & Personalization

SimBench: Benchmarking the Ability of Large Language Models to Simulate Human Behaviors
T. Hu, J. Baumann, L. Lupo, N. Collier, D. Hovy, P. Röttger. *ICLR 2026*.

iNews: A Multimodal Dataset for Modeling Personalized Affective Responses to News
T. Hu, N. Collier. *ACL 2025*.

Quantifying the Persona Effect in LLM Simulations
T. Hu, N. Collier. *ACL 2024*.

Can LLM Be a Personalized Judge?
Y. R. Dong*, **T. Hu***, N. Collier. *Findings of EMNLP 2024*.

When Personalization Meets Reality: A Multi-Faceted Analysis of Personalized Preference Learning
Y. R. Dong, **T. Hu**, Y. Liu, A. Üstün, N. Collier. *Findings of EMNLP 2025*.

Model Behavior & Human–Agent Interaction

Steer Model beyond Assistant: Controlling System Prompt Strength via Contrastive Decoding
Y. R. Dong*, **T. Hu***, Z. Hui*, N. Collier. *Preprint 2026*.

Value of Information: A Framework for Human–Agent Communication
Y. R. Dong, **T. Hu**, Z. Hui, C. Zhang, I. Vulić, A. Bobu, N. Collier. *ACL 2026*, ***Oral***.

Selected Publications (continued)

Calibration & Model Reliability

Position: Large Language Model Failures from Hallucination to Homogenization Are Different Facets of Miscalibration
T. Hu, C. Zhang, D. Hovy, N. Collier. *Preprint 2026*.

Navigating the Alignment–Calibration Trade-off: A Pareto-Superior Frontier via Model Merging
T. Hu, B. Minixhofer, N. Collier. *Findings of ACL 2026*.

Confidence Estimation for LLMs in Multi-turn Interactions
C. Zhang, R. Yang, X. Zhu, C. Li, **T. Hu**, Y. R. Dong, D. Yang, N. Collier. *Findings of ACL 2026*.

Decoupling the Effect of Chain-of-Thought Reasoning: A Human Label Variation Perspective
B. Chen, **T. Hu**, C. Zhang, R. Litschko, A. Korhonen, B. Plank. *Findings of ACL 2026*.

Social & Multi-Agent AI

Generative Language Models Exhibit Social Identity Biases
T. Hu[*], Y. Kyrychenko[*], S. Rathje, N. Collier, S. van der Linden, J. Roozenbeek. *Nature Computational Science 2024*.

Multi-agent AI Systems Outperform Human Teams in Creativity
T. Hu, Y. Jiang, H. Li, J. Hernández-Orallo, X. Xie, N. Collier, D. Stillwell, L. Sun. *Preprint 2026*.

TRACE: A Corpus of Team Creative Discussions
Y. Jiang, **T. Hu**, J. Hernández-Orallo, D. Stillwell, L. Sun. *ACL 2026*.

A Reporting Checklist for Large Language Models in Behavioural Science
S. Feuerriegel, C. Barrie, M. J. Crockett, et al. (including **T. Hu**). *Nature Human Behaviour 2026*.

AI Safety

Personalization up to a Point: Why Personalized Content Moderation Needs Boundaries, and How We Can Enforce Them
E. Moscato, **T. Hu**, M. Orlikowski, P. Röttger, D. Nozza. *EMNLP 2025*.

International AI Safety Report 2026
Y. Bengio, S. Clare, C. Prunkl, et al. (including **T. Hu**). *Technical report, 2026*.

Selected Honors & Funding

2025	**Apple Scholars in AI/ML Ph.D. Fellowship** — 1 of 21 scholars globally.
2022	**Gates Cambridge Scholarship** — full Ph.D. funding; 1 of 79 scholars globally.
2023–2024	**Research Grants**: Cambridge Language Sciences Incubator Fund (£4,167), Google PaliGemma Grant ($5,000), OpenAI Researcher Access ($2,000), Cohere For AI Grant ($2,000)

Invited Talks & Academic Service

2026	**Stanford University**: *From Individuals to Groups to Interactions: What Is Missing in LLM Simulations?*
2025	**Cambridge Judge Business School**: *LLM Social Simulation is a Promising Research Method*.
2024	**Outstanding Reviewer**, EMNLP; **Special Recognition for Outstanding Review**, CHI.
2023–present	**Reviewer**: NeurIPS, ICML, ICLR, ACL ARR, CHI, ICWSM, LREC-COLING, JMIR. **Co-organizer**: Cambridge Language Technology Lab Seminar.
2023	**Panelist**: EMNLP Widening Natural Language Processing Workshop.