

Jan-Philipp Fränken

janphilipp.franken@gmail.com
janphilippfranken.github.io

Research Experience

- Jan 2025 – Present **Research Scientist**, Google DeepMind (London)
- Post-training and RL, Gemini Team
- Mar 2023 – Nov 2024 **Postdoc**, Stanford University
- Post-training and evaluations (advised by Noah Goodman and Tobias Gerstenberg)
- Sept 2022 – Jan 2023 **Visiting Researcher**, UC Berkeley
- Concept and representation learning (advised by Steven Piantadosi)

Education

- Sep 2019 – Dec 2022 **Ph.D. in Psychology**, The University of Edinburgh
- Sep 2018 – Sep 2019 **M.S. in Cognitive Science**, University College London
- Sep 2015 – Jul 2018 **B.S. in Psychology**, Maastricht University

Selected Works

* Equal Contribution

- 2025 **Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities**
Google Gemini Team.
- 2025 **Towards System 2 Reasoning in LLMs: Learning How to Think With Meta Chain-of-Thought**
Violet Xiang, Charlie Snell, Kanishk Gandhi, ..., **Jan-Philipp Fränken**, Nick Haber, and Chelsea Finn.
- 2024 **Self-Supervised Alignment with Mutual Information: Learning to Follow Principles without Preference Labels**
Jan-Philipp Fränken, Eric Zelikman, Rafael Rafailov, Kanishk Gandhi, Tobias Gerstenberg, Noah Goodman. *NeurIPS 2024*
- 2024 **STaR-GATE: Teaching Language Models to Ask Clarifying Questions**
Chinmaya Andukuri*, **Jan-Philipp Fränken***, Tobias Gerstenberg, Noah Goodman. *COLM 2024*
- 2024 **Procedural Dilemma Generation for Evaluating Moral Reasoning in Humans and Language Models**
Jan-Philipp Fränken, Kanishk Gandhi, Tori Qiu, Ayesha Khawaja, Noah Goodman, Tobias Gerstenberg. *CogSci 2024 (Oral)*
- 2023 **Social Contract AI: Aligning AI Assistants with Implicit Group Norms**
Jan-Philipp Fränken, Samuel Kwok, Peixuan Ye, Kanishk Gandhi, Dilip Arumugam, Jared Moore, Alex Tamkin, Noah Goodman. *NeurIPS 2023 SoLaR Workshop (Oral)*
- 2023 **Understanding Social Reasoning in Language Models with Language Models**
Kanishk Gandhi*, **Jan-Philipp Fränken***, Tobias Gerstenberg, Noah Goodman. *NeurIPS 2023 Datasets & Benchmarks (Spotlight)*