

HAIJIANG YAN

✉ haijiang.yan@warwick.ac.uk

Education

PhD Psychology (candidate)

2022 – Expected Oct. 2026

University of Warwick, England, UK

Supervisors: Adam Sanborn (Psychology) and Nick Chater (Warwick Business School)

MSc Psychology

2019 – 2022

University of Chinese Academy of Sciences, Beijing, China

BSc Medical Technology & Psychology

2015 – 2019

Peking University, Beijing, China

Research experience

Large Language Models for Cognition

2023 – ongoing

Collaborators: Dr.Jianqiao Zhu (HKU) and Prof.Tom Griffith (Princeton)

- Through training a GPT model from scratch informed by human prior, we examined the feasibility of employing auto-regressive AI as a rigorous cognitive model;
- Grounded in psychology and machine learning, we recovered LLMs' symbolic mental representations from their behavior or hidden activation, with which we steered LLMs' decision-making causally;
- We built Bayesian models of token prediction to explain flaws in LLMs' generation.

Integrating Neural Networks in Bayesian Models of Cognition

2022 – ongoing

Advisors: Prof.Adam Sanborn and Nick Chater (Warwick)

- By integrating deep learning models into MCMC sampling with people, human mental representations can be identified at the individual level quickly and comprehensively;
- By training neural nets as normative front-ends of Bayesian Sampler, we built cognitive models that are stimulus-computable and generalizable across domains.

Heterogeneity in Gaming Addiction

2020 – 2022

Advisor: Dr.Guozhen Zhao (UCAS)

- Applied Multi-level Kernel Density Analysis to integrate the findings on abnormal functional neural circuits of addiction, provide the first meta-analytic evidence for large-scale dysfunctions;
- Proposed and verified a bi-axial Psychological structure of addiction (symptoms & problems) with evidence from behavior and brain activity.

Research outputs

Preprints

Yan, H., Zhu, JQ., Sanborn, AN. (2025). Large Language Models Can Take False First Steps at Inference-time Planning. [ArXiv](#)

Yan, H., Chater, N., Sanborn, AN. (2025). PriorProbe: Recovering Individual-Level Priors for Personalizing Neural Networks in Facial Expression Recognition. [ArXiv](#)

Yan, H., Tsvetkov, C., Chater, N., Sanborn, AN. (2024). Quickly recovering comprehensive individual mental representations of facial affect. Under Review. [OSF](#)

Publications

Zhu, JQ.*, **Yan, H.***, Griffith, TL. (2025). Steering Risk Preferences in Large Language Models by Aligning Behavioral and Neural Representations. Accepted by the 48th Annual Meeting of the Cognitive Science Society. [ArXiv](#)

Zhu, JQ., **Yan, H.**, Griffith, TL. (2025). Recovering Event Probabilities from Large Language Model Embeddings via Axiomatic Constraints. Accepted by the 48th Annual Meeting of the Cognitive Science Society. [ArXiv](#)

Zhu, JQ., **Yan, H.**, Griffith, TL. (2025). Language Models Trained to do Arithmetic Predict Human Risky and Intertemporal Choice. The Thirteenth International Conference on Learning Representations, ICLR 2025.

Sanborn, AN., **Yan, H.**, Tsvetkov, C. (2024). Combining meta-learned models with process models of cognition. Behavioral and Brain Sciences.

Zhu, JQ., **Yan, H.**, Griffith, TL. (2024). Recovering Mental Representations from Large Language Models with Markov Chain Monte Carlo. In Proceedings of the 46th Annual Meeting of the Cognitive Science Society.

Yan, H., Li, Q., Yu, K., and Zhao, G. (2021). Large-scale network dysfunction in youths with Internet gaming disorder: a meta-analysis of resting-state functional connectivity studies. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 109, 110242.

Conference

July, 2023 Exploring human representations of facial affect by integrating a deep generative model into Markov Chain Monte Carlo With People. Poster at the 56th Annual Meeting of the Society for Mathematical Psychology, Amsterdam, NL.

July, 2024 Recovering individual mental representations of facial affect using Markov Chain Monte Carlo with People and Gatekeepers. Talk at the 57th Annual Meeting of the Society for Mathematical Psychology, Tilburg, NL.

July, 2024 Recovering individual mental representations of facial affect using Markov Chain Monte Carlo with People and Gatekeepers. Poster at the 46th Annual Meeting of the Cognitive Science Society, Rotterdam, NL.

Software Package

AIOM,  2024 – ongoing. An automated online behavioral experimentation toolkit, with which researchers can design, configure, deploy and manage online data collection easily. Tech Stack: *Nodejs, shell, postgresql, docker*

Awards & Honors

Warwick Behavioral Spotlight Small Grant, University of Warwick, 2025

Fully funded doctoral scholarship, Warwick Chancellor's International Scholarship, 2022-2026

China National Scholarship, Institute of Psychology, CAS, 2021

University's Merit Student & First-class Scholarship, Peking University, 2018

Teaching & Mentoring

Teaching Experience

Graduate Teaching Assistant, Department of Psychology, University of Warwick

2023 – ongoing

- Methods & Analysis in Behavioural Science (tutorial sessions)
- Bayesian Approaches in Behavioural Science (R laboratory sessions)
- Research & Statistical Methods in Psychology (tutorial sessions)

Mentoring Experience

MSc student supervisor in Behavioral and Data Science, Department of Psychology, University of Warwick

2025

- Thesis topic: Integrating Neural Nets in Process Models of Cognition

Technical Skills

Programming: Python, JavaScript, R, MATLAB

Machine Learning: PyTorch, TensorFlow

Research Related: EEG, Dallerger, jsPsych, PsychToolbox, L^AT_EX

Web Development: Node, Flask, HTML/CSS, PostgreSQL, Linux, Git, Docker