Nick (Seungheon) Oh

nick.sh.oh@gmail.com

LINKS

Google Scholar, LinkedIn, GitHub, Substack

PROFILE

Researcher pursuing an interdisciplinary approach to intelligence, exploring the full landscape of cognitive solutions to understand whether human intelligence represents one path among many or a convergent optimum. Published 6 papers (2 conference proceedings, 4 workshops) at NeurIPS, COLM, ICWSM, and AIES, including recent work with Prof. Fernand Gobet on metacognitive architectures for LLM reasoning, and investigating the epistemic value of post-hoc explanations despite their imperfect factivity. Currently interested in two research approaches: (1) using neural networks trained on behavioural data to identify where cognitive theories fall short, with XAI techniques bridging pattern recognition and theoretical understanding; and (2) comparing where humans and machines independently converge or diverge when solving identical problems under matching constraints. Seeking doctoral training to develop rigorous behavioural experiments and formalise theories within rational metareasoning frameworks. Ultimately striving to develop systematic methods for using AI both as a lens to examine human cognition and as a parallel experiment in intelligence, revealing which cognitive principles are universal versus uniquely human.

EDUCATION

Apr 2022 — Sept 2023

MSc Data Science and Artificial Intelligence, University of London — London

Distinction

Sept 2016 — Jun 2020

BSc Government and Economics, London School of Economics & Political Science (LSE) — London

Upper Second Class Honours [Cumulative GPA: 3.5-7/4.0]

EMPLOYMENT HISTORY

Mar 2025 — Present

Defence AI researcher, Ministry of National Defence — Republic of Korea

Assigned to the C2 (Command & Control) Concept and Requirements Branch, Military Science and Technology Research Unit, Republic of Korea Army HQ. Expected military discharge in September 2026.

Jan 2023 — Present

Researcher (founder), socius labs — London

Founded socius labs (https://socius.org), building it from concept to LSE-backed research lab.

- Working with Prof. Fernand Gobet (LSE) on theoretical frameworks and metacognitive architectures for LLM reasoning, drawing from cognitive psychology research:
  - (1) Proposed a framework for machine decision-making inspired by how human experts coordinate intuitive and analytical processing through metacognitive regulation (*System 1.5*); and
  - (2) Formalised Flavell and Nelson & Narens' metacognitive theories into computational architecture (*Monitor-Generate-Verify*)
- Instantiated the Monitor-Generate-Verify (MGV) framework as computational implementations in production systems:
  - (1) Implemented Flavell's metacognitive monitoring model as zero-shot inference-time reasoning framework with Llama-3.1 and vllm, improving performance by up to 7pp over iterative self-refinement baselines on GSM8K (*Before you <think>, monitor*); and
  - (2) Explored parameter-efficient fine-tuning through pilot experiments using unsloth and trl, where branched curriculum multi-LoRA with 3B routing model outperformed monolithic 8B GRPO fine-tuning by 1.4pp on GSM8K
- Applied philosophy of science to XAI, arguing explanation methods can be "wrong" about models while revealing truths about reality (*In Defence of Post-hoc Explainability*);
- Developed research infrastructures (*sentibank* for sentiment dictionaries, *RedditHarbor* for collecting/storing/exporting Reddit data, and *PETLP* framework addressing privacy, ethics, and legal compliance in social media AI research pipelines)

Nov 2021 — Oct 2022 | Junior NLP engineer, Numen Capital | London

Contributed to Knowsis research team extracting trading signals from non-traditional data sources (social media) through NLP and machine learning. Built transformer-based ESG (Environment, Social, Governance) tweet classifier achieving 97% accuracy, implemented rule-based sentiment analyser with custom financial dictionaries, and integrated these NLP-derived features into deep learning models for market movement forecasting. Developed end-to-end pipeline from raw text to actionable trading signals.

CONFERENCE PROCEEDINGS

Oh, N., Vrakas, G. D., Brooke, S. J., Morinière, S., & Duke, T. (2025). PETLP: A Privacy-by-Design Pipeline for Social Media Data in AI Research. *Proceedings of the 8th AAAI/ACM Conference on AI, Ethics, and Society (AIES'25).* [**Paper**] [**Poster**]

Oh, N. (2024). Sentibank: A Unified Resource of Sentiment Lexicons and Dictionaries. *Proceedings of the 18th International AAAI Conference on Web and Social Media (ICWSM'24).* [**Paper**] [**Poster**]

WORKSHOPS

Oh, N., & Gobet, F. (2025). Monitor-Generate-Verify (MGV): Formalising Metacognitive Theory for Language Model Reasoning. *Workshop on Foundations of Reasoning in Language Models @ the 39th Annual Conference on Neural Information Processing Systems (NeurIPS'25).* [**Paper**]

Oh, N. (2025). Before you <think>, monitor: Implementing Flavell's metacognitive framework in LLMs. *Workshop on the Application of LLM Explainability to Reasoning and Planning @ Conference on Language Modeling (COLM'25).* [**Paper**] [**Poster**]

Oh, N., & Gobet, F. (2024). System 1.5: Designing Metacognition in Artificial Intelligence. *System 2 Reasoning at Scale Workshop @ the 38th Annual Conference on Neural Information Processing Systems (NeurIPS'24; **Spotlight Poster**).* [**Paper**] [**Poster**] [**Spotlight Talk (Video)**]

Oh, N. (2024). In Defence of Post-hoc Explainability. *Interpretable AI: Past, Present and Future Workshop @ the 38th Annual Conference on Neural Information Processing Systems (NeurIPS'24).* [**Paper**] [**Poster**]

MANUSCRIPT IN PREPARATION

Humphrey, J., **Oh, N.**, & Ross, D. (working paper). High-Intensity Ambivalence: How Reddit Occupational Forums Respond to Developments in Artificial Intelligence.

TEACHING & TUTORIALS

Scraping Reddit the Right Way: A Guide to Legal and Ethical Data Collection with RedditHarbor — Tutorial @ ICWSM'24, co-organised with the Open Data Institute (ODI)

RESEARCH INTERESTS

- Experimental studies of human expertise in strategic games (e.g., chess puzzles, Gomoku, 24)
- Resource-rational analysis of metareasoning and metacognitive control in human cognition
- Convergence-divergence patterns between human reasoning and algorithmic learning (e.g., (STaR, V-STaR, RaM)
- XAI methods as instruments for scientific theory discovery and refinement
- Computational representations of emotion, belief, and metacognitive states
- Naturalistic decision-making in cooperative and adversarial strategic contexts

RELEVANT TRAINING & SKILLS

LSE MA314: Algorithms and Programming (First Class)

LSE GV324: Applied Quantitative Methods for Political Science (First Class)

LSE GV249: Research Design in Political Science (Upper Second Class)

LSE Methodology & Quantitative Training: Causal Inference and natural experiments in the social sciences (regression, matching, panel data, diff-in-diff, instrumental variables, regression discontinuity)

Python: PyTorch, spaCy, NLTK, transformers, trl, unsloth, vllm, peft, PRAW, OpenAPI

HOBBIES

On the side, I'm a maker! I enjoy making abstract concepts tangible. Currently building a physical LED array that simulates LeNet-1 (Le Cun et al., 1989), with each LED representing a neuron firing in real-time [project link]. I'm also a passionate cook -- kitchen is my other laboratory, where I experiment with Korean and Italian cuisine.