John Russell Himawan

johnr.himawan@gmail.com | linkedin.com/in/johnrhimawan | github.com/johnrhimawan | johnrhimawan.com | 83587366

Education

National University of Singapore (NUS) Aug 2021 – Jul 2025

B.Comp. (Honours) in Computer Science; Minors in Statistics & Entrepreneurship

- **GPA:** 4.67/5.00 (First Class Honours); **Awards:** ASEAN Undergraduate Scholarship, Dean's List, NUS Overseas Colleges
- **Teaching Assistant:** Data Structures & Algorithms, Machine Learning, Big Data Systems, NLP
- **Leadership:** VP, NUS Entrepreneurship Society; VP, Google Developer Group

Work Experience

AI Engineer Jul 2026 – Present

UBS *Singapore, Singapore*

- Developed LLM guardrails to mitigate profanity, prompt injection, and hallucinations, reducing these failure cases by **30%** on average
- Built a hierarchical evaluation suite for LLM applications, assessing query routing, retrieval, and response generation to identify component-level failures
- Played a key role in architectural decisions to migrate a deterministic chatbot to an LLM-based system, shaping its design and evaluation strategy

Machine Learning Research Assistant Aug 2025 – Aug 2026

National University of Singapore *Singapore, Singapore*

- Co-first-authored **DIAL**, a framework for fine-grained steering of LLM instruction adherence
- Developed **Watershed**, a benchmark for membership auditing, source attribution and machine-unlearning evaluation
- Implemented watermarking and membership-inference algorithms; created **20+ datasets** and **40+ fine-tuned LLM checkpoints** on Llama/Gemma
- Explored control barrier functions for weight-free inference-time unlearning; evaluated on TOFU

Undergraduate Researcher (B.Comp. Dissertation) Aug 2024 – Jul 2025

National University of Singapore *Singapore, Singapore*

- Designed a RAG-based unlearning framework; adapted gradient-based unlearning for embedding models and trained LLMs with DeepSpeed on multi-GPU systems

Machine Learning Engineer Intern Jan 2024 – Jun 2024

Convier AS *Oslo, Norway*

- Built an LLM assistant for financial-crime investigations, reducing KYC/EDD compliance time by **80%** and enabling adoption by **8 Norwegian banks**
- Developed RAG and text-to-SQL workflows; accelerated LLM inference **10**× using vLLM, quantization and prompt engineering

Software Engineer Intern May 2023 – Aug 2023

One X Group *Singapore, Singapore*

- Delivered 7 React/Next.js client websites; improved organic traffic by **50%** through SEO

Publications

[1] Thanh Q. Tran*, **John Russell Himawan***, Arun Verma*, Kiwan Wong, Wei Xiao, Daniela Rus, Bryan Kian Hsiang Low. DIAL: Steerable Language Models with Distilled Instruction-Adherence Levels. *arXiv:2610.31438*, 2026.

[2] **John Russell Himawan**, Gregory Kang Ruey Lau, Bryan Kian Hsiang Low. Watershed: A Unified Benchmark for End-to-End Data Provenance Evaluation. *ICML AI4GOOD Workshop*, 2026.

[3] Xinyang Lu*, Xinyuan Niu*, Gregory Kang Ruey Lau*, Nhung Bui, Rachael Hwee Ling Sim, **John Russell Himawan**, Fanyu Wen, Chuan-Sheng Foo, See-Kiong Ng, Bryan Kian Hsiang Low. WaterDrum: Watermark-based Data-centric Unlearning Metric. *ICLR*, 2026.

* *Equal contribution.*

Projects

OpenVINO (Intel Open Source) 2025

- Merged a C++ RISC-V CPU JIT emitter for the `NotEqual` operation into Intel's OpenVINO inference toolkit

Multimodal Recommendation System 2025

- Integrated contrastive learning into SASRec, increasing ranking metric from **0.013 to 0.052** (**4**×)

Technical Skills

Programming: Python, C/C++, Java, SQL, JavaScript/TypeScript
AI/ML: PyTorch, Hugging Face, DeepSpeed, vLLM, FAISS, LangChain | **Tools:** Docker, AWS, Spark, Git, Linux