YEO **WEI JIE**

+65 9652 4660
yeow0082@e.ntu.edu.sg

An AI enthusiast whose interest lies in understanding and interpretating the behavior of complex AI systems.

EDUCATION

JAN 2022 – PRESENT

PHD IN COMPUTER SCIENCE

NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

- Research interest: Interpretable AI, Explainability Evaluation, LLMs, Mechanistic Interpretability, AI Safety

- Publications:
    1. NAACL 2024 - How Interpretable are Reasoning Explanations from Prompting Large Language Models? (1st author)
    2. ACL 2024 - Plausible Extractive Rationalization through Semi-Supervised Entailment Signal (1st author)
    3. EMNLP 2024 – Self-training Large Language Models through Knowledge Detection (1st author)
    4. AIRE journal - A comprehensive review on financial explainable AI (1st author)
    5. NAACL 2025 – SusGen-GPT: A Data-Centric LLM for Financial NLP and Sustainability Report Generation (Co-author)
- Under review:
    1. Interpreting and Correcting Spurious Bias in CLIP (Neurips 2025)
    2. Understanding refusal in LLM with SAEs (EMNLP 2025)
    3. Faithful NLP Explanations using activation patching (EMNLP 2025)
- On-going work:
    - Designing computationally feasible defense against prompt-based/parametric jailbreaking attacks.

AUG 2017 – MAY 2020

BACHELOR IN MECHANICAL ENGINEERING WITH HONOURS (DISTINCTION)

NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

- Obtained Cumulative Grade Point Average (CGPA) of 4.44/5
- Specialization in Energy and Environment
- Final Year Project: Model Predictive Control (MPC) Implementation in Building Simulation
    1. The objective is to develop a Building Automation and Control (BAC) system based on the Model-Predictive Control (MPC) concept and machine learning techniques to optimize building energy efficiency and occupant comfort.
    2. Participated in Engie x NTUitive Innovation Challenge and achieved improved thermal comfort in Ng Teng Fong hospital test site
    3. Achieved 32% energy savings during final demonstration day
    4. Emerged 1st with cash prize of S$3,000 and secured S$10,000 demonstration grant for FYP

JAN 2019 – JUN 2019

STUDENT EXCHANGE – GEM EXPLORER

TECHNICAL UNIVERSITY OF DENMARK, DENMARK

APR 2012 – MAY 2015

DIPLOMA IN AEROSPACE ENGINEERING

TEMASEK POLYTECHNIC, SINGAPORE

EXPERIENCE

JUN 2020 – JAN 2022

EQUIPMENT ENGINEER, QUALCOMM

- Oversee the semi-conductor manufacturing operations to ensure optimal production and troubleshoot to determine root cause of quality-related incidents
- Maintain uptime of several essential equipment for Wafer manufacturing process while preventing any potential delays or bottleneck.
- Coordinate with internal and external teams during day-to-day operations
- Involved in overseeing the expansion & additional of equipment in Qualcomm's new building

SKILLS

Language	English, Chinese (Mandarin)
Software	Python, Microsoft Office
Deep Learning Frameworks	TensorFlow, Pytorch, Huggingface, TransformerLens, NNsight

ACTIVITIES

- Working out/ Running
- Cooking
- Socializing with friends and family.
- Travelling
- Films