# Puah Yi Hao | Machine Learning Engineer | Singapore

http://www.linkedin.com/in/yi-hao-puah

https://strikenone.github.io/PortfolioWebsite2/

## EDUCATION

**Nanyang Technological University, Singapore** Aug 2020 – Dec 2024
**Bachelor of Science (Data Science & Artificial Intelligence)**

- Nanyang Scholarship (Highest Tier of Scholarship offered by NTU)
- Dean's List AY20/21, AY21/22
- Obtained 16 A+, 6 A, 5 A- across all graded modules
- Highest Distinction (**GPA: 4.91**)

**Queensland University of Technology, Australia** Jan 2023 – Jun 2023
**GEM Explorer semester exchange program**

- CGPA: **6.75/7**
- Club Activities: QUT Cliffhangers (Outdoor Climbing)

**Saint Andrew Junior College** Feb 2016 – Dec 2017

- Achieved **90 Rank Points**, AAA/A
- Honor Roll Student (Top 10%) for every Major Exam

## EXPERIENCE

**TikTok** Jul 2024 – Present
**Machine Learning Engineer**

- TikTok Engineering – Business Risk Integrated Control (Local Service Team)
- Developed and implemented scalable creator-level risk detection systems supporting TikTok Go Local Service onboarding across ID, TH, JP & US. Developed comprehensive risk coverage over 11 risk categories using 3-tier scoring system. Achieved real-time enforcement with a ~1.8% block rate and maintained 96% enforcement precision.
- Designed and implemented pipeline leveraging OpenAI embeddings for similarity search, integrated with LLM-based prompt evaluations, significantly enhancing the accuracy and scalability of business account detection.
- Built and deployed content and behavior-based ML models to proactively identify and mitigate irrelevant and spam location reviews on TikTok. Developed an end-to-end real-time inference pipeline from scratch, including defining Thrift IDLs, model development, deployment, and hosting on internal cloud infrastructure. Improved enforcement effectiveness, reducing advertising/spam Violative View Rate (VVR) by over 60% and irrelevant review incidence by 24%.
- Developed RAG capabilities through LangChain integration with LLM to enhance offline risk perception capabilities
- Built risk perception capability from ground up in the form of group mining (Leiden), target abuse (Fake view/follow metrics)
- Collaborated closely with Governance & Experience teams to create and implement dashboards tracking key metrics, including creator rejection rates, improving visibility and decision-making capabilities across teams.

**TikTok** Jul 2022 – Jan 2023

**Data Science Intern**

- TikTok Engineering – Business Risk Integrated Control (Transaction Integrity Team)
- Performing risk perception for the Transaction Integrity (TxN) team to mitigate financial loss caused by fraudulent activity on the platform.
- Analyze business and security data to understand user behavior to develop rules to block bad users. Generate dashboards to monitor the risk related metrics. Evaluation of work can be referenced to the bi-monthly objective key results set by the team.
- Integrating machine learning models based on analysis and features generated to the problem. Able to account for 69.25% recall of core region refunds.
- Optimize process in the TxN team daily investigation. Improved operation team members on-call efficiency by over 20%.

**Centre for Strategic Infocom Technologies** May 2022 – Jul 2022

**Data Science Intern**

- Operational Product Department
- Researched on current tactics, techniques and procedures employed by certain groups.
- Data Analytics on Network metadata. Involved collection of own simulated tunnelling data and pre-processing. Learned to handle datasets of large sizes (35 million) via alternative packages to pandas such as Dask
- Worked on Classification on network data to identify and flag out malicious threats. Explored the problem through a machine learning and neural network perspective.

**Defence Science Organisation** May 2021 – Jul 2021

**Computer Vision Intern**

- Robotics Department (Data Analytics in the world of UAVs)
- Worked on Computer Vision project to detect unique signal frequencies (codeword could be changed depending on input) on live video feed and to subsequently track them. (Object detection and tracking).
- Data Analytics on Time Series data. Involved data filtering on noisy data through using low pass filters. Important features extraction was also done using Decision Tree classifiers to look at the features chosen in the first few levels.
- Worked on Predictive Analysis on Time series data to identify and flag out anomalous data. Trained LSTM models for Time series classification and sequence predictions. Library used was Keras and Tensorflow. (Linux and Windows)

## ABOUT ME

Would describe myself as a curious person looking to challenge myself and hone my skillset in machine learning and data science