

Ayrton San Joaquin

TRUSTWORTHY ML RESEARCHER | DATA SCIENTIST | WRITER

✉ ayrton@u.yale-nus.edu.sg | 📍 Singapore | [in ajsanjoaquin](https://www.linkedin.com/in/ajsanjoaquin) | [a jsanjoaquin.github.io/values](https://github.com/ajsanjoaquin/ajsanjoaquin.github.io/values)

Education

Yale-NUS College

Singapore

BSc. (HONORS) IN DATA SCIENCE, MINOR IN PHILOSOPHY. **SCHOLARSHIP RECIPIENT.**

August 2018 - May 2023

Focus: Computer Vision (CV) and Natural Language Processing (NLP). **Semester Abroad at University of Copenhagen.**

Experience

Machine Learning Safety Scholars Program, Center for AI Safety

Palo Alto, United States

SCHOLAR

June 2022 - August 2022

- Led NLP research on analyzing Transformer models' adaptability to new word definitions using few-shot learning.
- Received a grant of *US\$4500* to complete the inaugural 2-month program.

Data Privacy and Trustworthy Machine Learning Lab, NUS

Singapore

UNDERGRADUATE RESEARCHER

May 2021 - August 2021

- Pitched and led an analysis on **Unlearnable Data** as a data protection method against ML training.
- Collaborated with Google Brain on privacy and adversarial machine learning research for my bachelor's thesis in a team across 4 time zones. Published as the youngest and only undergraduate co-author.

NExT++ Research Center

Singapore

RESEARCHER - DEEPPAKE DETECTION

May 2020 - August 2020

- Preprocessed 200,000 images from FaceForensics++ Dataset and trained various detector models (Based on EfficientNet and Xception Net) using a High Performance Computing Cluster.
- Adapted various robustness strategies against adversarial noises (e.g. Adversarial Training, Randomized Smoothing)

Arterys (Freelance)

San Francisco, United States

DEEP LEARNING ENGINEER

March 2020 - June 2020

- Created a COVID-19 Pneumonia classifier 4 days after pandemic declaration in collaboration with A.I. Singapore.
- Collaborated with Arterys to **deploy the model in their platform** for use by American hospitals and researchers. Model engineer in a team of 4 across 3 time zones.

Open-Source Projects & Contributions

Equitable Valuation of Data Using Shapley Values

Data Governance

- Implemented the training data valuation algorithm from *What is your data worth? Equitable Valuation of Data* (Ghorbani and Zou., 2019).

Explaining Neural Networks with Meaningful Perturbations

Explainable AI, CV

- For explaining an image classifier's prediction, I implemented the algorithm described in *Explanations of Black Boxes by Meaningful Perturbation* (Fong, et. al., 2018).

COVID-19 Pneumonia Classifier for Diagnosis Triage

Medical Imaging, CV

- Trained a Resnet-34 Convolutional Neural Network (CNN) on ~ 26,000 images with resampling to detect Pneumonia caused by COVID-19 on xray scans ultimately to triage patients for urgent diagnosis.

Miscellaneous

Machine Learning, NLP, CV

- Added new features for major machine learning projects including Pytorch, HuggingFace Transformers, and YOLOv4 (object detection model).

Publications

December 2022 **San Joaquin, A.,** Haroen, A. , Understanding How Model Size Affects Few-shot Instruction Prompting

[arXiv link](#)

November 2022 Tramer, F., ..., **San Joaquin, A.,** et.al. , Truth Serum: Poisoning Machine Learning Models to Reveal Their Secrets

[ACM CCS 2022 link](#)

March 2020 **San Joaquin, A.,** *Using Deep Learning to Detect Pneumonia caused by COVID-19*

[Towards Data Science \(Editor's Choice\) link](#)

Skills

Programming Languages: Python, Java, R

Machine Learning in Python: Pytorch, NLTK, Spacy, NumPy, Sickit-Learn, Tensorflow, Keras, Jax

Data Management: Pandas, SQL, PySpark

Business Intelligence: Dash, Streamlit

Application Deployment: Kubernetes, Docker, Google Cloud, Git, Singularity