TeYang Lau

ty_lau@outlook.com | +65 82336103
linkedin.com/in/teyang-lau | teyanglau.com | github.com/teyang-lau

WORK EXPERIENCE

Data Scientist | *Sleep & Cognition Laboratory, School of Medicine, NUS* **Apr 2018—Apr 2023**

- Collaborated with external government and industry partners (HPB, Oura) to explore best approach in mining datasets and structuring new projects for public health intervention studies, including a trial to test the effectiveness of incentive-based interventions for HPB's Health Insights Singapore (hiSG) study
- Mined HPB Fitbit longitudinal dataset, resulting in applications that enhance understanding of public health behavior, and improving study design methods. Trained ML and statistical models (K-Means, Hierarchical, ANOVA) to cluster and identify subgroups of adults differentially susceptible to unhealthy behaviours during COVID-19, leading to better public health responses and interventions for future pandemics by tailoring strategies to specific population subgroups
- Analyzed effects of lockdown stringency on sleep and heart rate on Oura wearable data from 20 countries using linear mixed models to advance scientific understanding of how public health policies that resulted in changes in environmental factors can impact population sleep and health patterns
- Employed UMAP for dimensionality reduction of adolescent sleep and steps data from Adolescent Brain Cognitive Development (ABCD) study followed by clustering techniques (K-Means, EM, HDBSCAN) to identify distinct group profiles, leading to better understanding and promotion of healthy sleep in adolescents
- Leveraged behavioral, cognitive and MRI techniques to investigate the effects of napping, which improves memory encoding by 20%, and demonstrated its benefits on brain functioning, vigilance and productivity
- Automated sleep visualization, computation and report generation that expedited processing time by 30-50%
- Co-led 2 nap research projects, co-authored 7 peer-reviewed manuscripts (3 selected for Editor's Choice)

Data Scientist Intern | *Programme Evaluation*, *Health Promotion Board* **Aug 2021—Dec 2021**

- Collaborated with Apple to drive HPB's mission in promoting healthy lifestyles among the continuously growing participant base of ~120k individuals of LumiHealth program
- Analysed effectiveness of A/B experiments including new features and wellness challenges/offers on the app to help improve LumiHealth program, user experience and engagement levels
- Conducted thorough analysis to extract valuable insights, synthesizing findings into impactful slide decks for presentation and actionable recommendations to stakeholders, enabling data-driven decision-making
- Built a preliminary workout recommender system using K-Means clustering for HPB's future health programs
- Ensured data integrity by performing regular checks on and integrating the 56 data tables of LumiHealth app, and conducting User Acceptance Test of data ingested into Singapore's Population Health Data Hub

EDUCATION

Master of IT (Artificial Intelligence) | *SMU* | GPA: 4.00/4.00 | (Dean's List) **2021—2022**

- Awarded the SMU AI Talent Development Grant and SMU MITB Scholarship
- Relevant Courses: Recommender System; NLP; Computer Vision; ML Engineering; Applied ML; Multi-agent Systems; Text Analytics; Cybersecurity; Social Analytics; Data Management

BA Psychology (Honours) | *Flinders University* | 1st Class Honours **2017**

BA Psychology & Management | *Murdoch University* | GPA: 3.75/4.00 **2014—2016**

- Awarded University Medal (Top 7 Graduates), Vice Chancellor's Commendation for Academic Excellence

Certificates: Machine Learning Engineering for Production (MLOps); Deep Learning Specialization; Tensorflow Developer; AI for Medicine; Big Data Analytics Using Spark

SKILLS

- **Languages & Software:** Python, R, SQL, MongoDB, Matlab, PySpark, Tableau, GCP, AWS (SageMaker), Git
- **Machine Learning:** regression, classification, SVM, bagging, boosting, clustering, dimensionality reduction
- **MLOps:** MLflow, GCP — Vertex AI, TensorFlow Extended, AWS — SageMaker, CI/CD
- **Computer Vision:** object detection, image classification, neural style transfer
- **Natural Language Processing:** large language models, text generation, chatgpt and whisper API
- **Recommender System:** collaborative & content-based filtering, matrix factorization, factorization machines, multimodal recommendations, graph neural networks
- **Libraries:** SK-Learn, Tensorflow, PyTorch, Microsoft Recommenders, Streamlit, OpenCV, Gensim, HuggingFace
- **Statistical Analysis:** T-Test, ANOVA, Non-Parametric Tests, Linear Mixed Models, Latent Class Mixed Models
- **Others:** hypothesis testing, A/B testing, experimental design

Selected AI/ML PROJECTS (link to all projects)

Auto-Trimming App for Diving Videos Using Object Detection Models | *Links: GitHub , Web App*

You Only Edit Once is an innovative tool for leisure divers to easily upload their diving videos and have them auto trimmed and ready for sharing. The YOEO app was presented at a Computer Vision application competition and won 3rd place honours. YOEO was also conferred the title of 'most investable project' by the panel of judges.

- Finetuned and trained YOLOX and YOLOv5 object detection models in Tensorflow and Datature
- Created an intelligent algorithm to remove uninteresting segments based on object detection model output to optimize content for an enhanced viewing experience, including auto-extraction of beautified images
- Optimized the object detection process by allowing users to specify how many frames per second to analyze
- Deployed the model onto the streamlit app for public usage. The app includes features such as speed optimization, trimming strictness, audio stitching etc to cater to users who require more customization
- Showcased the potential applications of the system for underwater cameras and marine research purposes

End-to-end MLOps Using MLflow | *Link: GitHub*

- Developed and implemented a MLOps pipeline for a ML problem, leveraging the capabilities of MLflow
- Designed and built a robust automated workflow that covers the entire ML lifecycle with CI/CD, including data validation, processing, model training, evaluation, validation and deployment
- Integrated monitoring and logging functionalities within the pipeline along with a ML metadata store, providing insights into model performance and ensuring data lineage and reproducibility

Auto-conversion of Lectures/Meetings Audio Files to Markdown Notes | *Links: GitHub , Web App*

You Only Listen Once is a productivity tool for students and workers to easily upload their lectures/meetings and have them auto converted into summarized notes in Markdown format, which can be copied into any note-taking app

- Leveraged Whisper and ChatGPT API for auto transcription and markdown-formatted note conversion
- Overcame the 4k token limit of ChatGPT3.5 using a chunking summarization technique prior to conversion
- Employed prompt engineering for creating varied styles of notes and enabled user customization features
- Demonstrated app's positive impact on student & worker productivity, accessibility, and learning outcomes

HDB Resale Prices Prediction with Model Explainability and Visualization| *Links: GitHub , Web App*

- Analyzed 800k resale transactions to predict and identify key drivers of resale price using linear regression and random forest, achieving > 0.9 R^2 model fit and precise predictions with MAE of $20k
- Feature engineered new contributors that captured the distance and number of amenities, such as schools, parks and MRTs around each flat through web-scraping using public APIs (Data.gov.sg, OneMap.sg)
- Deployed model onto web app to predict resale prices, visualize surrounding amenities & identify most important contributors to the prices using SHAP values, enabling users to make informed decisions

Writing Singapore's Next National Day Song Using AI | *Link: GitHub*

This project won 3rd place honours for the MITB 15th Anniversary Poster Competition

- Developed language models (Seq2Seq with attention, LSTM, GRU) and finetuned pre-trained transformer models (GPT-2) for generating national day song lyrics
- Used BLEU and ROUGE metrics to evaluate the performance of the generative models and used VADER & GLoVE to evaluate sentimentality of generated song lyrics
- Leveraged pretrained AI music generative models to generate melody and vocals on generated song lyrics

Developing Singapore's Coffee Joints Recommender System | *Link: GitHub*

- Created scraper using BeautifulSoup to scrape café reviews from Yelp.com
- Trained recommender system models including weighted matrix factorization, multimodal models (convolutional matrix factorization, collaborative deep learning to incorporate textual reviews), context aware models (factorization machines to incorporate context of users and items). Computed additional metrics beyond accuracy such as coverage, novelty, diversity, serendipity to measure model performance
- Developed a dashboard for providing model recommendations to users along with explanations