

Shaun Yap

Contact: shaunyap01@gmail.com | (44) 7733933549 | linkedin.com/in/shaunyap0122/ | shaunyap01.github.io

CORE COMPETENCIES

Extensively utilised R, Python and SQL for the past 5+ years; deep familiarity with model capabilities and limitations, and effective applications to various data science and statistical investigations.

Programming languages: R, Python, Julia, Tableau, Java, SQL, NoSQL, Node, React, CSS.

Software: Git, Github, Markdown, LaTeX, Microsoft Excel, Visual Studio Code.

Mathematics, Statistics, Data Science, Algorithms, Statistical Modeling, Machine Learning, Computational Modeling, Model Validation, Feature Engineering, Statistical Inference, Quantitative Analysis, Natural Language Processing (NLP), Deep Learning, Neural Networks, Large Language Models (LLMs), Database Management.

WORK EXPERIENCE

PhD Researcher (Visual System Vulnerability in Dementia) – University College London (UCL) London, UK	Sep 2025 – Present
4-year PhD Research Studentship fully funded by the EPSRC and The National Brain Appeal as part of the Tech4Health Doctoral Training Programme in Digital Health Technologies at University College London (UCL).	
<ul style="list-style-type: none">Working with the UCL Queen Square Institute of Neurology (UCL QS IoN) and the UCL Hawkes Institute, under the supervision of Dr Keir Yong and Prof Andre Altmann.Research on visual system vulnerability in dementia, combining neuroscience, data science and genomics to improve detection and understanding of visual-led Alzheimer's disease.International collaboration (UCL × University of Colorado Anschutz): Supported cross-cohort replication of posterior cortical atrophy (PCA) disease-staging analyses by applying and validating cognitive Event-Based Models (EBM) on new Colorado baseline data, benchmarking model-derived event sequences against the PCA/tAD/control reference dataset (Firth et al., 2019) and running sensitivity analyses to assess the impact of control-cohort missingness thresholds on staging robustness.	
Postgraduate Teaching Assistant (PGTA), Statistical and Geospatial Data Analysis	Jan 2026 – Present
<ul style="list-style-type: none">Led computer labs; formative marking; summative second marking; office hours.	
Data Scientist (R&D) – Bourn.ai London, UK	Jul 2025 – Oct 2025
<ul style="list-style-type: none">AI fraud detection research and implementation.	
Database Intern – EPIC Homes Kuala Lumpur, Malaysia	Jul 2019 – Aug 2019
<ul style="list-style-type: none">Oversaw EPIC's database management operations, contributed to strategic planning efforts for the construction of green Epic Homes. EPIC is an award-winning social enterprise based in Malaysia.	
Investment Banking Intern – AmBank Kuala Lumpur, Malaysia	Jan 2019 – Jan 2019
<ul style="list-style-type: none">Internship with AmBank's Investment Banking Treasury Group based in Malaysia. Focused on credit rating analysis of specific quasi-government bonds.	

EDUCATION

PhD Researcher Digital Health Technologies – University College London UK	Sep 2025 – Present
MSc Data Science & Statistics (Distinction, 80% Avg.) – University of Exeter UK	Sep 2023 – Nov 2024
Emphasis on statistical inference and Bayesian modeling , including MCMC-based inference, Bayesian Gaussian Processes and spatio-temporal modeling .	
<ul style="list-style-type: none">Dissertation “Evaluating Environment & Climate Truthfulness in Social Media using Deep Learning & Large Language Models” - received University MSc Project Award (Best dissertation in cohort).Focused on feature engineering, natural language processing (NLP), text vectorisation, traditional machine learning algorithms, deep learning architectures and large language models (LLMs).	
BSc (Hons) Data Science – University of Warwick UK	Sep 2020 – Jul 2023
Used R extensively for university coursework - Python coupled with Pandas, NumPy, Pytorch for Machine Learning (ML).	
<ul style="list-style-type: none">In dissertation project, contributed new methods, extensions & functionality (e.g. profile likelihood, regression modeling) to the M-estimation package using Julia, a high-level programming language which notably facilitates automatic differentiation. Leveraged dual numbers to support infinitesimals, enabling efficient calculation of complex gradient structures on large functions to optimise performance.	
A-Levels – Garden International School Malaysia	Jul 2018 – Jun 2020
Mathematics – A*, Further Mathematics – A*, Chemistry – A, Level 3 Extended Project – A.	
<ul style="list-style-type: none">UK Mathematics Trust, Senior Mathematical Challenge – Gold (top 10%), 2018 & 2019.	
CERTIFICATIONS	
IBM Machine Learning Professional Certificate, 2024	
<ul style="list-style-type: none">Developed practical skills in Supervised, Unsupervised, Deep Learning and Reinforcement Learning. Focus on Time Series and Survival Analysis through a 3-month program covering algorithms and best practices in machine learning.	