

# Yeung-Yeung LEUNG, Ph.D.

 [github.com/oyyleung](https://github.com/oyyleung)

 [mail@leungyeungyeung.com](mailto:mail@leungyeungyeung.com)

 [www.leungyeungyeung.com](http://www.leungyeungyeung.com)

 [linkedin.com/in/yeungyeungleung/](https://www.linkedin.com/in/yeungyeungleung/)



## SKILLS

### Programming

- Python
- R
- Matlab
- Unix
- SQL

### Statistical

- SVD, PCA, ICA, matrix factor analysis, mixed models, permutation test

### Machine learning general

- data augmentation, Bayesian optimization, grid/random search, XGBoost, gaussian process, SHAP

### NLP

- Web scrapping, transformer, BART, BERT, LLM, LangChain, RAG, attention unrolling

### Time serious

- fourier transform, dynamical system, ARIMA, VAR, LSTM, GRU, transformer, 1D-CNN

### Image

- image augmentations, noise filtering, alignment, morphological operations, segmentation, 3D rigid transform, VGG-16, VAE

## LANGUAGES

- Cantonese - Native
- Mandarin - Professional
- English - Professional
- German - Intermediate

## CERTIFICATES

- Grade 8 Piano
- Driving License German Class B

## NATIONALITY

- Hong Kong SAR
- UK British

## EXPERIENCE

### Data Scientist - Omics data - Machine learning application, GSK 2023 - present

- Develop ML pipelines for target nomination, causal network inference and *in silico* perturbation
  - Design and establish workflow that integrate different ML tools and multi-omics readouts for target nomination, *in silico* perturbation and mechanistic inference
- Design, Benchmark and optimization of statistical, LLMs and generative models
  - Dynamical system, transformer, VAE, GNN, LLMs evaluation of predictive output

### Research Associate, Imperial College London 2021 - 2023

- Annotation of patient disease profile via natural language processing of clinical documents
  - Optical character recognition, RegEx, pre-trained Bert text embedding
- Genome wide association study of neurodegenerative diseases
  - Screening for causal genetic variants using local genotype data consisting 2.6 millions Single Nucleotide Polymorphisms, public eQTL data and tissue transcriptomics (Linear Mixed Effect Model, LDSC, MAGMA, Bayes Factor colocalization, WGCNA)
  - Designed a browser application for users to query genotypes information (R shiny)

### Principal Investigator, The Alan Turing Institute Data Science Study Group 2021-2023

- Modelling Amyloid Beta Plaque Formation in Alzheimer's Disease using VGG-16 and VAE
  - Facilitate the project delivery of a team of 10 participants
  - Spatial transcriptomics and pathological image feature extraction
  - Image segmentation, pre-trained VGG-16 embedding, VAEs training and optimization

### AI society NLP group project, University College London 2020-2021

- Explore machine learning methods for crypto currency price prediction.
  - Web scrapping, social media sentiment analysis, TFIDF, DistilBART text summarization

### Director, London Postdoc Network 2020 - Present

- Website establishment, organization of medium-scale (300+audiences) virtual events

### Research Associate, University College London 2018 - 2020

- Identify the dynamic features in brain oscillation that influence motor-behavioral outcome
  - Time frequency spectral analysis, linear mixed effects models, permutation statistics
  - Generate recurrent neural networks (LSTM, GRU) and Bayesian optimized XGBoost models
  - Using SHAP approach to evaluate the key determinants that contribute model's decision
- Generated a data processing pipeline for aligning neurophysiological and behavioral data
  - Multi-sources analogue data conversion, alignment, filtering and noise removal
  - Neuronal spike sorting: filtering, SVD whitening, PCA, superparamagnetic clustering
- 3D registration of MRI brain image and stereotaxic coordinates for recording location estimation
- Performed brain implants and electromyography surgeries as the second surgeon

## EDUCATION

### Computational Neuroscience (PhD), University of Heidelberg, Germany 2010 - 2016

- Grade: "magna cum laude"
- Investigation of a novel molecular pathway in synaptic vesicle release
- Validation of a mathematical model of 3D calcium dynamics in synaptic boutons
  - Designed a high-throughput pipeline for quantitative time serious image analysis
  - image alignment, noise filtering, artifact removal, automatic ROI detection

### Physiology (MPhil), The Chinese University of Hong Kong 2008 - 2010

- Multi-electrode field potential measurement of memory formation in rat brain slices
- Teaching of 40-60 medical students at practical laboratory session

### Research Intern, Toronto Western Hospital, University of Toronto 2006 - 2006

- Genetic association study: PCR optimization and analysis of sequencing results

### University exchange student, University of Groningen, The Netherlands 2006 - 2006

### Molecular Biotechnology (B.S.), The Chinese University of Hong Kong 2004 - 2007

## SCHOLARSHIPS AND AWARDS

- Academic Merit, Shaw College, The Chinese University of Hong Kong 2006
- Swire University Scholarship, John Swire & Sons (Hong Kong) Limited 2004 - 2007