

Dinghao Luo

Data Scientist | Experimentation, Causal Inference & Applied Machine Learning | Python, SQL, R
dinghao.luo@outlook.com | dinghaoluo.github.io | github.com/dinghaoluo | linkedin.com/in/dinghaoluo/
Relocating to UK/EU · Available January 2027

Data scientist and neuroscience PhD candidate with six years of experience building reproducible analytical pipelines and machine-learning tools for high-dimensional time series and images. Developed open-source PyTorch software, leakage-safe ML evaluation workflows, and Python pipelines processing large-scale scientific data. Strong background in experimental design, causal inference and technical communication.

SELECTED PROJECTS

FibreSight · github.com/dinghaoluo/fibre-sight

- Built and open-sourced a PyTorch U-Net and Qt desktop application for automatic identification of neural fibres in microscopy images, trained on 1,296 curated ROIs across 53 imaging sessions; achieved 0.93 recall and 0.83 F2 across nine held-out sessions and reduced manual curation time from days to hours.
- Shipped reproducible training, evaluation, and prediction workflows with CLI tools, automated tests and continuous integration through GitHub Actions, example data, and a detailed model card.

Kairos · github.com/dinghaoluo/kairos

- Built a Python pipeline using OpenAlex data to test whether historical publication patterns could predict imminent breakthroughs in science (inspired by Park *et al.* [*Nature*, 2023]’s debate on scientific novelty).
- Reviewed candidate papers to build breakthrough labels; evaluated leakage-safe models using expanding window validation and temporal baselines; found a modest improvement beyond a year-only baseline.

EXPERIENCE

Graduate Researcher · Max Planck Florida Institute for Neuroscience

May 2021–Dec 2026

- Designed controlled interventions across more than 100 animal recording sessions, estimating treatment effects on neural population dynamics during goal-directed behaviour.
- Built reproducible Python pipelines processing more than 50 TB of electrophysiology and two-photon imaging data, including time-series analysis, dimensionality reduction, and population-level modelling.
- Applied GLMs, matching, and permutation inference to distinguish modulation of an ongoing computation from initiation of the computation, overturning the prior hypothesis.
- Open-sourced pipelines and documented the workflows for a first-author manuscript submitted to *Cell*.

Undergraduate Researcher · University of Cambridge, Paulsen Lab

July 2019–Apr 2020

- Designed electrophysiology experiments on dopamine and synaptic plasticity; First-Class thesis.

Research Intern · UCL Queen Square Institute of Neurology

July–Aug 2018

- Quantified biomarker activity in Down syndrome models; co-authored peer-reviewed paper.

CORE SKILLS

Programming	Python (pandas), SQL (PostgreSQL), R (tidyverse, survival), MATLAB
Machine learning	PyTorch, scikit-learn; computer vision, classification, clustering, dimensionality reduction
Statistics	statsmodels; GLMs, Bayesian decoding, time-series analysis, permutation inference
Experimentation	Controlled interventions, within-subject and repeated-measures designs, power analysis
Causal methods	Matching, propensity scores, difference-in-differences
Engineering	Git, automated testing, CLI and Python package development, GPU-accelerated workflows

EDUCATION

PhD Candidate, Systems Neuroscience · FAU / Max Planck Florida Institute for Neuroscience

2021–2026

- IMPRS Fellow; defence scheduled for October 2026.

BA Biological Natural Sciences, First-Class Honours · University of Cambridge

2017–2020

- Jardine Scholar (full merit funding); Bundy Scholarship.

OTHER

Translator of *Entangled Life* (Sheldrake, 2020), Chinese edition 《菌络万象》; Pingshan Book Award 2025.
Former science editor, *Scientific American China* (2020–2021) and *Neu-reality* (2017–2021).
Native Cantonese and Mandarin; fluent English; conversational German.