Xuelong An

anwangxuelong@outlook.com @awxlong
LinkedIn GitHub Class Notes aithos
https://awxlong.github.io

Synopsis

I'm a proactive deep learning researcher inspired leading advocate of digital medicine Dr. Eric Topol's assertion that "AI is here to assist, not replace, doctors so they can reconnect with patients." I have a versatile skillset encompassing large scale biomedical data analysis, software engineering addressing healthcare problems and competitive machine learning.

Education

2023 – 2024 **University College London - Master of Science with Distinction** in Artificial Intelligence for Biomedicine and Healthcare.
• **Thesis**: **Unveiling Colon Cancer Therapy Resistance through Multiple-Instance Learning in Histopathology Images**, supervised by Dr. Petru Manescu, Dr. Pan Shi and Dr. Maria Secrier (Grade: 80%).
• Synopsis: We built a robust classifier of therapy-resistant cells in colorectal cancer whole slide images, achieving a test AUROC of 0.829 and F1 score of 0.724.

2019 – 2023 **University of Edinburgh - Master of Arts with Honours with 2:1 Class Division** in Cognitive Science.
• **Thesis**: **Charting the Landscape of Neuro-Symbolic Reasoners**, supervised by Dr. Antonio Vergari (Grade: 78%).
• Synopsis: We critically reviewed the limitations of deep learning, including their black-box nature and lack of robustness, and how to address them via explicit knowledge representation and integration of inductive biases.
• **Outstanding modules**: Informatics 1 - Object Oriented Programming (76%), Intro. to Linear Algebra (84%), Biomedical Sciences 2 (77%), Informatics 2D - Reasoning and Agents (70%), Machine Learning Practical (77%), Reinforcement Learning (83%)

2013 – 2019 **American School of Guayaquil - Bachiller de Ciencias (Science Baccalaureate)**
• International Baccalaureate Diploma 38/45

Software

2025 — **YachayMed**: I built a multilingual transcription system for doctor-patient conversations to support my colleagues who study medicine streamline their paperwork. Audio-to-speech is done by OpenAI's Whisper. Corrections and diarization are handled by a quantized medical LLM called Apollo-7b. Named entity recognition of medical concepts is done with the semantic NLP tool called MedSpaner. Code is written with Python and interface with Flask.

2024-2025 — **PathoInsightMIL**: I extensively improved my supervisor's package, HistoMIL, for analyzing gigapixel whole slide images. It supports the use of foundation models from Hugging Face. Training, validation and testing are monitored using Weights & Biases. We showcase a demo with Python's Dash. Code is written in PyTorch Lightning.

InQuery: I built a text-to-SQL app to support clinicians unfamiliar with SQL navigate a medical database with a customizable schema. We use the GPT2-based, state-of-the art LLM for text-to-SQL translation called SQLCoder-7b. Code is written in Dart to support cross-platform (IOS/Android) compilation, widening accesibility.

Research & Engineering Experience

01/2025 – 03/2025 — **Machine Learning Hackathon**

• **CIBMTR @ Kaggle (ongoing):** Leading a team in building machine learning models to ensure racial equity in post-hematopoietic stem-cell transplantation survival prediction.
• Using XGBoost, CatBoost and LightGBM libraries to train powerful, robust ensembles in Kaggle's platform leveraging P100 and T4 GPUs.
• Bayesian optimization using Optuna for hyperparameter tuning.
• We achieved a stratified concordance index of 0.692 (top score is 0.701).

03/2024 - 05/2024 — **MEDIQA-M3G@ ClinicalNLP 2024**

• Developed a framework for collaborative visual question answering
• Employed multiple language models and performed parameter-efficient fine-tuning.
• Employed quantized Mistral-7b at inference time to summarize the responses of multiple language models.
• Code is open-sourced at https://github.com/awxlong/manifold-medvqa/tree/main

01/2024 - 03/2024 — **Machine Learning Research Project**

• Developed a comprehensive machine learning pipeline to analyze multimodal brain anatomical data to classify schizophrenia. This UCL module is taught by Dr. Janaina Mourao-Miranda.
• Extensive benchmarking through cross-validation and hyperparameter tuning via Bayesian optimization yielded a robust logistic regressor which can achieve a validation F1-score of 0.77 ± 0.07.
• Wrote a paper-style report which earned a perfect mark of 100%

Research & Engineering Experience (continued)

01/2024 - 02/2024 **Machine Learning Optimization Project**

• Greatly optimized the Metropolis-Hastings algorithm for Markov chain Monte Carlo (MCMC) for the multispecies coalescent model so that it can run fast on a laptop. This UCL module is taught by Dr. Thomas Flouris and Prof. Ziheng Yang.

• My implementation achieved posterior estimates for 20000 MCMC iterations for ≈ 14000 loci in under 1 minute. As a baseline, Prof. Yang claimed the MCMC under these settings should finish in ≈ 10 minutes

• Wrote a scientific report which earned a extremely high mark of 94%

01/2023 - 05/2023 **Robust Deep Learning Research Project**

• Led a team on benchmarking deep learning models against neurosymbolic variants on image classification. This Edinburgh module is taught by Dr. Hakan Bilen.

• We found that neurosymbolic models are robust to ubiquitous data complications such as noise, class imbalance, small training data and mislabels, compared to a deep learning baseline.

• Wrote a paper-style report which earned us a high-mark of 76%.

09/2021 – 12/2021 **Personal Blog**

• **Notes:RE**: Maintaining a multilingual personal blog on the applications, philosophy and education of AI. I focus on discussing the strengths and limitations of LLMs, NLP benchmarks and the intricacies of language.

Skills

Languages	Professional proficiency in Spanish, English and fluent in Chinese.
Programming	Python, Unix, R, Java, LaTeX, Dart, Jekyll, HTML, SQL, Haskell
ML/DL	PyTorch Lightning, high-performance computing (SLURM/Sun Grid Engine), probabilistic programming, Bayesian optimization
Python Libraries	`scikit-learn`, `pandas`, `numpy`, `seaborn`, `matplotlib`, `huggingface_hub`, `torch`