

YULONG LIN

UK / Singapore | +44 7493 876 287 / +65 9807 8969 | lin.yulong@gmail.com | yulonglin.com

EDUCATION

University of Cambridge	Bachelors and Masters, Computer Science	<i>Graduated 2023</i>
<ul style="list-style-type: none">Jardine Scholarship, selectively granted to 12 students across Asia (~\$275k). Also offered NUS Global Merit ScholarshipFirst Class with Distinction (Starred First), awarded to 0-2 students out of 130 in the cohort for consistently high scoresThesis ranked top 5 in cohort by creating a dynamic graph dataset that overcomes limitations of current datasetsIndependently managed Terabytes of noisy unstructured data with parallel data processing, data exploration, data selection, comprehensive testing via continuous integration, good code style and data/model versioning. Advanced the state-of-the-art in dynamic link prediction by 3% in AP and AUROC with a dynamic graph neural network		

WORK EXPERIENCE

Research Intern	Center for Human-Compatible AI (CHAI)	<i>Mar 2025–Present</i>
<ul style="list-style-type: none">Investigating compute trade-offs between attackers and defenders in LLM jailbreak scenarios to develop more realistic and efficient benchmarking methodologiesDesigning and implementing novel defence approaches that integrate strong white-box attackers capable of obfuscating activations against detection probesQuantifying robustness transfer between different attack vectors for various defence techniques, identifying generalisable defence characteristicsCollaborating with Scott Emmons to prepare findings for publication at ICLR 2026		
Machine Learning Engineer, LLMs	ByteDance	<i>May 2024–Apr 2025</i>
<ul style="list-style-type: none">Optimised LLM systems for one of the world's largest AI chatbot platforms (second only to ChatGPT by MAU)Developed agent scaffolding framework and ran experiments for LLMs and VLMs to align models with user preferences. Improved summarisation results in 63% of cases and increasing the daily average user count for summarisation by fivefoldOptimised and served large-scale CV and NLP systems across the company. Reduced LLM video call latency by 50% and services running on a particular GPU type by ~30-40%Automated ~20% of model deployment troubleshooting with long-context LLMs by using system information, logs and documentation		
Machine Learning Intern	Cohere	<i>Aug-Dec 2022</i>
<ul style="list-style-type: none">Increased the context length of generation and summarisation LLMs without sacrificing performance when most models only accepted 2k tokens. This involved preprocessing data, devising model architecture changes, and designing automated and human evaluationsMiscellaneous tasks, including reducing hallucination, implementing evaluations in internal tooling, and investigating LLM ensemblesWorked in a distributed team spanning the US, Europe and Asia by proactively communicating and working independently		
Research Intern	Center for Human-Compatible AI (CHAI)	<i>Jun-Sep 2021</i>
<ul style="list-style-type: none">Led the development of an image classification model without confident errors on selected classes for unrestricted adversarial examples (images manipulated without restrictions), by designing and implementing experiments (with Scott Emmons)Learnt to juggle multiple streams of a project simultaneously as experiments ran for days		
Software Engineering Intern	Amazon Web Services (AWS)	<i>Jun-Sep 2020</i>
<ul style="list-style-type: none">Reduced lead time for delivering new Kafka versions to customers from weeks to days (a key team goal) by developing a platform to automate ingestion of new version, which became a gold standard for other managed AWS servicesWorked across ~20 code repositories to design and implement changes, such as build logic involving multiple repositoriesSpearheaded project planning and learnt various technologies (e.g. several AWS services, Ruby, Bash, build tools), while in a significantly different timezone from all other team members		
Software Engineering Intern	Agency for Science, Technology and Research (A*STAR)	<i>Jun-Sep 2018</i>
<ul style="list-style-type: none">Created software IP for 3 patents filed, and published in top-ranked medical informatics journal JMIR		

PROJECTS

Machine Learning Implementations
<ul style="list-style-type: none">Conducted first study on graph transformer robustness, focusing on spurious features, domain shifts and structural noisesReproduced Sharpness-Aware Minimization (SAM), finding that SAM improves generalisability across various optimisers and tasks, and some evidence that SAM improves adversarial robustnessReproduced A Better Alternative to Error Feedback for Communication-Efficient Distributed Learning. Reproduced most experiment results, but found evidence contrary to some claims with additional experiments on other compressorsImplemented PyTorch API, GPT-2 (with RLHF and parallelism), BERT, RLHF, RL and interpretability algorithms at Redwood Research

SKILLS AND INTERESTS

- Technical skills: Deep learning (PyTorch/TensorFlow/JAX), Data (pandas, NumPy, scikit-learn, Matplotlib, Beautiful Soup), Shell scripting, CI/CD (CircleCI), Infrastructure as Code (CloudFormation, Ansible), Microservices (Docker, AWS Lambda, AWS Step Functions), Monitoring and Logging (Grafana), Build tools (Gradle), Databases (MySQL, DynamoDB, S3), Task scheduling (cron), Linear algebra, Calculus, Statistics
- Extracurriculars: Homeless aid (e.g. food distribution, befriending), Writing (tech/policy editor), Bouldering, Badminton (college captain)