Yunlong Jiao

Applied Machine Learning Research

London, UK (No Sponsorship Needed)
jyunlong@google.com
linkedin.com/in/yunlong-jiao/

ABOUT ME

I am currently a Research Engineer at Google DeepMind, where I focus on applied machine learning research to improve various Google products. With over 5 years of post-doctoral experience, I am skilled in machine learning model development and performance evaluation. I am proficient in Python and well-versed in several deep learning frameworks. My passion lies in building machine learning and AI solutions to create meaningful societal impact through real-world products and services.

EDUCATION

2013 – 2017 **Doctor of Philosophy**
Centre for Computational Biology, Paris Sciences et Lettres – PSL Research University, Paris, France

2012 – 2013 **Master of Science** (MENTION TRÈS BIEN)
Department of Mathematics, University of Paris-Saclay, Orsay, France

2008 – 2012 **Bachelor of Science** (FIRST CLASS HONOURS)
Department of Mathematics, University of Science & Technology of China, Hefei, China

WORK EXPERIENCE

Research Engineer
CURRENT, SINCE SEP 2023

Google DeepMind, London, UK

- **Key Skills:** Large Language Models, Multi-modality (Text/Vision/Audio/Video)
- Working on bringing next-generation machine learning and AI technologies to Google's wide portfolio of products.

Machine Learning Scientist
NOV 2019 – JUL 2023

Amazon, London, UK

- **Key Skills:** Natural Language Processing, Large Language Models, Deep Generative Models, Neural Text-to-Speech
- At Alexa Shopping Science team (Mar 2021 – Jul 2023), my main responsibilities were:
  1) Leading research on bias mitigation technologies in NLP to enhance fairness while preserving user privacy;
  2) Designing and implementing rigorous sampling and evaluation frameworks to assess the quality and limitations of LLMs at scale and across diverse tasks;
  3) Building and deploying ML solutions for customer satisfaction and unified user experiences for Alexa across devices;
  4) Supervising research internships and coaching and mentoring junior team members.
- At Alexa Text-to-Speech Research team (Nov 2019 – Feb 2021), my responsibilities included:
  1) Proposing and implementing a deep generative model for 'universal' speech synthesis regardless of voice identity or language, which can massively save cost in TTS production and speed up launch of new TTS voices;
  2) Collaborating with engineers to scale up Alexa TTS production and launch new voices in multiple regions and countries.

Postdoctoral Research Scientist
NOV 2017 – OCT 2019

University of Oxford, Oxford, UK

- **Key Skills:** Gaussian Processes, Time Series Forecasting, Multi-modality
- My responsibilities included:
  1) Leading Oxford research in a multi-organizational project (involving organizations in the UK and EU) and developing novel methods for longitudinal modelling of complex disease using multi-modal data integration;
  2) Supervising master's theses in the Department of Statistics.

Doctoral Researcher
SEP 2013 – SEP 2017

PSL Research University, Paris, France

- **Key Skills:** Kernel Methods, Representation Learning, Graph Learning, Sparsity Regularisation
- My PhD work contributed novel ML methods and advanced scientific discoveries in biology and cancer research:

1) My research focus was kernel methods and representation learning of non-tabular data, such as rank data and graphs.
2) My research outputs significantly improved prediction of breast cancer survival using genetic data, and guided interpretable biomarker discovery with the help of biological networks.
3) I published multiple first-authored papers in top ML conferences and journals (3 at ICML and 1 at IEEE TPAMI) and developed open-source toolkits.

■ A few colleagues and I participated, and finally placed 2nd, in the DREAM Toxicogenetics Challenge, a Kaggle-style community competition in data science that is aimed to advance computational methods in biology.

Data Scientist Intern APR 2015 – JUN 2015
Roche Diagnostics GmbH, Penzberg, Germany

■ **Key Skills:** Information Extraction, Feature Engineering, Large-Scale Unstructured Database
■ During the internship, I proposed a data pipeline to process large-scale unstructured machinery performance data and built a model to predict failure state for automated immunoassay analysers. My work demonstrated how maintenance efficiency can be greatly improved for one of Roche's core hardware products, and had a direct impact on strengthening customer trust. The innovative solution was patented by Roche.

TECHNICAL SKILLS

PROGRAMMING	Python (numpy, pandas, scikit-learn), Deep Learning (Jax, Tensorflow, PyTorch), R, C/C++, Bash
BIG DATA	Accelerated Computing (CUDA, HPC), Cloud Computing (AWS/SageMaker, GCP), Database (SQL)
DEVOPS	Git/Mercurial, Docker, Workflow Management (XManager, Airflow), Testing, CI/CD, Open Source

PROFESSIONAL SKILLS

COMMUNICATION	Experienced speaker at international conferences and workshops Confident in presenting project ideas and results to peers, leadership, and stakeholders Mentoring and coaching research interns and junior team members
WRITING	Proficient in academic writing and providing guidance to early-stage researchers in their writing Experienced in leading R&D proposals and producing technical reports on milestone deliveries
PROJECT MANAGEMENT	Knowledgeable in the principles of Agile development and skilled working in Scrum teams Efficient in planning, organising, and coordinating resources, tasks, and people to achieve goals
OPERATING APPROACH	Building trust through Radical Candor communications as the foundation of any relationships Capable tech lead and accountable owner in cross-functional team collaborations Prioritizing action and always seeking hands-on opportunities to learn how things work
LANGUAGES	Mandarin Chinese (native), English (bilingual), Spanish (conversational), French (reading)

SELECTED PUBLICATIONS AND PATENTS

F Tonolini, N Aletras, **Y Jiao**, G Kazai. "Robust Weak Supervision with Variational Auto-Encoders." *ICML*, 2023.

Z Shi, F Tonolini, N Aletras, E Yilmaz, G Kazai, **Y Jiao**. "Rethinking Semi-supervised Learning with Language Models." *Findings of ACL*, 2023.

Y Feng, **Y Jiao**, A Prasad, N Aletras, E Yilmaz, G Kazai. "Schema-Guided User Satisfaction Modeling for Task-Oriented Dialogues." *ACL*, 2023.

F Liu, **Y Jiao**, J Massiah, E Yilmaz, S Havrylov. "Trans-Encoder: Unsupervised Sentence-Pair Modelling Through Self- and Mutual-Distillations." *ICLR*, 2022.

A Gabryś, **Y Jiao**, V Klimkov, D Korzekwa, R Barra-Chicote. "Improving the Expressiveness of Neural Vocoding with Non-Affine Normalizing Flows." *Interspeech*, 2021.

Y Jiao, A Gabryś, G Tinchev, B Putrycz, D Korzekwa, V Klimkov. "Universal Neural Vocoding with Parallel WaveNet." *ICASSP*, 2021.

F Heinemann, S Kobel, S Dahlmanns, JP Vert, **Y Jiao**. "Failure State Prediction for Automated Analyzers for Analyzing a Biological Sample." *US Patent App.*, 2019.

Y Jiao, JP Vert. "The Weighted Kendall and High-order Kernels for Permutations." *ICML*, 2018.

Y Jiao, A Korba, E Sibony. "Controlling the Distance to a Kemeny Consensus without Computing It." *ICML*, 2016.

Y Jiao, JP Vert. "The Kendall and Mallows Kernels for Permutations." *ICML*, 2015.