

QINGFEI YOU

CONTACT INFORMATION

Phone: +44 7592 115813
Email: qingfeiyou@gmail.com
Homepage: <https://yhgzxxfz.github.io>
Linkedin: <https://www.linkedin.com/in/qingfeiyou/>

EXPERIENCES

JULY 2022 - present	Research Engineer (FAIR - Meta AI) <i>Meta, London, England</i> Reinforcement Learning <ul style="list-style-type: none">Co-led the research on Generalization of Offline Reinforcement Learning (Offline RL). Built data pipeline to collect expert and suboptimal data; trained and evaluated Offline RL algorithms; compressed (x14) and open-sourced data (70GB) and code. Large Language Models (LLMs) <ul style="list-style-type: none">Prepared code dataset for LLaMa and CodeLLaMa pretraining, by processing / cleaning 100TB raw data. Explored different approaches from the community to close the gap for every processing step.Transformed Jupyter Notebooks (160GB) into 2 datasets (Scripts, Dialogs) for CodeLLaMa pretraining and RLHF fine-tuning.Reduced inference latency of T5-like LLMs by 2x (3B) to 8x (small) with NVIDIA's FasterTransformer; Benchmarked with +10 approaches from HuggingFace, NVIDIA, PyTorch, MicroSoft and Tencent.Productionized a T5-like language model (PEER) of 3B parameters on an AWS EC2 instance. Built a pipeline from scratch to reproduce deployment with 1 single docker command. Brain AI <ul style="list-style-type: none">Onboarded public neuroimaging datasets (EEG) under a common API for data quality and quantity analysis.
MAR 2020 - JUNE 2022	Senior Software Engineer (Integrity Experience) <i>Meta, London, England</i> <ul style="list-style-type: none">Designed and implemented a content management system from scratch, and led a small team to migrate existing and onboard new use cases, reducing average onboarding cost by 99% (3 days to 15 mins).Led a small group of engineers to migrate and aggregate contents, removed 9.8k (70%) duplicates, and saved 3600 hours and \$2.4M for translation.
MAR 2018 - FEB 2020	Software Engineer (Community Integrity) <i>Facebook, London, England</i> <ul style="list-style-type: none">Increased overall appeal coverage by 160% by expanding appeal workflow to Spam violations and bad URLs on Facebook.Saved human review resources by half by migrating appeals to new workflow and removing redundant middle steps.
SEPT 2014 - DEC 2017	Software Engineer (E-Business Trading) <i>Société Générale CIB, Paris, France</i> <ul style="list-style-type: none">Expanded e-trading scope by 6x, from futures only to options, forwards, spots, swaps, averages and time-spreads.Reduced e-trading latency to be <200ms by adding cache and optimizing workflow.
MAR 2013 - AUG 2013	Research Intern (CTO's team - Wind Turbine Blade Inspection) <i>Cornis, Paris, France</i> <ul style="list-style-type: none">Designed and implemented multiple image correlation algorithms from scratch in C++ to construct the full blade automatically, reaching 82% accuracy compared to human level.Optimized algorithm latency by 14x using Fast Fourier Transform, convolution, and integral images.

PREPRINTS

- [1] Timo Schick, Jane Dwivedi-Yu, Zhengbao Jiang, Fabio Petroni, Patrick Lewis, Gautier Izacard, **Qingfei You**, Christoforos Nalmpantis, Edouard Grave, Sebastian Riedel. (2022). [Peer: A collaborative language model](#). *arXiv:2208.11663*.

PUBLICATIONS

- [1] Ishita Mediratta*, **Qingfei You***, Minqi Jiang, Roberta Raileanu. (2024). [The Generalization Gap in Offline Reinforcement Learning](#). In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*.
- [2] Ishita Mediratta*, **Qingfei You***, Minqi Jiang, Roberta Raileanu. (2023). [A Study of Generalization in Offline Reinforcement Learning](#). In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS)*.

PROJECTS

- DEC 2023 Andrej Karpathy's [Neural Networks: Zero to Hero](#)
- MAR 2024
 - [micrograd](#): Built a PyTorch-like engine with autograd to train neural networks.
 - [makemore](#): Implemented a set of language models (BiGram, MLP, WaveNet) to generate names.
 - [nanoGPT](#): Built GPT from scratch to generate Shakespeare.
 - [tokenizer](#): Built and trained Byte-Pair Encoding (BPE) algorithm.

EDUCATION

- AUG 2011 **M.S.E. in Computer Science**
- SEPT 2013 *Télécom ParisTech, Paris, France*
Department of Computer Science

- SEPT 2007 **B.Eng. in Electronics Science & Technology**
- JUNE 2011 *Southeast University, Nanjing, China*
Department of Electronics Science & Engineering
Advisor: Prof. Yiping Cui
GPA: 92/100 (4th of 184)

CERTIFICATES

- 2018 [Deep Learning Specialization](#) (Coursera)
- 2017 [Neural Networks for Machine Learning](#) (Coursera)
- 2016 [Machine Learning](#) (Coursera)

HONORS & AWARDS

- 2010 Meritorious Prize in International Mathematical Contest in Modeling (IMCM)
- 2009 National Scholarship (top 3%)
- 2008 National Scholarship (top 1%)