

KYLE GAO

Tel:+1 412-892-0599 Email:kyle.ygao@gmail.com

EXPERIENCE

BenevolentAI

May.2018 - Present

Director of ML and Data Engineering, NYC Office Site Lead

- As engineering lead, led and grew a cross-functional team from 5 to 40+ people to design and develop BenevolentAI's Target Identification platform, which includes data pipelines, ML models and systems, API services, and web applications.
- As a manager, hired and managed ~15 individual contributors and managers across US and UK, including software engineers, ML engineers, and AI researchers. Standardized engineering hiring workflow, composed ML Engineer career level framework, and established retention risk mitigation program during 2021.
- Created and rolled out a company-wide ML infrastructure by seeking leadership sponsorship, recruiting team, and introducing 3rd party solutions. The infrastructure now supports 50+ users in 3 product areas with model training, model inference, and notebook workspace services.
- Established a data strategy to create a company-wide data platform and data governance program so that the efficiency and quality of data engineering, data science, and machine learning activities are improved.
- **Methodologies:** Data-centric AI, Domain-driven design, Data Mesh, Microservice, Servant leadership
- **Technologies:** AWS, Gitlab, Kubernetes, Docker, Git, Spark, Kubeflow, Databricks, Tableau, Python, Airflow, gRPC, GraphQL, REST, PyTorch, RDBMS, ElasticSearch, MongoDB, RabbitMQ

IBM Research

Sept.2016 - May.2018

Research Software Engineer

- Worked at Reasoning System group and Knowledge Induction group on research projects and client solutions using deep machine learning based semantic parsing, knowledge base completion, and semantic graph.
- **Publications:**
 - Gao, Kyle Yingkai, et al. "Interpretable Drug Target Prediction Using Deep Neural Representation." IJCAI. 2018.
 - IBM/pytorch-seq2seq package, a sequence-to-sequence framework in PyTorch.

Carnegie Mellon University

Aug.2014 - Aug.2016

Research Assistant

- Explored through a prototype the possibility of non-textual data retrieval. Tables from arXiv publications were extracted and indexed. Studied novel representation and enrichment for both document and query sides, explored user interface specifically for scientific search, and built a dataset with careful user study.
- **Publications:**
 - **Gao, Yingkai**, Jamie Callan, "Scientific Table Search Using Keyword Queries." arXiv:1707.03423.
 - Ma, Xuezhe, **Yingkai Gao**, Zhiting Hu, Yaoliang Yu, Yuntian Deng, and Eduard Hovy. "Dropout with expectation-linear regularization." ICLR'2017.
 - Hu, Zhiting, Poyao Huang, Yuntian Deng, **Yingkai Gao**, and Eric P. Xing. "Entity Hierarchy Embedding." ACL'2015.

Microsoft Research Asia

Apr.2013 - Jan.2014

Intern Research Assistant

- Proposed a weakly-supervised clustering framework that (a) automatically collects various expressions on attributes from query log and (b) ground the expressions to the attributes defined in a knowledge graph.

EDUCATION

Master of Language Technologies, Carnegie Mellon University

Aug.2014 - Aug.2016

Language Technologies Institute

Advised by Jamie Callan

Bachelor of Science in Computer Science, Beihang University

Sept.2010 - Jul.2014

School of Computer Science & Engineering

PROFESSIONAL SERVICE

- Reviewer of DSHealth KDD, Briefings of Bioinformatics, ICTIR