
Skills

Programming Languages: Python, SQL, Java, C/C++, PHP

Frameworks: PyTorch, TensorFlow, Keras, XGBoost, Spark

Platforms: Amazon Web Services (AWS), Google Cloud Platform, Microsoft Azure, Databricks, Snowflake

Experience
Senior Data Scientist, GoPuff

Jan 2022 – present

- **Query Classification:** Applied transformers deep learning model to resolve query-indicated product categories. Extended search results with relevant products. Boosted average order volume by 0.8%
- **Query Entity Tagging:** Led company negotiation effort for 3rd party annotation partnership. Trained transformer-based entity tagging model subsequently. Achieved 88% f1 score on query traffic
- **Product Ranker:** Designed XGBoost model for personalized product ranking. Applied Spark to process terabytes of features. Increased product recall by 15%

Data Scientist II, LivePerson

Nov 2020 – Jan 2022

- **Multilingual Text Classification:** Built deep learning systems that understand human intents in eight languages. Outperformed Watson and DialogFlow by 5% on conversational data
- **Entity Recognition:** Applied transformer-based model to resolve general entities in conversations. Achieved 88% f1 score on more than 20 entities. Outperforming Spacy by 13% on conversational data
- **Few-Shot Learning:** Designed contrastive-learning methods to solve entity tagging tasks with only five labels per entity type. Outperform previous state-of-the-art by 2% on these data scarce tasks

Software Development Engineer, Amazon

Apr 2020 – Nov 2020

- **Machine Learning Platform:** Developed privacy-centric machine learning platform for model training and releasing with AWS serverless: AWS Lambda, SQS, S3, and DynamoDB

Natural Language Understanding (NLU) Scientist, LivePerson

Nov 2018 – Apr 2020

- **Text Classification:** Created deep learning NLU library for intent detection and text classification. Achieved fast inference speed by searching among model architectures. Increased accuracy by 5%
- **Anomaly Detection:** Improved anomaly detection algorithm to identify out-of-topic content. Replaced former algorithm in intent classification and increased accuracy by 2%

Awards

Gold Medal, Chinese Physics Olympiad (2011): 51 winners in China among thousands of competitors

Publications

Domain-Specific Knowledge Distillation Yields Smaller and Better Models for Conversational Commerce. Kristen Howell, **Jian Wang**, Akshay Hazare, Joseph Bradley, Chris Brew, Xi Chen, Matthew T. Dunn, Beth Ann Hockey, Andrew Maurer, Dominic Widdows. *e-Commerce and NLP (ECNLP)*, 2022.

Think Visually: Question Answering through Virtual Imagery. Ankit Goyal, **Jian Wang**, Jia Deng. *Association for Computational Linguistics (ACL)*, 2018.

Premise Selection for Theorem Proving by Deep Graph Embedding. Mingzhe Wang, Yihe Tang, **Jian Wang** Jia Deng. *Neural Information Processing Systems (NeurIPS)*, 2017.

Education
University of Michigan

Master of Science in Computer Science GPA: 3.87

Ann Arbor, MI

Sep 2015 – Aug 2018

Peking University

Bachelor of Science in Physics GPA: 3.73

Beijing, China

Sep 2011 – Jun 2015