Hongdao Meng

mycrofthd@gmail.com | Github | Linkedin | (718) 3063737| New York City

EDUCATION

New York University Sep. 2024 - May 2026 (Expected)
Master of Science in Computer Science | GPA: 3.8 New York, NY

Beijing University of Technology Sep. 2020 - Jul. 2024
Bachelor of Engineering in Information Security | GPA: 4.00 Beijing, CN

PROFESSIONAL & RESEARCH EXPERIENCE

TikTok, Recommendation System *May 2025 — Sep. 2025 & Jan. 2026 — Present*
Machine Learning Engineer Intern *Seattle, WA*
- Developed a Python-based **DAG-DSL** migration toolkit (leveraging **Protobuf** and **regex**) to auto-convert legacy configuration files of the **Bytedance Feature Service (BFS)**, a scalable feature extraction service powering TikTok's recommendation system, into standardized DSL definitions, boosting operator migration throughput by 85%.
- Designed and implemented a **Blade**-based **C++** operator migration pipeline for BFS, streamlining the workflow to improve operator migration speed by 83.3%, and personally migrated 20 operators.
- Architected a one-stop workflow for User Data Accessor covering service build, auto-generate **Python** operators, **DSL** compilation, and **RPC** request, enabling collaboration dozens of engineers and accelerating development by 73.3%.
- Led the migration of TTmall BFS, owning migration planning and dependency decomposition.
- Designed and executed pre/post-migration **Diff validation** and **A/B parity experiments**, achieving metric alignment.

C2SMART Center *Jan. 2025 — Present*
Machine Learning Engineer Intern *New York, NY*
- Led a 6-member team to develop a RAG-based chatbot system using **LangChain** and **Flask-React**, achieving 24.3% accuracy improvement on the MS MARCO (F1=0.86) and 33.7% faster response latency through query optimization.
- Built a hybrid retrieval framework with **MilvusDB** vector database and **BGE-M3** embeddings, improving search relevance by 21.7% and boosting query performance by 25.6% through a unified re-ranking architecture.
- Implemented a **Docker**-based data pipeline with **MongoDB** on **AWS EC2**, reducing deployment setup time by 15.7% and deploying **CI/CD** pipelines with **Jenkins** to ensure high availability.

DeepFake Detection Startup *Sep. 2024 — Dec. 2024*
Machine Learning Engineer & Founder *New York, NY*
- Led a 5-member team to develop core modules of a DeepFake detection web platform using **React** and **TypeScript**, enabling 1,200+ concurrent users and reducing client-side rendering latency by 21.3%.
- Fine-tuned **Vision Transformer** models from **Hugging Face** for image and audio deepfake detection, achieving 91.2% and 88.1% accuracy. Built datasets with **OpenCV** and **FFmpeg**, and deployed models on **AWS EC2**.
- Built a real-time communication layer using **Django** and **WebSocket**, reducing task completion time by 25.6%, and deployed backend services on **Kubernetes** with **AWS ELB**, achieving 99.5% availability under 5k RPM.
- Optimized **PostgreSQL** query execution via composite index tuning, reducing average response time by 18%.

Data Mining & Security Lab *Sep. 2022 — Jul. 2024*
Machine Learning Engineer Intern *Beijing, China*
- Led research on **Federated Multi-View Multi-Label Learning (FMVML)**. Published first-author paper in *IEEE Transactions on Big Data* (2025): "Federated Multi-View Multi-Label Classification" (DOI: 10.1109/TBDATA.2024.3522812).
- Proposed the FMVML framework enabling cross-view feature fusion and multi-label semantic classification, outper-forming state-of-the-art methods with +8.3% Average Precision and -14% One Error.
- Utilized **Python**/**PyTorch** for model development and **Matlab** for signal processing; implemented data pipelines with **Pandas**/**Scikit-Learn**; prepared publication-ready documents with LaTeX.

SKILLS

Languages: C/C++, Java, Python, Go, SQL, JavaScript, HTML/CSS, Shell, PHP, LaTeX
Machine Learning: PyTorch, LangChain, Tensorflow, Pandas, Scikit-Learn, VGG16, ViT, Multi-View, Multi-Label
Frameworks: Blade, React, Angular, Vue.js, Django, Flask, Node.js, Spring Boot, Hadoop, Spark
Database: MySQL, Redis, MongoDB, PostgreSQL, DynamoDB, Oracle, Firebase, RocketMQ, Elasticsearch, MilvusDB
Tools: Git, Docker, AWS, Azure, CMake, Postman, CI/CD, Jenkins, Nginx, FFmpeg, OpenCV, Jira, Figma

PUBLICATION

- First Author: "Federated Multi-View Multi-Label Classification." *IEEE Transactions on Big Data*, 2025.
- Co-Author: "Susceptibility genes identification and risk evaluation model construction by transcriptome-wide association analysis for salt sensitivity of blood pressure: the EpiSS study." *BMC Genomics*, 2024.