

# Hongdao Meng

✉ mycrofthd@gmail.com | 🌐 github.com/Mycroft-s | 🔗 linkedin.com/in/hongdao-meng-70222b306 | 📞 (718) 3063737

## EDUCATION

---

### New York University

Master of Science in Computer Science GPA: 3.8

Sep. 2024 - May 2026 (Expected)

New York, NY

### Beijing University of Technology

Bachelor of Engineering in Information Security GPA: 4.0

Sep. 2020 - Jul. 2024

Beijing, CN

## SKILLS

---

**Languages:** Java, Python, C/C++, Go, SQL, JavaScript, HTML/CSS, Shell, PHP,  $\LaTeX$

**Frameworks:** React, Angular, Vue.js, Django, Flask, Node.js, Spring Boot, PyTorch, TensorFlow, Pandas, Scikit-Learn

**Database:** MySQL, Redis, MongoDB, PostgreSQL, DynamoDB, Oracle, Firebase, RocketMQ, Elasticsearch, MilvusDB

**Tools:** Git, Docker, AWS, Azure, CMake, Postman, CI/CD, Jenkins, Nginx, LangChain, FFmpeg, OpenCV, Jira, Figma

## PROFESSIONAL & RESEARCH EXPERIENCE

---

### *Machine Learning Engineer Intern @ C2SMARTER Center, New York*

Jan. 2025 - Present

- Engineered an interactive chatbot using **Python, Flask, React, Docker, Redis, ChromaDB**, and **RAG**; led the finetuning of the **Mistral 7B** model to specialize it for ITS (Intelligent Transportation Systems) project content, enabling semantic search across domain-specific documents and contributing to a USDOT-aligned knowledge base
- Established a distributed system architecture that integrates LLM-based question answering with advanced retrieval pipelines, improving user query relevance and reducing response latency by **23.5%**
- Implemented **CI/CD** workflows on **AWS** (EC2, RDS) and containerized services with **Docker**, ensuring scalability and high availability for 2k+ daily active users

### *Machine Learning Engineer and Founder @ DeepFake Detection Startup, New York*

Sep. 2024 - Dec. 2024

- Led 5-member team to develop core modules of deepfake detection web platform using **React** and **TypeScript** for seamless real-time interaction, enabling 1,200+ concurrent users and reducing client-side rendering latency by **21.3%**
- Fine-tuned **Vision Transformer** and **VGG16** models from **Hugging Face** for image and audio DeepFake detection, achieving **91.2%** and **88.1%** accuracy, respectively. Constructed two custom datasets from the MP4 FaceForensics corpus by extracting video frames with OpenCV and audio files with FFmpeg, and deployed both models on **AWS EC2**
- Built a real-time communication layer using **Django** and **WebSocket** for robust middleware communication, reducing task completion time by **25.6%** with <180ms P95 latency, and deployed backend services on **Kubernetes (AWS EKS)** with **AWS ELB** load balancing and **HPA policies**, achieving **99.5%** availability under 5k RPM
- Optimized **PostgreSQL** query execution through composite index tuning, reducing average response time by 18% (320ms→262ms)

### *Machine Learning Engineer Intern @ QingTeng, Cloud Platform R&D Department*

Feb. 2024 - Aug. 2024

- Led 6-member team to develop **RAG**-based chatbot system using **LangChain** and **Flask-React**, achieving **24.3%** accuracy improvement on MS MARCO dataset (F1=0.86) with **33.7%** faster response latency through query optimization
- Built hybrid retrieval framework with **MilvusDB** vector database and **BGE-M3** embeddings, improving search relevance by **21.7%** and boosting query performance **25.6%** through unified re-ranking architecture
- Implemented a **Docker**-based data pipeline with **MongoDB** document storage on **AWS EC2**, reducing deployment setup time by 15.7% and deploying **CI/CD pipelines** with **Jenkins** to ensure high availability and automated testing, which resulted in an **18.7%** reduction in deployment cycles and supported a 2× increase in concurrent users from 5k to 10k.
- Implemented monitoring and observability with **Grafana** and **Prometheus**, providing real-time system insights and reducing mean time to resolution (MTTR) by **25.3%**
- Streamlined development and quality assurance processes with **Postman** for API testing and **GitLab** for version control, resolving 82% of integration issues pre-deployment and improving team collaboration efficiency by 22%

### *Machine Learning Engineer Intern @ Data Mining & Security Lab, Beijing*

Sep. 2022 - Jul. 2024

- Led research on Federated Learning and Multi-View Multi-Label Machine Learning, focusing on privacy-preserving feature fusion and multi-label classification. Published first-author paper in *IEEE Transactions on Big Data* 2025: "**Federated Multi-View Multi-Label Classification**" (DOI: [10.1109/TBDATA.2024.3522812](https://doi.org/10.1109/TBDATA.2024.3522812))
- Devised solutions to complex data privacy challenges by proposing and developing the **FMVML** framework, a federated learning method enabling cross-view feature fusion and multi-label semantic classification, which outperformed all state-of-the-art methods, improving Average Precision by **8.3%** and lowering One Error by **14%**
- Utilized **Python/PyTorch** for model development and **Matlab** for signal processing; implemented data pipelines with **Pandas/Scikit-Learn**; produced publication-ready documents with  $\LaTeX$