Tianshu Fu

+1 646 898 8948 | @ tf2502@columbia.edu

EDUCATION

- **Columbia University** — New York, NY
  Master of Statistics, GPA: 3.76 — *Sep. 2021 – Dec. 2022*
- **University of California, Berkeley** — Berkeley, CA
  Exchange Student in Data Science, GPA: 3.5 — *Aug. 2018 – Dec. 2018*
  Courses: *CS100, Stat 140*
- **Central China Normal University** — Wuhan, China
  Bachelor of Computer Science, GPA: 86.7/100 — *Sep. 2016 – Jun. 2020*
  Courses: *Data Structures, Algorithms, Databases*

EXPERIENCE

- **Meta** — Burlingame, CA
  Data Engineer — *Nov 2023 – Present*
  - **LLM-based Product Social Listening Tool**: Developed a **ChatGPT-style tool** that analyzes feedback for Meta Reality Labs (e.g., Ray-Ban, Quest) by aggregating bug reports, tickets, and public posts. **Featured in Meta's blog**: [Harnessing the Power of Customer Feedback](). This tool unlocked insights from qualitative feedback and has been used in various use cases such as real-time user sentiment tracking during Meta Connect 2024, identifying top reported issues for Meta Quest, gather top feature requests from social media, gather VR developer pain points
  - **LLM Taxonomy Tool**: Developed an **AI-powered tool** to assist with taxonomy management tasks. The tool automatically labels CSV records based on defined taxonomies and evaluates coverage gaps. It supported efforts like Meta Quest Store app genre classification, and enabled ML engineers to more easily prototype and test LLM prompts before formal experiments, saving 2–3 weeks of effort per half.

- **Knowtex** — San Francisco, CA
  Founding Software Engineer — *Jan 2023 – Nov 2023*
  - **AI-Powered SOAP Note System**: Built an AI-powered system to generate structured SOAP medical notes from doctor-patient conversations. Leveraged RAG (retrieval-augmented generation) to personalize note generation by referencing doctors' historical notes and learning their unique style and terminology.
  - **Customized LLM Model**: Fine-tuned an open-source LLM on production medical transcription data, achieving a 20% increase in SOAP note generation quality through domain-specific adaptation.
  - **App Development**: Built a cross-platform mobile app using **React Native** to enable doctors to record conversations and access generated notes. The iOS version is publicly available at: [Knowtex on App Store]().

- **Meta** — New York, NY
  Data Engineer Intern — *May 2022 – Aug 2022*
  - **Seller Retention Dashboard**: Built a data pipeline and dashboard to monitor seller retention trends and help PMs strategize for growth.

SKILLS

- **Languages**: Python, Hack, SQL, Go, Java, JavaScript
- **AI/ML Tools**: PyTorch, TensorFlow, RAG, LLaMA, HuggingFace Transformers, LangChain
- **Data Engineering**: Dataswarm, Presto, Spark, Hive, AWS, Docker, Kubernetes
- **Databases**: PostgreSQL, Tao, MySQL, DynamoDB, MongoDB, ElasticSearch, Neo4j
- **Web and Visualization**: React, React Native, Streamlit