# Minjiang (Ivy) Yang

Jacksonville, FL | your.email@example.com | (xxx) xxx-xxxx | LinkedIn/GitHub

## SUMMARY

Motivated Computer Science M.S. graduate (ASU, 2024) with one year of developer experience. Skilled in Python, data pipelines, and machine learning. Demonstrated ability to build and deploy clinical NLP + Knowledge Graph demos (Neo4j, FastAPI, Scikit-Learn). Passionate about biomedical data science and eager to contribute to impactful research at Mayo Clinic.

## TECHNICAL SKILLS

- Languages: Python, SQL, Java, C++  
- Data/ML: Pandas, NumPy, Scikit-Learn, TfidfVectorizer, Logistic Regression, NLTK  
- Knowledge Graph & Databases: Neo4j, py2neo, Cypher, PostgreSQL  
- APIs & Frameworks: FastAPI, Uvicorn, Docker  
- Tools: Git, Docker Desktop, Jupyter, Parquet, CSV, Windows PowerShell  
- Other: Data cleaning, feature engineering, basic frontend familiarity

## EDUCATION

M.S. Computer Science – Arizona State University, Online (May 2024)  
Coursework: Machine Learning, Data Mining, Distributed Systems, Database Systems, Software Verification & Validation

## PROFESSIONAL EXPERIENCE

Software Developer – Eth Tech (Intern/Contract) (2024–2025)  
- Developed distributed systems components using Python and Docker.  
- Integrated Kafka and Redis for data streaming pipelines.  
- Collaborated on backend microservices deployed to Azure AKS.  
  
Chinese & Math Teacher – Paxon School for Advanced Studies, Jacksonville, FL (2019–2024)  
- Led advanced Chinese IB curriculum and math instruction.  
- Built student data trackers and automated grading tools (Python, Excel).

## SELECTED PROJECTS

Clinical Knowledge Graph + NLP Demo (Showcased for Mayo Clinic) (2025)  
- Objective: Built a miniature clinical pipeline demonstrating data cleaning, knowledge graph enrichment, and note classification.  
- Data Cleaning: Converted labs\_raw.csv to labs\_clean.parquet using Pandas, ensuring proper formatting and types.  
- Knowledge Graph:  
 - Loaded lab/diagnosis relationships into Neo4j with py2neo.  
 - Created Cypher queries to link LOINC codes to ICD diagnoses.  
- Machine Learning:  
 - Trained a Logistic Regression model with TfidfVectorizer to classify short clinical notes.  
 - Implemented preprocessing with regex text cleaning and feature extraction.  
- API Development:  
 - Exposed endpoints (GET /dx\_by\_loinc/{loinc}, POST /classify\_note) via FastAPI and Uvicorn.  
 - Dockerized Neo4j instance and verified pipeline with Swagger UI.  
- Impact: Showed end-to-end integration of ETL, knowledge graph, and ML classification—demonstrating readiness for biomedical data tasks.  
  
Spatial Data Analysis on AWS EMR  
- Processed NYC taxi trip data using Spark-Scala on AWS EMR to identify trip density patterns and cost optimizations.  
  
Binary Fuzzer for Software Testing  
- Built a C++ binary fuzzer with coverage-guided mutation strategies to uncover hidden defects in compiled programs.

## RESEARCH INTERESTS

Biomedical data science, clinical NLP, knowledge graphs, healthcare informatics, and AI for precision medicine.