

# ZIRUI WANG

City: Syracuse, NY Phone: (517) 721-9616 Email: [wangzr926@gmail.com](mailto:wangzr926@gmail.com) Links: [LinkedIn](#), [GitHub](#)

## EDUCATION

<b>Syracuse University</b> , Master of Science in Computer Science, GPA: 3.9/4.0	2022 – 2025
<b>Michigan State University</b> , Bachelor of Science in Chemical Engineering, GPA: 3.5/4.0	2015 – 2019

## TECHNICAL SKILLS

**Programming Languages:** Bash/Shell, Python, C++, C, Java, JavaScript, HTML, CSS, SQL, NoSQL, R, Haskell, Swift, Dart, MATLAB

**Technologies:** Git, Docker, Kubernetes, Jenkins, AWS, BigQuery, Snowflake, Firebase, Airflow, Spark, Kafka, Vertex AI, SageMaker

**Frameworks:** Flask, Django, FastAPI, Streamlit, Spring Boot, React.js, Vue.js, Next.js, Node.js, Tailwind CSS, LangChain, LlamaIndex

## EXPERIENCE

<b>Data Scientist Intern</b>	May – Aug 2024
Regeneron Pharmaceuticals	remote, USA

- Collaborated with data engineering teams to automate the retrieval and processing of 40k+ publications from arXiv and PubMed using **Airflow** and **Docker**, resulting in a scalable ETL pipeline that stores model-ready data in **AWS S3**.
- Fine-tuned a **Hugging Face** LLM using Low-Rank Adaptation technique in **TensorFlow** and **Keras** to summarize clinical studies, enhancing real-world patient outcome extraction and reducing manual review time by 25%.
- Conducted document- and entity-level sentiment analysis on 100+ studies using Part-Of-Speech tagging, Named Entity Recognition in **NLTK**, and a RoBERTa-based model, generating actionable insights on patient experiences and treatment efficacy.

<b>Software Engineer Co-op</b>	May – Dec 2023
Regeneron Pharmaceuticals	Tarrytown, NY

- Built a **Flask**-based application with a RESTful API and a **Bootstrap** front-end UI to support parsing, conforming, and querying of lab data in **MongoDB** and **PostgreSQL** databases, improving research data accessibility and usability by over 80%.
- Enhanced software scalability and resilience by applying domain-driven design and implementing an event-driven architecture using **RabbitMQ** message broker, reducing technical debt by 30% and increasing system flexibility.
- Implemented unit and integration tests using **pytest**, increasing test coverage by 50% and reducing pre-release defects by 30%.
- Collaborated with cross-functional teams using **Bitbucket** for version control, and **Docker**, **Kubernetes**, and **Jenkins** for CI/CD processes, decreasing deployment time by 40% and improving delivery speed by 30%.
- Leveraged **Jira** and **Confluence** to facilitate Agile Scrum ceremonies, driving continuous improvement through backlog refinement, sprint planning, and reviews, resulting in a 30% reduction in blockers and faster delivery cycles.

<b>Associate ML Engineer</b>	Mar 2020 – Jul 2022
Institute for Quantitative Health Science and Engineering	East Lansing, MI

- Developed a BART-based transformer NLP model and a generative VAE in **PyTorch**, achieving a 24% improvement in protein stability (evaluated by FoldX) and expanding sequence library size by 300% while maintaining statistical consistency.
- Engineered a recommendation system in **Scikit-learn**, integrating unsupervised clustering with supervised classification models to streamline molecule screening, resulting in a 4-fold increase in lead yield.
- Led the development of data pipelines using **RDKit**, **PySpark**, **NumPy**, and **Pandas** to process and encode molecular and protein data, reducing preprocessing time by 40% and improving data accuracy.
- Managed file preparation, batch script generation, and data storage in **Python**, expediting 8k+ simulation runs on HPC clusters.

## PROJECTS

### AI Game Center [Link](#)

- Built a full-stack web application featuring interactive games that demonstrate AI algorithms including A\* search, Minimax, Markov Decision Process, and Gradient Descent, using a **React.js** frontend and **FastAPI** backend.
- Architected and deployed containerized services using **Docker Compose** with a **Nginx** reverse proxy on **AWS EC2**, enabling real-time gameplay and efficient API routing with static file serving.

### Textbook Exchange Platform [Link](#)

- Developed a microservices-based system using **Java** and **Spring Boot** with loosely coupled services communicating via **RabbitMQ**; managed dependencies with **Maven**, and applied OOD patterns like Singleton, Finite State Machine, Strategy, and Decorator.
- Implemented RESTful APIs and an asynchronous messaging system to handle user registration/login, book inventory operations (buy/sell/reserve), cart management, and payment processing with **MySQL** persistence.

### AI Agent Chatbot [Link](#)

- Designed and implemented a modular AI agent integrating **Google Gemini** with multiple Model Context Protocol servers (arXiv, ClinicalTrials, OpenFDA, PDB) using **Anthropic SDK**, advancing tool-augmented reasoning for life science research.
- Containerized **FastAPI** backend and **Streamlit** chatbot frontend with asynchronous context management and robust error handling to support real-time, multi-tool AI interactions.

## PUBLICATIONS

- Generative Models for Protein Sequence Modeling: Recent Advances and Future Directions. Briefings in Bioinformatics.
- Phytochemical drug discovery for COVID-19 using high-resolution computational docking and machine learning assisted binder prediction. Journal of biomolecular structure & dynamics.