

Wei Zhang Ph.D.

Phone: (205)567-4158 | Email: zwmc@hotmail.com | LinkedIn: [in/wei-zhang-76253523/](https://www.linkedin.com/in/wei-zhang-76253523/) | <https://github.com/vveizhang/Portfolio>

SUMMARY

Data scientist with 5 years of end-to-end Machine Learning experience mainly in biomedical and healthcare industry. Experienced in applying machine learning and NLP, with strong statistical analysis background.

SKILLS

Python, R, Pandas, Numpy, Matplotlib, ETL, EDA, Data Mining, Pytorch, Tensorflow, Scikit-learn, Machine Learning, clustering, decision tree, XGBoost, Deep Learning, time series modeling, NLP, Transformer, HuggingFace, LLM, OpenAI, ChatGPT, Llama2, LangChain, Retrieval Augmented Generation (RAG), Generative AI, SQL, Docker, Kubernetes, Azure, AWS (S3, EC2, SageMaker, Forecast, lambda, API Gateway), Git, Streamlit, Plotly, Flask, Tableau.

EXPERIENCE

Neo7 Bioscience

Remote, USA

Research Associate/ Data Scientist

02/2023 - Present

Document text summarizer for Automation of Patient Reports

- To help personalized precision medicine company automatically generate patients report, boost efficiency, built a large language model-based pdf summarizer.
- Use Langchain to summarize the information regarding gene and diseases from published scientific paper, pdf files, and generate text and report automatically. Saved 90% of the time for report building. ([Online demo](#) here)

Institute of Infectious Disease, HUST

Remote

Bioinformatics Scientist, Data Scientist

11/2021 - 02/2023

circular RNA Detector

- Circular RNAs are specific type of RNA which have important roles in diseases like cancer. It's important for researchers to know if a fragment of RNA contains circular RNA.
- Adapt transformer NLP model to analyze RNA sequence. Use circular RNA sequences from the circRNADb database to build and deploy a transformer-based model to predict if any given RNA sequences can form circular RNA, with F1 score of 0.93.
- Deploy the model as a [web APP](#) on AWS for fast and easy prediction within a few seconds.

The University of Texas Health Science Center at Houston

Houston, TX

Postdoctoral Researcher, Data Scientist

05/2018 - 06/2021

Bipolar Disorder Predictor

- About 2.8% of U.S. adults have bipolar disorder. The diagnosis is not easy.
 - Feature selection from blood RNA-seq data (over 20,000 genes), uncovered only 10 important genes as biomarker. Use grid search to choose the best model to predict bipolar disorder, got F1 score of 0.91.
 - This prediction methods only require the measurement of these 10 genes, save 80% time and funds comparing to large scale sequencing.
-

PROJECTS

Bitcoin Price Predictor Based on Reddit comments

12/2022 - 01/2023

- The price of cryptos are highly affected by social media. For personal interest I build a bitcoin price predictor.
 - Use pushshift API to automatically download everyday reddit comments regarding Bitcoin. Do sentiment analysis to get numbers of comments with different sentiment. Predict next day Bitcoin price using sentimental comment counts and historical price data. Build a daily-updating dashboard to visualize the prices.
-

EDUCATION

Chinese Academy of Science

PhD in Genetics

Wuhan, China

09/2007 - 06/2014