# John (Chenxi) Song

Pittsburgh, PA 15208 • chenxisong@hotmail.com • johncxsong.github.io

## TECHNICAL SKILLS

**LANGUAGES:**

Python 3, Java 18 & 1.8, R, JSX, XML, ANSI SQL, Bash

**PACKAGES & FRAMEWORKS:**

PyTorch, CUDA, Keras, Transformers, LangChain, Pandas, Numpy, Matplotlib, Tidyverse, Hugging-face

## EDUCATION

**University of Pittsburgh**, Pittsburgh, PA — Aug, 2022- Dec, 2023
*M. S. in Information Science, Big Data Analytics* — GPA: 3.6/4.0

**Reformed Presbyterian Theological Seminary**, Pittsburgh, PA — March, 2020- May, 2022
*Master of Theological Study in Biblical Counseling*

**Geneva College**, Beaver Falls, PA — Aug, 2016- Dec, 2018
*B. S. in Computer Science*
*B. S. in Engineering*

## EXPERIENCE

Reformed Presbyterian Theological Seminary — Feb, 2024- Present

**Network and System Administrator & Learning Management**

- Install, configure, and maintain the school's local area network (LAN), wide area network (WAN), operating systems (**Windows/Ubuntu/Linux Mint**), and physical and virtual servers (**Hybrid with Azure**).
- Perform system **monitoring network over 200 connections**, server resources and systems over 40 PCs.
- Analyze data from 300 students, implement visualizations to uncover key insights, optimized office software to **achieve $3,000 in annual saving**, and developed automated workflows to reduce processing time by 20%.
- Established an IT department inventory and ticketing system, centralizing asset management and streamlining issue tracking, resulting in improved resource allocation and enhanced accountability. (https://johncxsong.github.io/ITknowledgebase/)

School of Medicine, University of Pittsburgh — March, 2024- Present

**Prompting Engineer (ML Research Assistant)**

- **Supervised 2 summer internships**, providing coding guidance in Python for research projects.
- Led a Medical Image Processing project, managing 100GB of data for preprocessing, and utilizing a pre-trained large model **(Prov-GigaPath**) to extract features for lung cancer diagnosis.
- Guided the development of a project using a Large Language Model (LLM) to diagnose influenza, including designing **CoT prompt** templates for effective analysis.
- Deployed a **Llama3-8B** model locally on a 25GB GPU, ensuring data privacy and security for sensitive medical research.

School of Medicine, University of Pittsburgh — March, 2023- Dec, 2023

**Java Engineer (ML Research Assistant)**

- Implemented a machine learning project to predict and diagnose influenza, utilizing **supervised learning** techniques based on **Bayesian network graph structure** through Weka package in Java.
- Revamped a moderate-sized vanilla Java research project (Diagnosis Influenza) with 10 years history by repurposing variables and functions to follow Object-Oriented design pattern (OOP) for easy reusability and sustainability, use Maven as project management to save 90% time for auto setup project
- Preprocessed CSV file, including parsing, handling missing values, and transforming data like one-hot encoding through Python.
- Implemented a new feature (advanced multi-source domain data) derived from academic pseudocode, **increase 2% accuracy** comparing to existing single source model prediction.

## PUBLICATIONS

**Paper Published (IEEE)**

1. Yiming Sun, Yuhe Gao, Runxue Bao, Gregory F. Cooper, Jessi Espino, Harry Hochheiser, Marian G. Michaels, John M. Aronis, **Chenxi Song**, & Ye Ye. (2024). Online Transfer Learning for RSV Case Detection.

**Paper Under Review (AMIA 2024 Informatics)**

1. Yuhe Gao, Runxue Bao, Yuelyu Ji, Yiming Sun, **Chenxi Song**, Jeffrey P. Ferraro, & Ye Ye. (2024). Transfer Learning with Clinical Concept Embeddings from Large Language Models.

**Poster Under Review (AMIA 2024 Informatics)**

1. **Chenxi Song,** Yuhe Gao, RunXue Bao, Yiming Sun, Julians Tirado Alicia, & Ye Ye. (2024). Probabilistic Disease Surveillance Using Large Language Model.

## PROJECTS

**Probabilistic Disease Surveillance Using LLM**, University of Pittsburgh June, 2024- Sep, 2024
Skills: Large Language Model | CoT Prompting | NLP | PyTorch | Transformers | CUDA

- This study aims to examine the capability of LLMs to provide probability estimations. We compared different prompting strategies and evaluated an open-source LLM model, LLaMA 3, for detecting infectious disease cases from emergency department encounters.

**Exploring the Integration of Foundational Model**, University of Pittsburgh May, 2024- Aug., 2024
Skills: Pre-trained Model | Computer Vision | Image Embedding | PyTorch | OpenSlide | CUDA

- This study aims to drive insights by combining Prov-GigaPath's computational capabilities with Mesothelioma data Center. Potential replication of the model for other bio-repositories or research focuses.

**Bayesian Network Transfer Learning**, University of Pittsburgh March, 2023- Dec., 2023
Skills: Bayesian Network | Java | Supervised Learning | Tree Graph | Greedy Algorithm

- This project created the Bayesian Network Transfer Learning (BN-TL) algorithm to re-use of source model, such as influenza, learned from electronic medical record (EMR) data to predict the target data set.

## RELEVANT COURSES

- Information Storage & Retrieval
- Advanced Data Mining
- Artificial Intelligence
- Algorithm Design
- Machine Learning
- Data Structure