John (Chenxi) Song

Pittsburgh, PA 15208 • chs342@pitt.edu • johncxsong.github.io • Google Scholar

1.5+ years research assistant experience in **AI**, **machine learning,** and **bioinformatics**. Experience in implementing models, including a Java-bases machine learning **Bayesian Network** for flu diagnosis, and contributing to projects in large language models (**LLMs**) reasoning and pathology **images analysis**. Passionate about leveraging AI and machine learning to drive innovation across industries and create solutions that benefit society.

RESEARCH INTERESTS

- Machine Learning & Deep Learning
- Bayes' Theorem
- Artificial Intelligence (Large Language Model)
- Computer Vision (Pathology Slides)
- Robotics (Human/Medical Robotics)

EDUCATION

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

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

North Seattle College, Seattle, WA — April, 2015- June, 2016
Associate of Science Degree

Fundamental of Engineering Exam (FE) Mechanical — Nov. 2018
Credential ID 19-633-53

EXPERIENCES

Network and System Administrator, Reformed Presbyterian Theological Seminary — Feb, 2024- Present
Skills: Python, Git | Azure Cloud | Linux, Windows, Ubuntu, Mac OS | Bash | Hardware | ADDS | DNS

- The working experience aims to understanding network communication in large-scale web systems and data exchange. Gained skills in building systematic frameworks within domain networks to enhance visibility and manage a digital presence effectively.
- 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 10% 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/)

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.

- Guided the development of a project using a **Large Language Model (LLM) to diagnose influenza**, including designing **CoT prompt** templates for effective analysis.

Exploring the Integration of Foundational Model, University of Pittsburgh May, 2024- Aug., 2024

Skills: Computer Vision | Numpy | Pandas | PyTorch | OpenSlide | Matplotlib | Scikit-Learn

- 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.
- 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 Mesothelioma diagnosis.

Bayesian Network Transfer Learning, University of Pittsburgh March, 2023- Dec., 2023

Skills: Bayesian Network | Java 1.8 & 18 | 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.
- 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 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 1% accuracy** comparing to existing single source model prediction

PUBLICATIONS

Paper Published

Best Paper Awarded

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, June). Online transfer learning for RSV case detection. *In 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI) (pp. 512-521)*. IEEE.

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. *arXiv preprint arXiv:2409.13893.*

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.

HONORS/AWARD

Best Paper in Analytics Track, IEEE 12th ICHI, FL Aug. 2024

Winner, Geneva College Tower Scholarship, PA Aug. 2017 -Dec. 2018

Winner, College Hill Church Scholarship, PA Aug. 2017 -Dec. 2018