Xin Wang

Email: xwang314@binghamton.edu | **GitHub**: https://github.com/xin-wang-kr | **Webpage**: https://xin-wang-kr.github.io/

RESEARCH INTEREST

Representation Learning, Algorithm Fairness, Conversational AI, Natural Language Processing, Machine Learning, Computational Social Science, Network Science, Human-machine Interaction, Health Informatics, Geographical Information Analysis.

EDUCATION

Doctor of Philosophy (Ph.D.) in Systems Science (GPA: 3.96/4.0) Aug 2022 – May 2026

Department of Systems Science and Industrial Engineering, Binghamton University

- Advisor: Dr. Sadamori Kojaku

Oxford Machine Learning Summer School May 2024 – Jul 2024

- MLx Fundamentals
- MLx Representation Learning & Generative AI

Master of Science (M.S.) in Data Science (GPA: 4.0/4.0, Top 1) Aug 2016 – Dec 2018

Data Science Institute, Saint Peter's University

- **Capstone project:** Industry Classification on Company Annual Report using Machine Learning.
- Advisor: Dr Robert Finn, Dr Sylvain Jaume

Bachelor of Science (B.S.) in Industrial Engineering (Aeronautical Engineering direction) Sep 2010 – Jul 2014

Department of Aeronautical Engineering, Civil Aviation University of China

- **Dissertation:** Maintenance Plan Scheduling with Capability Constraints Based on Genetic Algorithm.
- Advisor: Dr Weigang Zhang

RESEARCH EXPERIENCE

Graduate Research Assistant Jan 2024 – Present

MAP-IT Lab, Department of Systems Science and Industrial Engineering, Binghamton University
Advisor: Dr. Sadamori Kojaku
Applying Conversational AI and Representation Learning to Detect and Mitigate AI Bias.

- Conducting the research on demographic bias in hospital reviews using word embedding approaches.
- Creating conversational AI chatbot prototype using LLaMa-based model and developing a human-bias-based user modeling framework for the chatbot application.
- Developing trust&bias-based conversational agent to support user's decision-making on healthcare providers.

Summer Research Intern Jun 2023 – Aug 2023

Research Foundation at SUNY
Advisor: Dr. Neha Patankar
Interactive Decision-making Toolkit Development

- Dealt with geographical features in New York State area through Geopandas and created different data selection controls, such as slider bar and scatter plot, using Plotly Express.
- Generated interactive geo map using Folium and developed web-based data visualization dashboard for decision-making on energy tradeoffs through Streamlit framework.
- Applied LangChain and OpenAI to build a conversational agent for user query to the visualization dashboard.

Summer Research Volunteer Jun 2023 – Aug 2023

Human-centered Computing Division, School of Computing, Clemson University
Advisor: Dr. Emma Dixon
Mobile App on Social Recall Network Data Collection for People with Dementia

- Created UI interface for the Android mobile app using Kotlin and xml.

COVID-19 Research Team Leader Jan 2020 – Apr 2022

BIG DATA & AI Lab, IntelligentRabbit LLC.

COVID-19 Anti-Asian Hate in Social Media Research Project

(1) Twitter Research

- Led the team to collect over 10 million tweets by tracking pandemic-related Twitter users between January 15 and October 15, 2020, through setting up Twitter API on Amazon AWS EC2 cluster.
- Created new features using VADER sentiment analysis. Labeled each twitter user as constantly like, constantly dislike, swing, or does-not-care towards Trump, Biden, Republican, and Democratic.
- Visualized geopolitical difference on heatmaps with the combination of geolocation data and election information.
- Evaluated feature importance by build a random forest algorithm and analyzed the model performance by using only sentiment-based features in DNN, which was trained on AWS EC2 cluster.

(2) YouTube Research

- Collected 1,452,373 comments from COVID-19-related news from mainstream media YouTube channels using YouTube Data API and Facepager.
- Extracted sample data based on Anti-Asian hate hashtags and keywords. Manually labeling the 3,756 YouTube comments to perform anti-Asian hate classification using SVM, Random Forest, LSTM, CNN models.
- Established anti-Asian hate indicator to portray the tendency of hate over time through the generated daily hate signal.

COVID-19 Fake News Monitoring and Misinformation Detection System Project

- Built a Python-based web scraper robot to collect 8,810 fake news during January 2020 to August 2020 from the CoronaVirusFacts/DatosCoronaVirus Alliance Database on Poynter.
- Collected and cleaned 1,673,354 real news by dealing with the semi-unstructured data from AYLIEN News API using Python programming.
- Performed LSTM, CNN, and DBN models on fake news detection and evaluated mainstream media credibility based on the algorithm-based ranking methods.
- Built COVID-19 fake news detection application with interactive interface based on Flask framework and set up on Heroku cloud-based infrastructure [link].

Research Assistant May 2019 – Jan 2020

BIG DATA & AI Lab, IntelligentRabbit LLC.

Bitcoin Prediction Research Project

- Collected historical data of S&P500, NASDAQ, and Dow Jones indices through Yahoo Finance API.
- Based on the predictive VAR model to analyze the impulse response between the parameters for stock and Bitcoin market.
- Utilized sliding window technique to optimize the prediction results in time series.

Smart Ship Recognition on Satellite Imagery

- Extracted pixel data from JSON object to perform image classification using Python programming.
- Built Support Vector Machine, Random Forest, Logistic Regression and CNN models to recognize ships in the bay and sea areas from satellite images.

More research and details are presented in Projects section at https://xin-wang-kr.github.io/projects/

INDUSTRY EXPERIENCE

Machine Learning Engineer Aug 2017 – Dec 2017

Tianjin Value Fair Technology Inc.

NLP-based Financial Software Application Development

- Worked with AI technology team to develop a financial software application based on Natural Language Processing.
- Built web crawler tools to collect annual reports text data from annualreports.com according to stock symbols.
- Constructed structure tree using tree kernel algorithm combines with chunking method to get named entity using NLTK.
- Vectorized text data by vector space modeling. Performed PCA algorithm to reduce dimensions for document vectors.
- Performed company industry clustering by K-means and Gaussian Mixture Models and obtained the satisfied 6 categories of company according to their business activity using Python Scikit-learn.
- Applied topic modeling to find out whether they have a common property in company's business activity make them in one cluster.
- Calculated TF-IDF scores in each category to rank and vote the key business information.

- Explored the relationship on business for companies in each category by calculating cosine similarity between companies' vectors.

PUBLICATIONS
[Google Scholar]

Journal Articles:

[J.7] **Wang X.**, Chen X., Bolian L. & Zhao P. (2021). Constructing an Anti-Asian Hate Indicator for Pandemic-related Comments from Mainstream Media YouTube Channels. *International Journal of Society Systems Science (IJSSS)*. *13*(4), 278-293. DOI: 10.1504/IJSSS.2021.124956

[J.6] Zhao, P., Chen, X., & **Wang, X.** (2021). Classifying COVID-19-related hate Twitter users using deep neural networks with sentiment-based features and geopolitical factors. *International Journal of Society Systems Science*, *13*(2), 125-139. DOI: 10.1504/IJSSS.2021.116373

[J.5] **Wang, X.**, Zhao, P., & Chen, X. (2020). Fake news and misinformation detection on headlines of COVID-19 using deep learning algorithms. *International Journal of Data Science*, *5*(4), 316-332. DOI: 10.1504/IJDS.2020.115873

[J.4] **Wang, X.**, Chen, X., & Zhao, P. (2020). The relationship between Bitcoin and stock market. *International Journal of Operations Research and Information Systems (IJORIS)*, *11*(2), 22-35. DOI: 10.4018/IJORIS.2020040102 (Acceptance rate: 42%)

[J.3] Xiao H., **Wang X.**, & Zhao, P. (2019). Satellite Image Recognition for Smart Ships Using A Convolutional Neural Networks Algorithm. *International Journal of Decision Science (IJDS)*,*10*(2), 85-91.

[J.2] Subramaniam V., Srungarapu G., Matijosaitiene I., Supe M., Agarwal A., Zhao P., **Wang X.**, Kwartler E. and Jaume S.(2016). Geospatial and Temporal Data Analysis on the New York City Taxi Trip. *International Journal of Data Analysis and Information Systems*, *8*(2), 63-73.

[J.1] **Wang X.** (2015). Study on Scheduling Algorithm Based on Capacity Constraints of Maintenance Plan. *Industrial Engineering Practice*. ISSN 2304-5337.

Conference Papers:

[C.1] **Wang X.**, Samer A., Grace B., & Stephanie J. (2024). Co-Designing an AI Chatbot to Improve Patient Experience in the Hospital: A human-centered design case study of a collaboration between a hospital, a university, and ChatGPT. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*. Accepted. (Acceptance rate: 24%) [link]

Workshop Papers:

[W.2] **Wang X.** & Kojaku S. (2024). User Trust Modeling in Conversational User Interface Based on Word Embedding Bias. In *CUI@CHI'24: Building Trust in CUIs – From Design to Deployment, May 11-16, 2024, HI, USA.* ACM, New York, NY, USA, 4 pages. [link]

[W.1] **Wang X.** (2023). Improving Patient Experience Feedback Collection for Healthcare Provider through Human-centered Chatbot Application. In *Proceedings of July 10-14, 2023 (ACM DIS'23 – Workshop "Designing for and Reflecting upon Resilience in Health and Wellbeing")*. ACM, New York, NY, USA, 5 pages. [link]

NEWS COVERAGE

- **BingUNews – Title:** Watson graduate students implement AI in healthcare practice as final project: Development team to present research at renowned conference — 2024
  Link: https://www.binghamton.edu/news/story/4765/watson-graduate-students-implement-ai-in-healthcare-practice-as-final-project
- **Scienceaq – Title:** La xénophobie liée au COVID-19 — 2021
  Link: http://fr.scienceaq.com/Autres/1001115239.html
- **Science Spot – Title:** COVID-19-related xenophobia — 2021
  Link: https://sciencespot.co.uk/covid-19-related-xenophobia.html
- **Techxplore – Title:** Editing out fake news — 2021
  Link: https://techxplore.com/news/2021-06-fake-news.html
- **Science Spot – Title:** Editing out fake news — 2021
  Link: https://sciencespot.co.uk/editing-out-fake-news.html
- **TechiLive.in – Title:** Editing out fake news — 2021
  Link: https://techilive.in/editing-out-fake-news/
- **News AZI – Title:** Editing out fake news — 2021
  Link: https://newsazi.com/editing-out-fake-news/

HONORS & AWARDS

- **Jim Geer Summer Scholarship** ($2,400) 2024
  Awarded by Thomas J. Watson College of Engineering and Applied Science, Binghamton University
- **Travel Award** ($500) 2024
  Awarded by Watson Professional Development Fund
- **ALife 2020 Student Scholarship** 2020
  Awarded by ALife 2020 Organizing Committee.
- **Outstanding Mobile App (Andriod mobile app - Java)** 2016
  Awarded by 2016 Google Developer StudyJams (100/400).
- **Third Prize of 2014 Cross-Strait Contest of Outstanding Bachelor's Degree Thesis** 2014
  Awarded by Chinese Institute of Industrial Engineering (10/15, among China Mainland, Hong Kong, Macao, and Taiwan).
- **Second Prize Scholarship** 2012, 2013
  For academic excellence at Department of Industrial Engineering at CAUC (for top 10).
- **Third Prize Scholarship** 2011
  For academic excellence at Department of Industrial Engineering at CAUC (for top 15).
- **Outstanding Student Award** 2011, 2012, 2013
  Awarded as an excellent student at Civil Aviation University of China.

MENTORING PROJECTS

Mentor, *BIG DATA & AI Lab, IntelligentRabbit LLC.* 2019 - 2020

- A New Solution of the Social Distancing and Face Mask Monitor Using Deep Learning Algorithms. [link]
- Fake News Monitoring and Anti-rumor System Using DL & Blockchain. [link]
- AI Drives Music Self-creation with Deep Learning in TensorFlow. [link]

Mentor, *AI HUB, GEC academy* 2020 - 2021

- Learn to Speak Like a Native: AI-powered Chatbot Simulating Natural Conversation for Language Tutoring.
- Pooling and Convolution Layer Strategy on CNN for Melanoma Detection.
- CNN-based Diagnosis System on Skin Cancer using Ensemble Method Weighted by Cubic Precision.
- Region-based Birdcall Recognition Using Signal Processing.
- Hurricane Damage Prediction through Satellite Imagery based on CNN.
- COVID-19 Fake News and Misinformation Detection using Transformer Learning.

TEACHING EXPERIENCE

Invited Lecture Guest, *Department of Systems Science and Industrial Engineering, SUNY at Binghamton*

- Course: Applied Soft Computing (Undergraduate and graduate level; Spring 2024)
- Lecture topic: Applying Embedding Approach to Chatbot Application
- Lecture content: (1) Text embedding and image embedding. (2) Streamlit chatbot framework. (3) Word2vec-based question and answer chatbot application. (4) Text-image search chatbot application.

Instructor, *AI Hub, GEC Academy*

- Intro to Java (Graduate level; Summer 2022)
- Academic writing (Graduate level; Fall 2021, Summer 2022 and Summer 2023)

Instructor, *BIG DATA & AI Lab, IntelligentRabbit LLC.*

- Mobile App Development: Python Kivy (Fall 2020)
- Python Programming in AI (Summer 2020)

Teaching Assistant, *Department of Systems Science and Industrial Engineering, SUNY at Binghamton*

- Introduction to Systems Science (Undergraduate and graduate level; Fall 2023)
- Probabilistic Systems (Undergraduate level; Spring 2023; teaching discussion section weekly)

Teaching Assistant, *GEC Academy*

- Machine Learning in Biomedical Monitoring (taught by Dr. Maarten De Vos at Katholieke Universiteit Leuven, Spring 2021)
- Applied Machine Learning (taught by Dr. Stephen Coggeshall at University of Southern California, Winter 2020)
- Algorithmic Foundations of Learning (taught by Dr. Patrick Rebeschini at Oxford University, Summer 2020)

- Introduction of Machine Learning (taught by Dr. Jovan Ilic at Carnegie Mellon University, Spring 2020)

Teaching Assistant, *Data Science Institute, Saint Peter's University*

- Data Mining (Graduate level; Spring 2018)
- Statistical Programming: R and SAS (Graduate level; Spring 2017)

SERVICES

- **Reviewer**

  CHI'24, CUI'24

- **Data Science Volunteer**, *DataKind Global* 2019 – 2022

  Evaluated data reliability and human-centered ML rules for the challenge projects.
  Served events: (1) Google AI Impact Challenge. (2) data.org Inclusive Growth and Recovery Challenge for COVID-19

- **Web Developer Volunteer**, *Marks JCH of Bensonhurst* 2017

  Developed class registration web application using Django, MySQL, HTML, CSS.

- **Data Analysis Volunteer**, *Dreamland Plus* 2015 – 2016

  Collected genealogy data from minority community groups. Generated data visualization dashboard and performed genealogy relationship analysis using Tableau.

CONFERENCES/ WORKSHOP ATTENDED

- Health Resilience Workshop, ACM Designing Interactive Systems 2023. Jul 2023
- Data+AI Summit 2022, Databricks Jun 2022
- Virtual Complex Systems & Data Science Seminar Series, University of Vermont Nov 2021
- Samsung AI Forum, Samsung Nov 2021
- Virtual DataDive Event 2021, DataKind Global Sep 2021
- 2020 IDEAS Global AI Conference, International Data Engineering and Science Association Oct 2020
- ALife 2020, International Society for Artificial Life Jul 2020

TECHNICAL SKILLS

- **Programming:** Python, R, SAS, C++, Java
- **Mobile/web Application:** Streamlit, Django, Flask, Kivy, HTML, XML
- **Deep Learning/ AI:** PyTorch, TensorFlow, Keras, OpenCV, NLTK
- **Machine Learning:** Scikit-learn, Pandas, Numpy
- **Data Visualization**: Matplotlib, Seaborn, Tableau, Minitab
- **Big Data:** Hadoop, Spark, Amazon AWS, Hive
- **Database:** MySQL, PostgreSQL

MEMBERSHIP

- ACM SIGCHI Feb 2024 – Now
- Alpha Pi Mu – Industrial Engineering Honor Society Mar 2024 – Now