Xuxin Tang

Blacksburg, VA • (857)230-2402 • tangxxwhu@gmail.com
LinkedIn • Github • Google Scholar • Personal Website

SUMMARY

Ph.D. student in Computer Science at Virginia Tech, focusing on Human-AI interaction, Data Visualization and Large Language Models. Engineer with 3 years of industry in full stack, big data, data science, and machine learning.

EDUCATION

- **Ph.D. in Computer Science and Application –** Virginia Tech (GPA: 4.00/4.00) *Aug*. 2022 – *Present*
  - Advised by Dr. Chris North (from Aug. 2023 – Present):
    - Area: Human-AI Interaction, Large Language Models
  - Advised by Dr. Yalong Yang (Aug. 2022 – Aug. 2023):
    - Area: Human-AI Interaction, Immersive Analytics
- **M.S. in Software Engineering –** Wuhan University *Sep*. 2015 – *Jun*. 2018
  - Advised by Dr. Zhijiang Li and Dr. Fan Zhang
- **B.A. in English (Minored) –** Wuhan University *Sep*. 2011 – *Jun*. 2015
- **B.S. in Engineering –** Wuhan University *Sep*. 2011 – *Jun*. 2015

WORK EXPERIENCE

Research Assistant - Virginia Tech – 2024 VIS Submitting *Aug. 2023 – Present*
Blacksburg, VA, United States

- Introducing LLMs for cognitively complex sensemaking tasks for analyzing multiple text documents.
- Make human Space to Think for the common ground for human continuous interactions and LLMs context updation, achieving seamless human-AI collaboration.
- Interactive web prototype for efficiency human-AI sensemaking.
- *LLMs, Sensemaking, Data Visualization, Human-AI Interaction, React, Javascript, Typescript, Docker*

Visiting PhD Student - MIT CSAIL *Jun. 2023 – Aug. 2023*
Cambridge, MA, United States

- Worked with Dr. Stefanie Mueller and Dr. Mustafa Doga Dogan in HCIE Group to augment static printed material with extendable interactive visualizations.
- Recognize invisible QR codes with ResNet-based DL method and achieve both offline/online in-context Augmented content on static printed material.
- User data collection and analysis. Model IR ink perception thresholds of participants by leveraging Gaussian and GBDT.
- *Unity, C#, Augmented Reality, Deep Learning, Graph Network, Study Design, User Study, Mobile Application*

Machine Learning Engineer - YY Business Unit, Baidu *Jul.* 2019 – *Mar.* 2021
Beijing, China

- Responsible for machine learning and data analysis for high-quality video recommendation services.
- Efficiently implement and scale advanced machine learning and deep learning techniques to provide robust services to a customer base in the billions.
- Tasked with architecting and developing the company's data analytics and data mining systems. This role involves strategic design and oversight of the entire system architecture, ensuring the maximization of data value, enhancing data integrity, and fostering the company's business growth through insightful data-driven strategies.
- *Python, C++, Hive, Hadoop, SPARK, GBDT, BERT, DeepFM, DSSM, DeepWalk*

Software Development Engineer - Oracle China R & D Center *Jul.* 2018 – *May.* 2019
Beijing, China

- Full-stack development for data visualization of Oracle SaaS cloud computing platform, providing customers with visual analytics solutions of integrated monitoring and management of computing resource.
- Obtained a good knowledge of *jQuery*, *JavaScript*, *OracleJET*, *requireJS*, *Knockout, HTML,* and other Oracle front-end development framework in short time, implementing frontend of a website in one week.
- Provided back-end technical support for new projects, involving *puppeteer*, *node.js*, *ApiGateway*, *Rest API*, and collaborate with transnational colleagues efficiently.
- *Frontend: jQuery*, *Java*, *JavaScript*, *OracleJET*, *requireJS*, *Knockout and Oracle front-end development framework*.
- *Backtend: puppeteer*, *node.js*, *ApiGateway*, *Rest API, Docker*.

PUBLICATIONS (76 citations now)

- Fan Zhang[1], **Xuxin Tang**[1], Xiu Li, S.U. Khan, and Zhijiang Li. Quantifying cloud elasticity with container-based autoscaling[J]. *Future Generation Computer Systems,* 2019, 98: 672-681. ([1]Co First Author, 40% newly added content)
- **Xuxin Tang**[1], Fan Zhang[1], Li X, S.U. Khan, and Zhijiang Li. "Quantifying Cloud Elasticity with Container-Based Autoscaling." *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE*, 2017. (Selected as special issue and extended)
- **Xuxin Tang**, Zhijiang Li, Yuhang Chen. "A Night Image Enhancement Algorithm Based on Guided Filtering." *Advanced Graphic Communications and Media Technologies*. Springer, Singapore, 2016.
- Pias, Tanmoy Sarkar, Yiqi Su, **Xuxin Tang**, Haohui Wang, Shahriar Faghani, and Danfeng Yao. "Enhancing Fairness and Accuracy in Type 2 Diabetes Prediction through Data Resampling." medRxiv (2023): 2023-05

ADDITIONAL PROJECT EXPERIENCE

NLP-powered Immersive Graph Sensemaking *Feb.* 2023 – *May. 2023*

Virginia Tech, *Blacksburg, VA.* Advisor: Dr. *Yalong Yang*

- Explored how multimodel (especially speech) interactions in knowledge graphassist human in complex sensemaking tasks.
- Costructed citation-based graph on academic papers for
- *Unity, C#, Oculus Unity Intergration, Graph Algorithms, Python, Graph, Prototyping, Study Design*

COMPETITION EXPERIENCES

Jigsaw Unintended Bias in Toxicity Classification (Kaggle) *May.* 2019 – *Jun.* 2019

Developed an NLP model for identifying toxicity in online conversations, Team Leader, Top 10% out of 3030 teams in Kaggle NLP competition, bronze medal

- Cleaned and regularized the data for model construction by using the deep model BERT and LSTM for a jointly learning and prediction.
- Learned the preprocessing of natural language in the course of the competition, learned *word2vec*, *glove*， and used embeddings to understand the principles of attention and transformer, memory.
- Used pytorch to train the model with two-layer BI-LSTM, and BERT to train the pre-trained model and predict the test dataset.

Ali Tianchi Intelligent Traffic Forecast Challenge *May.* 2017 – *Aug.* 2017

Developed a time sequence model in Python to analyze and predict the future traffic situation, Team Leader, top 3% out of 1716 teams

- Developed a time sequence model in order to estimated the average travel time from 7 AM to 8 AM based on the history of the daily travel time of each vehicle on 132 roads from March to May.
- Leveraged LSTM for prediction and parameter fine-tuning such as encoder and decoder layer, hidden units, batch size and dropout.

SKILLS

Languages: C| C# | C++ | Java | Python | MATLAB | HTML | JavaScript | SQL | Hive SQL | Spark
Libraries: jQuery | Bootstrap | TensorFlow | Keras | PyTorch | Scikit-Learn | Numpy | Matplotlib | Pandas | Pupeteer | Oculus SDK
Tools: Node.js | React | Flask | requireJS | Knockout | ApiGateway | Rest API | Unity
Machine Learning: CV, NLP, BERT, GPT, LSTM, CNN, Embeddings, Xgboost

HONORS & AWARDS

- Full Scholarship for Graduate students of Wuhan University, three times 2015 – 2017
- Graduate Scholarship of Wuhan University, two times 2015 – 2017
- Outstanding Student of Wuhan University, two times 2013 – 2015
- Outstanding Student Leaders of Wuhan University 2013
- 1st Prize of Summer Social Practice of Wuhan University 2012