Zijie Huang

Merchant Venturers Building, University of Bristol, BS8 1UB
E-mail: z.huang@bristol.ac.uk
Personal Webpage: http://zijiehuang.github.io

Education

University of Bristol, UK **09/2023 - 06/2024 (expected)**
Ph.D. in Electrical and Electronic Engineering
(Continued from University of Exeter)
Thesis: A Trustworthy Cycle System: Enabling the Interdependence between Machine Learning and Human.

University of Exeter, UK **09/2020 - 09/2023**
Partially Completed Ph.D. in Computer Science

Fujian Normal University, CN **09/2016 - 06/2020**
B.Eng. in Network Engineering (1st)

Professional Experience

Transport Research Lab Software, UK **01/2022 - 08/2022**
Research Intern via the Alan Turing Institute in Connected Vehicle Data Exchange project

- Responsible for the development of, and implementation of a state-of-the-art model interface and toolset in the PTV VISSIM simulation of connected vehicle data exchange.
- Support the progressing of the proposed methods for connected vehicle data exchange for traffic signal control and network management.
- Explore the consideration of ethical theories under the data exchange model, investigate the interpretable decision-making outcomes on assisting intelligent transportation system developers to improve model's efficiency.

University of Exeter, UK **05/2021 - 08/2023**
Master student research project supervisor

- Supervise projects on *'Explainable machine learning for firewall configuration'*, *'Explainable machine learning for DDoS Botnet attack on IoT devices'* and *'Mitigation of the Bias in Stable Diffusion'*.
- Assist with research projects development and students' assessment. Provide feedback for students' final thesis.

University of Exeter, UK **01/2021 - 06/2023**
Postgraduate Teaching Assistant

- Deliver workshops for multiple modules: 'Data Structure and Algorithm', 'Software Development', 'The C Family' and 'Enterprise Computing'.
- Assist with student coursework and exam assessment.

Fuzhou Yuke Information Technology Co.,Ltd, CN **09/2019 - 12/2019**
Software developer Intern in Front-end Developing

- Develop multi-pages applications leveraging Vue.js.
- Build dynamic navigation routes in Vue.js applications.
- Convert raw data from API and database calls to create unique and effective user experiences in the browser.

Technical Skills and Competences

- **Explainable machine learning**: extensive post-hoc explainable methods experience (e.g. LIME, SHAP, MAPLE); analysis explanation's accuracy, improve explainable methods' computational efficiency.
- **Reinforcement learning**: vast experience in building A3C, multi-agent RL, POMDPs for autonomous vehicles' path planning scenarios.
- **Digital Twin**: acquaintance in using Python (Simpy) to create DT.
- **Ethical theories**: interdisciplinary understanding of moral philosophy; interpret and formulate ethical theories into autonomous vehicles applications.
- **Trustworthy AI**: understanding of fairness metrics and algorithms; design human-value aligned AI system; knowledge of policies and procedures for governing AI systems.
- **Problem-solving**: innovative algorithm design; analyse and model novel solutions to address complex challenges.
- **Teamwork**: Successful internal and external collaborations led to several publications; taught new techniques to colleagues and collaborators.
- **Critical thinking**: identified current works limitations; proposed alternative and feasible perspectives; contributed original insights.
- **Attention to detail**: Accurate record keeping of experiments; use of checklists to organize my work and make sure that it is completed in a timely manner; peer reviewed scientific publications.

Research Interests

My research interests encompass the expansive field of trustworthy AI techniques. Specifically, my focus is on elevating the interpretability, fairness, and accountability of diverse machine learning models. This includes, but is not limited to, neural network, generative AI, and autonomous decision-making aligned with human values.

Research Projects

"Identifying tiger strips with supervised and unsupervised machine learning pipelines" 05/2023

A Data Study Group Project worked with the Environmental Investigation Agency (EIA) via the Alan Turing Institute. The project aims to create a user-friendly tool for recognizing individual tigers based on their stripe patterns and utilizing this information to enhance enforcement efforts and combat the trade in tiger skins, carcasses, and live animals.

- Design two pipelines: A *Vision Transformers* (Grounding DINO) with masking (Segment Anything Model), and Simple framework for contrastive learning of visual representations in tiger (*Sim-CLR*), to automated match images (match labelled (carcass/skin) to unlabelled images ('live' tigers)).
- Evaluate the proposed pipelines decision-making process via incorporating explainability: using *saliency maps* to visualize which parts of an image were more important for making the decision.

"Ethical decision-making for the inside of the autonomous bus moral dilemmas" 07/2022 - 04/2023

A proposed ethical decision-making system is designed to address ethical dilemmas that arise inside of the autonomous buses, offering moral route planning solutions.

- Design a two-stage system that involves *human-value aligned data representation* and an *ethical-compliant multi-objective thresholded lexicographic Deep Q-learning method*.
- Investigate different ethical principles, adopt, and interpret them into the design of proposed decision-making models.
- Validate the proposed system's accuracy and usability, position it as a valuable tool for ensuring ethical autonomy in real-world scenarios.

"An energy-efficient and trustworthy unsupervised anomaly detection framework (EATU) for IIoT" 02/2021 - 12/2021

An energy-efficient and trustworthy unsupervised anomaly detection framework is proposed to improve the self-diagnosing efficiency and infrastructures security in IIoT.

- Design an *Autoencoder-based* feature extraction and Efficient *DeepExplainer-based* feature selection.
- Evaluate the proposed framework has improved accuracy, trustworthiness, and energy-efficiency using real-world IIoT datasets with high-dimensional features.

Publications

[1] **An energy-efficient and trustworthy unsupervised anomaly detection framework (EATU) for IIoT**
Huang, Z., Wu, Y., Tempini, N., Lin, H., & Yin, H. (2022). ACM Transactions on Sensor Networks, 18(4), 1-18.

[2] **A Survey on Explainable Anomaly Detection for Industrial Internet of Things**
Huang, Z., & Wu, Y. (2022, June). In 2022 IEEE Conference on Dependable and Secure Computing (DSC) (pp. 1-9). IEEE.

[3] **Ethical Decision-making for the Inside of Autonomous Buses Moral Dilemmas**
To be published on IEEE Transactions on Artificial Intelligence Journal

[4] **Digital-twin assisted Ethical Decision-making for the Inside of Autonomous Buses Moral Dilemmas**
In submission to IEEE Transactions on Artificial Intelligence Journal

[5] **Unleashing Creative Potential: Nurturing Trustworthy Generative AI (Chapter in CUP Research Handbook on Generative AI and the Law)**
In submission to Cambridge University Press

Additional Experiences

GW4 Connect Programme: Cross-institutional mentorship programme for postgraduate researchers of colour 07/2023 - 03/2024

- Provide support for postgraduate researchers experiencing racism, discrimination, and other structural barriers and inequalities within and beyond the research arena.

Researcher Leadership Development Programme 04/2023 - 07/2023

- Enhance leadership skills and promote effective research management.
- Collaborated with fellow participants on group projects, fostering teamwork and the exchange of diverse perspectives.

Exeter Mathematics School A-level Mini-research Projects Leader 2021 - 2022

- Design mini-research project to challenge A-level students and give them an insight into the research methods.

Reviewer Duties 2021 - present

- Conferences: IEEE DSC, IEEE ICC, IEEE ICNC, CNSM XAI4Network, EAI BDTA
- Journals: IEEE Wireless Communications, IEEE Internet of Things Journal, IEEE Transactions on Network Science and Engineering, Computer Network, Scientific Reports

Awards

GW4 Open Research Prize 11/2023

Poster Category runner-up

- Autonomous Bus Ethical Decision Making for Moral Dilemmas