

ZHIXI CAI

☎ +61 432 154 888 | @ smczx@hotmail.com | 🔗 LinkedIn | 🐙 GitHub | 🌐 Profile | 📍 Melbourne, Australia

EDUCATION

Monash University <i>Doctor of Philosophy (Computer Vision, Artificial Intelligence)</i>	Melbourne, Australia 03/2021 – 02/2024 (Expected)
Monash University <i>Master of Data Science; WAM: 86, GPA: 3.88/4</i>	Melbourne, Australia 03/2019 – 12/2020
Ningbo University <i>Bachelor of Accounting (CPA Canada); WAM: 83, GPA: 3.20/4</i>	Ningbo, China 09/2014 – 07/2018

RESEARCH EXPERIENCE

DARPA Assured Neuro Symbolic Learning and Reasoning <i>Postdoc Research Fellow (Level B) - Supervised by Dr. Hamid Rezatofighi and, A/Prof. Yuan-Fang Li</i>	Monash University 01/2024 – Current
<ul style="list-style-type: none">...	
Content-Driven Multimodal Deepfake Generation and Temporal Localization <i>PhD Student - Supervised by Dr. Munawar Hayat, A/Prof. Abhinav Dhall, and Dr. Kalin Stefanov</i>	Monash University 03/2021 – 02/2024
<ul style="list-style-type: none">Deepfake uses advanced deep learning and computer vision for forgery, but poses risks despite its potential. Content-driven deepfakes, which fake only part of a video, have received less attention. The PhD project will create and detect content-driven deepfakes, preparing for their potential threat to society.	
EEG-Guided Explainable Deepfake Audio Detection <i>Research Assistant - Supervised by Dr. Mahsa Salehi, Dr. Kalin Stefanov and Dr. Ehsan Shareghi</i>	Monash University 06/2023 – 12/2023
<ul style="list-style-type: none">This project is about the correlation analysis between EEG and deepfake audio. The goal is to find the correlation between EEG and deepfake audio, and use the correlation to detect the deepfake audio.	
DARPA Computational Cultural Understanding <i>Research Assistant - Supervised by Dr. Hamid Rezatofighi and Dr. Munawar Hayat</i>	Monash University 03/2022 – 07/2023
<ul style="list-style-type: none">This program develops language technologies for the Department of Defense operators to improve cross-cultural understanding, situational awareness, and interactional effectiveness. It requires minimal training data and aims to assist negotiators and analysts in the field.As the team member in Subteam2: Culture-Aware Emotion Recognition, I mainly focus on two tasks: 1) Doing research and developing the better approach for the multimodal emotion valence arousal prediction task. 2) Deploy the approach as docker container to the DARPA websocket-based system.	
Data Generation Techniques for Automatic Facial Expression Recognition <i>Master Student - Supervised by A/Prof. Abhinav Dhall and Dr. Kalin Stefanov</i>	Monash University 12/2019 – 12/2020
<ul style="list-style-type: none">This project creates new face images with controlled expressions using neural networks, improving expression recognition and addressing bias towards certain cultures.	

SELECTED PUBLICATIONS

- Zhixi Cai**, Shreya Ghosh, Aman Pankaj Adatia, Munawar Hayat, Abhinav Dhall, and Kalin Stefanov. AV-Deepfake1M: A Large-Scale LLM-Driven Audio-Visual Deepfake Dataset. Submitted to ACM MM 2024. [[Paper](#)] [[Code](#)]
- Zhixi Cai**, Shreya Ghosh, Kalin Stefanov, Abhinav Dhall, Jianfei Cai, Hamid Rezatofighi, Reza Haffari, and Munawar Hayat. MARLIN: Masked Autoencoder for facial video Representation LearnINg. In CVPR 2023. [[Paper](#)] [[Code](#)]
- Zhixi Cai**, Shreya Ghosh, Tom Gedeon, Abhinav Dhall, Kalin Stefanov, and Munawar Hayat. Glitch in the Matrix: A Large Scale Benchmark for Content Driven Audio-Visual Forgery Detection and Localization. In CVIU 2023. [[Paper](#)] [[Code](#)]
- Zhixi Cai**, Kalin Stefanov, Abhinav Dhall, and Munawar Hayat. Do You Really Mean That? Content Driven Audio-Visual Deepfake Dataset and Multimodal Method for Temporal Forgery Localization. In DICTA 2022. [**Best Paper**] [[Paper](#)] [[Code](#)]
- Shreya Ghosh*, **Zhixi Cai***, Parul Gupta, Garima Sharma, Abhinav Dhall, Munawar Hayat, and Tom Gedeon. Emolysis: A Multimodal Open-sourced Group Emotion Analysis and Visualization Toolkit. Submitted to IUI 2024. *Equally contributed. [[Paper](#)] [[Code](#)]
- Simindokht Jahangard, **Zhixi Cai**, Shiki Wen, and Hamid Rezatofighi. JRDB-Social: A Multifaceted Robotic Dataset for Understanding of Context and Dynamics of Human Interactions Within Social Groups. In CVPR 2024. [[Paper](#)]
- Fucaï Ke*, **Zhixi Cai***, Simindokht Jahangard*, Teresa Wang, Pari Delir Haghighi, Hamid Rezatofighi. HYDRA: A Hyper Agent for Dynamic Compositional Visual Reasoning. Submitted to ECCV 2024. *Equally contributed. [[Paper](#)]

TEACHING EXPERIENCE

FIT5221 - Intelligent image and video analysis (2022 Semester 1) <i>Head Teaching Assistant - Student Satisfaction 93%</i>	Monash University 03/2022 – 07/2022
ITO5221 - Intelligent image and video analysis (2023 TP5) <i>OUC & OLA (Head Teaching Assistant)</i>	Monash University 09/2023 – 10/2023
ITO5221 - Intelligent image and video analysis (2022, 2023, 2024) <i>Course Designer</i>	Monash University 03/2022 – 02/2024

SKILLS

Programming: Python, R, JavaScript, Typescript, Java, Kotlin, Scala, Bash, C#, C
Tool: PyTorch, TensorFlow, Keras, Latex, Slurm, SQL, Spark, sklearn, skimage, pandas, markdown, git, matplotlib, D3.js, Vue.js, React, Svelte, seaborn, ggplot, ...
Software: Jupyter, JetBrains IDEs, Linux, Visual Studio, Microsoft Office, Docker, VirtualBox, SQLDeveloper, Tableau

RESEARCH ACTIVITIES

Reviewer: IEEE TAFFC, IEEE TMM, ACM TKDD, INFFUS, IEEE TAI, ACM MM, ACM ICMI, MBE, DICTA, CVIP, ICVGIP.
Program Committee: ICMI 2023, REACT 2023
Talks: DICTA Oral Presentation (2022), Workshop for JMSS Immersion Day (2022), Talk for IIT Ropar (2023), Talk for Voxel51 (2023)

SELECTED PROJECTS

- REACT2023 Workshop Website** [[Website](#)] [[Code](#)] 04/2023
• A website for the REACT2023 workshop. The website is built based on the Hugo framework and deployed on the GitHub Pages.
- Multi-Tangled (Mita)** [[Demo](#)] [[Code](#)] 02/2023 – Current
• A web application for monitoring runtime status of multiple machines. This application is composed by a server (based on Scala Play framework), a web UI frontend (based on Vue3), and a client (based on Python FastAPI). The server is wrapped as a docker image on [DockerHub](#). Users can use the client library in the Python code to sync the runtime status to the server, and can observe the status from multiple clients on the web UI.
- Deepfake Demo** [[Code](#)] 10/2022
• The deepfake demo used for Monash Open Day and John Monash Science School immersion day. The demo is built based on Vue3 and PyTorch. This web-based demo shows the deepfake generation (face reenactment based on Wav2Lip), voice cloning (based on SV2TTS) and deepfake detection (based on EfficientViT). The demo is wrapped as composed several docker images on [DockerHub](#).
- Food Recognition System on Android APP** 06/2022 – 09/2022
• Research Assistant. Responsible for building a food photo recognition system for an Android app. To overcome the shortage of suitable on-the-shelf datasets, the extra data is scraped online. The model is built with TensorFlow and deployed with the TensorFlow Lite framework.
- LightloC** [[Code](#)] 08/2021 – 09/2021
• A light-weight Inversion of Control (IoC) tools by Dependency Injection (DI) for Scala. The project is published in [Maven Central](#).
- Tencent Advertisement Algorithm Competition** [[Code](#)] 05/2021 – 07/2021
• We competed in a tag prediction contest for ad videos, using deep learning techniques like NextVlad, ResNet, BERT, and Transformer on multi-modal data. I led the team in designing the models, programming, and experiments. We ranked in the top 4%.
- Neural Engine Kompanion (tensorneko)** [[Code](#)] 07/2021 - Current
• A PyTorch util library for easier preprocessing, building, training, and evaluating PyTorch model. This library is published in [PyPI](#), and has been downloaded hundreds of times per month.
- COVID-19 Data Visualization** [[Demo](#)] [[Code](#)] 11/2020
• Using HTML/CSS/JavaScript and D3.js library to visualize COVID-19 data interactively. This narrative visualization combines data science and web development techniques to finalize the achievement. The webpage has been deployed on GitHub.
- Kaggle: Predicting author gender of Twitter text** [[Code](#)] 06/2020
• Applying natural language process techniques to perform preprocessing, modelling, and predicting author's gender based on Twitter text dataset. The accuracy of the result is 85.1% and ranked 6%.
- Forecast the area burned of the Australian forest fires** [[Code](#)] 05/2020
• Using R for data mining on forest fire dataset, including exploratory data analysis, data preprocessing and machine learning modelling prediction.
- War Thunder Data Project** [[Website](#)] [[Code](#)] 11/2019 – Current
• A web-based visualization application using Python and Selenium crawlers to collect time-series statistical data for a multiplayer online game War Thunder. And, aggregate and organize the data, and then use d3.js and typescript to build the interactive website. Currently, it has 565K visits.