AI X METAVERSE

AAAI's 2023 Summer Symposium Series

On behalf of the organizing committee, we wanted to take a moment to express our gratitude for attending the symposium that took place on the 17th – 18th of July 2023, in Singapore.

Over the two days, the symposium featured workshop training, sessions of talks for researchers and practitioners to present their work, a panel discussion, and a poster and networking session.

Your presence and active participation were appreciated and helped make future events possible. We have summarized symposium information along with sharing of recording.

Keynote Speaker

Dr. Aik Beng Ng

The regional manager of NVIDIA AI Technology Center (NVAITC), and his role includes being a technologist, engineer, researcher, evangelist, and ecosystem builder. His interests and work include Augmented Intelligence, Human AI collaboration, and technological innovations in general. Together with his team, they promote Artificial Intelligence within the ecosystem through strategic partnerships and collaborations with institutes of higher learning.

NVIDIA AI Technology Center (NVAITC)

[Recording](#)

[Slides](#)

Moderator

Prof Tan Chek Tien

Ph.D. in Computer Science, National University of Singapore, Singapore. Part researcher, part educator, mostly tinkerer in Immersive Technologies with Human-Computer Interaction and Machine Learning.

Director, Centre for Immersification at Singapore Institute of Technology (SIT)

Panel Discussion

Director, Centre for Immersification at Singapore Institute of Technology (SIT)	**Prof Tan Chek Tien** Moderator Ph.D. in Computer Science, National University of Singapore, Singapore. Part researcher, part educator, mostly tinkerer in Immersive Technologies with Human-Computer Interaction and Machine Learning.
NVIDIA AI Technology Center	**Dr. Aik Beng Ng** Panelist The regional manager of NVIDIA AI Technology Center (NVAITC), and his role includes being a technologist, engineer, researcher, evangelist, and ecosystem builder. His interests and work include Augmented Intelligence, Human AI collaboration, and technological innovations in general. Together with his team, they promote Artificial Intelligence within the ecosystem through strategic partnerships and collaborations with institutes of higher learning.
EDB	**Mr Philbert Gomez** Panelist Mr Philbert Gomez is always on the move as vice president and head of Digital Industry Singapore at the Singapore Economic Development Board (EDB). His job entails inking deals and forging partnerships with global tech companies that create jobs for Singaporeans, a duty that involves daily meetings with government personnel and clients.

Chulalongkorn University	**Prof Ekapol Chuangsuwanich** Panelist Ekapol Chuangwanich is a faculty member at the Department of Computer Engineering, Chulalongkorn University, Thailand. He did his undergraduate studies at Carnegie Mellon University and received his Ph.D. degree from MIT in 2016. At Chulalongkorn University, he leads the spoken language systems group working on Thai speech and NLP-related tasks. He also works on applied machine learning for the medical domain, including X-ray report generation, mental health assessment, and bioinformatics.
Nanyang Technological University	**A/P Holden Li** Panelist Bachelor of Engineering (Honors) from NUS. In 2000, Holden enrolled at Stanford University for his graduate studies under Professor Thomas Kenny. During his PhD studies, Holden was actively involved in MEMS process development and finding suitable packaging solutions for MEMS and BioMEMS devices. He was awarded his MSc and PhD in Mechanical Engineering in 2001 and 2005, respectively. In September 2005, Holden returned to Singapore and started to lead a research team in MEMS sensors research effort in the area of MEMS R&D and reliability study. Holden is currently holding a concurrent appointment as Research Director (Micro-Sensors) in Temasek Laboratories at NTU and Professor-In-Charge for MAE Innovation Lab. He is working closely with several senior faculties in the area of microelectronics, MEMS, and BioMEMS research at LKC School of Medicine, Asian School of the Environment, and Temasek Laboratories at NTU.

Playware Studios	**Mr Lim Yau Boon**
	Panelist
	Yau Boon, a seasoned entrepreneur, speaker, and consultant, brings a unique blend of expertise from three decades in corporate executive roles and start-ups. He guides corporations in building sustainable business models through disruption management, digital transformation, and embracing digitalization. Yau Boon held key roles at IBM, SAP, and Motorola and currently serves as an adjunct lecturer. As a co-founder of IMMRSIV Inc., specializing in metaverse and XR software, he focuses on transforming enterprises for future readiness.

Panel:

Dr. Aik Beng Ng (NVIDIA), Mr Philbery Gomez (EDB),
Prof Ekapol Chuangsuwanich (CU), A/P Holden Li (NTU),
Mr. Lim Yau Boon (Playware Studios)

Moderator: Prof Tan Chek Tien

[Panel session recording]()	[Slides]()

Guest Speakers

Ngoc Nguyen

Singapore University of Technology and Design (SUTD)

Ngoc Nguyen is a Ph.D. candidate in ISTD at the Singapore University of Technology and Design (SUTD), working under the supervision of Prof. Ngai-Man Cheung.

She holds a Bachelor's degree and a Master's degree in Computer Science, both earned from the University of Information Technology (UIT), Vietnam National University - Ho Chi Minh City, Vietnam.

Her research focuses on generative models, model inversion attacks, and model inversion defense.

Topic: Privacy threats for AI models in Metaverse

The metaverse involves the collection of users' data and personal information (facial features, fingerprints, voice patterns) to train AI models to improve user engagement and experience with the metaverse. However, this raises concerns about privacy and security. In this work, we focus on Model Inversion (MI) attacks, a type of privacy threat for AI models in the metaverse, where adversaries aim to reconstruct private data targeting individuals by abusing access to an AI model. We propose to address two critical issues. Firstly, we analyze the limitations of existing identity loss functions used in MI attacks and propose an improved loss function called Logit Maximization (LOM), which better aligns with the objectives of MI. Secondly, we introduce the concept of MI overfitting and propose a novel solution called Model Augmentation (MA) to mitigate this issue. Through extensive experiments, we demonstrate that our proposed solutions significantly improve the performance of the SOTA MI attacks, achieving more than 90% MI attack accuracy. Our findings highlight the potential threats associated with the leakage of private information from AI models in the metaverse.

| [Recording](#) | [Slides](#) |

Zhuoran Zhao & Yubo Pan

National University of Singapore (NUS)

Zhuoran Zhao is currently a master's student majoring in Computer Science (Artificial Intelligence Track) at the National University of Singapore.

Previously, she received her Bachelor's degree at SCUT, majoring in Software Engineering, and has gained valuable work experience at Tencent.

Yubo Pan is a master's student in Computer Science at the National University of Singapore, specializing in Artificial Intelligence.

She collaborated on mobile robotics projects during her time at SIIX-AGT, focusing on robotic systems and pathways.

Topic: Taming Diffusion Models for Music-driven Conducting Motion Generation

Generating the motion of orchestral conductors from a given piece of symphony music is a challenging task since it requires a model to learn semantic music features and capture the underlying distribution of real conducting motion. Prior works have applied Generative Adversarial Networks (GAN) to this task, but the promising diffusion model, which recently showed its advantages in terms of both training stability and output quality, has not been exploited in this context. This paper presents Diffusion-Conductor, a novel DDIM-based approach for music-driven conducting motion generation, which integrates the diffusion model into a two-stage learning framework. We further propose a random masking strategy to improve the feature robustness and use a pair of geometric loss functions to impose additional regularizations and increase motion diversity. We also designed several novel metrics, including Frechet Gesture Distance (FGD) and Beat Consistency Score (BC), for a more comprehensive evaluation of the generated motion. Experimental results demonstrate the advantages of our model.

Recording Slides

Prof Tong Rong

Singapore Institute of Technology (SIT)

Dr Tong Rong is an assistant professor at the Singapore Institute of Technology (SIT), Infocomm Technology (ICT) Cluster.

She is a recipient of the Inauguration Grant of the MOE Singapore Teaching and Academic Research Talent (START) Award 2022.

Currently is a senior member of IEEE and a council member of the Chinese and Oriental Languages Information Processing Society (COLIPS).

Research interests include multilingual speech and language processing, computer assistant language learning, automatic speech recognition, language and speaker recognition, and machine learning.

Topic: Multilingual Aphasia speech analysis with machine learning

Aphasia is an acquired language disorder that occurs after a brain injury, such as a stroke, head trauma, or tumor. People with aphasia (PWA) may have trouble speaking or understanding speech. If diagnosed early, aphasia is often treatable, and communication can be improved with speech therapy. Early detection and evaluation of aphasia is crucial for treatment and recovery. This paper reports a preliminary study of multilingual aphasia speech evaluation. In this study, the characteristics of speech from PWA and healthy controls are compared from both acoustic and linguistic perspectives. Multiple acoustic features are extracted from aphasic and normal speech to build a language-independent aphasic speech detection model. The model achieved good aphasic speech detection performance on both English and Mandarin test sets.

Recording	Slides

Leon Foo Cewei

Singapore Institute of Technology (SIT)

Leon Foo Cewei is a Research Engineer at the Singapore Institute of Technology, SIT. He has worked on a range of virtual reality-related projects that involve the use and creation of digital twin environments.

As a member of SIT's Centre of Immersification, he participates regularly in helping host and set up social events for knowledge sharing and networking for both academics and industry guests.

Topic: Progressive 3D reconstruction for collaborative construction of digital twins

Digital twins are increasingly useful in metaverse-related applications, but their construction process usually involves deep technical expertise and costly resources. This position paper demonstrates an early prototype of a web-based digital twin authoring system to enable untrained users to collaboratively build digital twin environments. The system aims to explore the combination of photogrammetry and GAN-based machine learning models to enable near-real-time collaboration between capture client (scanning objects using common smartphone cameras) and editing client (constructing 3D scenes on thin client devices) users. Symposium attendees will be able to experience an actual

hands-on demonstration of the current prototype and consequently engaging in meaningful exchanges on related areas of machine learning and digital twins.

Recording   Slides

Prof Immanuel Koh

Singapore University of Technology and Design (SUTD)

Dr Immanuel Koh is the Hokkien Foundation Chair Professor and Assistant Professor in Architecture and Sustainable Design (ASD) and Design & Artificial Intelligence (DAI) at the Singapore University of Technology & Design (SUTD).

He directs Artificial-Architecture -- a research laboratory with expertise in the design and development of deep learning models for generative architecture, AI art and aesthetics, neuroarchitecture, creative metaverse, and predictive urbanism.

Topic: Your Memory Palace in the Metaverse with AI

The metaverse is often conceptualized as an immersive environment where individuals congregate as avatars to perform activities in imaginary places, similar to how they would in reality. Although promising, it has been unsuccessful in maintaining novel experiences, with current renditions of the metaverse looking unappealing, devoid of human touch, and lacking in genius loci. The objective of this paper is to construct a framework for a metaverse that is built directly from human experiences such that it can become a deeply personalized and meaningful metaverse. This proposed framework leverages wearable technology and artificial intelligence (AI) to generate virtual spaces based on the conscious and subconscious experiences of individuals. The methodology demonstrates the feasibility of incorporating EEG (electroencephalogram), eye tracking, and photogrammetry technologies to capture and transfer human experiences into the metaverse. The eye tracker registers points of interest and generates gazemap signatures, while the EEG registers emotional valence and generates brainwave spectrograms, which in turn can be used to further validate the eye tracking outputs. Both gazemaps and spectrograms were utilized as inputs to train two image-to-image conditional generative adversarial network models (pix2pix). Once trained, the models were able to generate new virtual scenes with either the gazemaps or spectrograms alone as inputs. This proposed personalized metaverse can act as an ever-growing archive of memories and perceptions distinctly synthesized by its user, not unlike a memory palace.

[Recording](#)	Slides

Prof Alex Binder

Singapore University of Technology and Design (SUTD)

Prof Alexander Binder is currently an Associate Professor at SIT, Singapore. His research interests are in explainable deep learning methods such as LRP, deepfake detection, and aspects of machine learning in general.

Topic: A look into the capabilities of LRP explanations to identify relevant channels in top-randomized deep neural networks

We perform a preliminary analysis of the question of whether Layerwise Relevance Propagation is capable of finding channels in intermediate feature maps of deep neural networks that support regions in the input sample with positive relevance scores for the case when weights of the top-most layers in the neural network are randomized. We perform this analysis also for the output logits.

[Recording](#)

[Slides](#)

Yeo Jin Qi

NVIDIA AI Technology Center (NVAITC)

Jin Qi is an Assistant Solutions Architect of NVIDIA AI Technology Centre (NVAITC). Her team helps bring NVIDIA's AI research, with aspects of Digital Twins technology, within the ecosystem.

Her research interests and work lie in the intersection of Artificial Intelligence and VRAR, with a focus on NeRF representations for 3D perception. She is currently pursuing her PhD research on "NeRF Perception: Coordinate-Based Neural Representations for 3D

Perception", in collaboration with SIT.

Topic: PNP: A Novel Framework Towards Efficient NeRF Study and Research

Rapid advancements in techniques and applications of NeRF technologies have been seen in both academia and industry recently. It becomes apparent that there is a need to consolidate and organize the theoretical and practical knowledge on the end-to-end pipeline of using NeRF technology. To facilitate the study and research on NeRF technologies, we propose a novel review framework that identifies the key stages for NeRF-related technology, namely Pre-NeRF, NeRF, PostNeRF, or "PNP" in short. Each stage has four attributes: An overview of what and how NeRF technology is adapted at that stage, the common challenges faced, exemplary toolkits that are used to handle the tasks, and how it is applicable to the next stage in the framework.

[Recording](#)	[Slides](#)

Kavi Jayathunge

University of Glasgow (UOG)

Kavi Jayathunge holds an MEng in Electronics and Software Engineering from the University of Glasgow. He is working towards his PhD in Computer Vision, specifically an unsupervised approach to categorizing human emotions from audiovisual input.

Part of this research also involved investigating audio augmentation as a method for increasing the proportions of underrepresented emotions in audiovisual datasets, which led to the publication of this conference.

Topic: AdSMOTE: A Technique for "High Proportion" Audio Augmentation

Data augmentation is a practice that is widely used in the fields of machine and deep learning. It is used primarily for its effectiveness in reducing the generalization gap between training and validation, as well as to artificially increase available training data points. This is particularly relevant to audio datasets, which are usually smaller and suffer from imbalanced classes in some applications. This work presents adSMOTE (audio SMOTE), a novel sampling and augmentation strategy, and also compares it to Specaugment, one of the most effective augmentation strategies for audio data. We show that our method outperforms the latter by a considerable margin when the proportion of synthetic training samples is high. We also provide source code for the complete algorithm, which can easily be integrated into an existing model, enabling the rapid development of augmentation frameworks.

Recording

Slides

Dr Chee Kwang Quah

Singapore Institute of Technology (SIT)

Dr. Chee Kwang QUAH received his BEng from the University of Leeds, England, in 1996, an MSc from the University of Surrey, England, in 1997, and a PhD in Computer Engineering from Nanyang Technological University, Singapore, in 2008.

He is currently an academic staff at the Singapore Institute of Singapore (SIT). He has about 20 years of R&D experience working in the areas of computer vision, 3D human modeling and reconstruction, motion tracking and analysis, augmented reality, artificial intelligence, computer graphics, and applications to interactive media, smart cities, education, defense, and sports.

Topic: A Portable Vision-based Head Tracking Exergame Solution for Neck Rehabilitation

The current practice of neck rehabilitation is done through laser-pointing devices in a physical setup or Virtual Reality head-mounted devices, whereby the patient will control and navigate a laser beam to follow a specific pattern. These setups are cumbersome to the patient. We propose a portable computer vision-based marker-less head tracking exergame for neck rehabilitation. Our system retargets head postures to control a game object and considers neck mobility while incorporating movement modulation during posture overcompensation. This system would be tailored to the specific patient in the steps of automatic calibration and gameplay interaction. We used the head and body pose to calibrate the game scene dimensions and the upper body joint angles to feedback the posture overcompensation in the form of gameplay variables augmentation. From our testing, it has shown that this is a viable exergame solution that could be adopted for neck exercise.

Recording	Slides

Tan Teck Jun

Singapore Institute of Technology (SIT)

Tan Teck Jun is a passionate Computer Science undergraduate with a strong desire to make a mark in the global games industry. My journey started during Primary School when I first encountered programming, igniting a love for it that propelled me to pursue a diploma in Games Design and Development. Then, further honed his skills through his ongoing degree program.

Topic: Towards Decontamination Facility Training in VR

This paper presents the ongoing development of a Proof-of-Concept Virtual Reality (VR) system for staff training for Tan Tock Seng Hospital mass decontamination of Hazardous and Toxic Materials (HAZMAT) and/or radioactively contaminated casualties. As a Proof-of-Concept, the initial prototype focuses on allowing users to practice using a chemical scanner device that would be used to scan casualties for leftover chemical hazards. The preliminary development demonstrates the advantages of utilizing VR to simulate training scenarios for supplementing current decontamination training in a safe and economical way.

Recording | Slides

Prof Nisha Jain

Singapore Institute of Technology (SIT)

Dr. Nisha Jain received her M. Tech in Computer Technology and PhD in Computer Science from Indian Institute of Technology (IIT) Delhi. During her postdoc, she worked as a Research fellow at the Institute for Media Innovation at National Technological University (NTU), Singapore.

Currently, Dr Nisha is an Assistant Professor at the Singapore Institute of Technology (SIT) in the InfoComm Technology cluster.

Her research interests include Human-Computer Interaction, Virtual Reality, Augmented Reality, Mixed Reality, and Computer Graphics.

Topic: Extensible VR Emergency Preparedness Platform

Conventionally, emergency preparedness is carried out through a regimen of training - drills, and tabletop exercises. Emergency preparedness through the use of VR-based training has been a key application of immersive simulation. We propose an extensible immersive emergency preparedness platform that can be extended to various emergencies in different environments. We exemplify our framework for fire emergency preparedness and showcase how some parts of the platform can be tweaked to extend to other situations. Symposium attendees will be introduced to the different components of our Virtual Emergency Preparedness Platform with short videos, especially for fire emergencies.

| [Recording](#) | [Slides](#) |

Prof Ekapol Chuangwanich

Chulalongkorn University (CU)

Ekapol Chuangwanich is a faculty member at the Department of Computer Engineering, Chulalongkorn University, Thailand. He did his undergraduate studies at Carnegie Mellon University and received his Ph.D. degree from MIT in 2016.

At Chulalongkorn University, he leads the spoken language systems group working on Thai speech and NLP-related tasks. He also works on applied machine learning for the medical domain, including X-ray report generation, mental health assessment, and bioinformatics.

Topic: An Application for Mental Health Monitoring using Facial, Voice, and Questionnaire Information

Depression is a major societal issue. However, depression can be hard to self-diagnose, and people suffering from depression often hesitate to consult with professionals. We discuss the design and initial testings of our prototype application that performs depression detection using multi-modal information such as questionnaires, speech, and face landmarks. The application has an animated avatar asking questions concerning the users' well-being. To perform screening, we opt for a 2-stage method, which first predicts individual HAM-D ratings for better explainability, which may help facilitate the referral process to medical professionals if required. Initial results show that our system archives

0.85 Marco-F1 for the depression detection task.

Recording | Slides

Le Hoang Anh

Singapore University of Technology and Design (SUTD)

Le Hoang Anh is a master's student at the Singapore University of Technology and Design (SUTD), currently enrolled in the Masters of Science in Security by Design (MSSD).

He has obtained a BEng. in mechanical engineering from the University of Technology, Vietnam National University in Ho Chi Minh City.

His research interests are computer vision, artificial intelligence, and deep learning algorithms. His most recent research work is in the field of deepfake detection using foundation models.

Topic: Few-shot deepfake detection via foundation model

Deepfakes have the potential to significantly impact the metaverse, e.g., impersonation and spread of disinformation. To mitigate the negative impacts, it is important to detect deepfakes at the early stage. However, early detection of deepfakes under few shots has not been adequately studied. In this paper, we address this research gap and explore an approach based on vision foundation models (FMs), leveraging outstanding generalization capability of vision FMs. Our detection models are based on CLIP (Radford et al. 2021), and we explore designs for zero-shot and few-shot detection. We achieve encouraging results. Using only 16 real/ deepfake training samples, we achieved over 70% detection accuracy for a design based on CLIP ViT-B32. To the best of our knowledge, our work is the first study to explore vision foundation models for deepfake detection.

[Recording](#)

Slides

Summary and Thoughts

AAAI Symposium provides a Platform to drive ecosystem engagement on AI x Metaverse topics with a diverse and inclusive lineup in both agenda and representations from academia, industry, government (EDB Singapore), and entrepreneurs (working on active commercial metaverse projects, etc.).

Symposium over two days featuring: Workshop, Panel, 13 research and technical talks + 5 posters, 49 registered attendees locally, regionally, internationally from UK and US

Research talks touched on various aspects of topics from Generative AI to Metaverse and also on the intersection of AI and Metaverse.

Panel on the first day sparked a productive discussion at the intersection of AI and the metaverse. It featured diverse perspectives from platform and solution providers, academia, and governing agencies. Here were some top questions from the audience:

- What do you think about the conflict between AI fieldwork and human creators on intellectual property?

- What are your thoughts on the current ecosystem for AI or Metaverse development?

- What are some core competitive advantages of the AI/Metaverse industry in Singapore compared to the US or China?

A key takeaway from the panel was the unanimous recognition of the high demand for the convergence of these two domains. It was emphasized that we must prioritize the development of this capability to meet the growing demand effectively.

[Closing remarks](#)

[Symposium pictures](#)

[Event Link](#)

Committee Members

A/P Indriyati Atmosukarto (Singapore Institute of Technology)
Dr Simon See (Senior Director, NVAITC, NVIDIA)

Dr Aik Beng Ng (Regional Manager, NVAITC, NVIDIA)
Megani Rajendran (Solutions Architect, NVAITC, NVIDIA)

A/P Daniel Wang Zhengkui (Singapore Institute of Technology)
A/P Cheng Lock Donny Soh (Singapore Institute of Technology)
A/P Chek Tien Tan (Singapore Institute of Technology)