YUWEI SUN

yuwei_sun@araya.org | +81-8081165839
https://yuweisunn.github.io Tokyo, Japan

EDUCATION

The University of Tokyo Tokyo, Japan

Ph.D., Information Science and Technology (Hons.) GPA: 4.0/4.0 04-2021 ~ 03-2024

Minor: International Graduate Program of Innovation for Intelligent World

Thesis: Localized Learning and Generalization in Artificial Neural Networks with Properties of the Global Workspace (Best thesis award) Supervisor: Hideya Ochiai

Research Focus: NeuroAI, AI Security and Privacy

M.S., Information and Communication Engineering (Hons.) GPA: 3.84/4.0 04-2019 ~ 03-2021

Honor: Department Chair's Award

Thesis: Intrusion Detection Based on Distributed Trustworthy Artificial Intelligence

Research Focus: Multimodal Models, Decentralized Neural Networks

Post-Graduate Research Program, Graduate School of Information Science and Technology 10-2018 ~ 03-2019

Research Focus: Decentralized Neural Networks

North China Electric Power University Beijing, China

B.S., Computer Science and Technology (Hons.) 09-2014 ~ 08-2018

Thesis: Attacks on Deep Learning Systems Based on Generative Adversarial Networks

Research Focus: Computer Vision

EXCHANGE EXPERIENCES

Massachusetts Institute of Technology Cambridge, MA, US

Fellow of the Advanced Study Program, Graduate School of Engineering 02-2020 ~ 05-2020

Courses: Distributed neural circuits, Underactuated robotics, Blockchain

University of Pennsylvania Philadelphia, PA, US

Visiting Student 08-2019 ~ 10-2019

Waseda University Tokyo, Japan

Visiting Student 10-2016 ~ 08-2017

EMPLOYMENT

Araya Research Tokyo, Japan

Senior Researcher (Supervisor: Ryota Kanai) 04-2024 ~ Present

- Research on disentangled representation learning in Vision Transformers

Research Assistant, Moonshot Project 04-2023 ~ 03-2024

- Researched on Hopfield networks and associative memory for Transformers long-term memory

RIKEN Center for Advanced Intelligence Project Tokyo, Japan

Visiting Scientist, AI Security and Privacy Team (Supervisor: Jun Sakuma) 04-2024 ~ Present

- Research on the interpretability and security of multimodal models

PhD Student Researcher, AI Security and Privacy Team 04-2021 ~ 03-2024

- *Researched on adversarial attacks on neural networks*

The University of Tokyo Tokyo, Japan

Research Fellow (JSPS DC2), Japan Society for the Promotion of Science 04-2022~ 03-2024

Research Assistant, Graduate School of Information Science and Technology 04-2020 ~ 03-2022

- Researched on knowledge transfer and associative memory in multimodal models

United Nations University Tokyo, Japan
AI Consultant, Computing Centre (Supervisor: Ng Chong) 05-2021 ~ 06-2022
- Researched on domain adaptation in multimodal models

Research Intern 06-2020 ~ 12-2020
- Performed research on decentralized neural networks

RESEARCH GRANTS

Previous
- Microsoft Research Asia Collaborative Research Program (D-CORE 2023 with MSRA Beijing), JPY1270k, 2023-2024
- Japan Society for the Promotion of Science, Grant-in-Aid for JSPS Fellows, JPY1700k, 2022-2024
- Japan Science and Technology Agency, SPRING GX program, JPY340k, 2021-2022

SELECTED PUBLICATIONS

Journals
- **Yuwei Sun**, Hideya Ochiai, and Jun Sakuma. Attacking Distance-Aware Attack: A Semi-Targeted Poisoning Attack on Federated Learning. ***IEEE Transactions on Artificial Intelligence****. 2023.*
- **Yuwei Sun** and Hideya Ochiai. Homogeneous Learning: Self-Attention Decentralized Deep Learning. ***IEEE Access****, Vol.10, pp.7695-7703. 2022.*
- **Yuwei Sun**, Hideya Ochiai, and Hiroshi Esaki. Decentralized Deep Learning for Multi-Access Edge Computing: A Survey on Communication Efficiency and Trustworthiness. ***IEEE Transactions on Artificial Intelligence****, Vol.3, No.6, pp.963-972. 2022.*
- **Yuwei Sun**, Hideya Ochiai, and Hiroshi Esaki. Adaptive Intrusion Detection in the Networking of Large-Scale LANs with Segmented Federated Learning. ***IEEE Open Journal of the Communications Society****, Vol.2, pp.102-112. 2020.*

Conferences
- **Yuwei Sun**, Ippei Fujisawa, Arthur Juliani, Jun Sakuma, and Ryota Kanai. Remembering Transformer for Continual Learning. ***CVPR Workshop*** *2024*.
- **Yuwei Sun** and Hideya Ochiai. Bidirectional Contrastive Split Learning for Visual Question Answering. ***AAAI*** *2024.*
- **Yuwei Sun**, Hideya Ochiai, and Jun Sakuma. Instance-Level Trojan Attacks on Visual Question Answering via Adversarial Learning in Neuron Activation Space. ***IJCNN*** *2024*.
- **Yuwei Sun**, Hideya Ochiai, Zhirong Wu, Stephen Lin, and Ryota Kanai. Associative Transformer is a Sparse Representation Learner. ***NeurIPS Workshop*** *2023.*
- **Yuwei Sun**. Meta Learning in Decentralized Neural Networks: Towards More General AI. ***AAAI Doctoral Consortium*** *2023.*
- **Yuwei Sun**, Ng Chong, and Hideya Ochiai. Feature Distribution Matching for Federated Domain Generalization. ***ACML*** *2022.*
- **Yuwei Sun**, Hideya Ochiai, and Jun Sakuma. Semi-Targeted Model Poisoning Attack on Federated Learning via Backward Error Analysis. ***IJCNN*** *2022.*
- **Yuwei Sun**, Hideya Ochiai, and Hiroshi Esaki. Intrusion Detection with Segmented Federated Learning for Large-Scale Multiple LANs. ***IJCNN*** *2020.*

HONORS AND AWARDS

- Best PhD Thesis Award, The University of Tokyo, 2024
- WBAI Incentive Award, Whole Brain Architecture Initiative, 2023
  (Press release: https://wba-initiative.org/en/wbaiia_awardees)
- AAAI Student Travel Grant, 2023
- Department Chair's Award, The University of Tokyo, 2021
- Heiwa Nakajima Foundation Scholarship, 2021
- Graduate Student Scholarship, The University of Tokyo, 2019
- COMAP Mathematical Contest in Modeling, 2015

SKILLS

Programming: Python (Advanced), PyTorch (Advanced), Tensorflow (Advanced), OpenCV (Advanced), Linux commands (Intermediate), Git (Intermediate), Docker (Intermediate), SQL (Intermediate), HTML (Intermediate), JavaScript (Elementary), C++ (Elementary), Java (Elementary)
AI Research Computer: RAIDEN by Fujitsu in RIKEN Center for Advanced Intelligence Project
Languages: Chinese (native), English (TOEFL IBT 101/120), Japanese (JLPT N1 169/180)

OTHER ACTIVITIES

Invited Talks

- August 2024, RIKEN AIP – SJTU CS Joint Workshop on Machine Learning and Brain-like Intelligence, "Exploring Priors and Long-Term Memory in Transformers".
- April 2024, International Research Center for Neurointelligence (IRCN), the University of Tokyo, "Localized Learning and Generalization in ANNs with Properties of the Global Workspace".
- August 2023, Consciousness Research Network, "Localized Learning Through the Lens of Global Workspace Theory".
- April 2023, MBZUAI and RIKEN-AIP Joint Workshop on Intelligent Systems, "Meta Learning in Decentralized Neural Networks Through the Lens of Global Workspace Theory".
- Feb 2023, Victoria University of Wellington,
- "Meta Learning in Decentralized Neural Networks Through the Lens of Global Workspace Theory".
- Nov 2022, MIT Department of Brain and Cognitive Sciences, "Meta Learning and Modularity Towards Systematic Generalization".

Academic Services

- Reviewer: IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Artificial Intelligence, Neural Networks, Engineering Applications of Artificial Intelligence, ICLR, NeurIPS, CVPR, ICML, IJCNN, ACML, AISTATS
- Volunteer for NeurIPS 2021, ICLR 2023
- Organizer for the NeurIPS NeuroAI Social 2023 *(https://neurips.cc/virtual/2023/social/80638)*