* *Title of the proposed special session*.

**Simulating Edge Computing and Multimodal AI (SEAMAI): A Benchmark for Real-World Applications**

* *Name, affiliation, brief biography and contact information for each of the organizers.*

**Minh-Son Dao**

National Institute of Information and Communications Technology, Japan

He is a senior researcher and research manager of the National Institute of Information and Communications Technology (NICT), Japan. He received his Ph.D. from the University of Trento, Italy. He was a senior assistant professor (Universiti Teknologi Brunei), researcher (Trento University, Italy), and JSPS postdoc (Osaka University, Japan). His main interests include multimedia retrieval, event detection, video surveillance, data science, computer vision, and pattern recognition. He has led several national and international projects under the Society 5.0 framework. He is a PC member, (co)-chair, (co)organizer, and (co)editor in several international conferences/workshops and journals. He is a member of the MediaEval Community Council (2021), a well-known multimedia benchmarking challenge event. He is also one of the founders of the CheapFake Detection Challenge (MMSys2021, ACMMM2022, ICME2023, ICMR2024)

**Koji Zettsu**

National Institute of Information and Communications Technology, Japan

Dr. Koji Zettsu is a Director of Information Services Platform Laboratory at Universal Communication Research Institute of National Institute of Information and Communications Technology (NICT), Japan. He was a technical lead of knowledge cluster systems project, a distributed knowledge system on multi-site grid networks, from 2006 to 2010. He is a visiting associate professor of Kyoto University, Osaka University and Nara Institute of Science and Technology. He was also a visiting researcher of Christian-Albrechts-University Kiel, Germany in 2009. He received Ph.D. in Informatics from Kyoto University, Japan in 2005. He was in IBM Yamato Software Laboratory from 1992 to 2003. His research interests are information retrieval, databases, data mining, and software engineering. He was the technical editor of Value-creating Network sub-working group of New Generation Network Forum, Japan from 2009 to 2010.

* *A session abstract including significance justification and a brief overview of the state-of-the-art of the proposed special session topic.*
  + *Note: The session abstract should be in a format that can be copied directly to the conference web-page to advertise the session.*

Multimodal AI has attracted people for decades. The explosion of data processing demands from the Internet of Things (IoT) and future communication technologies (5G/6G) has brought the opportunity to understand complex events and phenomena more comprehensively. These events/phenomena exhibit characteristics that can be observed from different perspectives across varying spatial and temporal dimensions. Multimodal AI is specifically designed to process and analyze this kind of multimodal data. It aims to combine various types of data inputs to deliver comprehensive results. Traditionally, training multimodal AI involved gathering all data in a central location, such as cloud computing, for model training. This approach is well-suited for processing large volumes of non-time-sensitive data. However, cloud computing is not ideal for scenarios where data is constantly recorded, updated, and requires real-time processing on near-edge devices (e.g., low-latency IoT applications). Additionally, strict data privacy regulations create significant barriers to transferring data from local devices to other servers or clouds. Finally, the cost of large and powerful servers or cloud resources can be prohibitive for some organizations and individuals. Edge computing emerges as a solution that addresses these challenges. In edge computing, edge clients and servers act as nodes in a distributed network, sharing the training workload to alleviate the burden on a centralized server. In simpler terms, edge computing moves computational data, applications, and services away from centralized cloud servers and towards the network's edge. When combined with multimodal AI, this approach has led to the development of many valuable applications for various sectors, such as smart transportation, smart healthcare, and smart homes.

While the combination of multimodal AI and edge computing offers a promising solution for many applications, the lack of a robust simulation environment to evaluate their performance remains a significant obstacle. Not everyone has the resources to set up a full-scale edge computing environment for multimodal AI evaluation. Additionally, the absence of standardized benchmarks, where different solutions can be compared under controlled conditions, poses another challenge. To address this critical challenge, NICT, with the support of MIC, Japan, is creating a benchmark for edge computing and multimodal AI. This benchmark will include a simulation environment, along with the following components: 1) MM-sensing: A multimodal AI designed to process data from dashcams, Fitbits, and IoT devices, primarily for detecting traffic risks. 2) AOP: An Adaptive Offloading Point Approach in Federated Learning that can optimally determine the point at which data/parameters are uploaded from edge clients to the edge server, and 3) A dataset containing dashcam video, Fitbit data, and IoT device data.

The benchmark is open to participants who can utilize the simulation environment to explore the following research questions in the context of multimodal AI and edge computing:

1. Training process:
   1. Convergence of learning through iterative model transfer, update, and aggregation, and hyperparameter adjustment
   2. Improved performance of horizontal/vertical federated learning for heterogeneous client data (e.g., client selection, model aggregation improvement, local model correction by estimating the direction of updating the global model)
2. Infrastructure:
   1. Efficient handling of large amounts of traffic during model transfer and aggregation
   2. Efficient handling of opportunistic computing resources (e.g., differences in client performance, participation, and withdrawal)
3. Security:
   1. Poisoning of models and data by malicious clients
   2. Backdoor attacks by malicious clients
   3. Inverse estimation of client private data from models

Participants are encouraged to use the provided MM-sensing, AOP, and datasets to find the answer to any of the challenges mentioned above. However, participants can discuss with NICT, Japan, if they would like to utilize their own multimodal AI model and dataset.

Participants will be granted permission to utilize the simulation environment for two months with full support from NICT, Japan.

The benchmark schedule unfolds as follows:

* May 1st: Registration opens. Participants must register on the benchmark website to receive their account credentials for accessing the simulation environment.
* May 15th: The environment becomes accessible to all registered participants.
* July 15th: Participants submit their experiences and results.
* July 22nd: Submission deadline for special session papers. From this point onward, the calendar will align with the MMM2025 calendar.

The length of the paper submitted is expected to be 6 pages, in LCNS format, where participants express their discoveries and answers for any challenges mentioned above.

* *List of invited papers, if applicable, including for each paper: tentative title, author list, and preferably a short abstract*.
* *Proposal for arrangement of the reviews from session organizers (e.g. list of reviewers for the special session)*.

**Academic Reviewers**

Do-Van Nguyen, National Institute of Information and Communications Technology, Japan

Anh-Khoa Tran, National Institute of Information and Communications Technology, Japan

Cathal Gurrin, Dublin City University, Ireland

Michael Riegler, Oslo Medicine University, Norway

Duc-Tien Dang-Nguyen, Bergen University, Norway

Mianxiong Dong, Muroran Institute of Technology, Japan

Uday Kiran, Aizu University, Japan

Cong-Thang Truong, The university of Aizu, Japan

Takahiro Komamizu, Nagoya University, Japan

Hamza Rafik, Tokyo International University, Japan

The-Bao Pham, Saigon University, Vietnam

Thanh-Binh Nguyen, Vietnam National University in HCM City

Hanh-Nhi Tran, University of Toulouse Paul Sabatier, France

**Technical Support Chief**

Sadanori Ito, National Institute of Information and Communications Technology, Japan

* *Description of the session format (e.g. panel, technical talks, poster session).*

**The special session will follow a two-part format:**

* **Part 1: Oral Presentations:** Participants will deliver presentations detailing their experience and results obtained while participating in the benchmark. This will offer valuable insights into the effectiveness of the provided simulation environment.
* **Part 2: Demonstration Session:** Participants will have the opportunity to showcase their developed applications running live within the simulation environment provided by the benchmark. This interactive session allows attendees to witness the practical implementation of the presented solutions. [Optional]
* *Plans for advertising the special session (e.g. targeted distribution lists, projects, communities).*

The special session call for papers (CFP) will be disseminated through various channels, including those managed by the special session organizers and their colleagues (such as connectionist groups, multimodal mailing lists, computer vision and machine learning mailing lists), social networks (such as LinkedIn), the organizers' affiliated websites, and conferences attended by the organizers, where leaflets and brochures will be distributed.

* *If applicable, plans for exploitation of the results of the special session (e.g. summary papers).*

The overview paper for the benchmark will be authored by the organizers to consolidate the collective experiences of all participants, along with discussing the benefits and merits the benchmark can offer to participants. Additionally, promising applications and models developed using the benchmark resources will be considered for further development through collaboration with NICT.