All the papers accepted will appear in the proceedings with no distinctions except the tracks to wish they have been submitted. The conference is in-person. Although we understand some colleagues could not attend for different reasons, we have decided to give priority to authors with at least 1 author physically on site. For all the papers, the 5-minute videos submitted will serve as poster presentations.

Monday June 30: Workshops and Tutorials (9:00AM-6:00PM)

10:00AM-10:30AM   Coffee Break

3:00PM-3:30PM   Coffee Break

Workshops:

1.   **LSC'25: The 8th Annual ACM Workshop on the Lifelog Search Challenge**

Main Contact: Dr. Cathal Gurrin

Location: Dorin Forum Main Hall C

2.   **CEFSW'25: The 2nd Collaboration and Evolution of Foundation and Specialized Models Workshop**

Main Contact: Dr. Shengyu Zhang

Location: Dorin Forum Main Hall D

3.   **ICDAR'25: The 6th Workshop on Intelligent Cross-Data Analysis and Retrieval**

Main Contact: Dr. Takahiro Komamizu

Location: Dorin Forum Main Hall E

4.   **MAD'25: The 4th ACM International Workshop on Multimedia AI against Disinformation**

Main Contact: Dr. Dan-Cristian Stanciu

Location: Dorin Forum Main Hall F

<u>**Tutorials:**</u> **(Location: Dorin Forum Main Hall G)**

9:00AM-1:00PM: Visual Content Generation in the Era of Large Foundation Models

        Presenters: Leigang Qu, National University of Singapore, Singapore

        Fei Shen, Nanjing University of Science and Technology, Nanjing, China

        Zhenglin Zhou, Zhejiang University, Hangzhou, China

        Jiayi Lyu, University of Chinese Academy of Sciences, Beijing, China

        Wenjie Wang, University of Science and Technology of China, Hefei, China

        Lu Jiang, ByteDance, San Jose, USA

2:00PM-6:00PM: Hallucinations in Large Language Models and Large Vision-Language Models

        Presenters: Liqiang Jing, Yue Zhang, Xinya Du (University of Texas at Dallas, USA)

Tuesday July 1: Main Conference (Location: Dorin Forum Main Hall C)

8:45AM-9:00AM   Opening Remarks

9:00AM-10:00AM   Keynote 1

Unveiling Bias and Safety Issues in Generative Models

Nicu Sebe, University of Trento

10:00AM-10:30AM   Coffee Break

10:30AM-12:20PM   Full Paper Session 1：Neighborhood and 3D

ArtNVG: Content-Style Separated Artistic Neighboring-View Gaussian Stylization

Zixiao Gu (Fudan University), Zhenye Zhang (Fudan University), Mengtian Li (Shanghai University), Zhongxia Ji (Shanghai Theatre Academy), Ruhua Chen (Shanghai Theatre Academy), Zuo Hu (Shanghai Theatre Academy), Guangnan Ye (Fudan University)

PIG: Physically-Based Multi-Material Interaction with 3D Gaussians

Zeyu Xiao (Fudan University), Zhenyi Wu (Fudan University), Mingyang Sun (Fudan University), Qipeng Yan (Fudan University), Yufan Guo (Fudan University), Zhuoer Liang (Fudan University), Lihua Zhang (Fudan University)

OT-Talk: Animating 3D Talking Head with Optimal Transportation

Xinmu Wang (Stony Brook University), Xiang Gao (Stony Brook University), Xiyun Song (Futurewei Technologies), Heather Yu (Futurewei Technologies), Zongfang Lin (Futurewei Technologies), Liang Peng (Futurewei Technologies), Xianfeng Gu (Stony Brook University)

COMAE: COMprehensive Attribute Exploration for Zero-shot Hashing

Yuqi Li (Computer Network Information Center, Chinese Academy of Sciences), Qingqing Long (Computer Network Information Center, Chinese Academy of Sciences), Yihang Zhou (Computer Network Information Center, Chinese Academy of Sciences), Meng Xiao (Computer Network Information Center, Chinese Academy of Sciences), Ran Zhang (Computer Network Information Center, Chinese Academy of Sciences), Zhiyuan Ning (Computer Network Information Center, Chinese Academy of Sciences), Zhihong Zhu (Peking University), Xuezhi Wang (Computer Network Information Center, Chinese Academy of Sciences), Yuanchun Zhou (Computer Network Information Center, Chinese Academy of Sciences)

12:20PM-2:00PM   Lunch Break

2:00PM-3:00PM   Panel Session

3:00PM-3:30PM   Coffee Break

3:30PM-5:15PM   Short Paper Session 1：Multimedia Compressions, Modeling and Retrieval

Optimization of CLIP Models for Domain-Specific Video Search

Kazuya Ueki (Meisei University), Haruki Sato (Agoop Corporation), Yuma Suzuki (SoftBank Corporation), Takayuki Hori (SoftBank Corporation), Hiroki Takushima (SoftBank Corporation), Takumi Takada (SB Intuitions Corporation), Hayato Tanoue (SoftBank Corporation), Aiswariya Manoj Kumar (SoftBank Corporation), Hiroki Nishihara (SoftBank Corporation), Yuki Shibata (SoftBank Corporation)

Audio-Visual Driven Compression for Low-Bitrate Talking Head Videos

Riku Takahashi (HOSEI University), Ryugo Morita (HOSEI University), Jinjia Zhou (HOSEI University)

Mask-aware Text-to-Image Retrieval: Referring Expression Segmentation Meets Cross-modal Retrieval

Li-Cheng Shen (National Taiwan University), Jih-Kang Hsieh (National Taiwan University), Wei-Hua Li (National Taiwan University), Chu-Song Chen (National Taiwan University)

GroupAC: Inter-Group Context Modeling for Point Cloud Attribute Compression with RAHT

Guangjie Zhang (School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University), Chunyang Fu (City University of Hong Kong), Qiang Xu (School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University), Shan Liu (Media Lab, Tencent), Ge Li (School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University)

6:00PM-8:00PM   Welcome Reception (Location: Dorin Forum Main Hall C)

9:00AM-10:00AM   Keynote 2

The Power of "Why?" in Decision Making in Complex, Dynamic Systems

K. Selçuk Candan, Arizona State University

10:00AM-10:30AM   Coffee Break

10:30AM-12:20PM   Full Paper Session 2: Learning, Ranking and Prediction

Few-Shot Learning with Class-Number Non-Aligned Training and Cross-Scale Feature Differential Network for Hyperspectral Image Classification

Pan He (Xiangtan University), Bodong Li (Xiangtan University), Han Xiang (Xiangtan University), Bowen Xu (Xiangtan University), Chunhong Cao (Xiangtan University)

Label Ranker: Self-aware Preference for Classification Label Position in Visual Masked Self-Supervised Pre-trained Model

Peihao Xiang (Florida International University), Ou Bai (Florida International University)

KEGNN: Knowledge-Enhanced Graph Neural Networks for User Engagement Prediction

Ching-Hao Fan (New Jersey Institute of Technology), Hao Zhou (New Jersey Institute of Technology), Yao Sun (New Jersey Institute of Technology), Geovanny Palomino Roldan (New Jersey Institute of Technology), Olga Kokshagina (EDHEC Business School), Marc Santolini (Learning Planet Institute), Lijing Wang (New Jersey Institute of Technology)

Towards Interpretable User Intent Analysis with Deficient Evidence Fusion for Pseudo-Modalities

Chaochen Wu (Renmin University of China), Guan Luo (Institute of Automation, Chinese Academy of Sciences), Meiyun Zuo (Renmin University of China)

12:20PM-2:00PM   Lunch Break

2:00PM-4:05PM   Full Paper Session 3：Diffusion Models, Video search and Analysis

Evaluate the Generative Capability of Diffusion Models from a Discriminative Perspective

Yixuan Jiang (Cornell University), Hsiao-Dong Chiang (Cornell University), Yiqing Shen (Johns Hopkins University),

LLAUS: A High-Quality Instruction-Tuned Large Vision Language Assistant for UltraSound

Junhao Guo (Zhejiang University), XueFeng Shan (Xinjiang Medical University), Guoming Wang (Zhejiang University), Dong Chen (Zhejiang University), Rongxing Lu (University of New Brunswick), Siliang Tang (Zhejiang University)

RagMe: Retrieval Augmented Video Generation for Enhanced Motion Realism

Elia Peruzzo (University of Trento), Dejia Xu (University of Texas at Austin), Xingqian Xu (UIUC), Humphrey Shi (Georgia Tech | Oregon | UIUC | PAIR), Nicu Sebe (University of Trento)

Robust Relevance Feedback for Interactive Known-item Video Search

Zhixin Ma (Singapore Management University), Chong-Wah Ngo (Singapore Management University)

TourMLLM: A Retrieval-Augmented Multimodal Large Language Model for Multitask Learning in the Tourism Domain

Hiromasa Yamanishi (The University of Tokyo), Ling Xiao (The University of Tokyo), Toshihiko Yamasaki (The University of Tokyo)

4:05PM-4:35PM   Coffee Break

4:35PM-5:25PM   Brave New Idea

Inverse Farthest Point Sampling (IFPS): A Universal and Hierarchical Shell Representation for Discrete Data

Nayu Ding (Donghua University), Yujie Lu (Donghua University), Yao Huang (Donghua University), Long Wan (Donghua University), Yan Zhao (Donghua University), Zhijun Fang (Donghua University), Shen Cai (Donghua University), Lin Gao (Institute of Computing, Chinese Academy of Sciences)

Knowledge Discovery in Fuzzy Linguistic Triadic Context: Mining Data Hidden under Conditions

Yujie Cao (Shandong jianzhu University), HongPing Liu (Shandong jianzhu University), Benshuai Wang (Shandong jianzhu University), Li Zou (Shandong Jianzhu University)

6:30PM-9:30PM   Conference Banquet (boarding time 6:30pm)

Location: Navy Pier – City Cruise (600 E Grand Ave, Chicago, IL 60611, USA)

We will provide one-way shuttle service from the conference center (leaving time at 5:45 PM) to the banquet venue at Navy Pier. Alternatively, you may choose to travel to the banquet on your own.

9:00AM-10:00AM   Keynote 3

Three and a Half Generations of Video Generation Models
Sergey Tulyakov, Snap Inc

10:00AM-10:30AM   Coffee Break

10:30AM-12:20PM   Full Paper Session 4:

GraphDC: Detecting and Confusing in Node Injection Attack

Jialong Wang (University of Macau), Shilong Zhang (University of Macau), Zhiguo Gong (University of Macau)

SSD-Poser: Avatar Pose Estimation with State Space Duality from Sparse Observations

Shuting Zhao (Academy for Engineering & Technology, Fudan University), Linxin Bai (Academy for Engineering & Technology, Fudan University), Liangjing Shao (Academy for Engineering & Technology, Fudan University), Ye Zhang (College of Vocational and Technical Teacher Education, Shanghai Polytechnic University), Xinrong Chen (Academy for Engineering & Technology, Fudan University)

Multimodal and Multilingual Fact-Checked Article Retrieval

Stefanos-Iordanis Papadopoulos (ITI-CERTH), Ivana Beňová (KInIT), Sebastian Kula (KInIT), Michal Gregor (KInIT), George Karantaidis (ITI-CERTH), Tomáš Javůrek (KInIT), Marian Simko (KInIT), Symeon Papadopoulos (ITI-CERTH)

Towards Comprehensive Legal Document Analysis: A Multi-Round RAG Approach

Wutong Zhang (Shanghai Jiao Tong University), Hefeng Zhou (Shanghai Jiao Tong University), Qiang Zhou (Shanghai Data Group), Yunshen Li (Shanghai Data Group), Yuxin Liu (Shanghai Jiao Tong University), Jiong Lou (Shanghai Jiao Tong University), Chentao Wu (Shanghai Jiao Tong University), Jie Li (Shanghai Jiao Tong University)

12:20PM-2:00PM   Lunch Break

2:00PM-3:00PM   Reproducibility Track

Reproducibility Companion Paper: AdOCTeRA: Adaptive Optimization Constraints for Improved Text-guided Retrieval of Apartments

Ali Abdari (University of Udine, University of Naples Federico II), Alex Falcon (University of Udine), Giuseppe Serra (University of Udine)

Reproducibility Companion Paper: u-LLaVA: Unifying Multi-Modal Tasks via Large Language Model

Jinjin Xu (OPPO research institute), Xilu Wang (University of Surrey), Liwu Xu (OPPO research institute), Yunfan Hu (Amazon), Yuzhe Yang (OPPO research institute), Xiang Li (OPPO research institute), Fanyi Wang (OPPO research institute), Yanchun Xie (OPPO research institute), Yi-Jie Huang (OPPO research institute), Yaqian Li (OPPO research institute)

Reproducibility Companion Paper: Learning Differentiable Particle Filter on the Fly

Jiaxi Li (University of Surrey), Xilu Wang (University of Surrey), Yunfan Hu (Amazon)

3:00PM-3:30PM   Coffee Break

3:30PM-4:50PM   Short Paper Session 2: Multimedia Modeling and Analysis

DSSM-KG: Dual-Stream State-Space Modeling with Adaptive Knowledge Injection for Video Captioning

Haoying Sun (Beijing University of Technology), Shuyi Li (Beijing University of Technology), Zeyu Xi (Beijing University of Technology), Bowen Zhang (Beijing University of Technology), Lifang Wu (Beijing University of Technology)

AI Got Your Tongue? Analysing the Sounds of Audio Deepfake Generation Methods

Karla Schäfer (Fraunhofer SIT, ATHENE)

Multimodal Contrastive Learning for Music with Incomplete Modalities

Koto Nakata (Hiroshima University), Koji Eguchi (Hiroshima University)

Latent Sensor Fusion: Multimedia Learning of Physiological Signals for Resource-Constrained Devices

Abdullah Ahmed (University of Massachusetts Amherst), Jeremy Gummeson (University of Massachusetts Amherst)