ACMMM 2026 Grand Challenge Proposal: Machine-oriented Visual Media Quality Assessment

Chunyi Li[1,2], Zicheng Zhang[1], Chris Wei Zhou[3], Guangtao Zhai[1,2], Weisi Lin[4], Roger Zimmermann[5]
Shanghai Artificial Intelligence Laboratory[1], Shanghai Jiao Tong University[2], Cadiff University[3],
Nanyang Technological University[4], National University of Singapore[5]
lichunyi@pjlab.org.cn, zhangzicheng@pjlab.org.cn

Abstract

With the development of Embodied AI, machines have replaced humans as the main consumers of visual media, yet existing Image Quality Assessment (IQA) metrics remain focused on human and overlooked machine preference (a domain where quality depends on task utility rather than perceptual fidelity alone). To address this gap, we introduce the Machine-oriented Image Quality Assessment (MoIQA) Challenge, emphasizing simulation comprehension and Real-world exectution. The challenge comprises two tasks: 1) MoIQA-Sim, primarily evaluating the consistency between IQA models and the performance of mainstream Vision-Language Models (VLMs) in simulation software; and 2) MoIQA-Real, focusing on whether IQA models match the results of the latest Vision-Language-Action Models (VLAs) in the Real-world. This challenge aims to advance reliable image understanding for machine preference and support robust Embodied AI applications.

Keywords

Multimedia Signal Processing, Image Quality Assessment, Embodied AI, Vision-Language-Action Model.

ACM Reference Format:
Chunyi Li[1,2], Zicheng Zhang[1], Chris Wei Zhou[3], Guangtao Zhai[1,2], Weisi Lin[4], Roger Zimmermann[5]. 2026. ACMMM 2026 Grand Challenge Proposal: Machine-oriented Visual Media Quality Assessment. In *Proceedings of the 34th ACM International Conference on Multimedia (MM '26), November 10–14, 2026, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3746027.3758239

1 Introduction

The rapid ascend of Embodied Intelligence and the flourish of the Internet of Things (IoT) over the past decade have fundamentally transformed the framework of the application end. According to the Cisco white paper, the number of Machine-to-Machine (M2M) connections first exceeded that of Machine-to-Human (M2H) in 2023, reaching 147 billion. Machine have gradually replaced Human Intelligence and become the primary consumers of visual data. In the field of visual signal processing, the primary goal is to enhance the quality of processed images, ensuring they conform to human perception. However, 'What images do machines prefer' remains an open question. Since there is a huge gap between Human Visual System (HVS) and Machine Visual System (MVS), humans are sensitive to distortions such as noise and compression, which do not affect the downstream tasks of machines. Brightness and contrast are opposites, which have little impact on the Quality-of-Experience (QoE) of human, but can easily cause errors in Embodied AI during Manipulation and Navigation. Past human-oriented signal processing methods cannot fit the characteristic of Embodied Intelligence. Therefore, for the next generation of visual streaming media, a new Image Quality Assessment (IQA) paradigm is becoming a challenging and demanding topic. In this challenge, based on our machine preference datasets and benchmarks, we hope to explore how to implement Machine-oriented Image Quality Assessment (MoIQA) with international scholars. Thus promoting the evolution of the multimedia information processing community, from a human-centered to a machine-centered approach.

2 Relevance to the MM community

Current MM community (e.g. ACM SIGMM, IEEE SPS) are at the forefront of multimedia research, with recent years witnessing significant advances in human-oriented visual media processing. The community has established influential benchmarks for image/video quality assessment and actively contributes to international standards through ITU-T SG21. However, current standardization efforts and research initiatives remain predominantly human-centric, focusing on perceptual fidelity and QoE for human consumers. As Embodied AI and autonomous systems proliferate, this human-centric paradigm increasingly misaligns with the actual consumption patterns of visual media, creating both a critical gap in the community's research portfolio and an urgent standardization challenge. The MoIQA Challenge addresses a critical inflection point in the evolution of multimedia systems: the transition from human-centric to machine-centric visual consumption. This paradigm shift carries profound implications for the ACM Multimedia community, industry stakeholders, and broader society over the next 3–5 years.

Redefining Quality for the Machine-Centric Era: By 2027, over 75% of visual data will be consumed by machines rather than humans—from autonomous vehicles to robotic systems. However, conventional human-oriented IQA metrics correlate poorly with machine task performance (Spearman < 0.3), creating urgent demand for new evaluation frameworks. This challenge compels the MM community to address foundational questions: What constitutes "quality" when the consumer is a Machine? How should we optimize codecs for robotic vision rather than human perception? MoIQA provides the structured benchmarks necessary to develop these machine-centric multimedia systems.

MM '26, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3758239

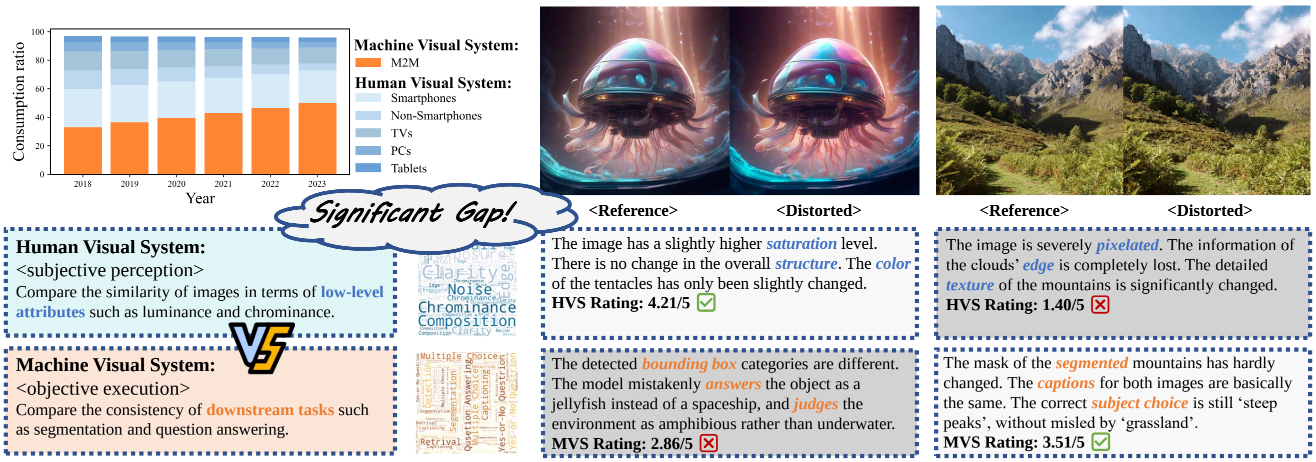

Figure 1: The significant gap between the well-explored Human Vision System (HVS) and the emerging Machine Vision System (MVS). Their different perception mechanisms leading to images that are subjectively satisfactory to humans not being applicable to the machines downstream tasks such as detection and question answering, vice versa.

Bridging Multimedia and Embodied AI: The MM community possesses deep expertise in signal processing and compression increasingly vital to Embodied Intelligence. Yet a significant gap exists between multimedia researchers and roboticists deploying VLMs/VLAs in the Real-world. MoIQA serves as a crucial bridge, establishing common benchmarks that translate visual quality optimization into measurable robotic task improvements. Success here will position multimedia researchers at the forefront of the autonomous systems revolution.

Industrial and Societal Imperatives: Industrial applications span autonomous driving, smart manufacturing, and cloud robotics—sectors where efficient, machine-optimized visual streaming directly impacts system reliability and operational costs. As global standards bodies formulate AI-oriented multimedia guidelines, MoIQA datasets will provide essential technical foundations. Societally, reliable machine visual comprehension is critical for safe autonomous systems in healthcare and public spaces, while optimized machine-oriented codecs offer pathways to more sustainable AI deployment by reducing computational overhead.

In Conclusion, introducing this topic at ACMMM 2026 positions the multimedia community at the forefront of this transformation, fostering cross-disciplinary collaboration between signal processing, computer vision, and robotics researchers. The dual-track structure ensures accessibility while maintaining scientific rigor, cultivating expertise for the machine-centric multimedia future.

3 SOTA Technique Review

Human-oriented IQA has been a cornerstone theme at ACMMM for decades, with the community achieving remarkable progress in human-centric quality evaluation. Traditional methods, from pixel-based metrics like PSNR and SSIM to deep learning approaches such as NIMA and MUSIQ, have established robust correlations with human perceptual scores, often exceeding Spearman correlations of 0.9 [1-5] on standard benchmarks like LIVE and KADID-10k. These advances have directly informed international standards including ITU-T recommendations and JPEG/HEVC coding optimization. The human-oriented IQA paradigm is now considered well-solved for most practical applications, with mature toolchains deployed across streaming media, photography, and broadcast industries.

However, for Machine-oriented IQA, these well-established human-oriented techniques fundamentally fail when applied to machine consumers. The divergence stems from radically different perceptual mechanisms: the HVS is sensitive to structural distortions, color fidelity, and semantic coherence, whereas MVS [6-9] prioritize task-relevant features, robustness to compression artifacts, and compatibility with network architectures. For instance, noise and block artifacts that severely degrade human QoE have negligible impact on VLM performance; conversely, subtle brightness variations imperceptible to humans can cause catastrophic failures in robotic navigation. Current human-oriented metrics show poor correlations with VLM/VLA task accuracy, rendering them unsuitable for the emerging machine-oriented multimedia ecosystem.

Accelerating Machine-oriented MM Research: The academic community has increasingly recognized this critical gap — the MachineIQA work reached **Full-Marked Ratings in CVPR2025** [6] as a pressing open problem, yet no comprehensive solutions or benchmarks exist. Meanwhile, industry standardization efforts such as **MPEG's Image and Video Coding for Machines** (I/VCM) initiative have begun addressing machine-oriented compression, but these approaches remain limited to simplistic downstream tasks (e.g., object detection, segmentation) following a fragmented "one-task-one-model" paradigm. They lack unified representations of holistic machine preferences that generalize across diverse Embodied tasks and model architectures. The MoIQA Challenge directly addresses these limitations by establishing the first large-scale benchmarks aligned with both VLMs in simulation and VLAs in Real-world execution. By providing standardized datasets and evaluation protocols, we will accelerate foundational research toward next-generation IQA systems, inform emerging I/VCM standards with comprehensive machine preference models, and position the MM community at the forefront of this technological transition.

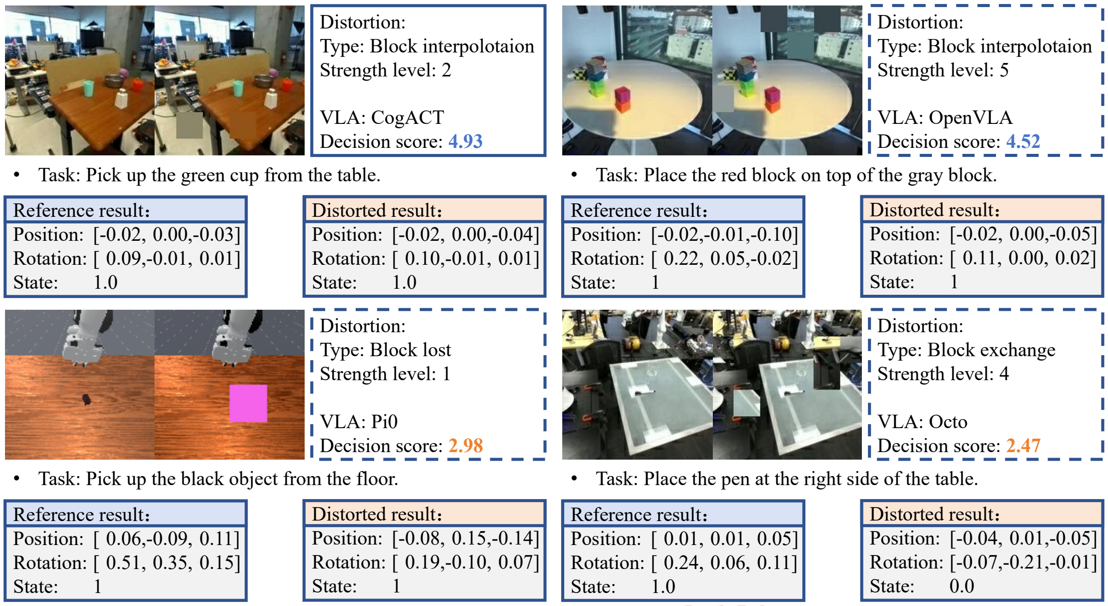

Figure 2: Positive and negative cases in MoIQA-Sim/Real Dataset in this challenge. Slight distortion may significantly affect the inference result of Embodied AI, while severe distortion may not. Emphasizing the significance of Machine-oriented IQA.

4 Challenge Task

Our challenge consists both Simulation and Real-world Tasks, and awards will be given for each task separately.

4.1 Simulation Comprehension IQA (Task1)

Task 1 focuses on evaluating IQA models' alignment with VLM perception in simulated environments. We will construct a comprehensive dataset comprising reference images and their distorted counterparts, covering diverse distortion types including compression artifacts, noise, blur, brightness/contrast variations, and color shifts relevant to Embodied AI scenarios. These images will be processed by at least 10 advanced VLMs (e.g., GPT-4o, Gemini, LLaVA, QwenVL, InternVL series) on standardized visual question answering and scene understanding tasks.

Ground Truth: For each distorted image, we aggregate VLM inference results across multiple models and prompts. An image receives higher scores if distortions minimally impact VLM response accuracy, and lower scores if performance degrades significantly. The final quality score is computed as the average accuracy retention ratio across all VLM-task combinations, producing a continuous scale from 0 to 5 representing machine-perceived utility.

Participant Objective: Teams will develop IQA models trained on our released training set. During evaluation, their models must predict quality scores for test images that maximize correlation with the hidden VLM-based ground truth. Success requires capturing distortion patterns that specifically affect machine comprehension rather than human perception.

4.2 Real-world Execution IQA (Task2)

Task 2 advances from simulation to physical embodiment, assessing whether IQA models predict Vision-Language-Action (VLA) execution success in real robotic systems. Due to substantial costs associated with Real-world robotic testing—including hardware deployment, maintenance, and execution time—**participation is restricted to the top 10 teams from Task 1.**

Ground Truth: Task success is determined by objective completion metrics: pick&place success rates for manipulation, and push&pull accuracy. Each visual observation is labeled with binary success/failure outcomes aggregated across multiple execution trials. The pipeline captures visual observations, processes them through participants' IQA models for quality assessment, and executes manipulation or navigation tasks using at least 10 advanced VLAs (e.g., OpenVLA, RT-2, Pi Series).

Participant Objective. Each qualifying team submits their trained IQA model weights and inference code. We deploy these models directly into our embodied intelligence pipeline featuring Franka Panda manipulators and Unitree Go2 navigation platforms. Unlike Task 1's continuous scores, Task 2 emphasizes decision-critical quality assessment—distinguishing images that enable successful physical execution from those causing failures. This requires modeling not merely VLA perception but the full perception-action loop, including temporal and dynamics-aware quality evaluation.

5 Evaluation Criteria

Task 1: MoIQA-Sim Evaluation. The MoIQA-Sim dataset comprises approximately 20,000 images with VLM-based quality annotations, partitioned into 70% training, 10% validation, and 20% testing. Participants may develop their IQA models using the official MoIQA-Sim training set supplemented by our previously released MachineIQA [6] dataset. Online validation is supported through the Codabench platform, allowing teams to monitor validation set performance during development. Final evaluation employs Spearman Rank

Correlation Coefficient (SRCC) and Pearson Linear Correlation Coefficient (PLCC) as $Eval_{sim}$:

$$Eval_{sim} = \frac{\text{SRCC}(S_{VLM}, S_{IQA}) + \text{PLCC}(S_{VLM}, S_{IQA})}{2}, \quad (1)$$

between S_{IQA} predicted scores and S_{VLM} perception ground truth. These correlation metrics are linearly mapped to a 0-100 scale to produce the Task 1 Score, with higher scores indicating stronger alignment with machine perception.

The MoIQA-Real dataset consists of approximately 300 standardized real-world environments spanning diverse objects, backgrounds, and distortion types, designated entirely as a hidden test set. Qualifying teams (top 10 from Task 1) may fine-tune their models using our EmbodiedIQA [7] dataset before submitting final weights and inference code. Given the binary nature of task success/failure outcomes in robotic execution, evaluation employs F1-Score computed on positive (success) and negative (failure) samples. The F1-Score is subsequently mapped to a 0-100 scale using the following formula to generate the F1-Score :

$$Eval_{real} = 2 \cdot \frac{\text{Precise}(S_{VLA}, S_{IQA}) \cdot \text{Recall}(S_{VLA}, S_{IQA})}{\text{Precise}(S_{VLA}, S_{IQA}) + \text{Recall}(S_{VLA}, S_{IQA})}, \quad (2)$$

where S_{VLA} denotes the binary task success labels from VLA execution (1 for success, 0 for failure), and S_{IQA} denotes the binarized quality predictions from the submitted IQA model (thresholded at the median score to produce binary labels). This measures the model's ability to discriminate between images that enable successful physical task execution versus those causing failures.

6 Link to Relevant Sites

Our MoIQA dataset is ready and the testing pipeline is already running on Codabench, so the competition can be started immediately. The relevant links are as follows:

- Codabench server: https://www.codabench.org/competitions/13658/#/pages-tab
- Task 1 Additional Dataset (MachineIQA [6]): https://hf-mirror.com/datasets/lcysyzxdxc/MachinePreference
- Task 2 Additional Dataset (EmbodiedIQA [7]): https://huggingface.co/datasets/xiaojiahao/Embodied-IQA
- Group Website: https://opencompass.org.cn/embodied-intelligence
- Official MoIQA Dataset: Released after acceptance

7 Commitment and Future Plan

Commitment to Long-Term Maintenance. The organizing team commits to publishing and maintaining a dedicated website for the MoIQA Grand Challenge at Codabench for a minimum of three years following ACMMM 2026. This website will host comprehensive challenge documentation, including detailed task descriptions, dataset download links, evaluation protocols, baseline code implementations, and leaderboard results. All datasets released through this challenge—MoIQA-Sim, MoIQA-Real, and previous MachineIQA/EmbodiedIQA—will remain publicly accessible with persistent identifiers and clear licensing terms (Creative Commons Attribution-NonCommercial-ShareAlike). We further commit to providing technical support through community forums and issue tracking systems, ensuring reproducibility and continued utility for follow-up research. Upon challenge conclusion, we will release complete training/validation/test splits with VLM/VLA ground truth annotations to facilitate future benchmark development.

Three-Year Roadmap for Research Advancement. Our future plan extends beyond the 2026 challenge cycle to foster sustained research momentum in machine-oriented multimedia quality assessment:

- Year 1 (2027): Publish challenge findings in a top-tier multimedia journal. Release MoIQA-Sim v2.0 with expanded distortions and VLMs. Establish an annual workshop at ACMMM/ICME/ICASSP/MMsys.
- Year 2 (2028): Launch MoIQA 2027 with video quality assessment and multi-modal evaluation. Collaborate with MPEG or AVS working groups to inform I/VCM standards.
- Year 3 (2029): Release an open-source toolkit for machine-oriented quality optimization integrated with ROS. Publish standardized test protocols for industry certification.

We will work closely with ACMMM organizers to publicize tasks through conference channels. We sincerely hope this competition can find new application scenarios for the classic IQA task beyond traditional multimedia service, and **promote the integration of multimedia and robotics communities.**

8 Important Dates

Our competition will follow the timeline below. For any changes, please follow the official requirements of the ACMMM 2026:

- 2026/04/07: Registration website opened; MoIQA-Sim Training dataset released.
- 2026/04/10: MoIQA-Sim Validation server onlined.
- 2026/06/01: Registration deadline.
- 2026/06/25: Submission deadline for simulation IQA code, model weights, and documentation.
- 2026/07/01: Real-world candidate teams announced, evaluation on the MoIQA-Real set begins.
- 2026/07/15: Paper submission deadline.
- 2026/07/30: Final results announced.

9 Previous Experience

Members of the organizing team have successfully hosted several high-impact challenges [10-12] at top-tier conferences, providing direct organizational foundations for Human-oriented IQA:

- NTIRE 2023 Quality Assessment of Video Enhancement Challenge, CVPR 2023. This challenge evaluated video enhancement methods in terms of fidelity, providing robust evaluation-pipeline that essential for executing IQA at scale.
- NTIRE 2024 Quality Assessment of AI-Generated Content Challenge, CVPR 2024. This challenge assessed the perceptual and semantic quality of AI-generated images, informing MoIQA emphasis on evaluating task utility.
- NTIRE 2025 XGC Quality Assessment Challenge: Methods and Results, CVPR 2025. This challenge evaluated the quality of cross-generated visual content using multimodal assessment models, offering insights from Embodied AI working environment that align closely with the goals of MoIQA.

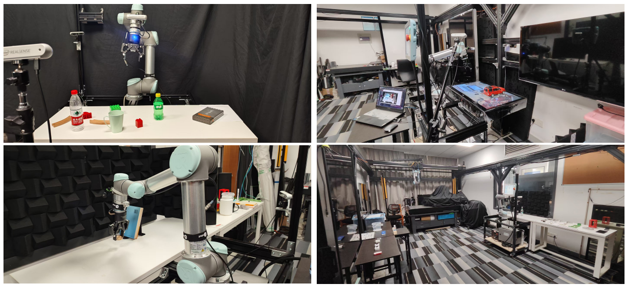

Figure 3: Real-world manipulation validation platform.

These events demonstrate the organizers' strong track record in designing datasets, building evaluation pipelines, and running large-scale competitions within the multimedia community. In addition, our organizing team will also host several events at multimedia-related conferences, including but not limited to:

- ICCV 2025 Workshop: Visual Quality Assessment Competition (VQualA), Lead: Chris Wei Zhou
- ICME 2026 Grand Challenge: Scientific Image Quality Assessment Challenge (SIQA), Lead: Zicheng Zhang
- ICME 2026 Special Session: Benchmark Evaluation and Quality Assessment in the Generative AI Era, Lead: Chunyi Li
- ISCAS 2026 Tutorial: Visual Signal Processing from Human to Embodied Intelligence, Lead: Chunyi Li

Beyond the experience in the fields of multimedia, team possesses leading capabilities in Embodied AI research and evaluation, with core contributions to the OpenCompass AI evaluation platform—one of the world's most influential open-source benchmarking systems launched at the 2025 World Artificial Intelligence Conference. The laboratory maintains an advanced infrastructure including over 20 high-performance servers and 10+ mobile robotic platforms spanning industrial manipulators, quadruped robots (Unitree Go2, B1), humanoid robots (Unitree H1), wheeled and tracked vehicles, drones, biomimetic systems, and specialized inspection robots for pipelines and rail systems. This diverse robotic fleet enables comprehensive real-world evaluation across manipulation, navigation, and human-robot interaction scenarios. With these substantial technical foundations and material resources, the organizing team is fully equipped to deliver rigorous simulation benchmarks and execute complex real-world robotic evaluations, ensuring the MoIQA Challenge sets a definitive standard for machine-oriented visual quality assessment research.

10 Organizer Bio

Our organizing team consists of **diverse genders, countries, and research institutions**, including young scholars, young faculty members, IEEE Fellows, and ACM Distinguished Member. All six members have actively participated in **ACM SIGMM, IEEE IVMSP, and ITU/ISO** events. The team has extensive experience in the multimedia society, who had multiple papers selected for the Oral Section [13-15] in ACMMM over the past five years, with one of them being nominated for Best Paper Award.

Chunyi Li, Shanghai AI Laboratory, China.
Email: lichunyi@pjlab.org.cn

Bio: Chunyi Li is leading the embodied evaluation group in the Center of AI Evaluation, Shanghai AI Laboratory. His research focuses on embodied visual signal processing and evaluation. He has published 10+ top-tier papers with 2,000+ citations. He serves as an area chair of ICME, a regular reviewer for top journals such as IEEE TPAMI/JSAC/TIP. His research has been recognized by the China Institute of Electronics Funding Program, and student travel grants from IEEE CASS and ACM SIGMM. He is a tutorial organizer in ICME/ISCAS. In the field of multimedia, his work is implemented into Chinese national (SAC/TC28/SC29 and SC42) standards.

Zicheng Zhang, Shanghai AI Laboratory, China.
Email: zhangzicheng@pjlab.org.cn

Bio: Zicheng Zhang is leading the large language model evaluation group in the Center of AI Evaluation, Shanghai AI Laboratory. He received his PhD degree from Shanghai Jiao Tong University, awarded by NSFC and CAST funding. His research focuses on perceptual quality assessment and multimodal large language models. He has published 20+ top-tier papers with 4,000+ citations. His research has been recognized by the annual best paper award of the IEEE BTS, and the best paper nomination of the ACM Multimedia.

Chris Wei Zhou, Cardiff University, United Kingdoms.
Email: zhouw26@cardiff.ac.uk

Bio: Wei Zhou is an Assistant Professor at Cardiff University, United Kingdom. Dr Zhou was a Postdoctoral Fellow at University of Waterloo, Canada. Dr Zhou is a Senior Member of IEEE and a committee member of British Standards Institution (ISO/IEC JTC 1/SC 29). Wei has led and participated in several funded projects from the USA, UK, Canada, and China, and has been an active participant in JPEG light field and point cloud standardization. She is an Associate Editor of 8 journals such as TNNLS, TOMM, PR, etc. Wei has also served as the Area Chair for 6 international conferences, such as ACM MM, ICASSP, ICME, etc., the Topic Editor and Guest Editor for many journals, such as IEEE TCSS, IEEE JBHI, etc., the Lead Special Session Chair for IEEE QoMEX, IEEE MMSP, etc., the Special Session Co-Chair (as well as the regular Session Chair) for IEEE ICIP, and the tutorial/workshop organizer for ACM MM, IROS and BIBM. She is the General Chair for QoMEX 2026.

Guangtao Zhai, Shanghai Jiao Tong University, China.
Email: zhaiguangtao@sjtu.edu.cn

Bio: Guangtao Zhai is the Distinguished Professor at Shanghai Jiao Tong University, and an adjunct Professor at the Shanghai AI Laboratory. His research interests include multimedia signal processing and Large Language Model evaluation. He has published 400+ top-tier IEEE/ACM Trans with 30,000+ citations, recognized as a highly cited researcher by Elsevier and Clarivate since 2021. He has received over 20 international awards, including the PCS Best Student Paper Award, the IEEE ICME Best Student Paper Award, the IEEE TMM Best Paper Award, and the IEEE TBC Best Paper Award. He serves as the Editor-in-Chief of the Elsevier Displays, an editorial board member of Science China: Information Sciences, and the president of the Shanghai Image and Graphics Association. He is a Fellow of IEEE.

Weisi Lin, Nanyang Technological University, Singapore.
Email: wslin@ntu.edu.sg

Bio: Weisi Lin is the President's Chair Professor and Associate Dean (Research) at the College of Computing and Data Science, Nanyang

Technological University. His research interests include image processing and perceptual signal modeling, in which he has published more than 200 journal articles, more than 230 conference papers, filed seven patents, and authored two books. He is an Honorary Fellow of Singapore Institute of Engineering Technologists. He is the General chair of the ICIP 2027, the General Co-Chair of the IEEE ICME 2025, and the Technical Program Chair of the IEEE ICME 2013, PCM 2012, QoMEX 2014, and IEEE VCIP 2017. He has been an Associate Editor of IEEE TIP/TMM/TCSVT/TNNLS/SPL. He has been a Highly Cited Researcher since 2019, and a Distinguished Lecturer for IEEE CASS (2016–2017). He is a Fellow of IEEE/IET.

Roger Zimmermann, National University of Singapore, Singapore.
Email: dcsrz@nus.edu.sg
Bio: Roger Zimmermann is a Full Professor in the School of Computing at the National University of Singapore. He is a Co-PI with the Grab-NUS AI Lab. From 2011 to 2021 he was a Deputy Director with the NUS Smart Systems Institute and from 2010 to 2016 he co-directed the Centre of Social Media Innovations for Communities, a research institute funded by the National Research Foundation of Singapore. Prior to joining NUS he held the positions of Research Area Director with the Integrated Media Systems Center and Research Assistant Professor at the University of Southern California. He is also an associate editor for the ACM TOMM, and the IEEE TMM. Previously, he served as Secretary of ACM SIGSPATIALand as associate editor for IEEE MultiMedia. He was a member and director of the IEEE Multimedia Communications Technical Committee (MMTC) Review Board. He has served on the conference programme committees of many leading conferences and as reviewer of many journals. Recently he was the general chair of the IEEE MIPR 2023, the IEEE BigMM 2019, the IEEE ISM 2015 and the ACM Multimedia Systems 2014 conferences, and TPC co-chair of the ACM Multimedia 2020, the IEEE ICME 2019 and the ACM TVX 2017 conferences. He is a Senior Member of IEEE and a Distinguished Member of ACM.

Reference

[1] **Z Zhang***, H Wu*, E Zhang, **G Zhai**, et. al., Q-bench: A benchmark for multi-modal foundation models on low-level vision from single images to pairs, IEEE TPAMI, 2025.
[2] **Z Zhang***, T Kou*, S Wang, **C Li**, et. al., Q-eval-100k: Evaluating visual quality and alignment level for text-to-vision content, CVPR, 2025.
[3] **Z Zhang***, H Wu*, Y Zhou, **C Li**, et. al., Lmm-pcqa: Assisting point cloud quality assessment with lmm, ACMMM, 2024.
[4] **Z Zhang***, Z Jia*, H Wu, **C Li**, et. al., Q-Bench-Video: Benchmark the Video Quality Understanding of LMMs, CVPR, 2025.
[5] **C Li**, **Z Zhang**, H Wu, W Sun, et. al., Agiqa-3k: An open database for ai-generated image quality assessment, IEEE TCSVT, 2024.
[6] **C Li***, Y Tian*, X Ling*, **Z Zhang**, et. al., Image Quality Assessment: From Human to Machine Preference, CVPR, 2025.
[7] **C Li***, J Xiao*, J Zhang*, F Wen, et. al., Image Quality Assessment for Embodied AI, ICLR, 2026.
[8] **C Li***, J Zhang*, **Z Zhang**, H Wu, et. al., R-Bench: Are Your Large Multimodal Model Robust to Real-World Corruptions?, IEEE JSTSP, 2025.
[9] J Zhang*, **C Li***, L Yuan, **G Zhai**, et. al., Embodied Image Quality Assessment for Robotic Intelligence, IEEE TCSVT, 2026.
[10] X Liu, X Min, **G Zhai**†, et. al., NTIRE Quality Assessment of Video Enhancement Challenge, CVPR Workshop, 2023.
[11] X Liu, X Min, **G Zhai**†, et. al., NTIRE Quality Assessment of AI-Generated Content Challenge, CVPR Workshop, 2024.
[12] X Liu, X Min, **G Zhai**†, et. al., NTIRE Quality Assessment of X-Generated Content Challenge, CVPR Workshop, 2025.
[13] T Kou, X Liu, **Z Zhang**, **C Li**, et. al., Subjective-Aligned Dateset and Metric for Text-to-Video Quality Assessment, ACMMM, 2024. (Oral, 3%)
[14] **C Li**, H Wu, **Z Zhang**, H Hao, et. al., G-refine: A general refiner for text-to-image generation, ACMMM, 2024. (Oral, 3%)
[15] **Z Zhang**, H Wu, Y Zhou, **C Li**, et. al., Lmm-pcqa: Assisting point cloud quality assessment with lmm, ACMMM, 2024. (Best Paper Nomination, 0.2%)
*Equal Conrtibution, †Corresponding Author