Quality of Experience Evaluation: Human, Data, and Agent

Guangtao Zhai
Shanghai Jiao Tong University
Shanghai AI Lab

饮 水 思 源 • 爱 国 荣 校

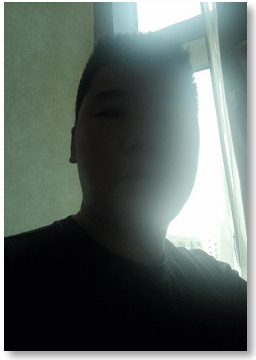
Quality of Experience (QoE) SUBJECTIVE VS. OBJECTIVE

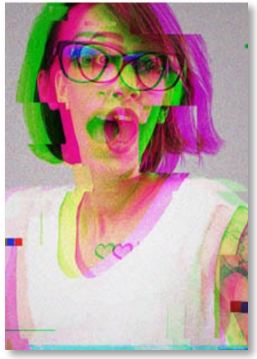
Slow
Expensive
Accurate

MOS : Mean Opinion Score

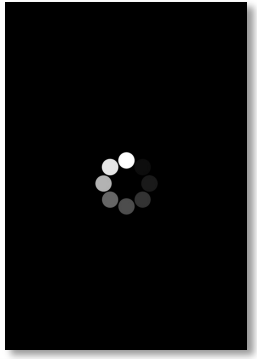








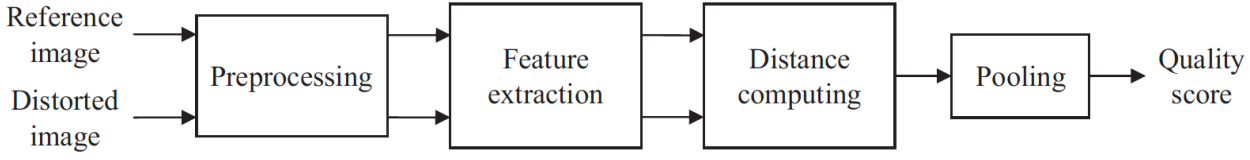




rates	clear	too dark	ghost	blur	freezing

Fast
Economical
Automatic

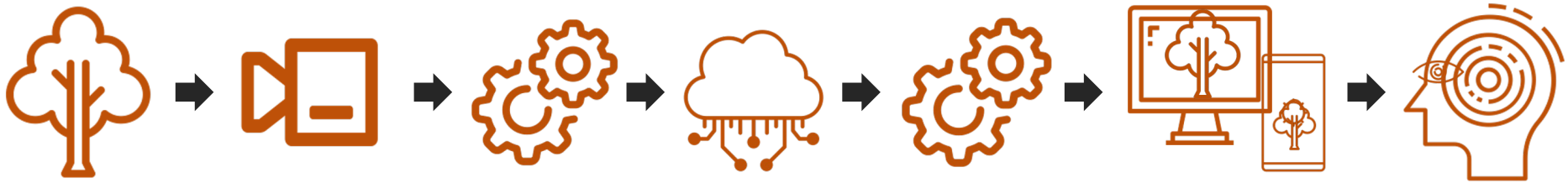




Generic Diagram

Objective quality assessment is important to visual signal processing at scale

Visual Communication System



Lighting Distance Moving etc.	Noise blur Color etc.	Quantization Spatiotemporal resolution etc.	Channel fading Package loss Bad Connection etc.	Decoder error Spatiotemporal upscaling etc.	Color Spatio- temporal resolution etc.	Vision Neuron. Environ. etc.

All steps are subject to visual distortions

Part I
Human Centric Subjective QoE Research

The Brain Mechanism for QoE

Patterns

CellPress
OPEN ACCESS

Article

Neural mechanisms of visual quality perception and adaptability in the visual pathway

Yiming Zhang,[1,3] Yitong Chen,[1,3] Ying Hu,[2,3] Xu Han,[2] Zhenhui Xie,[2] Xingrui Wang,[2] Yan Zhou,[2,*] Xiongkuo Min,[1,*] and Guangtao Zhai[1,4,*]

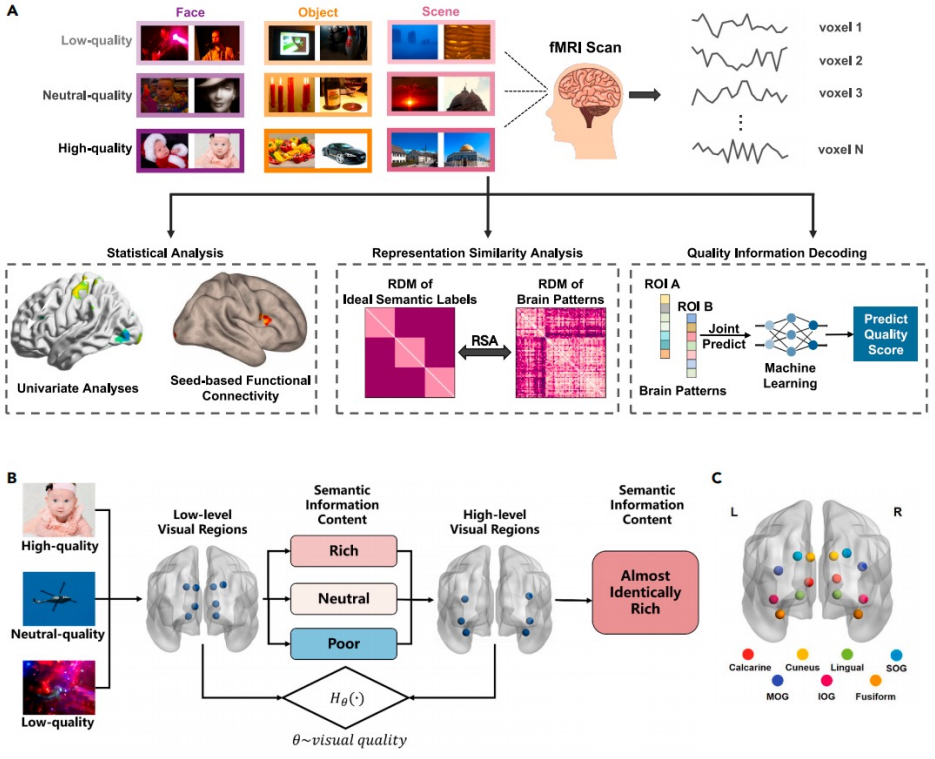
[1]Institute of Image Communication and Network Engineering, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
[2]Department of Radiology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China
[3]These authors contributed equally
[4]Lead contact
*Correspondence: clare1475@hotmail.com (Y.Z.), minxiongkuo@sjtu.edu.cn (X.M.), zhaiguangtao@sjtu.edu.cn (G.Z.)
https://doi.org/10.1016/j.patter.2025.101368

THE BIGGER PICTURE Visual quality assessment (VQA) plays a crucial role in real-world applications such as multimedia content distribution, video streaming, and computer vision, where ensuring the optimal quality of visual data directly impacts user experience and system performance. In this context, understanding the neural mechanisms behind visual quality perception is vital for enhancing the effectiveness of VQA models. This research bridges a gap in VQA by investigating how the human brain processes visual distortions of varying quality using functional magnetic resonance imaging (fMRI). The findings reveal that visual quality is not locally encoded in any single brain region but emerges from the interaction between lower-level and higher-level visual areas. This understanding challenges traditional VQA methods, which typically rely on static features, by highlighting the brain's compensatory mechanisms in handling low-quality signals.
Beyond the scope of VQA, this study holds significant implications for cognitive neuroscience and the field of human visual perception. It provides new insights into how the brain processes degraded visual stimuli, shedding light on the adaptive mechanisms that allow humans to maintain semantic understanding even when image quality is compromised. These findings deepen our understanding of the hierarchical structure of the visual system and the complex cognitive processes involved in perceiving and interpreting low-quality visual information. The study's integration of cognitive theories, such as the free energy principle (FEP), further strengthens its relevance to both neuroscience and AI. It offers a theoretical basis for the development of brain-inspired AI systems capable of mimicking human visual perception. These advancements could enhance AI's ability to process distorted or noisy data while maintaining high performance on clean images, thus providing a more accurate and adaptive approach to VQA.

SUMMARY

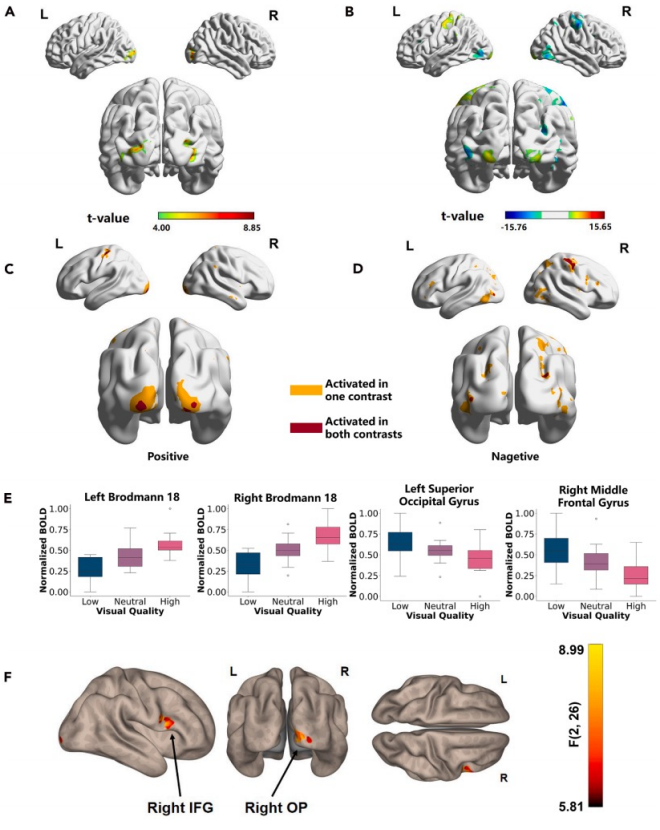
Visual quality assessment (VQA) is indispensable in multimedia for evaluating algorithm effectiveness and optimizing systems, yet its neurobiological mechanisms remain poorly understood. Using functional magnetic resonance imaging (fMRI), we investigate how the brain processes varying image qualities, revealing specialized mechanisms for handling low-quality stimuli. Results show that low quality significantly impacts semantic encoding along the visual pathway: low-level regions exhibit only 35.20% of the semantic information seen in high-quality condition, while higher-level regions compensate adaptively to maintain understanding. Visual quality is not locally encoded but emerges from inter-regional information gaps, with perception arising from this hierarchical discrepancy. Leveraging this compensatory mechanism, we decode quality from fMRI and propose a neural network feature fusion strategy, boosting ResNet's VQA performance by 14.29% on the BID dataset (586 instances). Our findings provide neurobiological evidence for degraded visual processing, addressing a gap in perception neuroscience and offering theoretical foundations for improving VQA models.

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

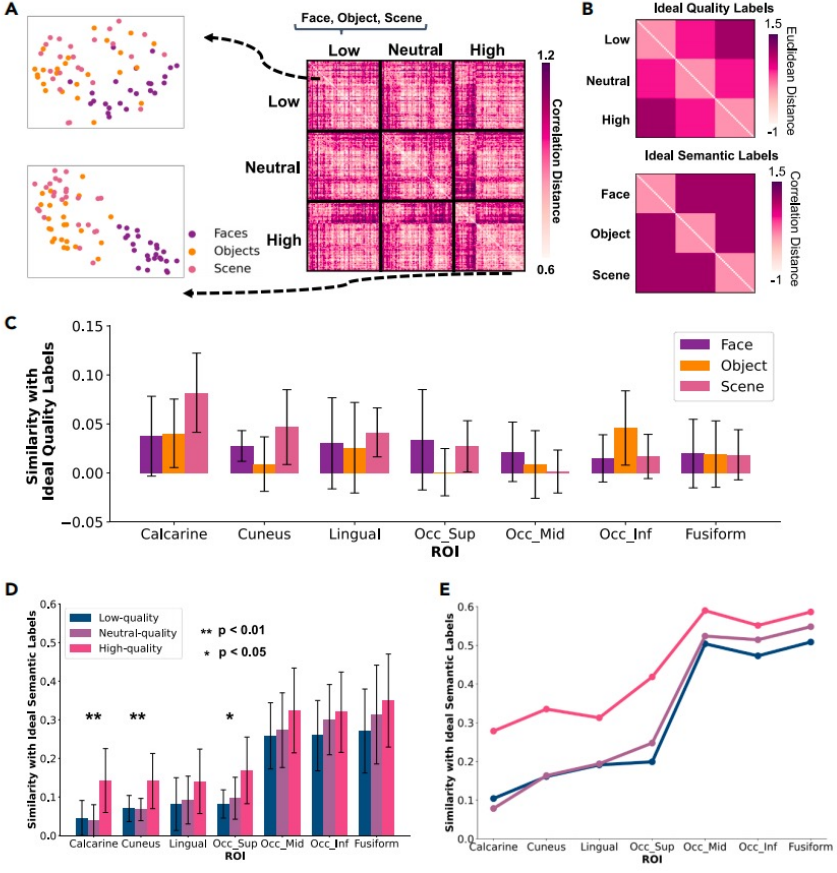


(A) fMRI signals are collected from participants observing images of varying quality
(B) The brain's processing mechanisms for images of varying visual quality

The Brain Mechanism for QoE



Statistical analysis

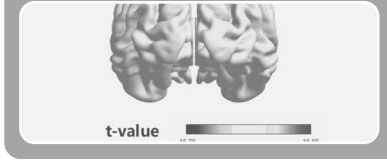


Representational similarity analysis

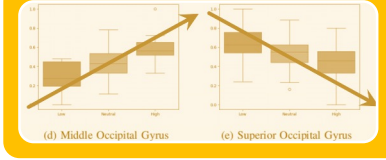
The Brain Mechanism for QoE



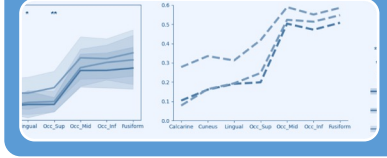
QoE Eval is not a low-level task, no specific brain area



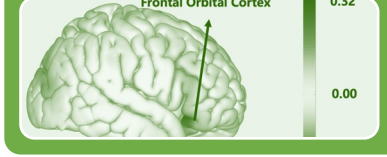
Low-quality input causes higher energy consumptions



The impact of distortions reduces in higher regions



Quality cues can be captured via indentifying info gap

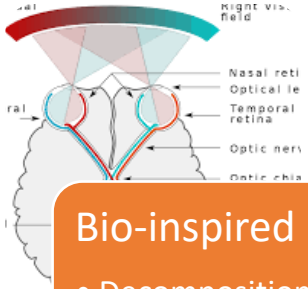


Higher brain regions compensates low quality visual input

Part II
Human Centric Objective QoE Research

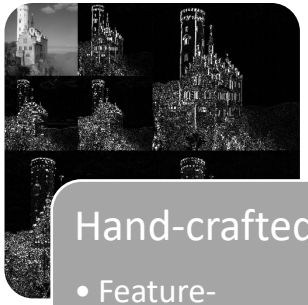
Analytic-tools Evolution

-00' 00'-10' 10'-20' 20'-



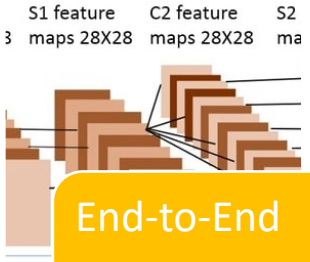
Bio-inspired

- Decomposition
- Weighting
- Synthesis



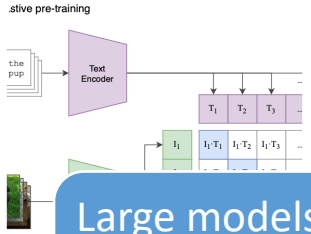
Hand-crafted

- Feature-extraction
- Optimal-combination



End-to-End

- Adaptive-sampling
- DNN structure variation



Large models

- Pre-trained
- SFT/RL
- Qualitative vs. Quantitative

Knowledge-driven Data-driven

Human Imitation

Feature Experimentation

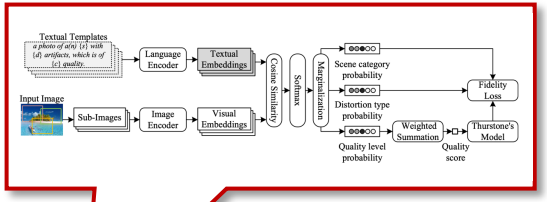
Network Adaptation

Data Curation

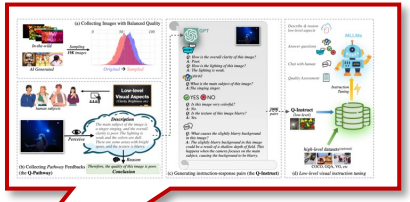
Quality assessment research method evolution

LMMs for QoE

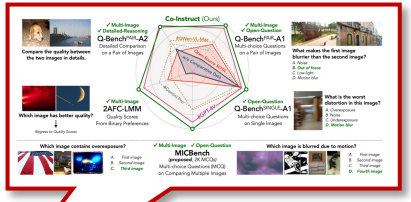
LIQE: Language-Image Quality Evaluator



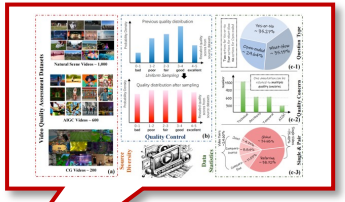
Q-Instruct: SFT for VQA



Co-Instruct: Multiple Comparison



Q-Bench-Video: VQA



ICLR '24

CVPR '23

ICML '24

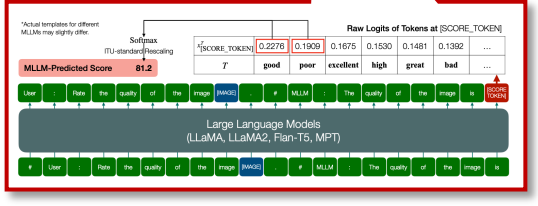
CVPR '24

TPAMI '24

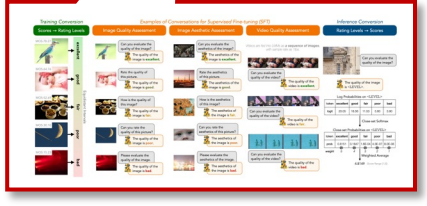
ECCV '24

CVPR '25

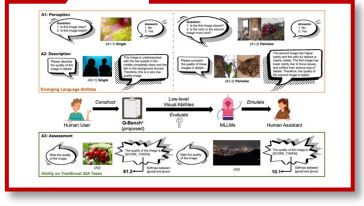
CVPR '25



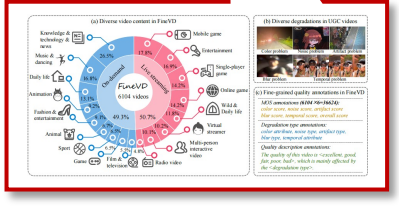
Q-Bench: Benchmarking LMMs for VQA



Q-Align: Using Discrete Levels

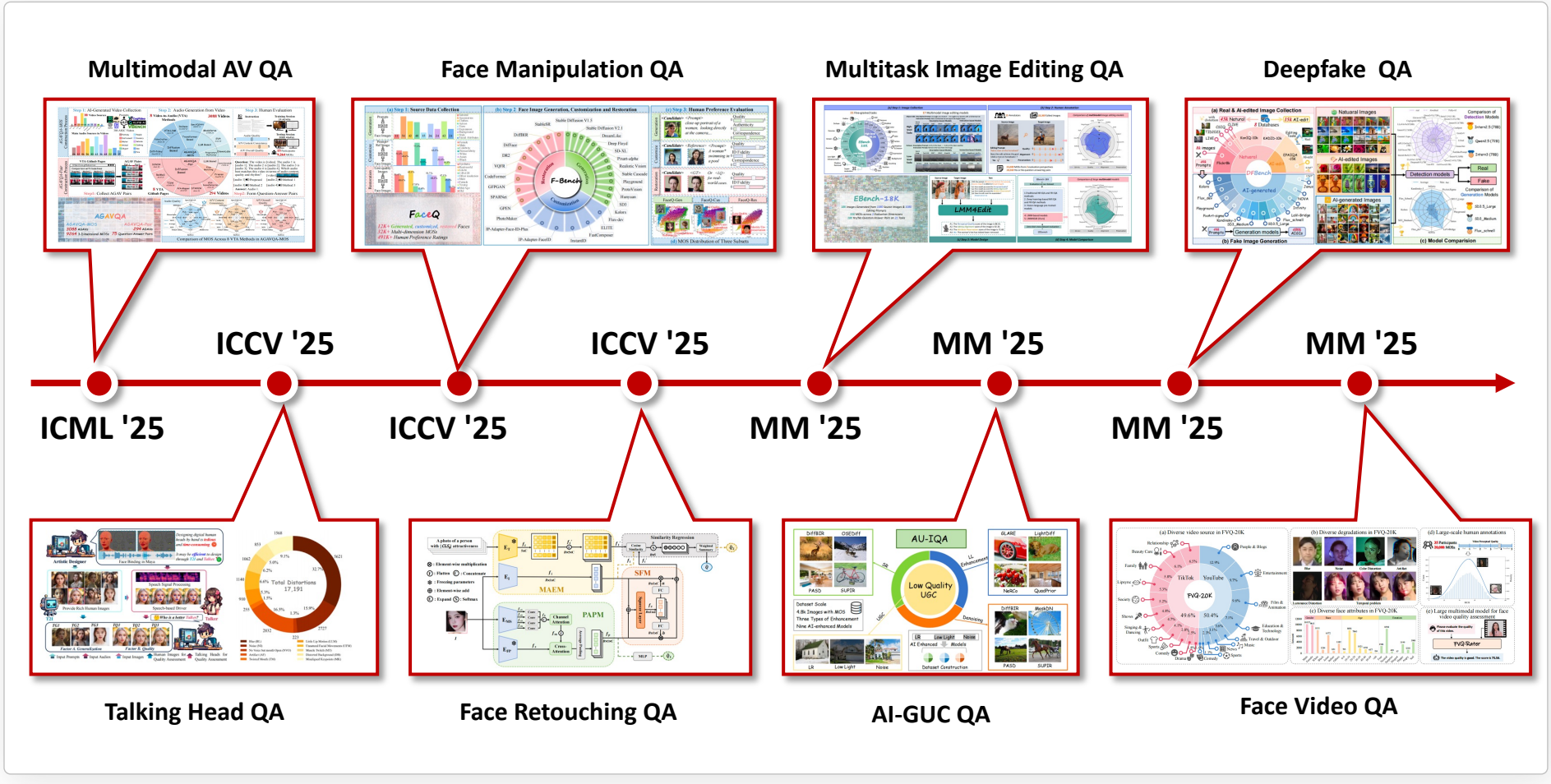


Q-Bench+: Paired QA



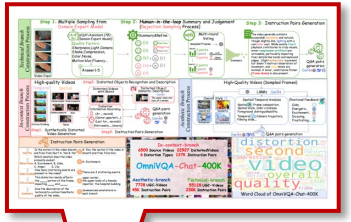
Fine-VQ: Fine-Grained VQA

LMMs for QoE cont'd

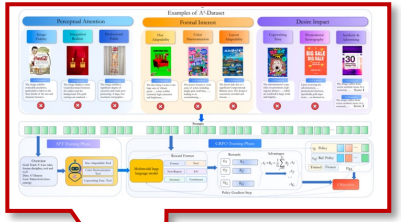


LMMs for QoE cont'd

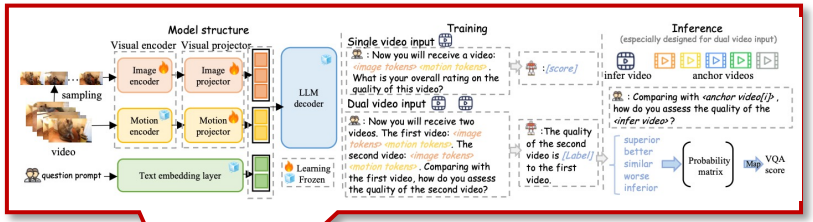
Scaling-up VQA to 400k



Ad Aesthetic QoE Assessment



Self-supervised gradual Post-raining for QoE Eval



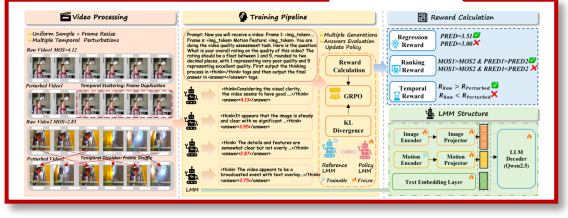
AAAI '26

CVPR '26

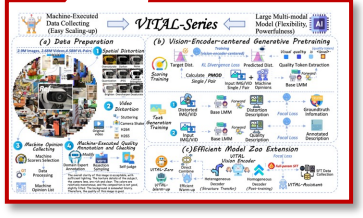
CVPR '26

AAAI '26

CVPR '26

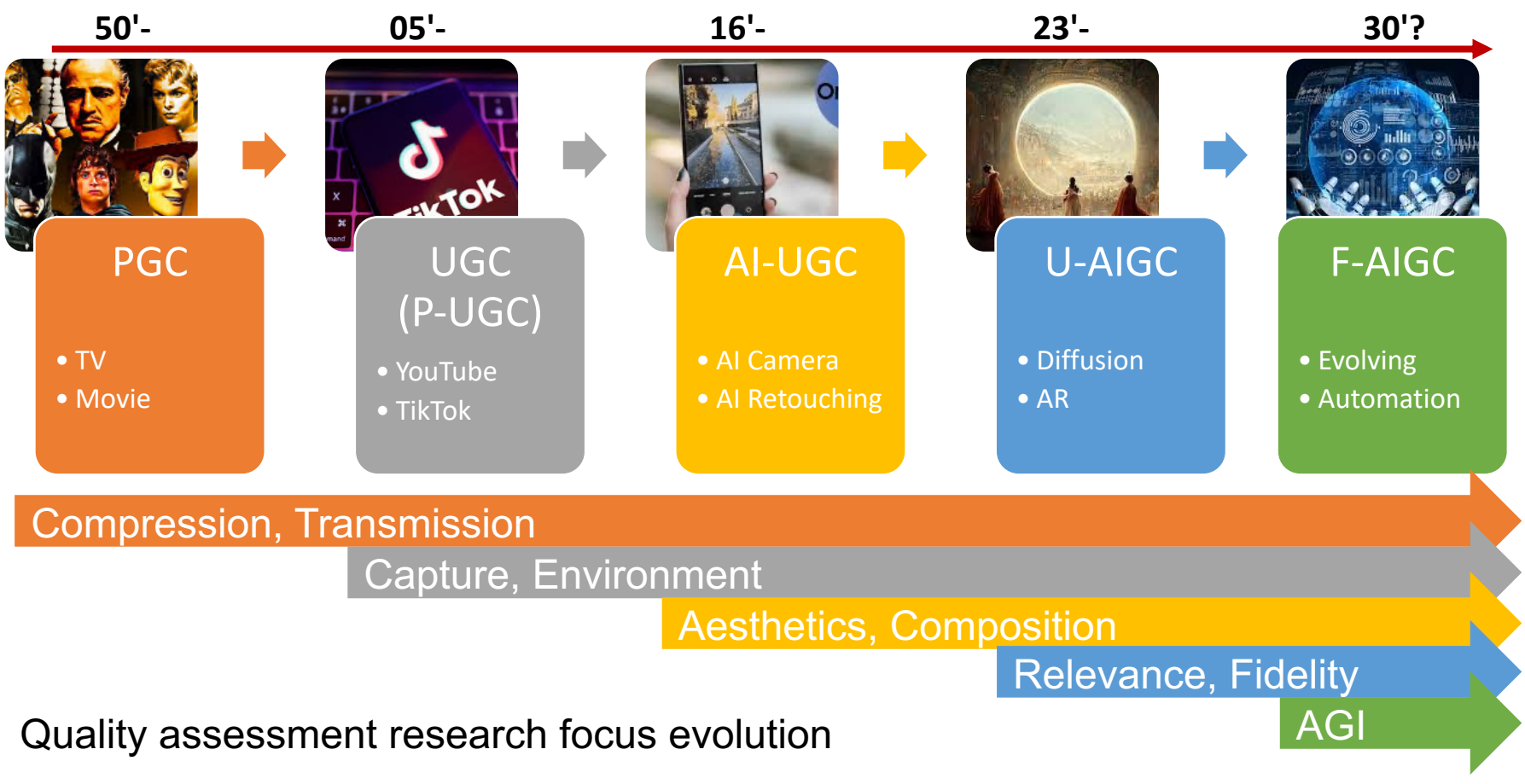


VQA Thinker, Reasoning for QoE Eval

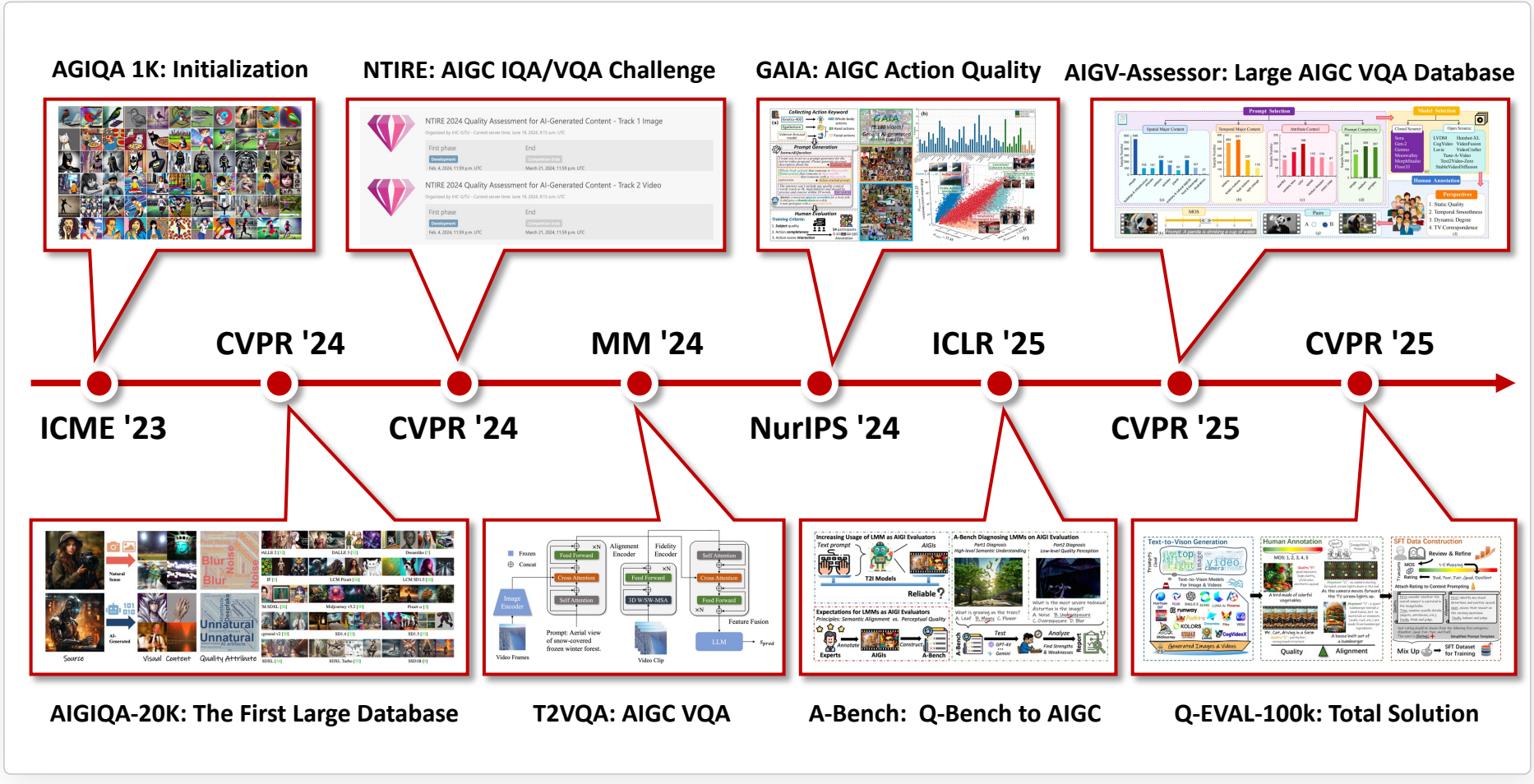


Pretraining for QoE Eval

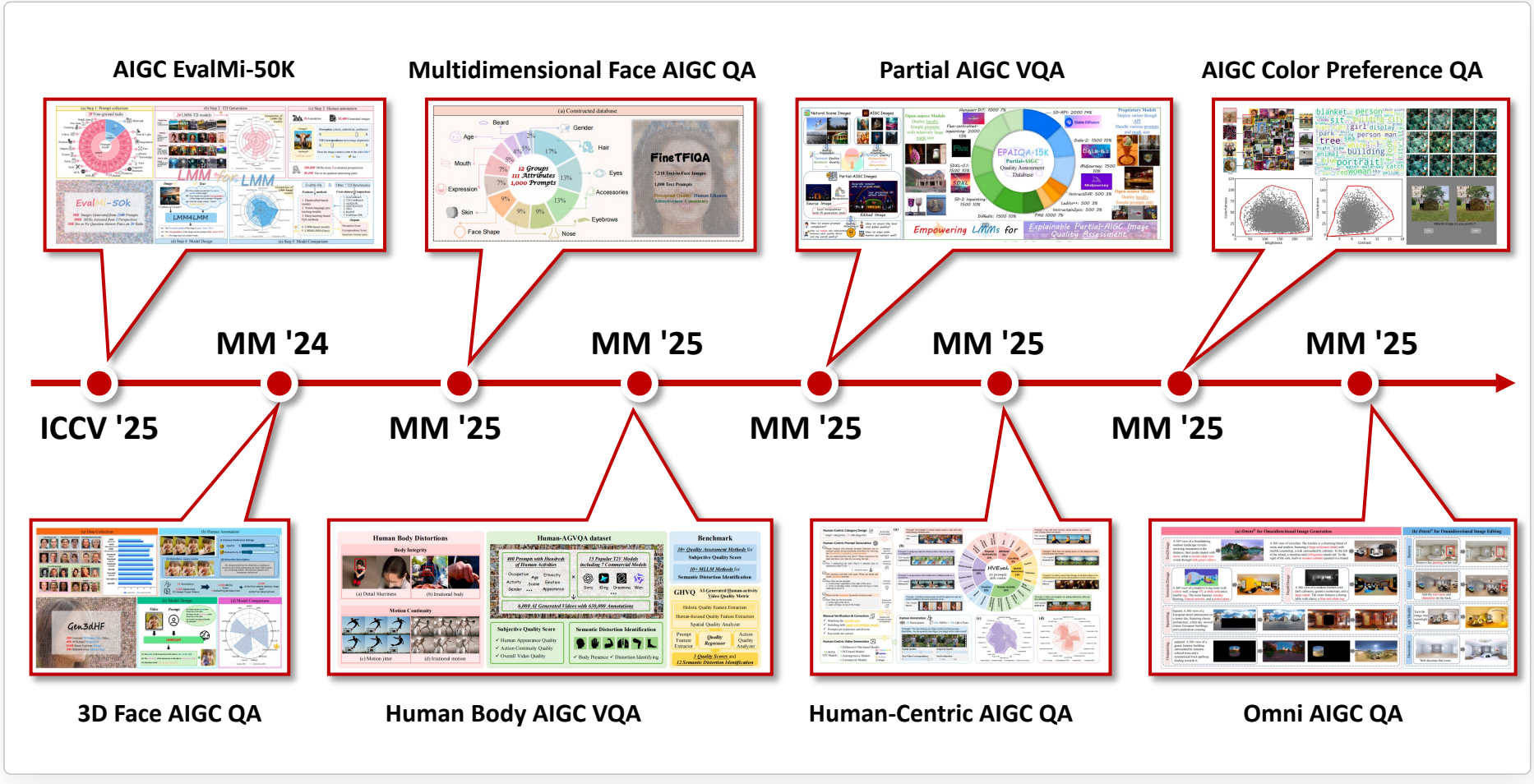
Display-contents Evolution



AIGC QoE


AGIQA 1K: Initialization
NTIRE: AIGC IQA/VQA Challenge
GAIA: AIGC Action Quality
AIGV-Assessor: Large AIGC VQA Database
CVPR '24
MM '24
ICLR '25
CVPR '25
ICME '23
CVPR '24
NurIPS '24
CVPR '25
AIGIQA-20K: The First Large Database
T2VQA: AIGC VQA
A-Bench: Q-Bench to AIGC
Q-EVAL-100k: Total Solution

AIGC QoE cont'd

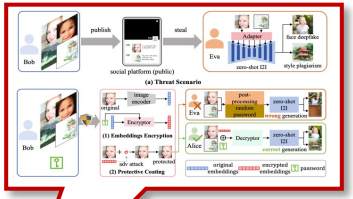


AIGC QoE cont'd

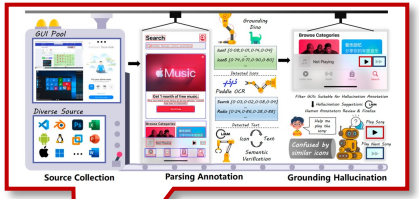
AIGC Editing QoE



Authentication for I2I AIGC



GUI Grouding QoE



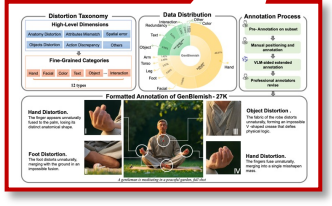
CVPR '26

CVPR '26

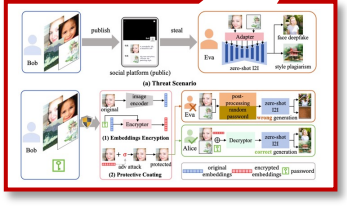
CVPR '26

CVPR '26

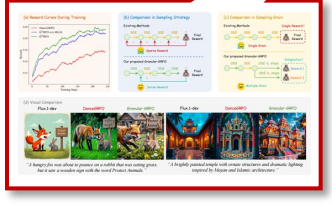
CVPR '26



QoE Driven Retoucher

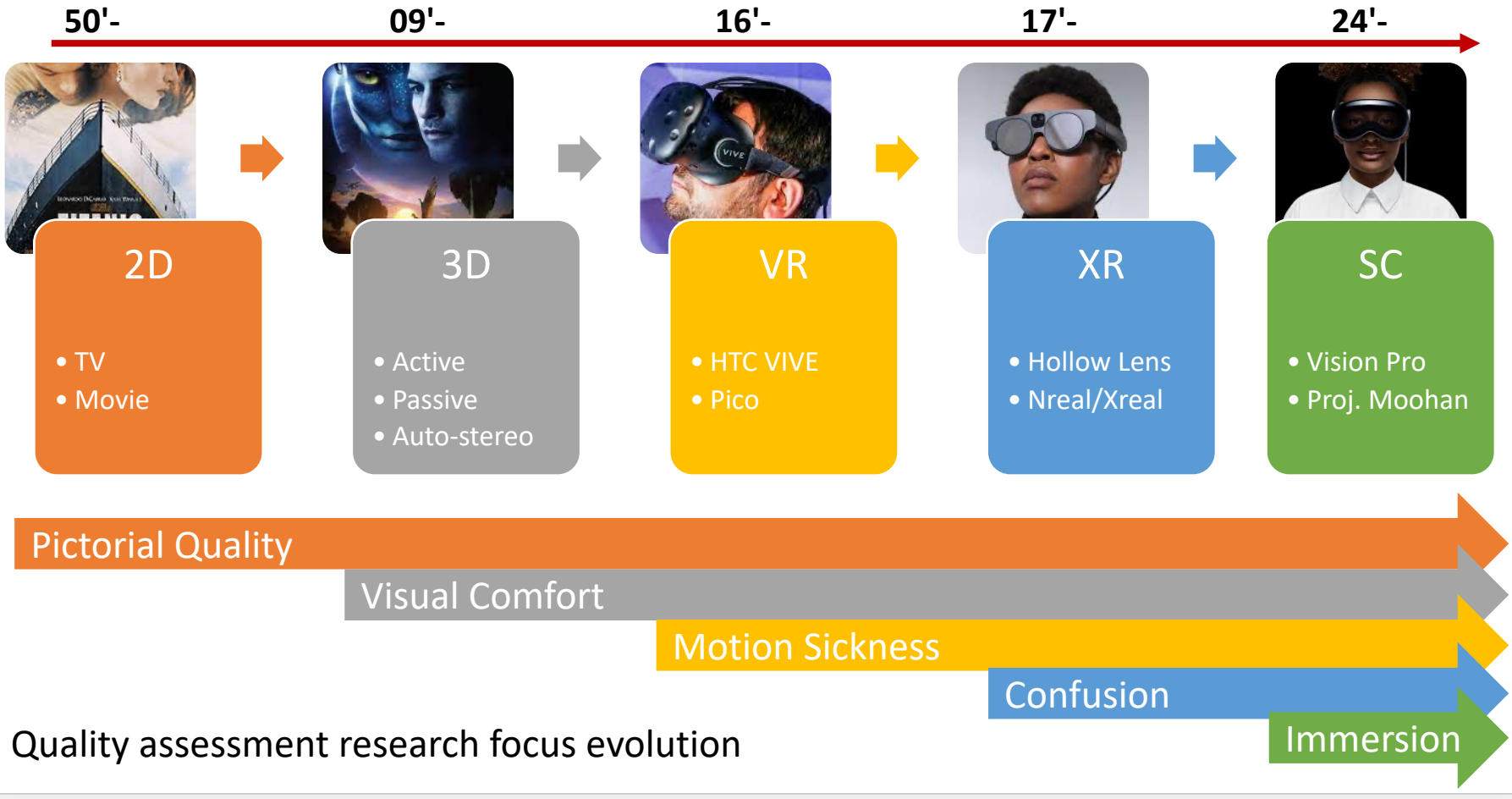


Authentication for I2I AIGC



QoE for GRPO in AIGC

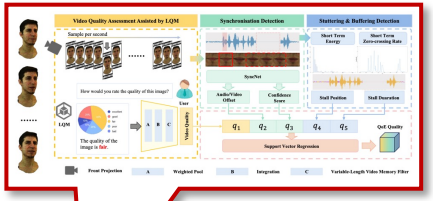
Display-mode Evolution



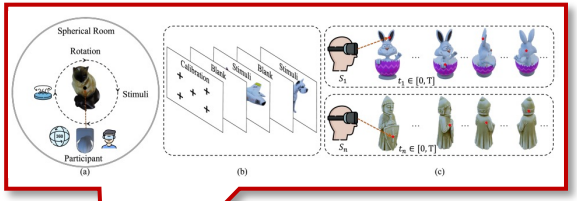
Quality assessment research focus evolution

LMMs for QoE VR/Immersion

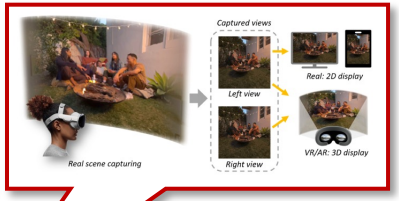
THQA-3D VQA



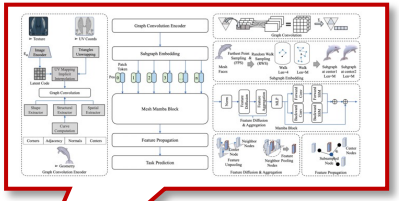
Textured Mesh Saliency



ESIQA: Vision-Pro VQA



Mesh-Mamba Saliency



MM '24

VR/TVCG '25

CVPR '25

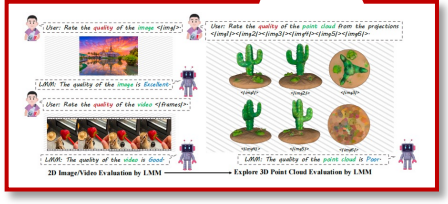
MM '25

MM '24

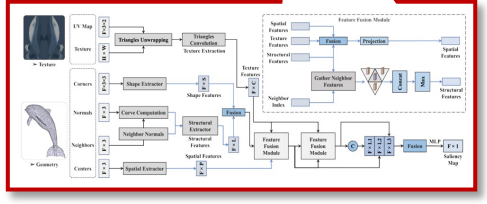
AAAI '25

VR/TVCG '25

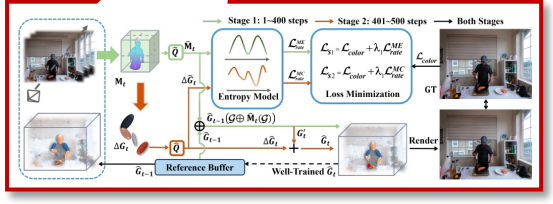
CVPR '25



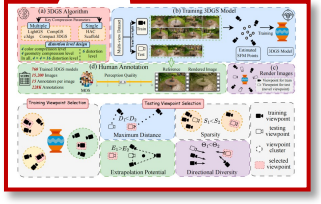
LMM-PCQA: Point Cloud QA



Textured and Non-T Meshes Saliency



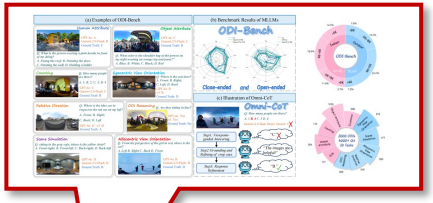
4DGC: Quality Optimized Coding



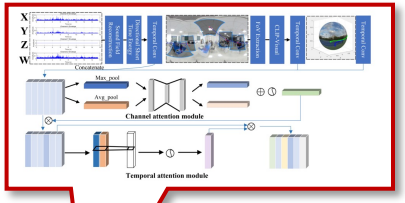
3DGS QA

LMMs for QoE VR/Immersion

3D Understanding Bench



Head Move Prediction



AAAI '26

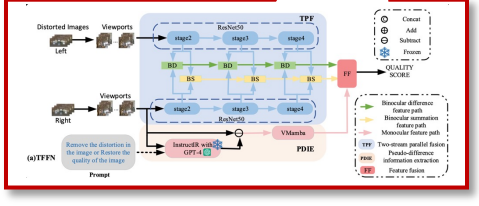
TMM '26

ICLR '26

OJID '26

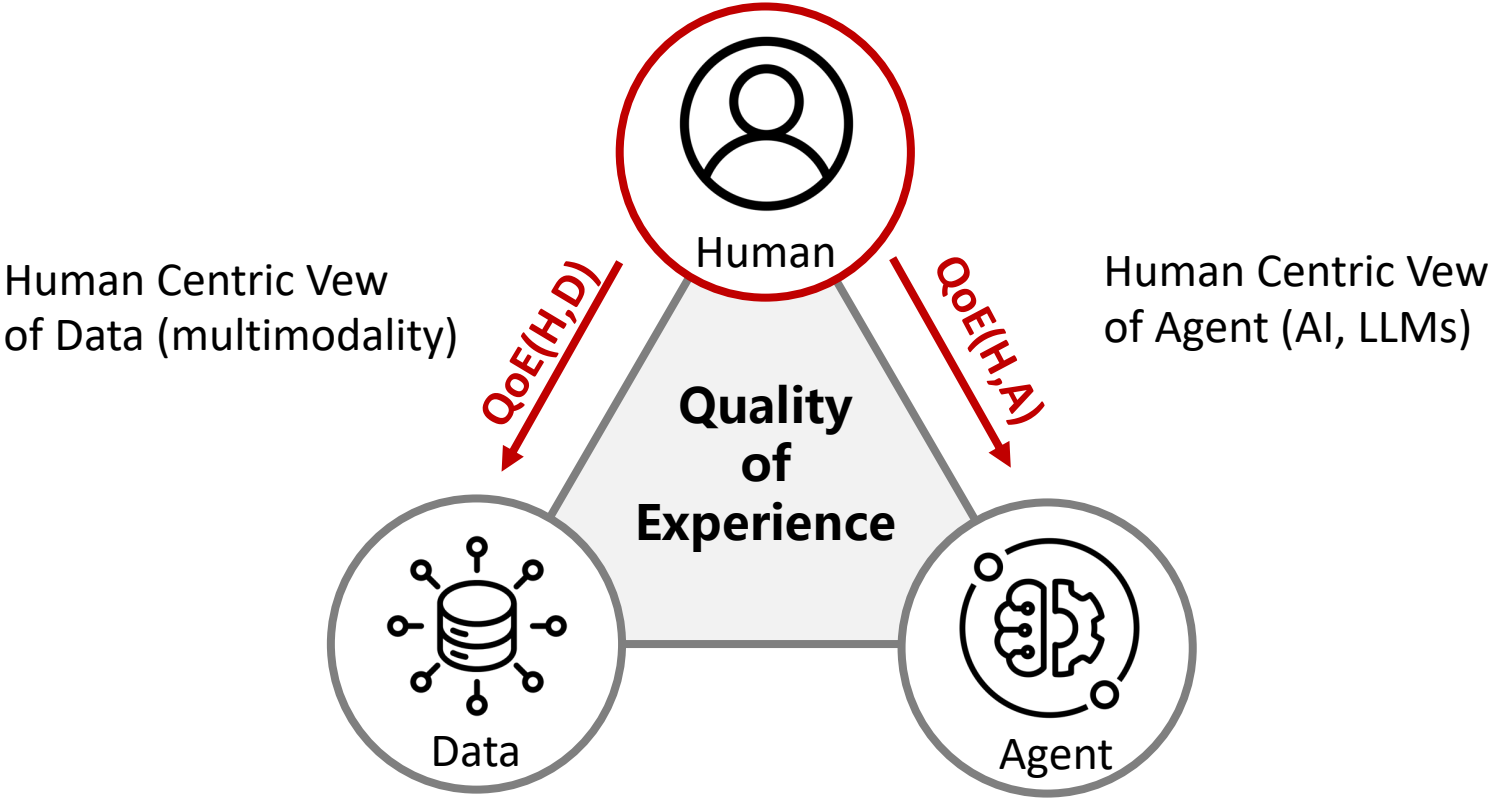


GeoX-Bench



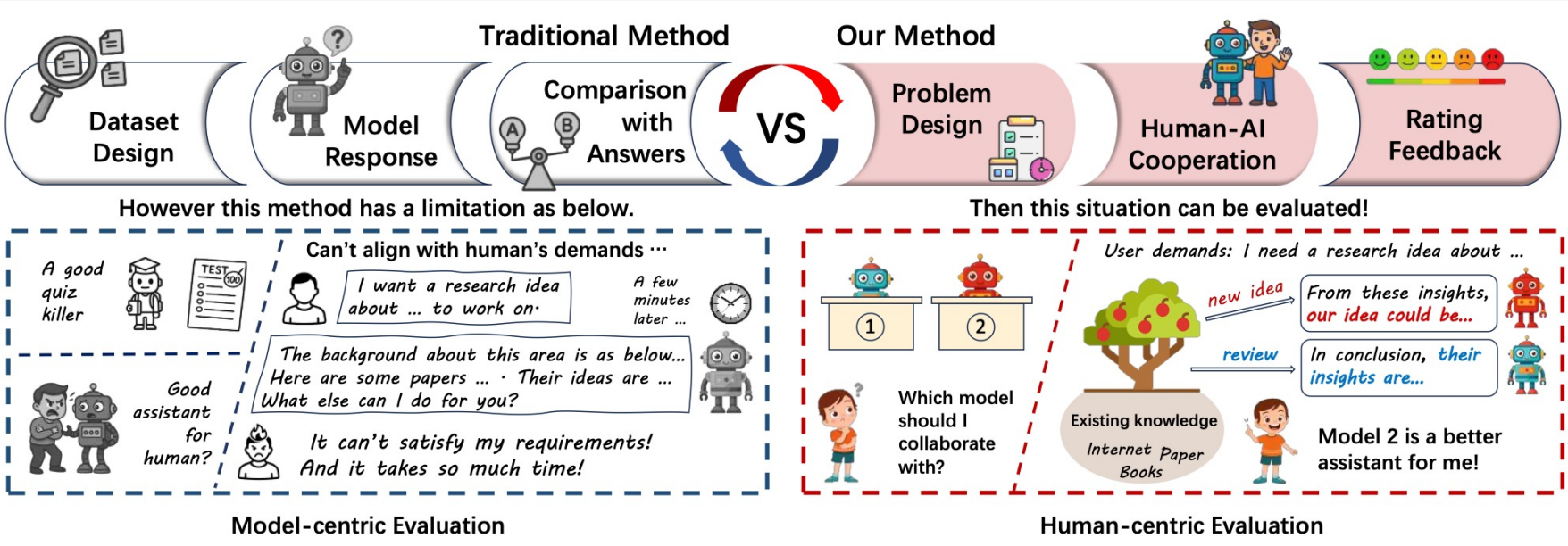
Omni Image QoE

QoE, connecting Human, Data and Agent



Part III
Human Centric AI QoE Research

Human Subjective QoE of AI



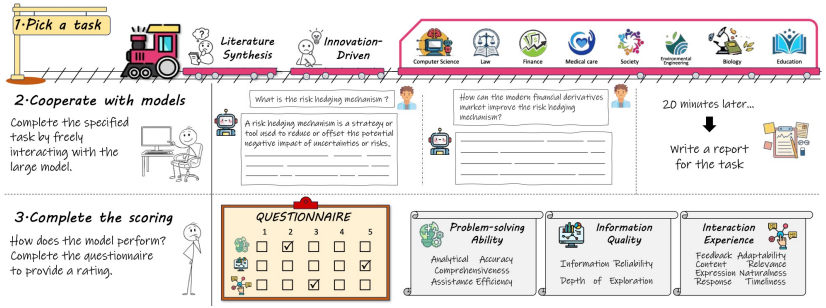


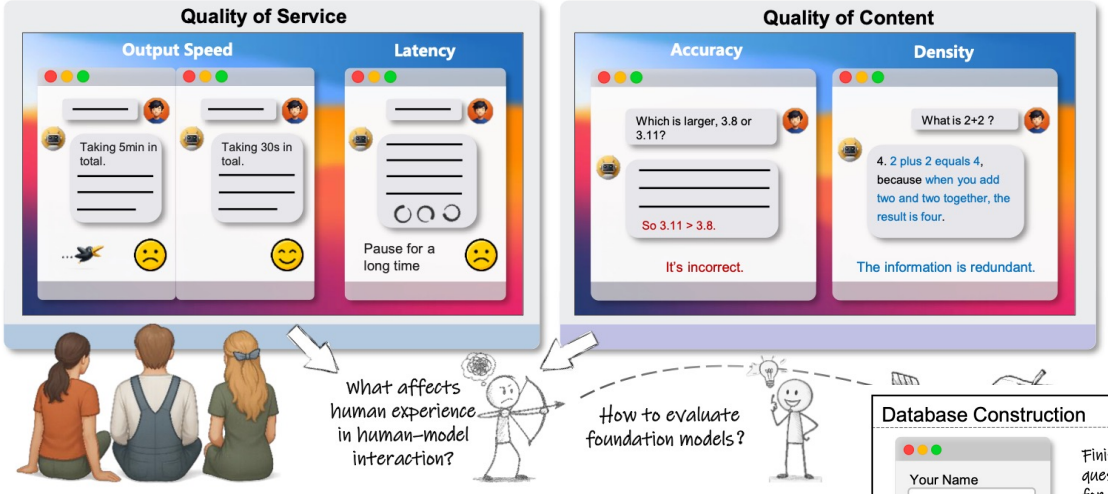
Table 1: Compared to prior work, the proposed HCE framework simultaneously addresses comprehensive evaluation dimensions, multimodal capabilities, diverse disciplines, and extensive application coverage.

Method	Time	Type	App. Cov.	Evaluation Dimension			Disciplines		Multi Modal
				PS.A.	Info.Qua.	Inter. Exp.	Sciences	Humanities	
MMLU	2020/09	Objective	✗	✔	✗	✗	✔	✔	✗
SWE-bench	2023/10	Objective	✗	✔	✔	✗	✗	✔	✗
GPQA[18]	2023/11	Objective	✗	✔	✗	✗	✗	✔	✗
Chatbot Arena	2024/03	Arena	✔	indirectly	indirectly	✔	✗	✗	✗
HLE[17]	2025/01	Objective	✗	✔	✗	✗	✔	✔	✔
Our method	–	Subjective	✔	✔	✔	✔	✔	✔	✔

App. Cov. refers to **Application Coverage**. PS.A to **Problem-Solving Ability**, Info. Qua. to **Information Quality**, and Inter. Exp. to **Interaction Experience**.

Human-Centric Evaluation for Foundation Models, https://arxiv.org/pdf/2506.01793

Human Objective QoE of AI

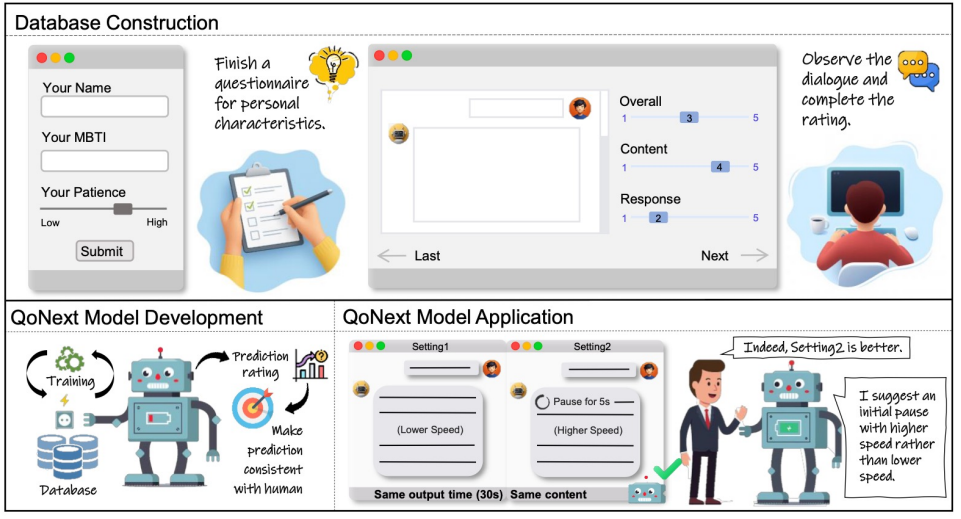


Type	Dimension	Description	Symbol	Abbr.	Range	Param ID
Content Quality	Info. Density	Conciseness and focus of Info.	ρ	CD	$(0,1)$	D0, D1
	Content Accuracy	Correctness and logicality	α	IA	$(0,1)$	A0, A1
Quality of Service	Output Speed	Rate of generating responses	v	OS	$(0.01, 0.05, 0.1)$	S0.01, S0.05, S0.1
	Latency Position	When delay occurs in output	l_{pos}	LP	$(0, 0.25, 0.5, 0.75)$	P0.0, P0.25, P0.5, P0.75
	Latency Duration	How long the delay lasts	l_{time}	LD	$(3, 5, 7)$	T3, T5, T7

Table 1: Dimension Design for Content Quality and QoS.

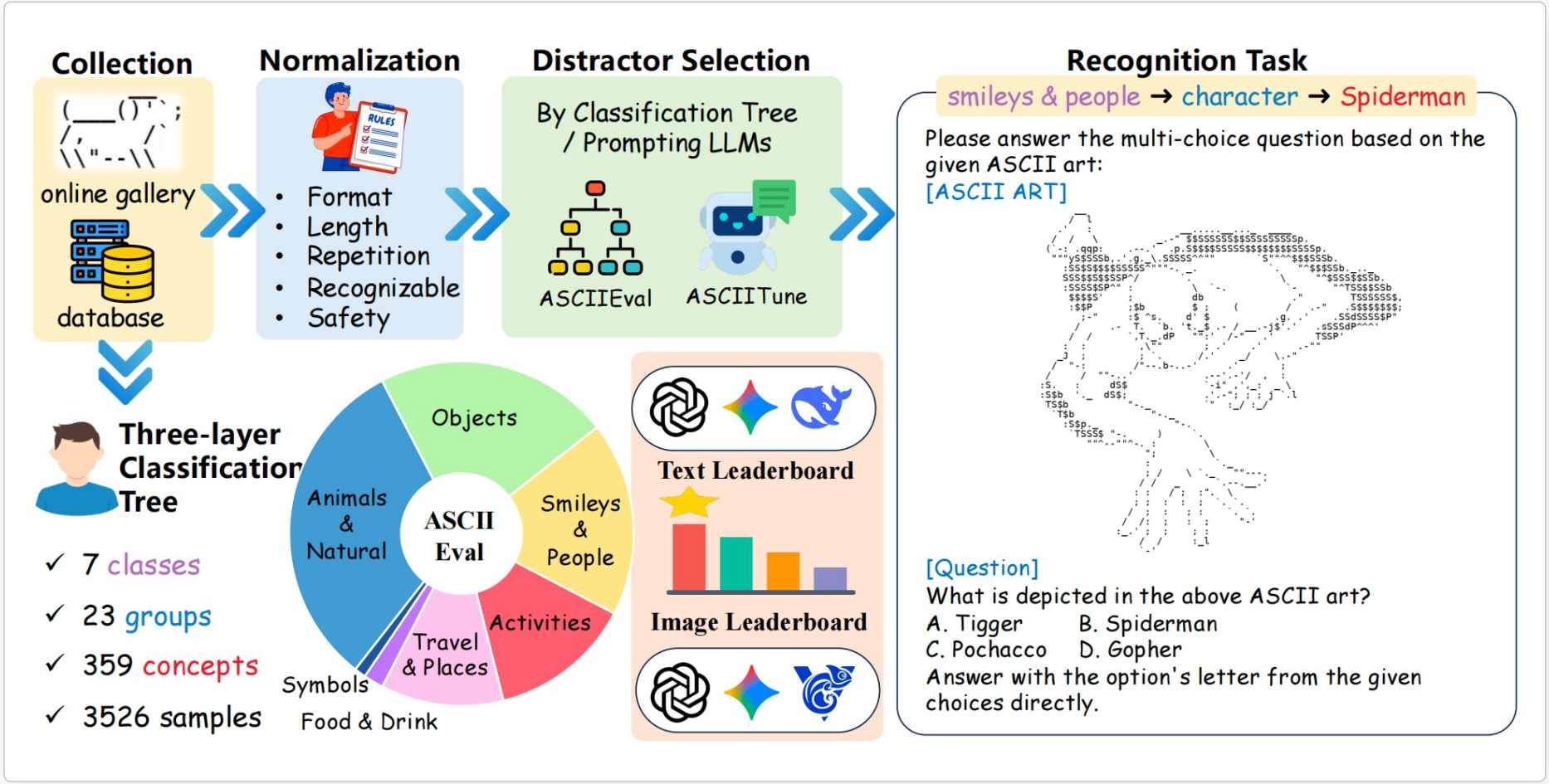
Model	Dimension					Overall			
	ρ	α	v	l_{pos}	l_{time}	SRCC↑	PLCC↑	KRCC↑	RMSE↓
Original (All Params)	✓	✓	✓	✓	✓	**0.7874**	**0.8322**	**0.6486**	**0.8746**
w/o Content Accuracy	✓	×	✓	✓	✓	0.5907	0.6392	0.4693	1.1663
w/o Information Density	×	✓	✓	✓	✓	0.7805	0.8278	0.6365	0.8905
w/o Latency Position	✓	✓	✓	×	✓	0.7842	0.8295	0.6396	0.9695
w/o Latency Duration	✓	✓	✓	✓	×	0.7809	0.8296	0.6365	0.9821
w/o Output Speed	✓	✓	×	✓	✓	0.7362	0.8176	0.5816	0.9392

Table 2: Performance of QoNext models and ablation study results. The mark ✓ in the **Dimension** columns indicates the corresponding dimension is used. The metric values for each method are obtained as weighted averages across the results of all regression models. Dimension Symbol can be found in Table 1. Best in bold.



QONEXT: Towards next-generation QoE for Foundation models, https://arxiv.org/pdf/2509.21889

Objective Eval of AI



Objective Eval of AI



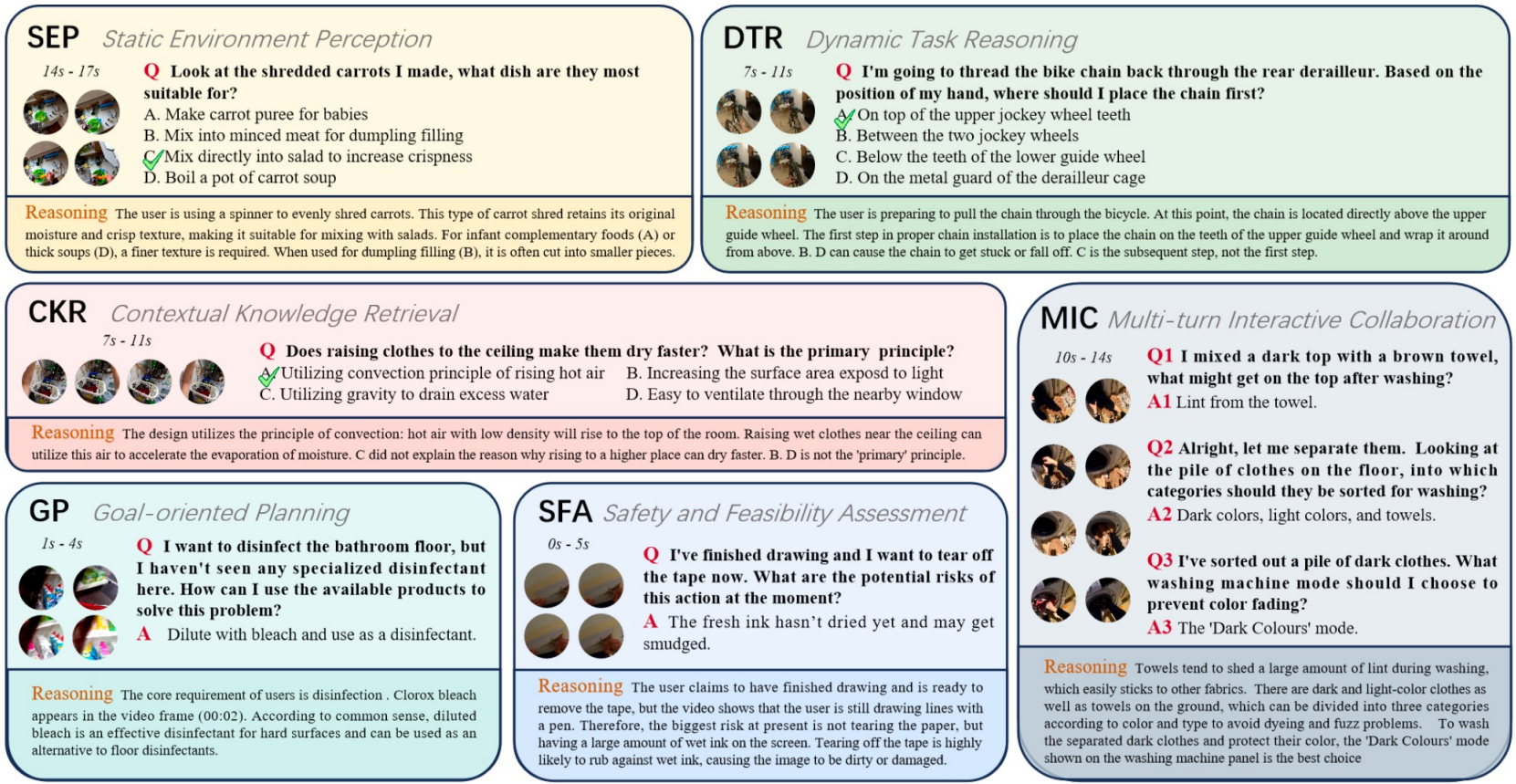
(a) All the information contained in a User Generated Video (UGC)

(b) User-like Thinking Task

(c) User-like Mimicking Task

MIMIC-Bench: Exploring the User-Like Thinking and Mimicking Capabilities of Multimodal Large Language Models
https://openreview.net/forum?id=J7wc4G6woS , ICLR 2026

Objective Eval of AI



GeoX-Bench: Benchmarking Cross-View Geo-Localization and Pose Estimation Capabilities of Large Multimodal Models, https://arxiv.org/abs/2511.13259, AAAI 2025

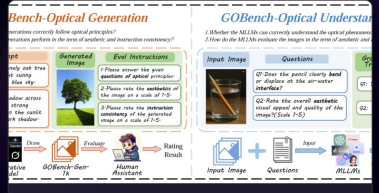
AIBENCH an AI Eval portal

AIBENCH Home Rank Insight Benchmarks

AIBENCH Benchmarks

Cover AI reliability, scientific reasoning, multimodal understanding, and safety evaluation as core research directions. Support researchers and developers with a systematic and traceable entry point to frontier research through continuously updated benchmarks, leaderboards, and accompanying resources.

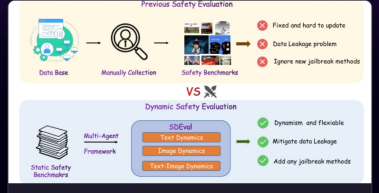
ALL LLM Vertical Domain Safety Embodied Intelligence



GoBench

A leaderboard for optical domain, including Optical Understanding and Optical Generation.

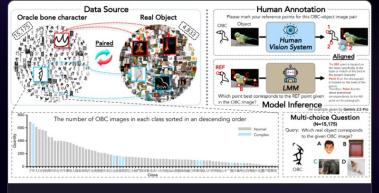
2026–03–10 Vertical Domain



SDEval

A benchmark for MLLMs' safety performance evaluation in real attack scenarios.

2026–03–10 Safety



PictOBI–20k

A benchmark for the real capabilities of large models in visual identification, object correspondence, and...

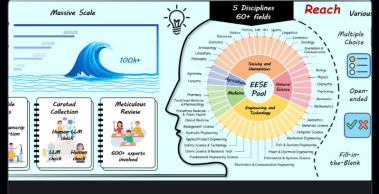
2026–03–10 Vertical Domain



EmbodiedIQA

A benchmark for Image Quality Assessment (IQA) in Embodied AI, evaluating the performance of IQA...

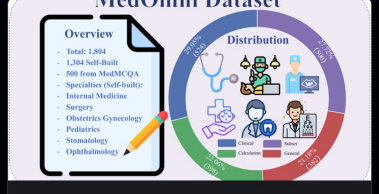
2026–03–09 Embodied Intelligence



EESE

EESE is a dynamic benchmark designed to evaluate foundation models' capabilities in scientific fields.

2025–11–20 LLM

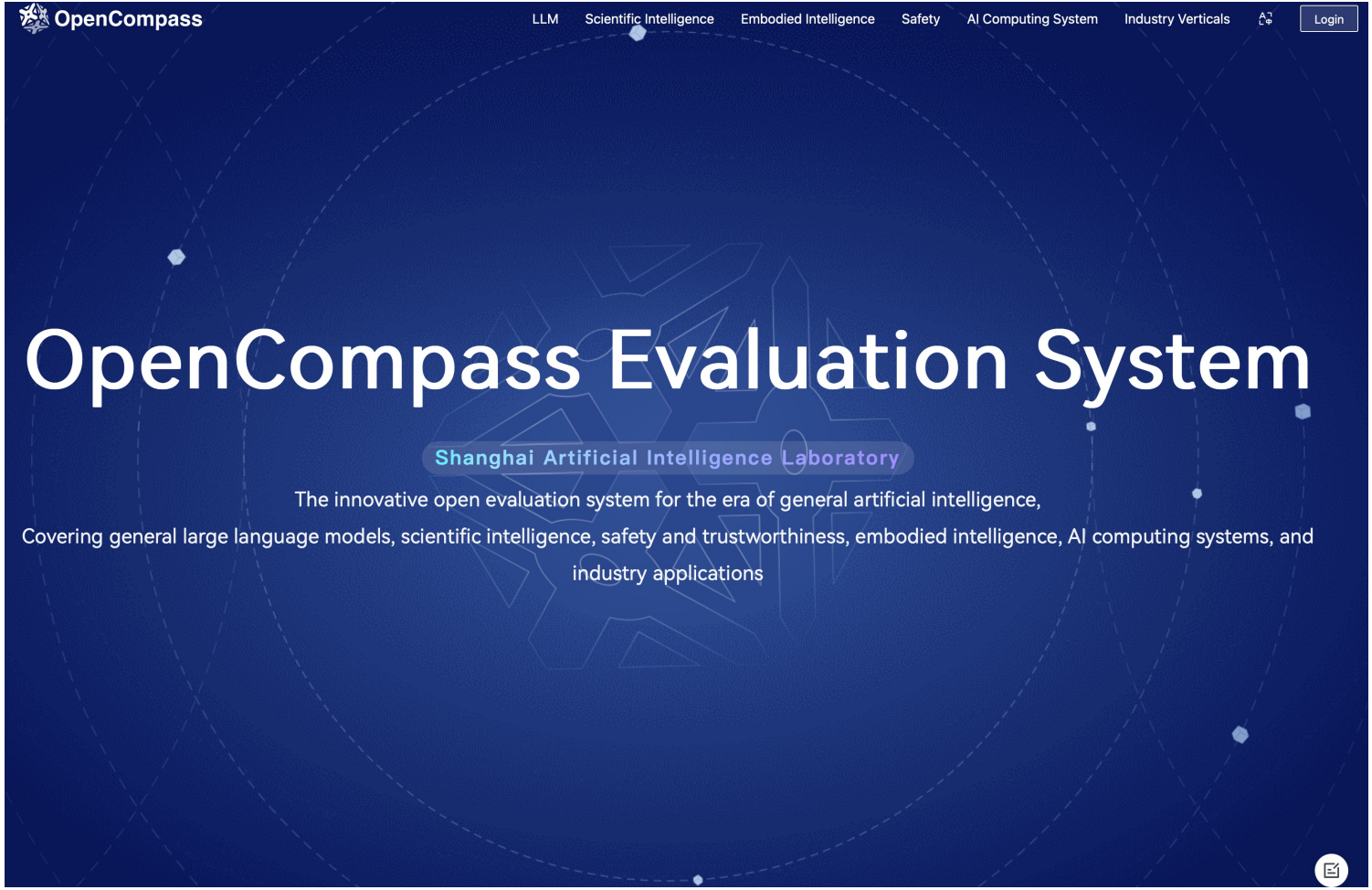


MedOmini–45°

A Safety–Performance Benchmark for Reasoning–Oriented LLMs in Medicine

2026–03–10 Vertical Domain

OpenCompass.org.cn



The Era of Experience

Welcome to the Era of Experience

David Silver, Richard S. Sutton*

Abstract

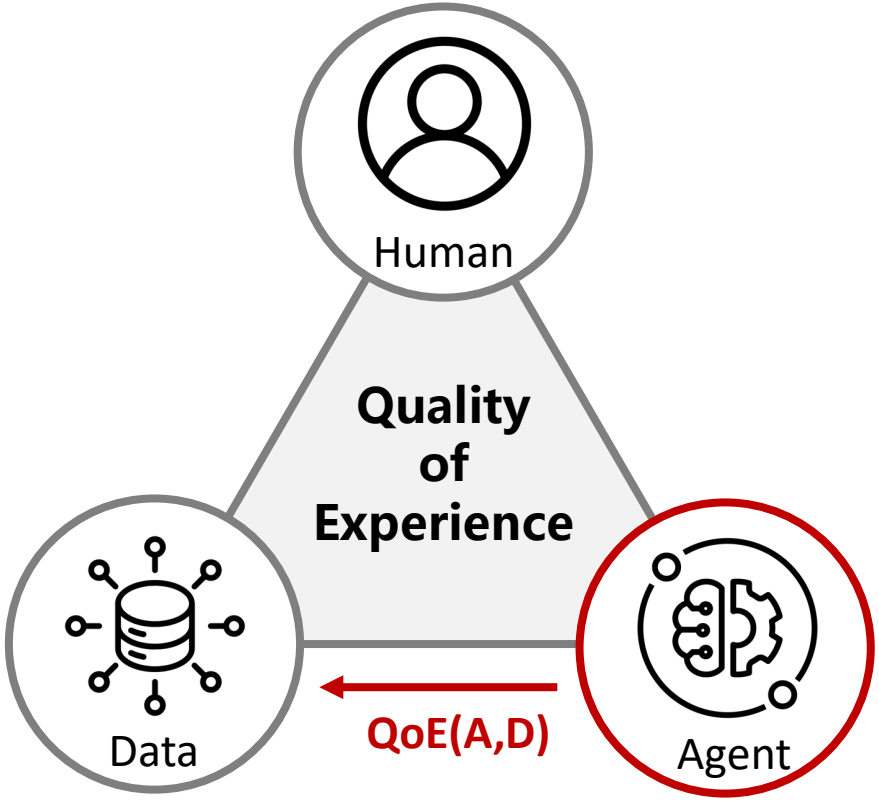
We stand on the threshold of a new era in artificial intelligence that promises to achieve an unprecedented level of ability. A new generation of agents will acquire superhuman capabilities by learning predominantly from experience. This note explores the key characteristics that will define this upcoming era.



Richard Sutton



Agent Centric View of Data



Part IV
Agent (AI) Centric QoE Research

Machine Generated Contents and AI QoE



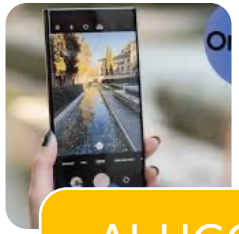
PGC

- TV
- Movie



UGC (P-UGC)

- TikTok Kwai
- Youtube Bilibili



AI-UGC

- AI Camera
- AI Retouching



U-AIGC

- Diffusion
- AR



F-AIGC

- AGI ?



MGC

- Monitor Cam
- EAI/Agents
- Auto. System

- Since 2023, more visual media were consumed by machines than humans
- From human to machine preference
- Task performance-based QoE
- Quality metrics for EAI, RLVR, …

Machine Preference Prediction

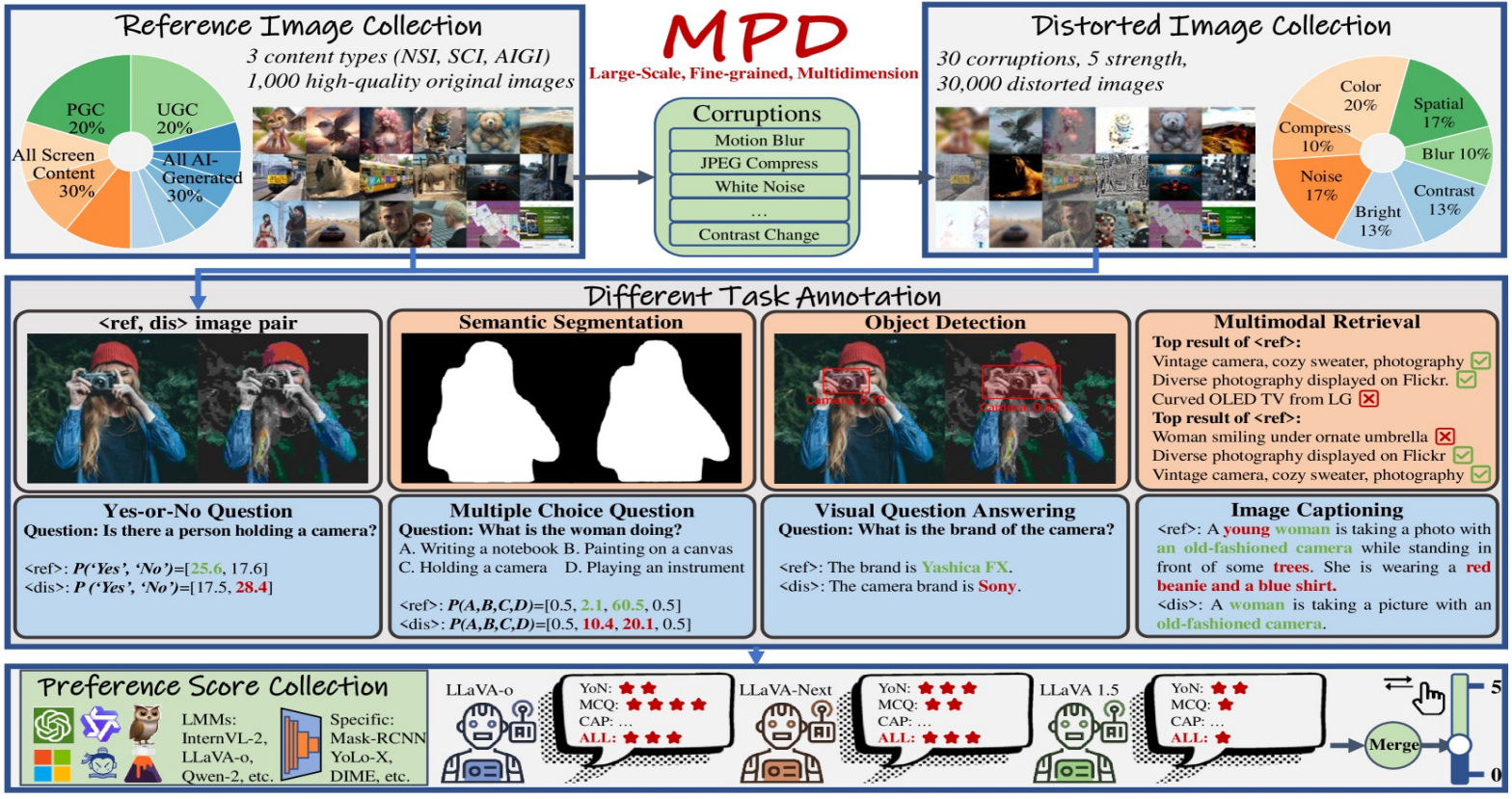
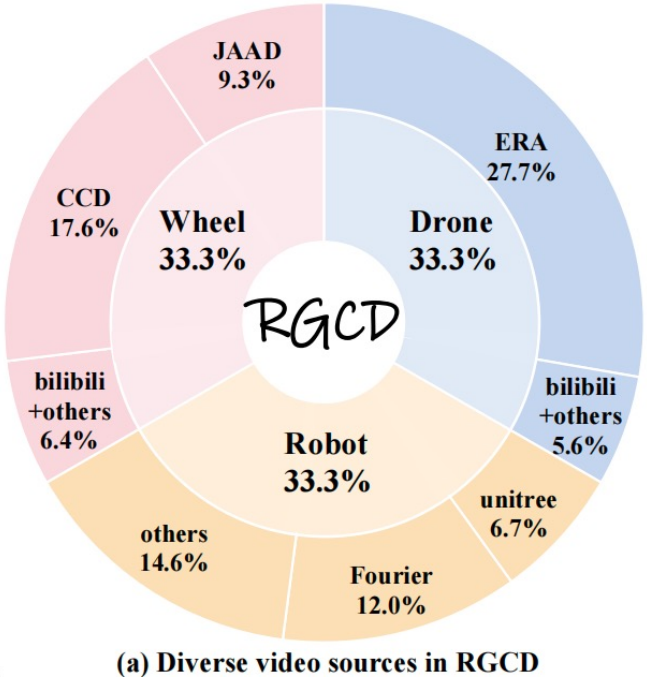
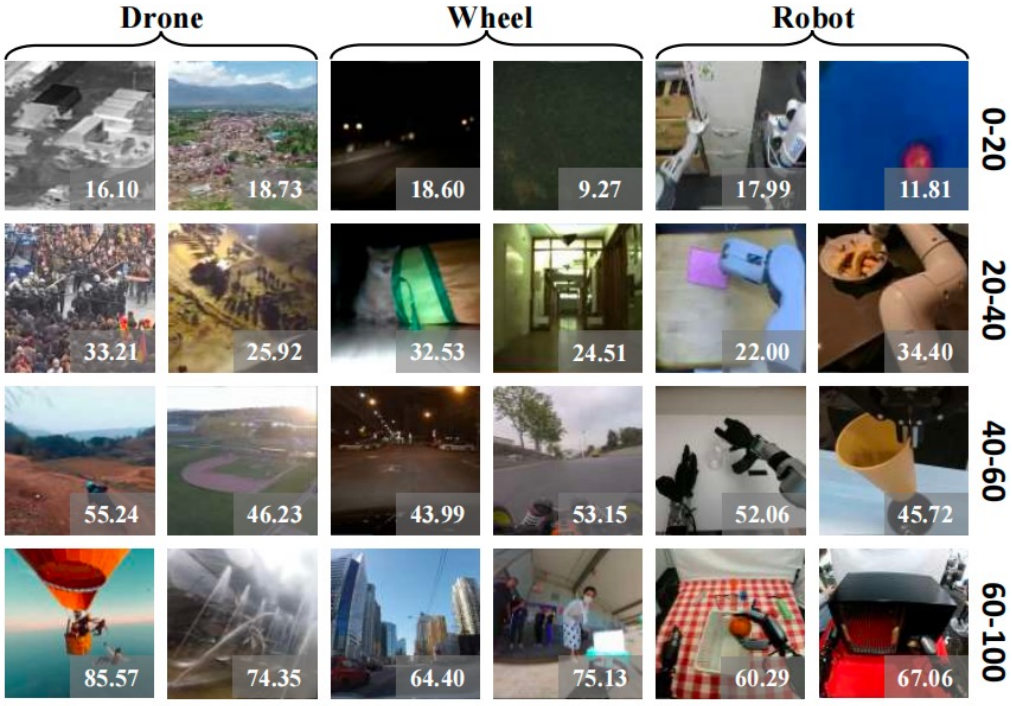


Image Quality Assessment: From Human to Machine Preference, https://arxiv.org/abs/2503.10078, CVPR, 2025

RGC/MGC QoE Database, with MOS



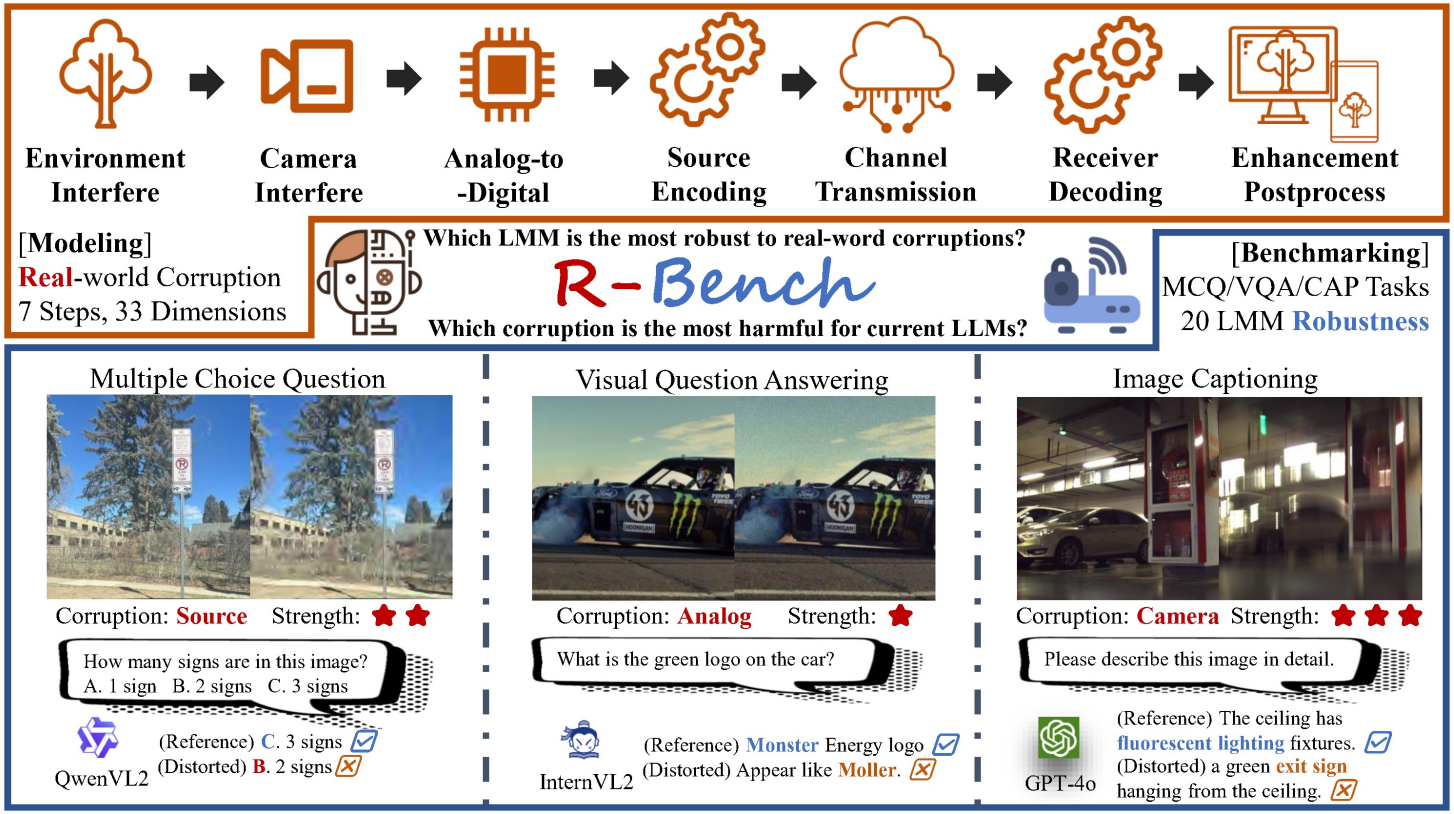
(a) Diverse video sources in RGCD



(b) Sample video frames and corresponding MOSs

2100 RGC videos with MOS

Thorough Benchmarking of LLMs for Robotic Vision Tasks



R-Bench: Are your Large Multimodal Model Robust to Real-world Corruption?, IEEE J-STSP, 2025

Embodied AI QoE, Database and Benchmarking for real tasks

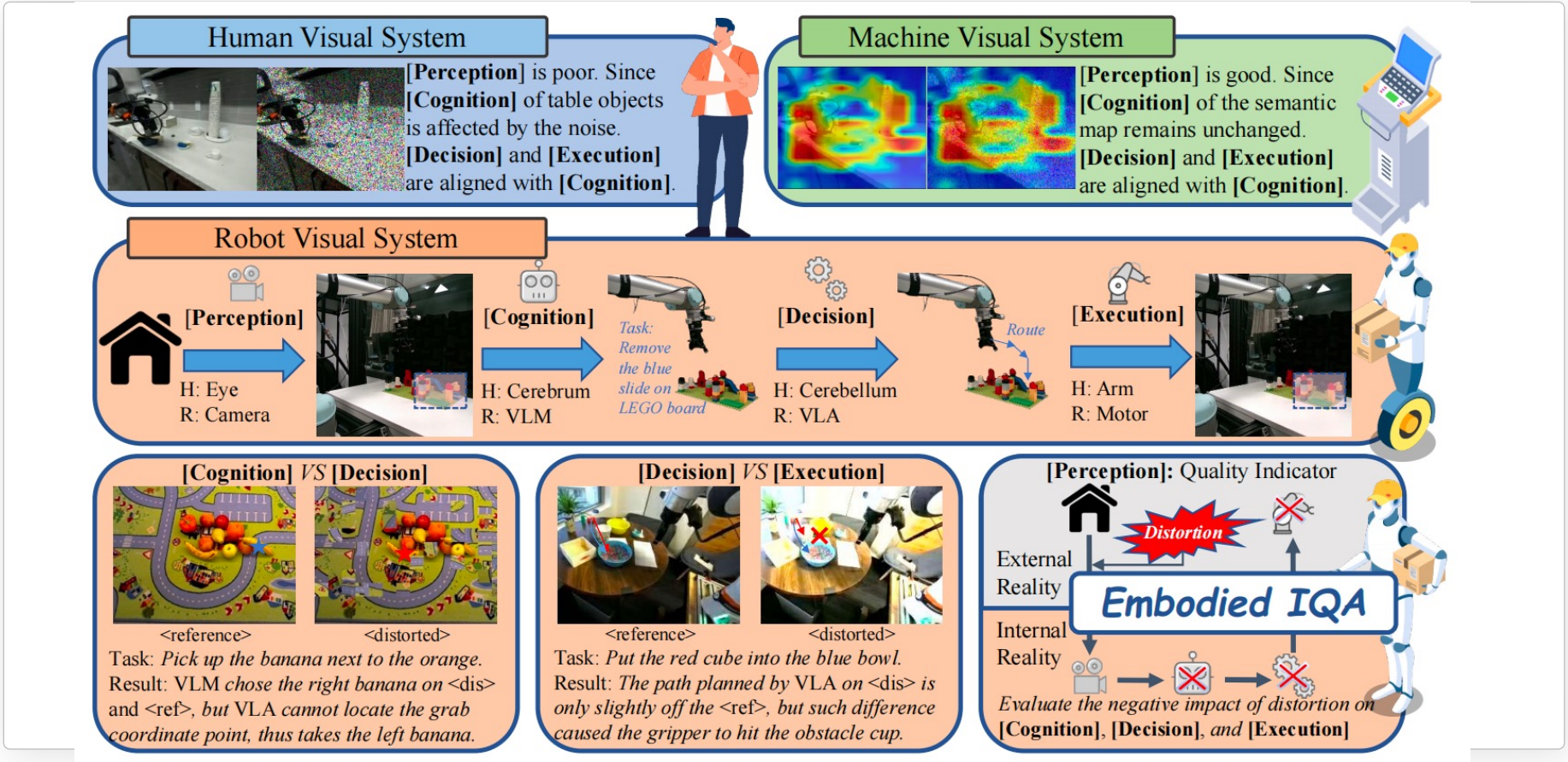
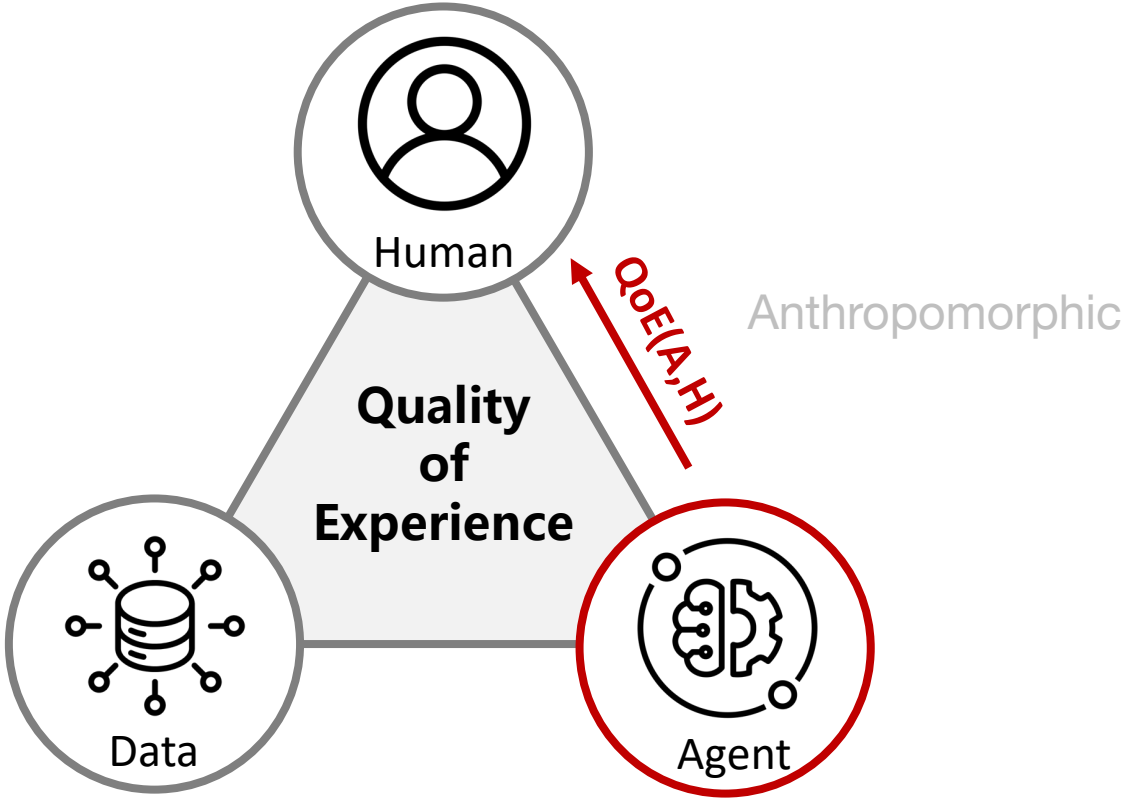


Image Quality Assessment for Embodied AI, https://arxiv.org/abs/2505.16815 , ICLR, 2026

Agent QoE of Human





Thanks

zhaiguangtao@sjtu.edu.cn