

**ICTS**

INTERNATIONAL
CENTRE *for*
THEORETICAL
SCIENCES

TATA INSTITUTE OF FUNDAMENTAL RESEARCH

ICTS Seminar

Title : Evaluating and Advancing Cultural Competence in Multimodal Large Language Models

Speaker : Shachi Dave

Date : Friday, 18 September 2026

Time : 11:30 AM (IST)

Abstract : Multimodal foundation models have achieved remarkable visual fidelity and technical problem-solving capabilities, yet their real-world deployment reveals severe blind spots in representing and reasoning across the world's diverse cultures. This talk examines the challenge of evaluating and building culturally competent multimodal systems across both image generation and video reasoning. First, we investigate text-to-image generation using CUBE, a large-scale benchmark of over 300,000 cultural artifacts spanning eight global regions. We introduce formal metrics grounded in ecological diversity (the Vendi Score) to quantify the Pareto trade-off between aesthetic realism and representational diversity, demonstrating how modern alignment often accelerates "aesthetic collapse" and cultural homogenization. We then discuss mitigation strategies, including contextualized Vendi Score guidance in diffusion sampling. Second, moving from static perception to dynamic reasoning, we present CURVE, a multicultural, multilingual long-video QA benchmark spanning 18 locales and featuring human-authored native-language reasoning traces. Evaluating frontier models reveals a stark ~50% performance gap compared to human baselines. Through causal evidence DAGs and automated error attribution, we show that ~80% of model failures stem from cultural perception and localization rather than downstream logical reasoning. The talk concludes with open directions for data curation, representation alignment, and culturally grounded multimodal evaluation.

Venue : Emmy Noether Seminar Room

Zoom Meeting: <https://icts-res-in.zoom.us/j/95864105490?pwd=5dv9hEsaHLxp6kx7zKD3TkzewNdQS.1>

Meeting ID: 958 6410 5490

Passcode: 665055