OSMO: Open-vocabulary Self-eMOtion Tracking

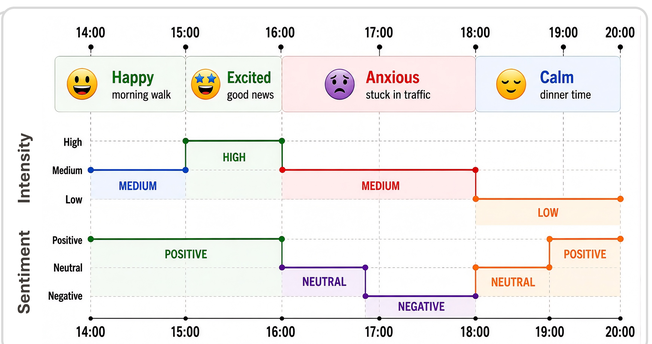
Mohamed Abdelfattah (OS), Bugra Tekin, Fadime Sener, Necati Cihan Camgoz,
Eric Sauser, Shugao Ma, Alexandre Alahi, Edoardo Remelli

CVPR JUNE 3-7, 2026 DENVER COLORADO

Goal

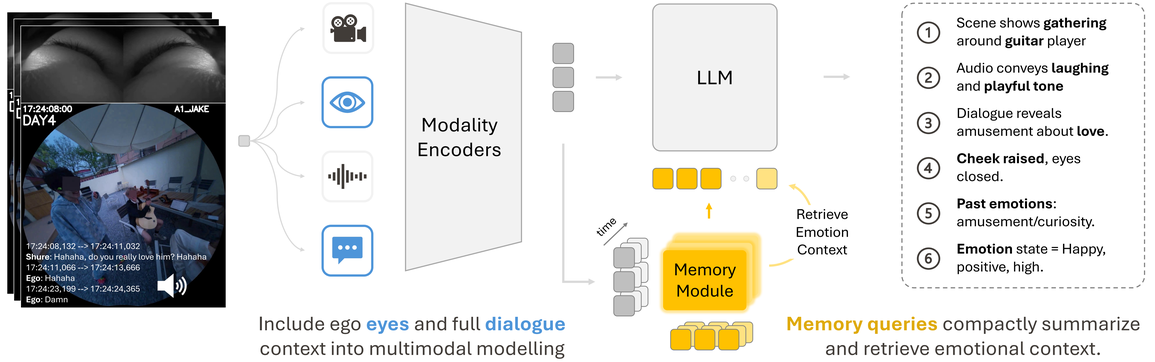
Input

Output: Emotion timeline



Our goal: Automate continuous *emotion* tracking using *egocentric* smart-glasses data.

OSIRIS Multimodal LLM



Include ego **eyes** and full **dialogue** context into multimodal modelling

Memory queries compactly summarize and retrieve emotional context.

OSMO Dataset

Person-wise *egocentric* timelines annotated with open-vocab **emotion**, **sentiment**, **intensity**, and **reasons.**

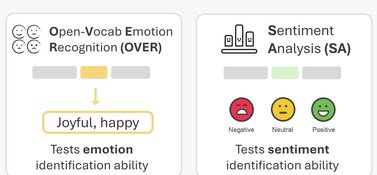
Affective labels

- Open-vocab **Emotions** (e.g., joyful, content)
- **Sentiment** (e.g., positive)
- **Intensity** (e.g., high)
- **Reason** (e.g., paper accepted)

110 hours	**116K** annotations
288 subjects	**1000s** of emotions

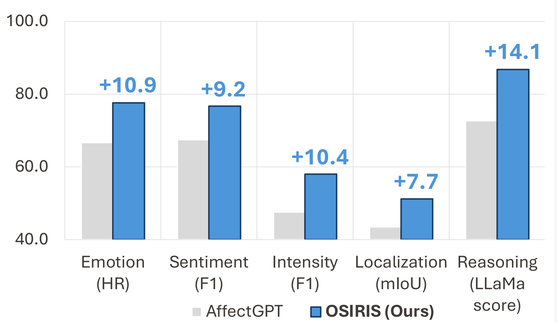
OSMO Benchmark

Five benchmark tasks



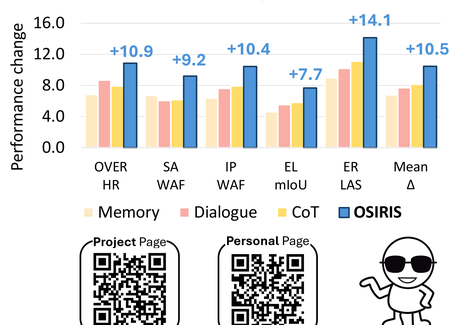
Results

Outperforms previous SOTA on **all** benchmark tasks



Ablations

Order of usefulness: **CoT**, Dialogue context, and memory



Project Page

Personal Page