

The Problem

Emotion Recognition (ER) systems infer emotion from facial or vocal cues alone — ignoring the **situation** in which the emotion occurs. This makes them **context-blind**, causing systematic errors in real-world settings.

Motivating example: The same “crying” face signals *opposite* emotions depending on context — sadness at a funeral vs. joy at a wedding. Behavioral cues are insufficient without situational grounding.

Research Question

“What categories of situation context cues appear in emotion perception research, and how do they modulate perceived emotions?”

Goal: Define a structured taxonomy of *Situation Context*, scoped to exclude:

- *Perceiver* Context (beliefs, mood, culture)
- *Sender* Context (expression, body language)

Why? “Context” is a buzzword in both Psychology and CS with no agreed-upon structure, limiting CAER systems and agentic AI.

Theoretical Scope

Perceiver	Situation	Sender
beliefs, culture, mood	event, social, environment, conversation	expression, body language

This work focuses exclusively on the **Situation** column.

Methodology Overview

Reporting standard: PRISMA 2020

SLR Process:

1. **Literature Identification** (*Scopus search*)
2. **Literature Screening** (*AI Pipeline*)
3. **Quality Assessment** (*manual reading*)
4. **Data Extraction / Coding** (*manual reading*)
5. **Synthesis** (*taxonomy creation*)

Augmented at the screening step with a **custom AI pipeline** (see right) to handle the volume of retrieved records.

Study Selection

Search & Screening Results:

- **Identified (Scopus):** $N = 294$
- **Excluded (Non-English):** $N = 5$
- **AI Screened (Gemini):** $N = 289$
- **Excluded by AI (Score 1–4):** $N = 274$
 - Sender Confounds: 129
 - Perceiver Confounds: 70
 - Non-Empirical Formats: 47
 - Tech/Other Outcomes: 28
- **Fully Coded & Synthesized:** $N = 15$

Example Search Query

- emotion recognition, affect recognition, emotion perception
- C1 Emotion:** emotion recognition, emotion perception
- C2 Context:** situation context, scene context, location, social context
- C3 Perception:** emotion perception, emotion recognition, emotion judgment
- C4 Modality:** audio-visual, multimodal, visual, face, voice

Inclusion & Exclusion Criteria

Inclusion: Empirical human research on how situational factors modulate emotion *perception*.

Exclusion:

- Focus on emotion *expression* (not perception)
- Focus on verbal emotion perception only
- Reviews or surveys (primary evidence only)
- Non-neurotypical or developmental populations as focus

AI Relevance Screening

We deployed an automated screening pipeline to screen the 289 abstracts:

Pipeline Configuration: Gemini-3.1-flash-lite configured at $T = 0.0$ to ensure deterministic scoring.

Screening Rubric (Score 1–5):

- **Score 5:** Isolates situation context and measures its perception on human/machine outcomes. Surfaced for coding.
- **Score 4:** Mentions situation context, but secondary to other cues.
- **Score 1–3:** Excluded (clinical focus, sender cues like body posture, or non-empirical formats).

The Proposed Taxonomy

Our systematic analysis of the 15 core papers revealed a three-pillar taxonomy of situation context cues:

- **Objective Environment:** Non-human physical settings (e.g., location, room layouts, background lighting) and temporal timing.
- **Social Array:** Relational frame of secondary actors (bystanders, crowds, cross-species arrays) modulating the central target.
- **Semantic/Narrative Information:** Linguistic event-based descriptions and vignettes explaining the story behind the emotion.

Key Findings

We mapped how cues within each taxonomic pillar systematically bias observer emotion judgments (RQ2):

- **Objective Environment Bias:** Context masking degrades real-time tracking speed and accuracy. Spatiotemporal timing delays affect processing speed.
- **Social Array Bias:** Background expressions act as contrast or congruence cues. Decoded background emotions feed back to modulate target authenticity.
- **Semantic/Narrative Bias:** Vignettes and event framing systematically shift categorical emotion labels (e.g., classifying a smile as polite social compliance rather than happy amusement).

Conclusions & Future Work

- Establishes a standard taxonomy to guide context-aware computational models.
- **Future:** Expand to multiple databases, implement dynamic attention models, and develop context-aware bias auditing frameworks.

References

- [1] L. Barrett et al. Emotional expressions reconsidered. *Psych. Sci. Public Interest*, 2019.
- [2] J. Russell. Reading emotions from a face. *The psychology of facial expression*, 1997.
- [3] J. Israelashvili et al. When emotions run high. *Emotion*, 2019.
- [4] B. Dudzik et al. Context in human emotion perception. *IEEE ACII*, 2019.
- [5] R. Kosti et al. Context-aware emotion recognition. *IEEE TPAMI*, 2019.
- [6] Z. Chen et al. Inferential tracking of continuous emotion. *Nat. Hum. Behav.*, 2020.
- [7] R. Mui et al. Contextualising smiles. *Cogn. Emot.*, 2020.
- [8] S. Namba et al. Social context and culture. *J. Cross-Cult. Psychol.*, 2020.