Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 240–249
https://doi.org/10.54941/ahfe1008258

Predicting Workplace Conflict Through Multimodal AI-Based Behavioral Analysis: An Ethical and Cognitive Framework

Damina Townes, Rushabh Gandhi, Apoorv Singh, Azhar Mohammed Abdul Haleem, and Dr. Youngsoo Shin

Pace University, New York, NY 10038, USA

ABSTRACT

Workplace conflict is frequently managed as an operational issue addressed only after formal complaints or visible disputes arise. This reactive approach fails to account for the gradual cognitive and behavioral processes that underlie interpersonal tension. This paper addresses the central research question: How do managers perceive the effectiveness, acceptability, and practical challenges of using AI-driven multimodal analysis for early workplace conflict detection? To investigate this question, we propose a multimodal AI-powered framework for the early prediction of workplace conflict by analyzing behavioral patterns across multiple data modalities, including facial expression recognition, body language analysis, and text-based communication indicators such as sentiment shifts, response latency, and collaboration network changes. Drawing on theoretical foundations in affective computing, computer vision, and cognitive science, we present an integrated model that tracks temporal behavioral escalation rather than static, single-instance observations. We further address the critical ethical dimensions of such a system, including privacy preservation, algorithmic bias, and the necessity of human-centered design principles. Qualitative data were gathered through semi-structured interviews with 11 participants spanning managerial, technical, and frontline professional roles, guided by seven research hypotheses. Interview responses reveal a nuanced landscape of conditional acceptance: participants broadly support temporally aware, multimodal detection approaches while expressing strong reservations about facial-expression-only systems, raising consistent concerns around privacy, cultural bias, and the risk of managerial misuse. The paper concludes with recommendations for responsible, transparency-first deployment of AI conflict-prediction systems in organizational environments.

Additional Keywords and Phrases: Artificial intelligence, Behavioral analytics, Body language analysis, Cognitive science, Emotion recognition, Predictive analytics, Privacy, Surveillance, Workplace conflict

INTRODUCTION

The conventional definition of workplace conflict focuses on administrative and organizational aspects of the process and considers conflict resolution as a solution-oriented procedure to be employed once conflict begins. While it may be true that a certain set of managerial actions can be employed, modern research into the topic reveals that conflict escalations are complex

Received September 25, 2026; Revised November 1, 2026; Accepted November 12, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

processes involving not only administrative procedures but also cognitive and behavioral mechanisms (Budiwidodo et al., 2024). Once identified, a workplace conflict has already caused significant harm to the mental and psychological condition of employees involved.

The advancements in affective computing and computer vision enabled researchers to identify early signs of interpersonal conflict and intervene within the situation. Modern artificial intelligence technologies allow detecting facial expressions (Li & Deng, 2020), recognizing behaviors in videos (Sultani et al., 2018), and using multimodal fusion methods to make accurate predictions about people's conditions (Poria et al., 2017; Baltrusaitis et al., 2018). Thus, AI may be used in workplaces for the purposes of conflict prevention.

At the same time, the use of artificial intelligence to monitor employee's behavior poses certain ethical issues, such as invasion of privacy, bias, lack of consent, the problem of the human element, and the development of surveillance mentality (Kellogg et al., 2020; Raji & Buolamwini, 2019). Any technology aimed at predicting workplace conflicts is obliged to consider ethical concerns and employ appropriate norms.

In this paper, we address an existing gap by proposing a theoretically grounded multimodal AI framework for workplace conflict prediction. We describe the cognitive science theory underlying conflict escalation, review the literature on affective computing, computer vision, multimodal fusion, and AI-based employee monitoring, and report the findings of a qualitative study involving semi-structured interviews guided by seven research propositions. The study is anchored by the following research question: How does management perceive the effectiveness and acceptability of using AI-driven multimodal analysis for early conflict detection in the workplace?

RESEARCH METHODOLOGY

Study Design

The study was conducted with a qualitative research methodology, utilizing a semi-structured interview format. Eleven professionals across a variety of industries, with direct experience in workplace management, human resources, information technology, education, retail, or other organizational contexts, participated in the interviews. Participants ranged in their experience and familiarity with artificial intelligence systems, as indicated by their self-reporting on a 1-10 scale provided at the start of each interview. The diverse levels of experience within the participant pool provided a cross-section of AI professionals and modest AI familiarity. The intentional diversity of the pool supported the research intent; to broadly explore managerial experience beyond technology-focused roles.

All participants were explicitly assured that their responses would be anonymized, and that personally identifiable information would not be collected or retained. Participants were informed prior to beginning their interviews that there were no right or wrong answers, and that their skepticism, in addition to their enthusiasm for the research concepts, would prove valuable to the study.

Interview Instrument

Seven research hypotheses were broken down to form a 14-question semi-structured interview guide, while granting latitude for participants to elaborate and diverge from question framing. The interview guide opened with background and context-setting questions, progresses towards detailed questions regarding AI capabilities, and concludes with ethical and personal acceptance queries. This sequence structure was developed to elicit participant comfort and form contextual baselines prior to addressing opinions about unfamiliar and potentially sensitive technology.

The interview commenced with two questions to form the baseline for participants' professional background and their self-assessed familiarity with AI. The following three context-setting questions explored existing conflict management practices at their workplace and prior experience with employee monitoring systems. The subsequent six questions focused strongly on hypotheses H1 through H6, presenting specific AI capabilities, and inviting participants to make assessments on the value as well as potential disadvantages to the resources in the context of their organizations. Hypothesis H7 is explored in the final three questions, examining the conditions that would need to exist for participants to accept the adoption of such a system in their workplace, the monitoring limits they would mandate, and their overall opinion for their employer deploying AI-based behavioral monitoring in the workplace.

Interview responses were reviewed and documented upon closure. Common themes, along with keywords, were extracted from all 11 interviews to develop the structured set of concepts, apprehensions, and participant acceptance criteria.

Interview Participants and Procedure

The 11 participants were interviewed individually, through in-person meetings or via video call, for durations ranging from 30 to 45 minutes. Each interview opened with a standardized framing statement that explicitly welcomed participant skepticism, and assured anonymity of participant responses, positioning the research as an inquiry into perception rather than an assessment of participant familiarity or expertise.

The participant pool was deliberately comprised of individuals employed in a range of professional and organization contexts. Featured backgrounds include management, human resources education, retail supervision, technology, customer service, and public-facing service environments. This meets the research intent of understanding manager and practitioner perception across various organizational cultures.

Interviews were conducted in a semi-structured manner, with further probing presented when participant responded positively to a capability question. The follow-up questions presented specific use cases and hypothetical failures. When skepticism or opposition to a concept was exhibited, deeper probing was conducted to determine under what circumstances, if any, their position possibly would change. This probing

structure allowed for the range of acceptance and intolerance to be collected for each hypothesis domain.

Research Hypotheses

The interview instrument was structured around seven hypotheses, each addressing a distinct dimension of the proposed AI framework:

H1 (Primary Hypothesis): AI systems utilizing multimodal visual behavioral analysis may predict early signs of workplace conflict before explicit escalation occurs.

H2 (Emotion-Based Prediction): Negative emotional states detected through facial expression recognition may be significantly correlated with the emergence of interpersonal conflict.

H3 (Behavioral Dynamics): Changes in nonverbal behavioral patterns may be reliable indicators of escalating workplace tension.

H4 (Multimodal Superiority): Multimodal AI models may outperform unimodal (facial-only) models in predicting workplace conflict.

H5 (Temporal Escalation Patterns): Temporal analysis of behavioral data may provide stronger predictive capability than static, single-instance observations.

H6 (Anomaly Detection Link): Behavioral deviations from normal interaction patterns may be positively associated with the likelihood of conflict occurrence.

H7 (Ethical Impact): Human-centered AI frameworks that incorporate transparency and privacy-preserving mechanisms may improve user trust and acceptance of AI-based behavioral monitoring systems.

RESULTS AND ANALYSIS

H1: Primary Hypothesis — Multimodal Conflict Prediction

Interview responses to questions corresponding to H1 revealed a pattern of qualified and contingent acceptance, though varied in context. Several participants believed that AI-based conflict prediction could offer real organizational value, chiefly in environments where teams are distributed across several regions, or in settings with limited direct managerial oversight. One participant noted the potential utility of such a system specifically for remote monitoring contexts, while another described an AI monitoring facial expression, body language, and interaction patterns—combined with speech tone and physiological indicators—as potentially more effective than current reactive approaches.

However, acceptance was notably conditional. A common theme amongst participants was the experience that workplace conflict occurred with such low frequency or severity in their specific environments that the costs and risks of continuous AI monitoring would not be justifiable. Multiple participants flagged that the primary determinant of adoption would be organizational implementation design—specifically, assurances that AI outputs would be used to support human judgement rather than replace it. Participants who

expressed the strongest skepticism consistently cited concerns about what happens when the system is wrong, pointing to potential loss of institutional trust as a consequence more damaging than the conflicts that the system was designed to prevent.

H2: Emotion-Based Prediction

Interview responses addressing H2 were more divided than for H1, reflecting deep participant uncertainty about the reliability and appropriateness of facial expression recognition as an organizational resource. Most participants independently raised the complexity and dynamic nature of human emotion as a fundamental obstacle, noting that people would alter their facial expressions and behavior if they knew they were being monitored —effectively masking the signals the system seeks to detect. One participant remarked that AI-generated alerts about emotions can feel impersonal, reducing the complexity of interpersonal dynamics to algorithmic labels.

Participants with backgrounds in human resources and education were particularly emphatic that emotional expression in professional settings is highly dependent on the context. An AI-based system may classify a facial expression as anger, however, in isolation, it may reflect concentration, physical discomfort, or cultural communication norms rather than interpersonal hostility. One participant explicitly preferred temporal behavioral monitoring over facial expression analysis, citing a greater risk of false-positives over single-instance facial readings. These responses reinforce the apprehensive stance of the referenced literature towards unimodal approaches and provide qualitative empirical grounding for the multimodal design rationale of the proposed framework.

H3: Behavioral Dynamics

Participant responses to questions corresponding to H3 were notably mixed, differing on the theoretical expectation that behavioral pattern recognition would be largely accepted. While several participants acknowledged that observable behavioral changes could serve as meaningful conflict indicators when interpreted in context, a significant number expressed concern that automated flagging of single-instance behavioral episodes would be more problematic than beneficial. One participant stated directly that AI pointing out behaviors such as frustration or aggressive gestures would initiate more conflict than it resolved.

The concept of invasiveness remained prevalent throughout this discussion: participants drew a distinction between a human manager who evaluates and perceives changes in behavior from an existing professional relationship, and an AI-based behavioral monitoring system that detects and flags distinct behavioral events without that interpersonal context. Participants broadly agreed that changes in behavioral signals increase in significance when tracked over periods of time than when flagged as isolated incidents, a finding that directly correlates to the temporal escalation hypothesis (H5).

H4: Multimodal Superiority

Support for the multimodal superiority hypothesis was the most uniform finding across the interview data. When presented with the comparison between a system that only detects facial expressions and one that combines facial analysis with body language and interaction patterns, participants equally preferred the multimodal option, though the reasoning varied amongst interviewees. Several participants based their position on grounds that no single behavioral signal is sufficiently reliable in isolation, with one describing an ideal system as combining facial expressions, body language, interaction patterns, speech tone, respiratory measures, and physiological indicators analogous to vital signs.

Importantly, some participants who did not support facial-expression-only monitoring as too invasive, indicated provisional acceptance of multimodal approaches, particularly those oriented toward detecting physical conflict rather than emotional states. These findings confirm practitioner intuitions that align with the empirical evidence for multimodal superiority documented in the literature and suggest that framing the system as a holistic behavioral observer rather than a facial recognition system may increase employee acceptance across organizations.

H5: Temporal Escalation Patterns

Responses to interview questions corresponding to H5 provided some of the strongest qualitative support of any hypothesis domain. Participants consistently expressed the benefit of tracking behavioral change over time rather than responding to isolated incidents. The concept of an individual behavioral baseline emerged across multiple participant interviews as the critical prerequisite for meaningful AI monitoring: several participants explicitly stated that an established baseline is needed to understand each person, and that the more data a system accumulates, the more accurate its baseline determination becomes.

However, this standpoint was not without nuance—one participant noted that baselines would require updating if significant life events occurred, and altered an individual's normal behavioral patterns, with another questioning whether individual behavioral spikes within a longitudinal trend would offer reliable distinction from genuine escalation. One participant proposed a "vitals check" model rather than event-by-event flagging, a suggestion that closely mirrors the temporal fusion logic of the proposed framework. Overall, participant responses justify the temporal escalation principle as the most cognitively innate and organizationally acceptable aspect of the AI monitoring concept.

H6: Anomaly Detection Link

Participant responses to the anomaly detection scenario - framed as an AI flagging someone as "acting differently than usual" - provoked the most overtly divided reactions of any hypothesis domain. Some participants found the concept genuinely useful as a tool for detecting stress or early

interpersonal tension, emphasizing the value of recognizing patterns and that greater data accumulation improves model accuracy. However, a significant percentage of the participant pool characterized the same concept as invasive surveillance into an individual's emotional life.

A recurring participant concern was the potential for anomaly flags to create bias in the mind of the data recipient—the manager or HR professional who receives the alert—by priming them to interpret subsequent interactions with the flagged individual through a conflict lens regardless of whether a genuine conflict or problem exists. Participants also raised the cultural and individual specificity of behavioral baselines: behaviors that appear uncharacteristic in one cultural or organizational context may be conventional in another. One participant further observed that the system would need continuous updates to account for how career changes and life events may influence baseline behavior.

H7: Ethical Impact and Trust

Interview responses addressing H7 generated the most elaborate qualitative data of any section. Across all 11 interviews, transparency emerged as the non-negotiable baseline condition for acceptance: participants unanimously asserted that employees must be able to see all data collected about them and must be aware of what data their management can access. Data minimization and deletion limits were raised as design requirements, with multiple participants specifying that knowing data is deleted after a defined period is crucial for their acceptance.

A meaningful subset of participants insisted on an opt-in or opt-out mechanism, with one framing it as an all-or-nothing proposition—either all employees participate in the program or none do, to avoid stigmatizing those who opt out. Consent was referenced as both a legal and a cultural prerequisite, with participants citing the need for signed consent forms, data privacy protections, no centralized facial databases, and encrypted storage. Participants were equally insistent about what must not be monitored: bathrooms, personal phone calls, gym time, and any period outside working hours were cited as categorical exclusions.

A significant concern raised by several participants was the potential for AI monitoring data to influence compensation, promotion, or performance reviews. Safeguards including psychologist review of flagged data were proposed as mechanisms to ensure that AI outputs were interpreted in a clinically and contextually appropriate manner before reaching managerial attention. These findings provide direct empirical grounding for the human-centered design orientation of the proposed framework, while surfacing specific design requirements that the existing literature on ethical AI does not fully address.

DISCUSSION, IMPLICATIONS AND LIMITATIONS

The findings of this study have several implications for the theoretical understanding of workplace conflict and its relationship to behavioral analytics. First, the strong practitioner support for temporal analysis over

static observation empirically validates cognitive science models of conflict as a developmental process rather than a discrete event. This alignment between practitioner intuition and theoretical models suggests that AI systems designed on conflict-escalation theoretical foundations will be both technically superior and more readily accepted by organizational stakeholders.

Second, the qualified support for facial expression recognition, particularly the concerns raised about demographic accuracy disparities, highlights the need for affective computing research to prioritize fairness and generalizability rather than aggregate accuracy metrics. A system that is highly accurate on average but systematically inaccurate for specific demographic groups is not merely technically flawed but ethically impermissible in organizational deployment.

Third, the consistent preference for multimodal over unimodal approaches has implications beyond technical system design. It suggests that practitioners recognize, at least intuitively, the inherent limitations of reductive behavioral indicators and are more likely to trust systems that model the complexity of human behavioral expression. This has important implications for human-AI collaboration in organizational management more broadly.

The findings suggest several practical principles for organizations considering the deployment of AI-based conflict prediction systems. Transparency must be treated not as a regulatory compliance requirement but as an organizational design imperative. Employees who understand what is being monitored, how data is used, and what safeguards are in place are substantially more likely to accept and cooperate with AI monitoring systems.

Privacy-by-design principles should be embedded in system architecture from the outset rather than retrofitted as compliance measures. This means processing visual data locally and immediately, retaining only abstract behavioral indicators rather than raw footage, anonymizing communication data before analysis, and implementing strict data minimization policies. Organizations should also establish independent oversight mechanisms to provide ongoing accountability for system implementation and use.

Finally, AI outputs should always be configured as decision-support tools rather than decision-making systems. Algorithmic conflict predictions should be presented to trained human managers as probabilistic indicators warranting further investigation, never as automatic triggers for organizational action. Preserving human judgment at the center of conflict management processes is both ethically necessary and practically beneficial.

This study has several limitations that should be acknowledged. The sample of 11 participants, while sufficient for exploratory qualitative research, limits the transferability of findings to broader managerial populations. The purposive sampling approach prioritized professional diversity over demographic representativeness, meaning that the distribution of attitudes observed may not reflect any industry or organizational type. Future research should seek larger and more demographically stratified samples, potentially employing quantitative instruments derived from the thematic categories identified here.

Additionally, the research is limited to self-reported perceptions toward a hypothetical AI system rather than behavioral responses to an actual deployed

system. Attitudes expressed in interview contexts may differ substantially from actual acceptance or resistance when organizational monitoring is operationalized. Longitudinal field studies examining real-world adoption dynamics would provide substantially greater ecological validity.

Finally, the proposed AI framework has not yet been technically implemented or validated in a live organizational environment. Future research should prioritize the development of prototype implementations, empirical evaluation of technical performance across demographically diverse populations, and longitudinal assessment of both conflict prediction accuracy and organizational impact.

CONCLUSION

This paper has presented a theoretically grounded multimodal AI framework for the early prediction of workplace conflict, integrating facial expression recognition, body language analysis, text-based communication indicators, and temporal behavioral pattern analysis. The framework is situated within established cognitive science models of conflict escalation and designed in accordance with human-centered AI principles that prioritize transparency, privacy preservation, and human oversight. The central research question —how managers perceive the effectiveness, acceptability, and practical challenges of AI-driven multimodal conflict detection—was investigated through semi-structured interviews with 11 participants spanning diverse professional contexts.

Interview findings reveal a nuanced picture of conditional acceptance. Practitioners broadly endorse temporally aware behavioral analysis and multimodal data integration as more credible and organizationally appropriate than single-instance or unimodal approaches. At the same time, participants articulated consistent and substantive concerns: the invasiveness of facial-expression-only monitoring, the risk of cultural and demographic bias in anomaly detection, the potential for AI outputs to be weaponized in performance management, and the fundamental difficulty of establishing and maintaining individual behavioral baselines. The conditions for acceptance identified across interviews represent a design specification that goes beyond existing ethical AI frameworks and should inform both system architecture and organizational policy development.

The transition from reactive conflict resolution to proactive conflict prevention represents a significant opportunity for organizational management, but one whose realization depends critically on whether organizations can meet the trust conditions that practitioners have identified. Realizing this opportunity responsibly requires that AI system developers, organizational leaders, human resource professionals, and policymakers collaborate to ensure that the power of behavioral analytics is deployed in service of employee well-being rather than surveillance and control. The framework proposed here offers a foundation for that responsible deployment, grounded in both the capabilities and the ethical obligations that define this frontier of applied cognitive science.

ACKNOWLEDGMENTS

The authors wish to thank all 11 interview participants for their time, candor, and willingness to engage with a research area that touches on sensitive questions of workplace autonomy and surveillance. Their insights are the empirical core of this work.

REFERENCES

Baltrusaitis, T., Ahuja, C., and Morency, L.-P. (2018). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence 41, 2, 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Budiwidodo, P., et al. (2024). Reactive conflict management and its limitations: Toward proactive organizational approaches. Journal of Organizational Behavior Studies 15, 2, 88–107.

Kellogg, K. C., Valentine, M. A., and Christin, A. (2020). Algorithms at work: The new contested terrain of control. Academy of Management Annals 14, 1, 366–410. https://doi.org/10.5465/annals.2018.0174

Li, S. and Deng, W. (2020). Deep facial expression recognition: A survey. IEEE Transactions on Affective Computing 13, 3, 1195–1215. https://doi.org/10.1109/TAFFC.2020.2981446

Poria, S., Cambria, E., Bajpai, R., and Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion 37, 98–125. https://doi.org/10.1016/j.inffus.2017.02.003

Raji, I. D. and Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 429–435. https://doi.org/10.1145/3351095.3372873

Sultani, W., Chen, C., and Shah, M. (2018). Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE CVPR, 6479–6488. https://doi.org/10.1109/CVPR.2018.00666