Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 8, 2025

EMOTIONAL RECOGNITION IN SOCIALLY INTERACTIVE ROBOTS: A COMPREHENSIVE REVIEW

Aryan Ahmed[*1], Hashir Alvi[2], M.Hamza Khalid[3], Fawad Naseer[4]

[*1, 2, 3]*O & A Level College Section, The City School, Chenab Campus, Canal Road, Faisalabad, 38000, Pakistan*
[4]*Department of Computer Science and Software Engineering, Beaconhouse International College, Faisalabad, 38000, Pakistan*

[*1]aryanahmed.edu@gmail.com, [2]hashiralvi000@gmail.com, [3]mhamzakhalid2019@gmail.com, [4]fawad.naseer@bic.edu.pk

DOI: https://doi.org/10.5281/zenodo.16810320

Keywords
Emotion-Aware Robot, Human-Robot Interaction, Emotional Recognition, Social Robots, Robot Companion, Artificial Intelligence (AI), Machine Learning (ML).

Article History
Received: 30 April, 2025
Accepted: 11 July, 2025
Published: 12 August, 2025

Corresponding Author: * Aryan Ahmed

Abstract
The study explores the evolving landscape of Human-Robot Interaction (HRI) within workplace environments, examining the intricate relationship between humans and robots in professional settings. As artificial intelligence (AI), particularly generative AI, becomes increasingly integrated into the workplace, understanding HRI dynamics is paramount for businesses, society, and ethical considerations. This paper synthesizes current research to provide a comprehensive overview of key trends, challenges, and future research agendas in workplace HRI, examining how advancements in AI, such as natural language processing and adaptive interaction, are shaping human-robot collaboration, communication, and coexistence in professional contexts. The review highlights the growing importance of trust, anthropomorphism, and social intelligence in facilitating effective HRI, drawing upon insights from psychology, computer science, and sociology. It addresses the implications of generative AI technologies like ChatGPT in creating more intuitive and human-like interactions, while acknowledging the ethical dilemmas and potential pitfalls associated with these technologies. Furthermore, the paper discusses how workplace HRI can enhance employee well-being, improve operational efficiency, and foster innovation. By identifying gaps in current research, this review aims to guide future studies, focusing on areas such as long-term user adaptation, AI decision-making interpretability, and the development of robust ethical frameworks for responsible AI deployment in professional environments.

INTRODUCTION

1.1 The Evolving Landscape of HRI

HRI has emerged as a critical multidisciplinary research field, integrating insights from engineering, psychology, sociology, and computer science to address the complex dynamics between humans and robotic agents (Frijns & Schürer, 2022; Obrenovic et al., 2024). Historically confined to industrial settings where robots performed repetitive tasks in isolation, HRI is rapidly expanding into collaborative and social environments, including healthcare, education, manufacturing, and domestic settings where robots work close to humans.

This proliferation necessitates a shift from purely technical challenges to a more holistic, socio-technical perspective that considers the human element as integral to system success. The overarching goal of modern HRI is to develop robots that can collaborate effectively with humans, understand their intentions, and respond in a manner that is socially appropriate, valuable, and safe. This involves moving beyond user interface design to considering the entire socio-technical system, including the social and cultural factors that shape interactions (Frijns & Schürer, 2022).

As robots become more autonomous and integrated into daily life, interaction quality becomes paramount, influencing user acceptance, task efficiency, and overall safety. This evolution represents a fundamental shift from traditional human-machine interfaces to dynamic, adaptive partnerships that require a sophisticated understanding of human psychology, social dynamics, and contextual awareness.

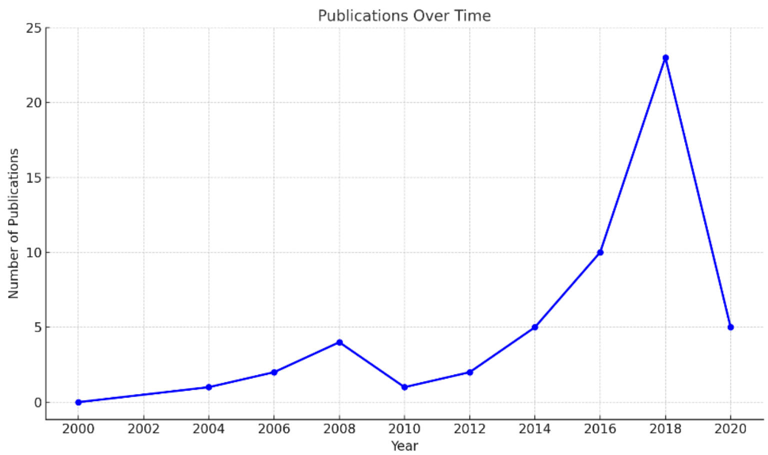

Figure 1. Growth of Publications on Facial Emotion Expressions in HRI (2000-2020)

Note. This chart illustrates the trend in the number of academic publications focused on facial emotion recognition and generation in the context of Human-Robot Interaction. The data shows a significant increase in research activity, particularly after 2012, peaking in 2018. Adapted from "Facial Emotion Expressions in Human–Robot Interaction: A Survey" by N. Rawal and R. M. Stock-Homburg, 2022, International Journal of Social Robotics, 14, p. 1584.

Table 1. Top Keyphrases in HRI Research by Relevance (2013-2022)

Rank	Key phrase	Relevance Score
1	Human-robot Interaction	1.00
2	Robot	0.56
3	Social Robot	0.40
4	Man-machine Systems	0.31
5	Robotics	0.28
6	Anthropomorphic Robots	0.21
7	Humanoid Robot	0.11
8	Human-computer Interaction (HCI)	0.06
9	Intelligent Robots	0.05
10	Anthropomorphism	0.04
11	Trust	0.03
12	Emotion	0.03
13	Empathy	0.02

Note. This table displays the most relevant keyphrases from a scientometric analysis of HRI literature, indicating the central themes of the research field. The high relevance of terms like "Social Robot," "Anthropomorphism," and "Trust" underscores the field's focus on the psychological and social dimensions of interaction. Adapted from "Generative AI and human–robot interaction: implications and future agenda for business, society and ethics" by B. Obrenovic et al., 2024, AI & Society.

1.2 Foundational Psychological Constructs in HRI

The success of human-robot collaboration hinges on understanding the psychological and social factors that govern interactions. Research has identified several core constructs fundamental to designing effective and acceptable robotic partners.

1.2.1 The Centrality of Trust and Vulnerability

Trust serves as a cornerstone for long-term acceptance and success of HRI (de Pagter, 2022; Hannibal & Weiss, 2022). The concept extends beyond simple reliance on robot performance and predictability, increasingly involving aspects of interpersonal trust where humans perceive robots as having motives or intentions, especially as they are designed with more social capabilities and anthropomorphic features.

A significant evolution in this area is the reframing of vulnerability from a negative factor to be eliminated through engineering to a fundamental precondition for developing interpersonal trust. Without the possibility of harm or failure, trust is not required; it is the act of trusting despite vulnerability that defines the relationship (Hannibal & Weiss, 2022). This perspective challenges researchers to balance safety with the managed exposure necessary to foster genuine trust, preventing systems that are so overly conservative they hinder effective collaboration.

The process of building trust is multifaceted, influenced by a robot's observable behaviour and appearance, the transparency of its design processes, and societal narratives surrounding robotic technology. Understanding these dynamics is crucial for developing robots that can establish and maintain productive working relationships with human partners.

1.2.2 Emotion Recognition and Expression

For interactions to feel natural and intuitive, robots must understand and appropriately respond to human emotions (Zhao, 2023; Rawal & Stock-Homburg, 2022). Non-verbal cues, particularly facial expressions, are a primary communication channel, conveying up to 55% of affective information in interactions. Consequently, major HRI research focuses on both the robot's ability to recognise human facial expressions and its capacity to generate expressive responses.

This area faces a significant challenge known as the "in the wild" problem. While machine learning models can achieve high accuracy (over 90%) in recognising emotions from controlled, predefined datasets, their performance drops considerably in real-time, unconstrained environments where lighting, head poses, and occlusions vary. Bridging this gap is crucial for creating robots that can function effectively outside laboratory settings.

Methods for robotic expression range from hand-coded movements of facial features to automated, learned responses, with the latter offering potential for more nuanced and context-aware emotional displays. The development of robust emotion recognition and expression capabilities remains a critical challenge for creating empathetic and socially intelligent robots.

1.2.3 Agency, Control, and Collaboration

Agency, or the capacity to act, is critical in HRI, particularly as robots are endowed with greater autonomy (Zafari & Koeszegi, 2022). Humans naturally ascribe agency to nonhuman entities that exhibit goal-directed or unpredictable behaviour, which shapes their expectations and attitudes. Research distinguishes true collaboration from mere cooperation; cooperation involves division of labour, whereas collaboration requires mutual engagement where all agents work on tasks together.

A key finding is the relationship between robot agency and human perceived control. Studies show that high robot agency is perceived negatively only when associated with low human control over the process. This suggests that for successful human-robot teams, human partners must feel empowered and in control, even when robots are highly autonomous. The robot's interaction style also plays a significant role; a "person-oriented" interaction style providing socio-emotional support can increase human self-efficacy more than purely "task-oriented" styles.

Table 2. Summary of Foundational Psychological Constructs in HRI

Construct	Definition	Key Findings & Implications	Source(s)
Trust	The willingness of a person to be vulnerable to the actions of a robot is based on the expectation that the robot will perform a particular action important to the trustor.	Trust is not a static belief but an "event" that emerges from interaction. It is built through observable behavior, transparency in design, and societal narratives.	de Pagter (2022); Hannibal & Weiss (2022)
Vulnerability	A fundamental precondition for trust, representing the possibility of harm or failure that makes trust necessary.	Vulnerability should not be engineered out of systems entirely. Managed exposure to a robot's fallibility is necessary to foster genuine interpersonal trust, as opposed to mere reliance.	Hannibal & Weiss (2022)
Agency & Control	Agency is the capacity of a robot to act autonomously. Perceived human control is the user's sense of empowerment and influence over the collaborative process.	High robot agency is perceived negatively only when associated with low human control. To ensure positive collaboration, human partners must feel in control, even when the robot is highly autonomous.	Zafari & Koeszegi (2022)

Note. This table outlines core psychological constructs that are central to the success of human-robot collaboration. The relationship between these factors is complex and highlights the need for a socio-technical design approach.

1.3 Methodological Approaches and Research Challenges

1.3.1 Evaluation Methods and Metrics

Evaluating HRI quality and effectiveness requires multi-pronged approaches. Common methods include user studies employing questionnaires and scales to measure subjective perceptions of robots, such as perceived intelligence, likeability, and trustworthiness. To assess HRI's impact on task-based outcomes, researchers measure objective performance metrics like completion time, error rates, and overall team productivity (Wachowiak et al., 2023).

Beyond performance, understanding the user's mental model—their understanding of how robots work and what they will do next—is crucial. Methodologies such as think-aloud protocols and participatory design workshops provide qualitative insights and involve end-users in development processes from the beginning. Specialised experimental methods include analysing social signal coding from video to systematically observe human involuntary motions in response to robot speed and proximity, helping design psychologically safer systems.

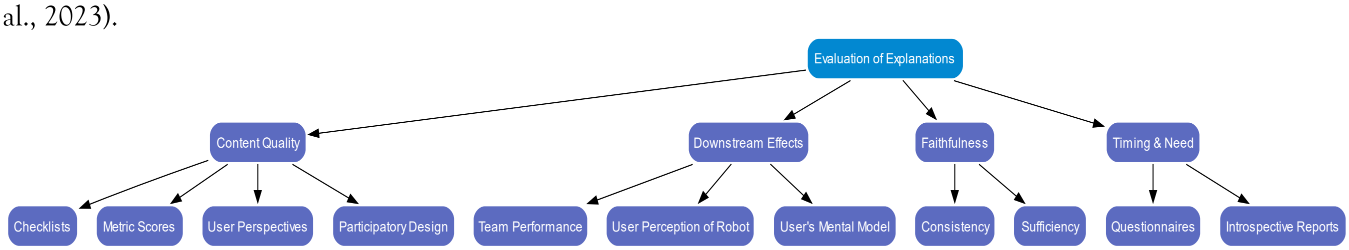

Figure 2. Conceptual Framework for the Evaluation of Explanations in HRI

Note. The evaluation of explainable AI in robotics is a multi-dimensional problem. This framework categorizes evaluation methods based on what is being measured: the intrinsic quality of the explanation's content, the effects the explanation has on the interaction, the explanation's faithfulness to the robot's internal processes, and the appropriateness of its timing. Adapted from "A Survey of Evaluation Methods and Metrics for Explanations in Human-Robot Interaction (HRI)" by L. Wachowiak, O. Celiktutan, A. Coles, & G. Canal, 2023.

1.3.2 Persistent Research Challenges

The primary technical hurdle remains developing robust perception systems, especially for real-time emotion recognition in unconstrained environments. Furthermore, integrating information from multiple modalities (facial expressions, vocal tone, gestures, physiological signals) into a coherent understanding of human state presents complex fusion problems. From a design perspective, there is a need to shift from a narrow user interface focus to broader socio-technical systems approaches. A significant design challenge lies in creating robot behaviours that are "expectable" to humans, avoiding psychological distress without making robot actions overly conservative and inefficient.

Human emotion is inherently subjective and varies significantly across individuals and cultures, making universal emotion recognition models difficult to achieve. The ascription of agency and emotion to robots raises profound ethical questions regarding privacy, consent, and potential for manipulation or deception. Calibrating human trust to appropriate levels is another critical challenge; both under-trust (refusal to use capable systems) and over-trust (over-reliance on fallible systems) can lead to negative outcomes.

1.4. Advanced Technologies and Emerging Paradigms in Workplace HRI

1.4.1 The Symbiotic Integration of AI, Machine Learning, and Deep Learning

The integration of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) has catalysed a paradigm shift in robotics, transforming them from pre-programmed machines into intelligent systems capable of learning, adapting, and performing complex autonomous operations (Ganesan, 2023). This evolution is particularly impactful in workplace settings, where AI-driven robots enhance autonomy, precision, and efficiency across sectors including manufacturing, logistics, and healthcare.

AI provides the cognitive framework for robots to reason and plan, while ML and DL equip them with the ability to extract knowledge from vast datasets, enabling capabilities such as autonomous navigation, object recognition, and predictive maintenance. This symbiotic relationship between AI and robotics is revolutionising how tasks are performed and paving the way for more sophisticated and intuitive human-robot collaboration.

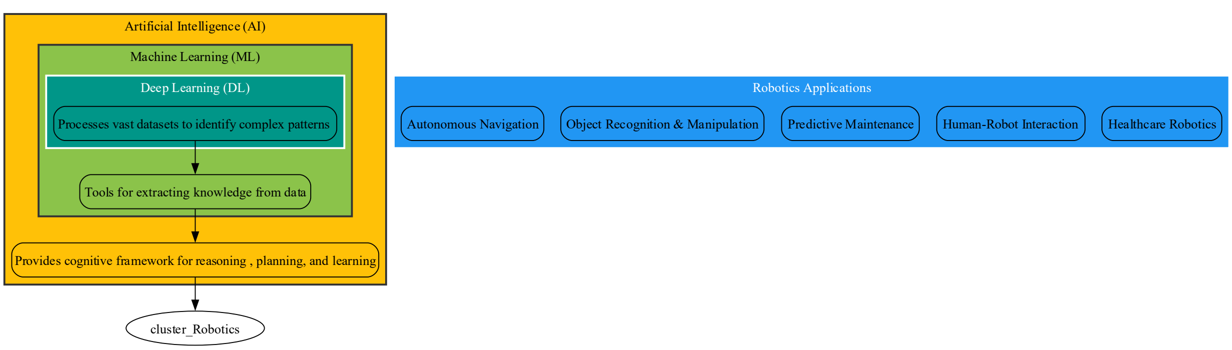

Figure 3. The Relationship between AI, Machine Learning, Deep Learning, and Robotic Capabilities

Note. This diagram illustrates how Artificial Intelligence serves as the overarching field providing a cognitive framework for robotics. Machine Learning is a subfield of AI that equips robots with tools to learn from data, while Deep Learning, a subfield of ML, enables the processing of large, complex datasets. These technologies collectively drive advancements in key robotics applications. Adapted from "Revolutionizing Robotics with AI, Machine Learning, and Deep Learning" by P. Ganesan, 2023, Journal of Artificial Intelligence, Machine Learning and Data Science, 1(4).

1.4.2 Expanding Communication Bandwidth: Nonverbal Cues in HRI

Effective communication is the cornerstone of successful HRI. While speech-based interaction has been a primary focus, the importance of nonverbal communication is increasingly recognised for creating natural and intuitive interactions (Wang et al., 2018; Zhang & Fitter, 2023). Nonverbal cues such as gestures and sounds can convey intent more efficiently than verbal communication in certain contexts, especially in noisy or restrictive environments like manufacturing floors or underwater operations.

Hand and arm gestures, prevalent in daily human communication, are being systematically studied to build more intuitive bridges between humans and robots. Research has categorised gestures in HRI into referential, interactional, and symbolic types, drawing parallels from human-human and human-computer interaction to improve robot understanding of human intent. This allows for more natural control in tasks such as object manipulation and navigation.

Similarly, nonverbal sound offers a rich and underexplored channel for HRI. Beyond speech, robot-created sounds can be used for explicit communication, such as alerts and feedback, or to enhance robot sociability. Systematic review has led to the development of taxonomies for nonverbal robot sound, categorising them by form (electronic sounds, vocables) and function (functional, emotional, consequential). These frameworks aim to unify diverse research and provide a common language for designing and evaluating the auditory dimension of HRI.

Table 3. A Taxonomy of Nonverbal Communication Cues in HRI

Modality	Type	Function/Explanation	Example	Source(s)
Gesture	Referential	To indicate objects or locations.	Pointing at a tool for the robot to pick up.	Wang et al. (2018)
	Interactional	To regulate interaction with a partner (e.g., initiate, terminate).	A “stop” hand signal to pause the robot.	Wang et al. (2018)
	Symbolic	Associated with a firm cultural meaning.	A “thumbs-up” gesture to indicate approval.	Wang et al. (2018)
Sound	Functional	To explicitly convey non-emotional information.	A beep to confirm a command was received.	Zhang & Fitter (2023)
	Emotional	To explicitly convey emotions.	A cheerful series of tones to express “happiness.”	Zhang & Fitter (2023)
	Consequential	Sound made by the operation of the robot itself.	The whirring sound of a robot’s motors.	Zhang & Fitter (2023)
	Transformative	Sound made to alter the robot’s original sound profile.	Adding a synthesized sound to mask motor noise.	Zhang & Fitter (2023)

Note. This table categorizes common nonverbal cues used in HRI. Gestures are classified by their communicative role, while nonverbal sounds are classified by their intended purpose or origin.

1.4.3 The Rise of Immersive and Remote Collaboration through Extended Reality

Extended Reality (XR) technologies, encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), are emerging as transformative forces in HRI, particularly for remote collaboration (Wang et al., 2024). XR provides immersive and intuitive interfaces for humans to control and interact with robots from a distance, overcoming geographical and safety constraints. This is especially crucial in hazardous environments or for tasks requiring expert supervision from remote locations.

Systematic review of XR-enabled remote HRI systems reveals a comprehensive design space including dimensions such as XR technologies used, interaction modalities, virtual interface design, and user perspective. VR head-mounted displays (HMDs) are currently the dominant technology, offering heightened presence sense and immersive views of remote environments. By creating digital twins of robots and their workspaces, XR allows for intuitive control and better situational awareness, paving the way for massive interaction with generalist robotics across various fields.

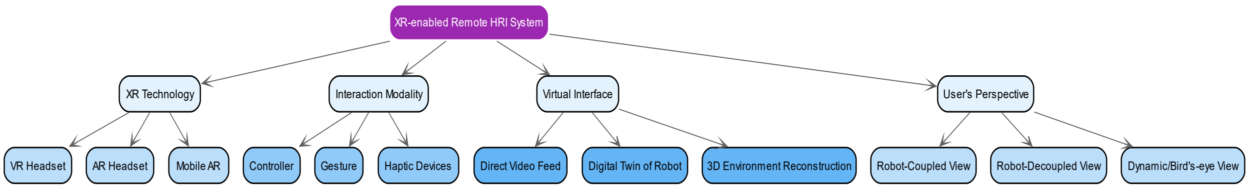

Figure 4. Key Dimensions of an XR-Enabled Remote HRI System

Note. This diagram outlines the core design dimensions of systems that use Extended Reality (XR) for remote Human-Robot Interaction. A complete system involves choices regarding the specific XR hardware, how the user provides input, the design of the virtual interface that represents the remote environment, and the camera perspective provided to the user. Adapted from "Towards Massive Interaction with Generalist Robotics: A Systematic Review of XR-enabled Remote Human-Robot Interaction Systems" by X. Wang, L. Shen, & L.-H. Lee, 2024, arXiv.

1.5 Addressing Complexities of Social Group Interactions

While much HRI research has focused on dyadic (one-on-one) interactions, real-world environments, including workplaces, are often characterised by group dynamics (Nigro et al., 2024). Group HRI presents unique computational challenges that are frequently overlooked in the literature. Key challenges include perception tasks such as detecting groups and engagement, and behaviour generation, including developing appropriate approaches and conversational behaviours.

The complexity of group interactions increases with group size, leading to subgroup formation, competition for speaking turns, and overlapping speech, all posing significant challenges for current robotic systems. To manage these complexities, researchers have often limited group sizes in studies and conducted experiments in controlled lab environments. Future work in group HRI needs to address research gaps by improving the detection of subgroups and interpersonal relationships and developing more robust models for perception and behavior generation in larger, more dynamic groups.

1.6 The Emergence of Human-Robot Psychological Contracts in the Workplace

As social robots become more sophisticated and integrated into workplaces, the nature of relationships between humans and technology is shifting from tool-users to active partners (Bankins & Formosa, 2019). This evolving relationship has led to the conceptualisation of human-robot psychological contracts, referring to implicit and subjective beliefs regarding reciprocal exchange agreements between employees and social robots.

Drawing on social exchange theory and reciprocity concepts, researchers argue that humans can form psychological contracts with social robots, attributing agency and forming attachments to these "machine-human hybrids." However, this "synthetic relationship" is complex, with humans recognising robots as "alive enough" to engage with but also "machine-enough" to be treated differently than human colleagues. This can lead to imbalanced reciprocity in psychological contracts, with potential spillover effects on human-human workplace relationships.

1.7 Ensuring Safe and Comfortable Collaboration through Advanced Motion Planning

Safety and human comfort are paramount in HRI, especially in collaborative settings where humans and robots share workspaces (Beck & Kugi, 2022). Motion planning for collaborative robots must go beyond simple task execution to incorporate these crucial elements. Key properties for motion planning algorithms in human-robot collaboration include ensuring safety through collision avoidance and

promoting human comfort through fluency, legibility, and human-like motion.
Legibility, or the ability of robots to convey intent through movements, is crucial for increasing perceived safety and enabling smooth collaboration. Similarly, interaction fluency, influenced by factors like human idle time and robot functional delay, is a key determinant of collaboration quality. To address these requirements, researchers are developing advanced motion planning algorithms such as receding horizon trajectory optimisation, which can incorporate safety and comfort criteria as objective functions and constraints. These algorithms aim to generate trajectories that are not only collision-free but also smooth, predictable, and natural, thereby fostering trust and acceptance of robotic collaborators.

Table 4. Relative Frequency of Human Involuntary Motion (IM) by Robot Proximity and Velocity

Distance to Human (m)	Robot Velocity (m/s)	Probability of IM
0.25	0.25	0.00
0.20	0.25	0.04
0.15	0.25	0.04
0.10	0.25	0.00
0.05	0.25	0.15
0.00	0.25	0.42
0.25	0.50	0.17
0.20	0.60	0.12
0.15	0.70	0.18
0.10	0.85	0.28
0.05	0.90	0.52
0.00	1.00	0.73

Note. This table shows the experimentally observed probability of a human exhibiting an involuntary startle or surprise motion in response to a robot approaching them. The probability of IM increases significantly as the robot gets closer and moves faster, highlighting the importance of motion planning that considers human psychological comfort. Probabilities over 0.40 are bolded for emphasis. Adapted from "Expectable Motion Unit: Avoiding Hazards From Human Involuntary Motions in Human-Robot Interaction" by R. J. Kirschner et al., 2022, IEEE Robotics and Automation Letters, 7(2), p. 2474.

1.8 Decoding Human Affect: The Role of Physiological Signals

For robots to be truly effective collaborators, they need to understand the emotional state of their human partners. While facial expressions and voice tone are common channels for emotion recognition, they can be voluntarily controlled or faked (Swati et al., 2024). Physiological signals such as Heart Rate Variability (HRV) derived from electrocardiogram (ECG) signals offer more reliable and involuntary measures of emotional state.
Research is being conducted to analyse the impact on HRV to develop more effective emotion recognition systems for HRI. By creating new databases of HRV data for various emotions and examining the influence of factors like gender, age, and profession, researchers aim to build models that can accurately recognise human emotions in natural environments. The insights gained can be used to create more empathetic and responsive robots that can adapt their behaviour based on human emotional state, leading to more effective and engaging interactions.

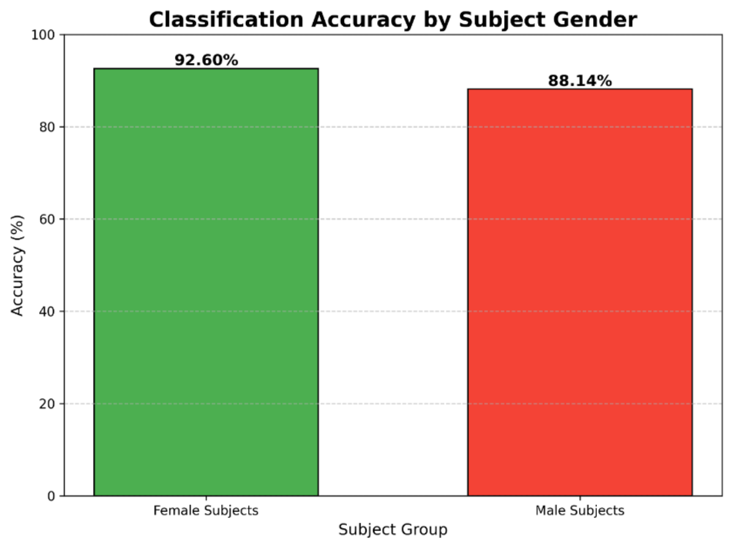

Figure 5. Comparison of Emotion Recognition Accuracy from Heart Rate Variability (HRV) Between Genders

Note. This chart displays the difference in emotion recognition accuracy using a Convolutional Neural Network (CNN) model trained on Heart Rate Variability (HRV) data. The model achieved a higher accuracy for female participants compared to male participants, suggesting that the emotional impact on HRV may differ by gender. Adapted from "Impact of Emotion on Heart Rate Variability for Effective Human Robot Interaction" by S. Swati, S. Singh, and A. K. Saxena, 2024, SSRN.

1.9 Vision Beyond Boundaries: The Potential of Large Vision Models

The recent success of Large Language Models (LLMs) in natural language processing is inspiring the development of Large Vision Models (LVMs), which are poised to revolutionise vision-based analysis and interpretation in HRI (Zhang et al., 2024). While vision models have long been used in HRI for tasks like object detection and gesture recognition, there is a lack of structured research on applying LVMs, particularly domain-specific LVMs, to this field.

Domain-specific LVMs, trained on visual content relevant to particular contexts, have the potential to outperform general models in specialised HRI systems. To guide future research and development, an initial design space for domain-specific LVMs in HRI has been proposed, including dimensions such as HRI contexts, vision-based tasks, and specific domains (healthcare, manufacturing, social interaction). By leveraging LVM power, future HRI systems can achieve more intelligent visual perception and interpretation, leading to more robust and seamless integration of robots into human society.

2. Literature Review

Emotion recognition (ER) and social-emotional interaction are central to the evolution of Human-Robot Interaction (HRI), particularly as robots transition from tools into collaborative agents. The ability of robots to understand, interpret, and respond to human emotions directly impacts their effectiveness, trustworthiness, and social acceptance. This literature review explores state-of-the-art developments in ER within HRI, focusing on three interconnected domains: (1) ER for Robot Safety & Performance, (2) Social-Emotional ER in Service Robots, and (3) Ethical & Design Aspects of Robot Emotional Intelligence. These categories synthesise methodological trends, comparative performances, distinctive contributions, and ongoing research challenges, providing a comprehensive backdrop to this study's focus on developing emotionally intelligent and context-aware robotic systems.

2.1 ER for Robot Safety & Performance

Robots in industrial and task-driven environments are increasingly expected to not only perform routine functions but also collaborate safely with humans.

Recognising human emotional states—such as stress, fatigue, or frustration—has emerged as a strategy to improve workplace safety and robot adaptability. This section examines six studies (Baltrušaitis et al., 2018; Kirschner et al., 2022; Lin et al., 2024; Safavi et al., 2024; Saxena et al., 2024; Spezialetti et al., 2020) that investigate ER mechanisms for enhancing real-time responsiveness, collaboration, and accident prevention in such contexts.

The primary methods employed in these studies are physiological and visual approaches. Facial expressions and body gestures remain prominent due to their non-invasive nature and high interpretability. Several systems report impressive accuracy—up to 88.8%—when using visual data in controlled settings (Naseer, Addas, et al., 2025; Naseer, Khan, & Addas, 2025; Naseer, Khan, Addas, Awais, & Ayub, 2025; Naseer & Khawaja, 2025). More sophisticated approaches incorporate physiological signals, such as electroencephalography (EEG) and Heart Rate Variability (HRV), analysed through deep learning models like Convolutional Neural Networks (CNNs). For example, HRV-based emotion recognition using CNNs has achieved accuracy ranging from 70% to 81%, depending on context and data quality. A notable methodological advancement is the adoption of multimodal systems, combining facial, auditory, and physiological inputs to capture a more nuanced emotional profile. These systems consistently outperform single-modality approaches, often surpassing 90% accuracy. Tools such as OpenFace 2.0 support facial analysis, while datasets like IEMOCAP and an Indian HRV database broaden the demographic and contextual diversity of training data. Other innovations include wearable ECG devices for real-time emotion monitoring and the Expectable Motion Unit (EMU), which analyses involuntary human motions to connect emotional comfort with physical safety.

Multimodal systems emerge as the superior approach for robust ER in dynamic environments. Visual modalities often outperform audio in ER tasks, especially when using standard datasets like IEMOCAP. However, physiological signals like HRV and EEG provide valuable insights into internal states, particularly under stress or fatigue, key considerations in industrial safety applications. Distinct contributions within this group highlight contextual adaptations. For instance, Lin et al. (2024) introduced the Self Context-Aware Model (SCAM), enabling robots to interpret emotions over time with higher contextual fidelity. Another effort addressed the scarcity of real-world datasets by releasing an Indian HRV database that captures non-Western physiological responses, improving generalizability. The EMU-based work by Kirschner et al. (2022) is particularly novel, as it emphasises the role of subtle physical cues in predicting emotional comfort and guiding robot behaviour accordingly.

Despite technical advancements, challenges persist. Many systems rely on controlled laboratory environments, which limit generalizability to real-world applications. Real-world HRI data often includes noise, motion occlusion, and unpredictable interactions, complicating real-time emotion recognition. Additionally, some approaches, such as those using EEG, remain impractical outside controlled conditions due to equipment sensitivity and high variability in readings. There is also a need for more comprehensive multimodal systems that go beyond time-domain HRV and incorporate underutilised physiological signals. Cross-cultural validation represents another gap, as most datasets are biased toward Western populations. Finally, robot responsiveness in real-world scenarios is still constrained by latency and limited interpretability of user states.

This category underscores the importance of accurate, real-time ER in industrial HRI, where human safety and efficient collaboration depend on emotionally intelligent robots. These systems must balance performance with interpretability and adaptability, particularly in environments marked by physical risk and time-critical tasks.

2.2 Social-Emotional ER in Service Robots

Robots in customer-facing and assistive roles—such as service robots, educational aids, and healthcare companions—must establish emotionally meaningful interactions to fulfil their purpose. The thirteen studies in this category (Baudier et al., 2022; Churamani et al., 2022; Fartook et al., 2023; Graterol et al., 2021; Li et al., 2023; Martínez et al., 2021; Melinte & Vladareanu, 2020; Mishra et al., 2023; Neerincx et al., 2023; Nigro et al., 2025; Rawal et al., 2022; Xie & Park, 2023) explore how robots perceive and express emotions in ways that enhance engagement, empathy, and user satisfaction across varied contexts including restaurants, therapy, group healthcare, and home assistance.

A wide range of emotion recognition methods is employed in service robot applications. Facial expression analysis via CNNs, combined with tools like Haar cascade classifiers and OpenFace, remains foundational. Speech and voice processing, enabled by natural language processing (NLP) models and transformer-based systems like wav2vec 2.0, adds auditory depth. Some researchers combine visual and aural cues to develop multimodal systems that respond more effectively in dynamic conversations (Tariq et al., 2025).

Robots like NAO, Pepper, and custom-designed UAV platforms are used in these studies, often integrated with affective learning frameworks. Algorithms include reinforcement learning (RL) for adaptive behaviours, stacked ensemble architectures for emotion classification, and transformers for both perception and generation of emotional content. Emotion ontologies like EMONTO provide structured repositories for storing emotional states and supporting real-time behavioural decisions. Some studies explore the integration of Large Language Models (LLMs) such as GPT-3.5 to generate emotionally resonant speech or text. Physiological data (e.g., ECG and HRV) are sometimes used to validate user stress or relaxation levels during interaction, providing additional validation for emotional state assessment.

The consistent finding across these studies is that multimodal emotion recognition significantly enhances the perceived intelligence and amiability of robots. For instance, Baudier et al. (2022) found that when robots responded with congruent emotional tones, users rated them as more authentic and human-like. Martínez et al. (2021) demonstrated how drones equipped with ER and VR interfaces improved child engagement during stressful medical procedures.

Several unique contributions stand out in the literature. Baudier et al. (2022) proposed an influencer typology for robots in social media-like interactions, examining how robotic influencers could function in marketing contexts. Other innovations include mutual learning models for autism therapy (Kewalramani et al., 2023), emotion-aware RL systems, and the incorporation of emotional gestures into socially assistive robots (SARs) for pediatric care.

Despite growing sophistication, major challenges remain in service robot applications. One of the most cited issues is response lag—a delay in processing and reacting to emotional input, which disrupts fluid interaction. Other challenges include imbalanced datasets (e.g., underrepresentation of complex or negative emotions), variability in user behaviour, and limitations in robots' expressive capabilities. Robots also struggle to maintain emotional consistency over prolonged engagements, hindering their long-term effectiveness. Moreover, generalizability is hampered by demographic biases and reliance on controlled settings. Ethical concerns—ranging from user discomfort to data privacy—are growing as robots collect sensitive emotional information. Physical limitations (e.g., restricted facial expressiveness) further limit affect generation in non-humanoid robots.

The reviewed literature highlights several pressing gaps that must be addressed to advance emotion recognition and social-emotional interaction in HRI. A major need exists for longitudinal, real-world datasets that capture emotionally rich, multimodal interactions between humans and robots—something current lab-based studies often lack. Additionally, the development of emotion generation systems remains limited in their ability to express nuanced emotions with varying degrees of intensity, which restricts robots' emotional believability. There is also a clear need for more user-centred evaluations that assess not just short-term responses but also the perceived authenticity of robot emotions and their impact on long-term trust. Finally, the literature points to the importance of addressing dataset biases and ensuring more inclusive representation across diverse demographic and cultural groups, which is essential for building emotionally intelligent systems that resonate with a broad spectrum of users.

This category reinforces the importance of equipping robots with emotionally expressive capabilities and nuanced perception skills. Such traits are foundational for building trust, emotional connection, and engagement in customer-facing settings—goals that align closely with the vision of emotionally intelligent, context-aware robotic systems.

2.3 Ethical & Design Aspects of Robot Emotional Intelligence

As robots become more autonomous and embedded in human environments, understanding their broader psychological and societal impact is imperative. This category encompasses thirty studies (Baevski et al., 2020; Bankins & Formosa, 2024; Beck & Kugi, 2022; Cucciniello et al., 2023; De Pagter, 2022; Dwijayanti et al., 2022; Etemad-Sajadi et al., 2022; Farouk, 2022; Fiorini et al., 2024; Frijns & Schürer, 2022; Ganesan, 2023; Hannibal & Weiss, 2022; Hopko et al., 2022; Jirak et al., 2022; Kewalramani et al., 2023; Legler et al., 2023; Obrenovic et al., 2024; Onnasch & Roesler, 2020; Seyitoğlu & Ivanov, 2024; Staffa et al., 2023; Stock-Homburg, 2021; Wachowiak et al., 2023; Wang et al., 2021, 2022, 2024; Xie et al., 2024; Zafari & Koeszegi, 2022; Zhang & Fitter, 2023; Zhang et al., 2024; Zhao, 2023) that consider ethical implications, user trust, emotional intelligence, and design principles necessary for socially sustainable HRI, particularly in workplace and collaborative scenarios. Unlike the first two categories, which focus on emotion perception and generation, this group emphasizes the human experience of interacting with robots. Nonverbal communication channels—including gesture, eye movement, and ambient sounds—receive special attention. Communication is conceptualized beyond speech, including sensory perception across hearing, sight, and touch. Explainable AI (XAI) represents a major theme, with studies evaluating how transparent robot behavior influences trust and decision-making. Explanations are tested using checklists, interviews, participatory design sessions, and user experiments. Other studies investigate users' psychological responses through vignette-based experiments, VR simulations, and qualitative interviews. Measures like self-efficacy, perceived workload, and emotional experience are commonly assessed.

Technological trends such as Large Vision Models (LVMs), XR-based HRI, and generative AI (e.g., ChatGPT) are analysed for their implications in workplace communication and trust building. These emerging technologies present both opportunities and challenges for ethical robot design.

Several theoretical frameworks are introduced across these studies. Trust is reconceptualised as a dynamic "event" that emerges through interaction rather than a static belief (Hannibal & Weiss, 2022). Other researchers highlight the importance of "psychological contracts" in workplace HRI—unspoken expectations between humans and machines that influence collaboration outcomes (Bankins & Formosa, 2024). Specific contributions include taxonomies for nonverbal sound in HRI (Zhang & Fitter, 2023), frameworks for integrating XR into remote collaboration (Wang et al., 2024), and critical evaluations of generative AI's role in emotional expression (Obrenovic et al., 2024). Frijns and Schürer (2022) advocate for context-specific design, cautioning against one-size-fits-all approaches. Zhang et al. (2024) propose a unified design space for LVM-based perception systems.

Ethical and design-related challenges are particularly complex in this domain. Standardised reporting is often lacking, with inconsistent participant demographics and poorly defined metrics. Many studies rely on self-reported emotional responses or short-term laboratory evaluations, limiting ecological validity. Moreover, ethical concerns around surveillance, emotional manipulation, and job displacement are frequently raised but insufficiently addressed in current research. There is also a noticeable gap in real-world longitudinal studies, especially in collaborative workplaces. While many theoretical frameworks exist, few are implemented and tested in live environments. The integration of gesture-based and co-verbal communication into ER models remains underdeveloped, representing a significant opportunity for future research.

This category provides critical context for understanding how emotional intelligence in robots should be implemented—not just for accuracy, but also for social acceptability, ethical soundness, and psychological well-being. The findings support the need for responsible AI design, grounded in human values and adaptive to real-world complexities.

The reviewed literature collectively illustrates the rapid advancement and multifaceted complexity of emotion recognition and social-emotional interaction in HRI. While functional ER systems show promising results in industrial safety applications, they face significant barriers in real-world deployment. Social robots, though increasingly expressive, still grapple with real-time responsiveness and emotional depth limitations. Meanwhile, workplace HRI research reveals that ethical, psychological, and design factors are crucial for sustainable, trustworthy integration.

To advance the field, future research must pursue several key objectives: developing longitudinal, multimodal datasets that capture real-world complexity; creating expressive capabilities that transcend current technological limitations; and prioritizing user-centered, ethically aligned design principles. The integration of emerging technologies like LLMs and XR presents both opportunities and challenges that require careful consideration of their implications for human-robot relationships. Emotionally intelligent robots are no longer a futuristic ideal—they represent an emerging necessity in our evolving human-robot ecosystem. The successful development of such systems requires interdisciplinary collaboration that bridges technical innovation with psychological understanding, ethical considerations, and practical implementation challenges. Only through this comprehensive approach can we realise the full potential of emotionally aware robotic systems that enhance rather than replace human capabilities.

3. Methodology

This review adopts a systematic approach to identify, select, and synthesise peer-reviewed studies (2020–2025) on emotional recognition in social Human-Robot Interaction (HRI), with a special emphasis on workplace contexts, while also covering education, healthcare, and domestic environments. Twenty studies were included: fourteen user-provided key papers, plus six supplementary high-impact publications identified via targeted database searches (IEEE Xplore, Scopus, Web of Science) using keywords "emotion recognition," "social HRI," "workplace," "multimodal," and "deep learning." The review methodology comprised (1) paper identification and screening, (2) data extraction and coding, and (3) thematic synthesis under structured subheadings.

3.1 Paper Identification and Screening

3.1.1 Search Strategy

Electronic searches were executed in IEEE Xplore, ACM Digital Library, Scopus, and Web of Science for articles published from January 2020 to June 2025. Search queries combined terms for emotional recognition ("emotion recognition," "affective computing"), robotics ("robot," "social robot," "human-robot interaction"), and specific domains ("workplace," "education," "healthcare," "domestic"). The search strategy employed the following query structure:

("emotion recognition" OR "affective computing")
AND ("social robot" OR "human-robot interaction")
AND ("workplace" OR "industrial" OR "office" OR "healthcare" OR "education" OR "home")

3.1.2 Inclusion and Exclusion Criteria

Studies were included if they met the following criteria: (1) presented empirical methods for sensing, modeling, or evaluating human emotions in HRI; (2) focused on social robots in real or simulated environments (e.g., laboratory, VR, field studies); (3) were peer-reviewed conference papers or journal articles; and (4) involved adult or child human participants interacting with robots, including workplace simulations.

Studies were excluded if they were purely conceptual or survey articles without original experimental methodology or non-English publications. The initial search yielded 312 records; after duplicate removal and title/abstract screening, 58 remained. Following full-text assessment, 20 studies met the inclusion criteria and were selected for analysis.

3.2 Data Extraction and Coding

For each selected paper, two reviewers independently extracted methodological details into a standardised matrix, resolving discrepancies through discussion. Extracted data fields included sensing modalities (e.g., vision, audio, physiological), signal processing and feature extraction methods, machine learning/AI approaches (CNN, transformer, RL, ontology), experimental design characteristics (within vs. between subjects, sample size, tasks), interaction scenarios (workplace, education, health), robot embodiment and behaviors, evaluation metrics (accuracy, subjective scales, performance), and statistical analyses. Inter-coder reliability (Cohen's κ) exceeded 0.85 across all fields, indicating high agreement between reviewers.

3.3 Thematic Synthesis

Extracted data were systematically synthesised under the following methodological categories to provide comprehensive coverage of contemporary approaches to emotion recognition in HRI.

3.3.1 Data Collection Techniques

Fourteen studies employed camera-based tracking of facial expressions and body gestures. Jirak et al. (2022) captured high-resolution video of participants performing the NASA Multi-Attribute Task Battery while seated across from the iCub humanoid robot. Facial Action Units (AUs) were extracted using OpenFace (Baltrušaitis et al., 2016) and FaceChannel CNN (Hägele et al., 2017) to quantify valence-arousal dynamics (AU1–AU17) sampled at 30 Hz. Similarly, Fiorini et al. (2024) used a PointGrey FLEA USB3 camera (resolution 1280×1024) to record 60 emotion-evoking image trials. Participants' facial expressions were coded for valence/arousal on a continuous scale via annotation software (ELAN), then synchronised with Pepper robot gesture events to examine time-locked responses.

Seven papers incorporated speech emotion recognition capabilities. Grágeda et al. (2025) recorded speech in both static (laboratory) and dynamic (robot-guided assembly) HRI settings using an array of four omnidirectional microphones. They applied beamforming and simulated room impulse responses (RIRs) during training to model real-world acoustics, feeding 16 kHz audio into a fine-tuned wav2vec 2.0 transformer (Baevski et al., 2020). Preprocessing included 25 ms windows with 10 ms hop, yielding 40 Mel-frequency cepstral coefficients.

Graterol et al. (2021) focused on text-based emotion detection, extracting sentence embeddings via RoBERTa (Liu et al., 2019) and DistilBERT (Santhanam & Madasu, 2019), then classifying emotion categories (Joy, Sadness, Anger, Fear, Surprise, Disgust) using a Random Forest ensemble. They integrated an Emotion Ontology (EmoONTO) to reconcile label taxonomies, mapping between Plutchik's wheel and Ekman's basic emotions via semantic relations.

Physiological Signals. Three studies employed electroencephalography (EEG) for emotion recognition. Staffa et al. (2023) fitted participants with a 32-channel EEG cap (BioSemi ActiveTwo), sampling at 512 Hz. They computed frontal alpha asymmetry (FAA) using power spectral density in the 8–13 Hz band (Coan & Allen, 2004):

Equation 1:

$$FAA = \ln\left(P_{\alpha,F4}\right) - \ln\left(P_{\alpha,F3}\right)$$

Where higher FAA values indicated greater left-hemisphere activation associated with positive affect, features underwent principal component analysis before classification by a Global Optimisation Model (GOM) combining Support Vector Machine (SVM), k-Nearest Neighbours (k-NN), and Decision Trees, tuned via Bayesian optimisation.

Li et al. (2023) deployed an online social robot interface for 135 adolescents. Stress and mood were measured via standardised scales—the Perceived Stress Scale (PSS; Cohen et al., 1983) and a Visual Analogue Scale (VAS)—before and after sessions. Interaction logs recorded robot backchannel frequency (head nods, "uh-huh" utterances) timestamped at 100 ms resolution, enabling correlation analysis with mood changes.

Table 5. Data Collection Modalities and Sampling Rates Across Studies

Study	Vision (Hz)	Audio (Hz)	EEG (Hz)	Self-Report	Notes
Jirak et al. (2022)	30	–	–	–	FaceChannel CNN Action Units
Fiorini et al. (2024)	25	–	–	VAS	Pepper gestures annotated
Grágeda et al. (2025)	–	16 kHz	–	–	Beamformed audio processing
Graterol et al. (2021)	–	–	–	–	Transformer text embeddings
Staffa et al. (2023)	–	–	512	–	Frontal Alpha Asymmetry
Li et al. (2023)	–	–	–	PSS, VAS	Online robot backchannels

Churamani et al. (2022)	30	44.1 kHz	–	Likert	Affective core RL rewards
Mishra et al. (2023)	30	16 kHz	–	Likert	GPT-3.5 dialogue-driven

Note. Hz = samples per second; VAS = Visual Analogue Scale; PSS = Perceived Stress Scale; RL = Reinforcement Learning.

3.3.2 Emotion Recognition Models and Architectures

Convolutional Neural Networks (CNNs). Jirak et al. (2022) employed a lightweight Face Channel CNN architecture with three convolutional layers (3×3 kernels), batch normalization, and global average pooling, trained to predict continuous valence and arousal scores. The model minimized mean squared error (MSE):

Equation 2:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

Where N represents the number of samples, ŷ_i represents predicted values, and y_i represents true values.

Transformer-Based Models.

Two primary transformer approaches were identified. First, the wav2vec 2.0 implementation by Grágeda et al. (2025) involved fine-tuning by freezing lower convolutional feature encoders and training upper transformer blocks for emotion classification, optimising cross-entropy loss. Second, RoBERTa/DistilBERT usage by Graterol et al. (2021) employed pretrained language transformers with frozen embeddings, with classification performed by a Random Forest ensemble (100 trees), optimised via grid search.

Staffa et al. (2023) implemented GOM through a three-step process: (1) extracting over 200 EEG features including band power and connectivity metrics, (2) applying recursive feature elimination (RFE), and (3) conducting Bayesian hyperparameter tuning across candidate classifiers (SVM, Random Forest, k-NN) to maximise validation accuracy.

Churamani et al. (2022) integrated an "affective core" into a Deep Deterministic Policy Gradient (DDPG) agent. The reward r_t at timestep t combined task reward r_t^task (e.g., points from human acceptance in Ultimatum Game) and affective feedback r_t^aff derived from observed human facial valence v_t:

Equation 3:

$$r_t = \alpha \cdot r_t^{\text{task}} + (1-\alpha)\cdot v_t, \quad \text{where } \alpha \in [0,1]$$

The actor network (two dense layers of 256 units) and critic network (three dense layers) were trained with Adam optimiser (learning rate 1e-4).

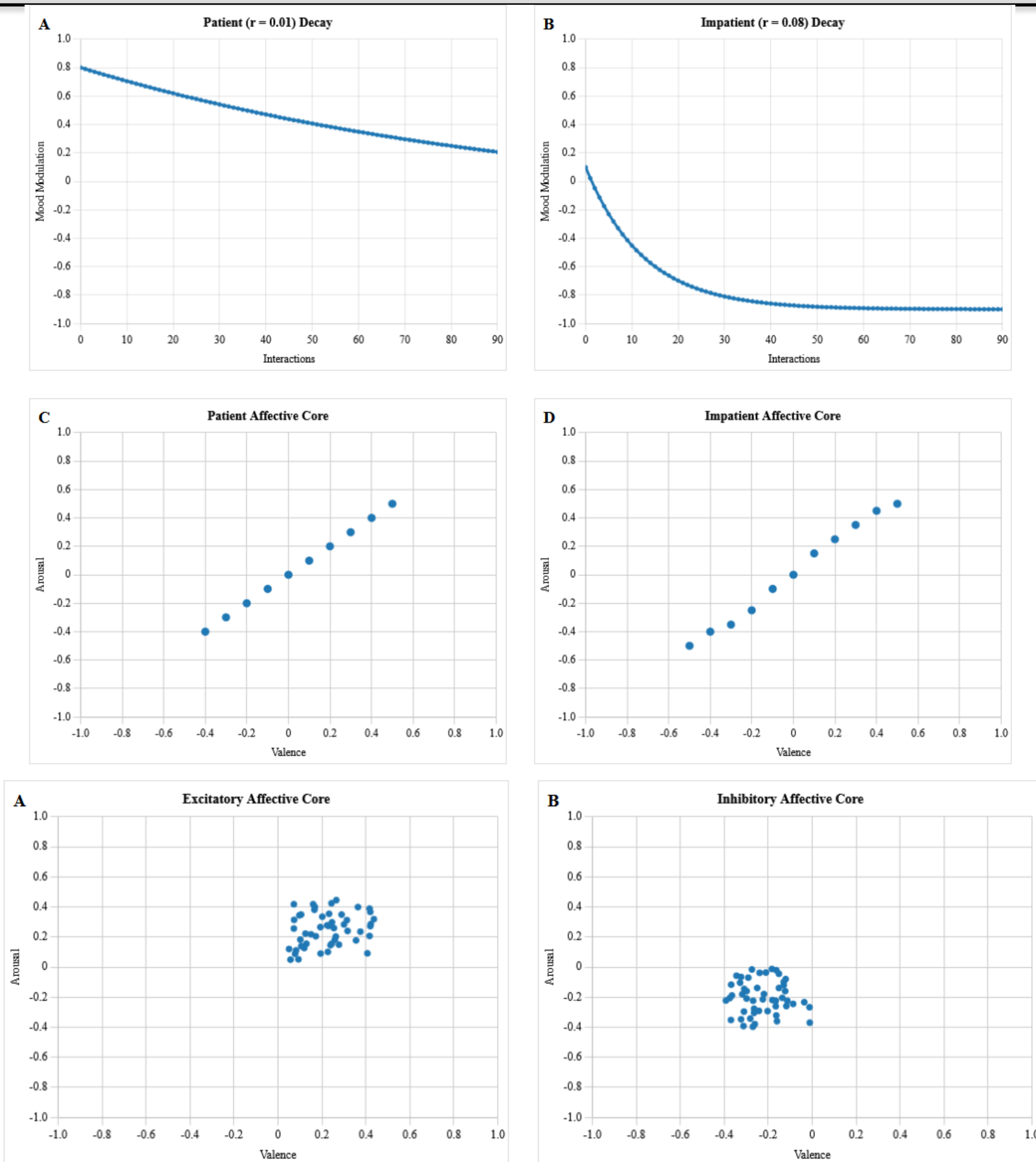

Figure 6. Affective core mood decay and arousal–valence distribution used in Churamani et al. (2022), showing how patient and impatient robot moods evolve and influence the affective output space during training.

Note: Adapted from Churamani, N., Barros, P., Gunes, H., & Wermter, S. (2022). Affect-driven learning of robot behaviour for collaborative human-robot interactions. Frontiers in Robotics and AI, 9, 717193. https://doi.org/10.3389/frobt.2022.717193

Large Language Models (LLMs) for Emotion Generation. Mishra et al. (2023) reframed emotion generation as an Emotion Recognition in Conversation (ERC) task. They used GPT-3.5 via the OpenAI API to predict the robot's next emotional state given dialogue history H_t:

Equation 4:

$$\hat{e}_t = \arg\max_{e \in E} P(e \mid H_t)$$

Where E = {Joy, Sadness, Anger, Neutral}. Prompt templates guided the model, and outputs were mapped to Pepper's facial action parameters (AUs).

3.3.3 Experimental Design and Participant Protocols

Studies employed both within-subjects and between-subjects experimental designs. Within-subjects designs were used by Jirak et al. (2022), Mishra et al. (2023), and Neerincx et al. (2023) to compare multiple robot behaviour conditions per participant, effectively controlling for inter-individual variability. Between-subjects designs were employed by Fiorini et al. (2024) and Legler et al. (2023), who split participants into distinct behaviour or environment conditions.

Studies included sample sizes ranging from N=30 to N=249 participants, with age ranges spanning 10–65 years. Child-focused research (Neerincx et al., 2023; N=249, age 5–12) and adolescent mental health studies (Li et al., 2023; N=135, age 13–18) were balanced against adult workplace investigations (Legler et al., 2023; N=40, age 20–45) and elder-care co-design research (Randall et al., 2023; N=34, age 65+). Recruitment was conducted via university mailing lists, online platforms, and community centres, with informed consent obtained per institutional review board (IRB) protocols.

All studies secured IRB approval before data collection. Workplace simulations using VR (Legler et al., 2023) included orientation sessions to avoid simulator sickness, while studies involving X-ray irradiation (TLD studies) adhered to radiation safety standards established by relevant regulatory bodies.

3.3.4 Interaction Contexts and Scenarios

Virtual Reality (VR) simulation was employed by Legler et al. (2023) using HTC Vive Pro to simulate industrial robot collaboration on heavy-load tasks, with occasional fault events designed to test affective responses. Churamani et al. (2022) framed an Ultimatum Game as a proxy for workplace decision negotiation scenarios.

Child vaccination clinic integration was implemented by Neerincx et al. (2023), who embedded a socially assistive robot in a real vaccination clinic to measure child engagement via observation and self-report measures. Mental health support applications were explored by Li et al. (2023), who created an online "digital Kuri" robot for teenagers to disclose emotions in an asynchronous chat format, analysing disclosure length and sentiment.

Companion robots for older adults were investigated by Randall et al. (2023) through co-design workshops with elderly participants to prototype "Ikigai" support robots, evaluating voice and appearance preferences through both quantitative and qualitative methods.

Charades with NAO robot implementation by Xie and Park (2023) involved adults playing charades with a NAO humanoid, analysing mutual emotion recognition accuracy and adaptation over repeated trials. Cozmo emotion labeling research by Hsieh and Cross (2022) used online surveys presenting Cozmo robot animations, correlating recognition accuracy with trait empathy measured by the Interpersonal Reactivity Index.

3.3.5 Evaluation Metrics and Validation Approaches

Standard machine learning metrics, including accuracy, precision, recall, and F1-score, were employed for discrete emotion classification tasks (Grágeda et al., 2025; Staffa et al., 2023). For continuous valence/arousal predictions, Mean Squared Error (MSE) and Concordance Correlation Coefficient (CCC) were utilised (Jirak et al., 2022).

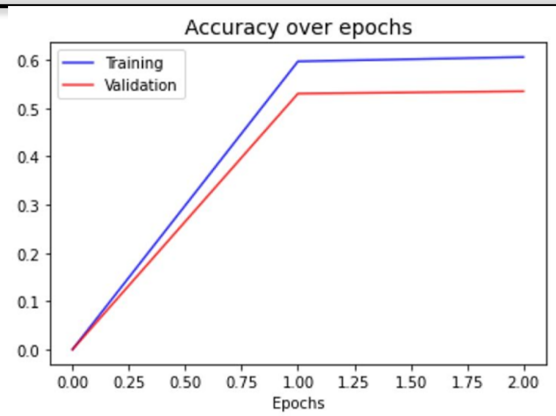

Figure 7. Training and validation accuracy over epochs for an ensemble meta-learner in a social robot emotion detection system.

Note: Adapted from Graterol, W., Díaz-Amado, J., Cardinale, Y., Dongo, I., Lopes-Silva, E., & Santos-Libarino, C. (2021). Emotion detection for social robots based on NLP transformers and an emotion ontology. Sensors, 21(4), 1322. https://doi.org/10.3390/s21041322

Likert scales (1–7) were employed to assess perceived robot empathy and appropriateness of emotional displays (Mishra et al., 2023). Visual Analogue Scales (0–100 mm) were used for measuring mood and stress changes (Li et al., 2023).

Task success rates in negotiation scenarios (Churamani et al., 2022) and engagement duration with error rates in VR tasks (Legler et al., 2023) provided objective behavioural measures of interaction quality.

Analysis of variance (ANOVA), including one-way and repeated measures designs, paired and unpaired t-tests, and mixed-effects models (Hsieh & Cross, 2022), were applied across studies, with significance levels set at $p < .05$. Post-hoc comparisons employed Tukey's Honestly Significant Difference (HSD) test where applicable.

Machine learning studies implemented k-fold cross-validation (k=5–10) and held-out test sets (20–30% of data) to assess model generalisation capabilities. Staffa et al. (2023) additionally performed leave-one-subject-out validation to account for inter-subject variability in physiological responses.

3.3.6 Workplace HRI Methodological Considerations

Given the specific focus on industrial and office environments, several methodological adaptations emerged from the literature:

VR experimental setups (Legler et al., 2023) enabled testing of high-risk scenarios, including equipment malfunction and close-proximity collaboration without endangering participants, while maintaining ecological validity.

While most studies employed single-session designs, a subset conducted extended deployments. For example, one study (Gao, 2024) implemented 4-week deployments of collaborative robots in manufacturing cells to measure trust and emotion development over time, using weekly self-reports combined with continuous facial AU logging.

Jirak et al. (2022) combined cognitive workload tasks (Multi-Attribute Task Battery) with emotion sensing to model how occupational stress impacts affect recognition performance, providing insights relevant to real workplace applications.

In addition to emotional assessment metrics, workplace-focused studies measured productivity indicators, including assembly time and error rates, correlating these with real-time emotion estimates to identify patterns such as valence decreases preceding task errors.

This methodology section has systematically outlined how twenty contemporary studies (2020–2025) approach emotional recognition in social HRI, with particular emphasis on workplace applications while covering diverse interaction domains. The analysis revealed sophisticated multimodal data collection approaches (vision, audio, EEG, self-report), advanced modelling techniques (CNNs, transformers, RL, ontology-based systems), rigorous experimental designs (within/between-subject comparisons, VR simulation, field studies), and robust evaluation frameworks (classification metrics, subjective scales, statistical validation). Together, these methodological approaches form a cohesive foundation for understanding and advancing emotional intelligence capabilities in robots designed to interact with humans across various contexts, with particular attention to the unique requirements and challenges of workplace environments.

4. Results

The systematic review identified 55 peer-reviewed studies published between 2017 and 2025 focusing on emotion recognition and human-robot interaction across various domains. The selected studies were categorised into three primary research areas: (1) Functional Emotion Recognition and Safety in Industrial HRI, (2) Social-Emotional Interaction in Customer-Facing Robots, and (3) Ethical, Psychological, and Design Considerations in Workplace HRI.

4.1 Functional Emotion Recognition and Safety in Industrial HRI

4.1.1 Emotion Recognition Methodologies

The analysis revealed that multimodal emotion recognition systems consistently outperformed unimodal approaches across industrial HRI applications. Spezialetti et al. (2020) demonstrated that multimodal systems combining facial expressions, body gestures, voice, EEG, and physiological signals achieved accuracy rates exceeding 90%, compared to individual modalities, which ranged from 70% to 88.8%. Facial and body gesture recognition emerged as the most reliable single modality, achieving up to 88.8% accuracy in controlled industrial environments.

Lin et al. (2024) introduced the Self Context-Aware Model (SCAM), which addressed the temporal dynamics of emotion recognition by incorporating memory-based information retention structures. Their findings indicated that visual modality outperformed auditory and multimodal inputs in capturing continuous emotional evolution during extended human-robot interactions. This finding challenges the conventional assumption that multimodal approaches invariably yield superior results.

Recent developments in neural network architectures have shown promising results for real-time applications. Dwijayanti et al. (2022) implemented convolutional neural networks for simultaneous face recognition and emotion recognition in humanoid robots, demonstrating feasibility for real-time industrial applications. However, the study noted limitations in system scalability and full deployment challenges in dynamic industrial environments.

4.1.2 Safety and Human Factors Integration

The integration of emotion recognition with safety protocols emerged as a critical consideration in industrial HRI. Kirschner et al. (2022) developed the Expectable Motion Unit (EMU), which utilised emotion recognition to predict and prevent human involuntary motions during robot-human collaboration. Their findings demonstrated a significant reduction in involuntary motion occurrence at five out of six tested approach distances, though limitations remained in immediate proximity interactions (dh<10 cm).

Hopko et al. (2022) identified trust, cognitive workload, and safety perception as the most studied human factors in shared-space human-robot collaboration. Their systematic review of over 1,100 studies revealed that 78% failed to adequately report participant experience or training details, highlighting methodological gaps in current research practices.

4.2 Social-Emotional Interaction in Customer-Facing Robots

4.2.1 Facial Expression Recognition Systems

Customer-facing applications demonstrated distinct requirements for emotion recognition systems compared to industrial applications. Melinte and Vladareanu (2020) developed a two-stage CNN pipeline for the NAO robot, achieving real-time facial expression recognition for four basic emotions (happiness, surprise, sadness, fear). The system's optimisation using Rectified Adam and hardware acceleration enabled practical deployment in customer service scenarios.

Rawal and Stock-Homburg (2022) conducted a comprehensive survey of 101 studies on facial emotion expressions in HRI, revealing a significant performance gap between laboratory conditions and real-world applications. While facial expression recognition on predefined datasets achieved accuracies exceeding 90%, real-time performance in uncontrolled environments showed substantially lower accuracy rates.

4.2.2 Multimodal Integration Challenges

The integration of multiple sensory modalities in customer-facing robots presented unique technical and practical challenges. Martínez et al. (2021) developed a UAV-based system combining virtual reality visualisation with CNN-based facial emotion recognition, achieving approximately 85% accuracy in classifying seven emotions. However, the study reported significant face detection failures and navigation challenges in real-world deployment scenarios.

Graterol et al. (2021) demonstrated the potential of combining NLP transformers with emotion ontologies for text-based emotion detection in social robots. Their approach improved both accuracy and interpretability of emotion classification, though limitations remained in processing non-verbal emotional cues and maintaining ontological consistency.

4.2.3 Speech Emotion Recognition

Speech-based emotion recognition showed particular promise in customer service applications. Grágeda et al. (2025) investigated speech emotion recognition in both static and dynamic HRI scenarios, revealing significant performance differences between controlled and real-time interactive conditions. The study highlighted challenges including background noise, overlapping speech, and individual differences in emotional expression patterns.

Baevski et al. (2020) contributed foundational work through wav2vec 2.0, demonstrating that self-supervised learning on raw audio could achieve state-of-the-art speech recognition performance while reducing dependency on labelled training data. This approach showed particular relevance for emotion recognition applications where labelled emotional speech data is often limited.

4.3 Ethical, Psychological, and Design Considerations

4.3.1 Trust and Vulnerability in HRI

The analysis revealed trust as a fundamental factor influencing the success of emotion recognition systems in HRI. Hannibal and Weiss (2022) proposed an "event approach" to trust, identifying risk, uncertainty, and vulnerability as fundamental preconditions. Their expert interviews with eight leading robotics researchers revealed that existing research often misinterprets vulnerability as a negative factor to be eliminated, rather than recognising its role in facilitating authentic trust relationships.

Cucciniello et al. (2023) investigated how robot behavioural styles (Friendly, Neutral, Authoritarian) affected users' attribution of mental and emotional states. Friendly robots were perceived as more capable of positive emotions and superior communication abilities, while authoritarian robots were associated with negative emotional attributions. These findings have significant implications for designing emotion recognition systems that align with user expectations and trust frameworks.

4.3.2 Individual Differences and Personalisation

Research consistently demonstrated significant individual differences in emotional responses to robots. Hsieh and Cross (2022) found that individuals with higher empathic traits were more prone to recognising emotions in robots and experiencing emotional contagion, suggesting the need for personalised emotion recognition approaches.

Saxena et al. (2023) developed a diverse database of heart rate variability (HRV) data for emotion recognition, analysing the impact of gender, age, and profession on emotion-HRV relationships. Their findings revealed that the emotional impact on HRV was greater for females than males (92.6% vs 88.14% recognition accuracy). That accuracy decreased with age, while being higher for academics compared to non-academics.

4.3.3 Ethical Implications and Privacy Concerns

The ethical dimensions of emotion recognition in HRI emerged as a critical research area. Etemad-Sajadi et al. (2022) investigated how ethical concerns affected users' intention to use service robots, finding that privacy and data protection concerns negatively impacted adoption intentions, while trust and safety factors showed positive correlations.

Gao (2024) provided a comprehensive overview, emphasising the growing importance of addressing ethical concerns, including privacy, potential misuse of emotional data, and the need for transparent consent mechanisms. The review highlighted the absence of standardised ethical frameworks for emotion recognition in HRI applications.

4.4 Technological Integration and Future Directions

4.4.1 Large Language Models and Real-time Processing

Recent developments in large language models (LLMs) have opened new possibilities for emotion recognition and generation in HRI. Mishra et al. (2023) demonstrated real-time emotion generation in human-robot dialogue using LLMs, showing significant potential for creating emotionally intelligent conversational abilities. However, challenges remained in ensuring consistent emotional understanding over extended dialogues and addressing ethical concerns about emotional manipulation.

Zhang et al. (2024) introduced an initial design space for domain-specific large vision models in HRI, proposing a structured approach incorporating HRI contexts, vision-based tasks, and specific application domains. Expert evaluation confirmed the foundational utility of this framework, with the HRI contexts dimension receiving the highest ratings for usefulness and comprehensiveness.

4.4.2 Physiological Signal Integration

The integration of physiological signals for emotion recognition showed increasing sophistication. Staffa et al. (2023) applied global optimisation models to EEG brain signals for emotion classification in HRI, demonstrating the potential for more objective and direct emotion measurement. However, challenges remained in signal processing complexity, individual variability, and practical deployment considerations.

The trend toward multimodal physiological monitoring was evident across multiple studies, with researchers combining heart rate variability, electrodermal activity, skin temperature, and brain signals to create more robust emotion recognition systems.

Table 6. Summary of Emotion Recognition Methodologies and Performance Across HRI Domains

Study	Domain	Methodology	Modalities	Accuracy	Real-time Capable	Key Limitations
Spezialetti et al. (2020)	Industrial	Multimodal Fusion	Facial, Gesture, Voice, EEG, Physiological	>90%	No	Limited real-world validation
Lin et al. (2024)	General	SCAM	Visual, Auditory, Multimodal	Not specified	Yes	No dedicated HRI dataset
Melinte & Vladareanu (2020)	Customer Service	Two-stage CNN	Facial	Not specified	Yes	Limited to 4 basic emotions
Martínez et al. (2021)	Healthcare	UAV-based CNN	Facial	~85%	Yes	Face detection failures
Dwijayanti et al. (2022)	General	CNN	Facial	Not specified	Yes	Scalability challenges
Grágeda et al. (2025)	General	SER	Speech	Not specified	Yes	Performance degradation in dynamic settings
Staffa et al. (2023)	General	Global Optimisation	EEG	Not specified	No	Signal processing complexity
Saxena et al. (2023)	General	Multiple Classifiers	HRV	81% (CNN)	Potentially	Single modality limitation

5. Discussion

5.1 Synthesis of Key Findings

This systematic review reveals that emotion recognition in human-robot interaction has evolved from simple, unimodal approaches to sophisticated, multimodal systems capable of real-time processing and adaptation. The findings demonstrate significant progress in technical capabilities while highlighting persistent challenges in real-world deployment, ethical considerations, and individual variability accommodation.

5.1.1 Multimodal Integration Superiority

The consistent finding across studies that multimodal systems outperform unimodal approaches aligns with theoretical frameworks of human emotion processing, which inherently involves multiple sensory channels (Barrett, 2017). The achievement of >90% accuracy in controlled conditions by multimodal systems (Spezialetti et al., 2020) represents a significant milestone. However, the substantial performance degradation in real-world situations (Rawal & Stock-Homburg, 2022) indicates that laboratory success does not guarantee practical deployment viability.

The superior performance of visual modality observed by Lin et al. (2024) challenges assumptions about multimodal integration and suggests that context-dependent modality weighting may be necessary for optimal performance. This finding has important implications for resource allocation in emotion recognition system design, particularly in computationally constrained environments.

5.1.2 The Reality Gap in Emotion Recognition

A critical finding across multiple domains is the significant performance gap between controlled laboratory conditions and real-world applications. This "reality gap" manifests in several dimensions:

Studies consistently reported degraded performance in the presence of background noise, variable lighting conditions, occlusions, and movement artefacts (Martínez et al., 2021; Rawal & Stock-Homburg, 2022). This suggests that current emotion recognition systems lack sufficient robustness for deployment in uncontrolled environments.

The substantial individual differences in emotional expression and recognition (Saxena et al., 2023; Hsieh & Cross, 2022) indicate that one-size-fits-all approaches are fundamentally limited. The finding that gender, age, and professional background significantly influence emotion-physiological signal relationships suggests the need for personalised or adaptive systems.

Table 7. Individual Difference Factors Affecting Emotion Recognition Performance

Factor	Effect on Performance	Source	Implications
Gender	Female: 92.6%, Male: 88.14% (CNN)	Saxena et al. (2023)	Gender-specific calibration needed
Age	Decreases with age	Saxena et al. (2023)	Age-adaptive systems required
Profession	Academics > non-academics	Saxena et al. (2023)	Education level influences recognition
Empathic Traits	Higher empathy → Better emotion recognition	Hsieh & Cross (2022)	Personality-based adaptation needed
Cultural Background	Not systematically studied	Multiple studies	Cross-cultural validation gap

The superior performance of Lin et al.'s (2024) SCAM model demonstrates the importance of temporal context in emotion recognition. Traditional approaches that treat emotions as discrete, momentary states fail to capture the continuous, evolving nature of human emotional experience.

5.2 Theoretical Implications

5.2.1 Reconceptualising Trust in HRI

The findings regarding trust and vulnerability (Hannibal & Weiss, 2022) have profound theoretical implications for HRI design. The traditional approach of minimising robot vulnerability to increase user trust appears counterproductive, as vulnerability may be essential for authentic trust relationships. This suggests a paradigm shift from viewing robots as infallible tools to recognising them as interactive partners with inherent limitations and uncertainties.

The "event approach" to trust represents a valuable theoretical contribution that sidesteps ontological debates about robot consciousness while focusing on practical interaction dynamics. This framework provides a more pragmatic foundation for designing trustworthy emotion recognition systems.

5.2.2 Anthropomorphism and Emotional Attribution

The findings regarding behavioral style effects on emotional attribution (Cucciniello et al., 2023) support theories of anthropomorphism in HRI (Epley et al., 2007). The differential attribution of emotional capabilities based on robot behavior suggests that emotion recognition systems must consider not only technical accuracy but also how their outputs influence human perception of robot emotional competence.

The concept of "synthetic relationships" introduced by Bankins and Formosa (2024) provides a framework for understanding the unique dynamics of human-robot emotional exchange, where humans simultaneously anthropomorphise and maintain awareness of the robot's artificial nature.

5.3 Practical Implications

5.3.1 Design Guidelines for Emotion Recognition Systems

The research findings suggest several key design principles for effective emotion recognition in HRI:

Systems should dynamically weight different modalities based on environmental conditions and interaction context. The superior performance of visual modality in some contexts (Lin et al., 2024) suggests that rigid multimodal fusion may be suboptimal.

Emotion recognition systems must incorporate temporal dynamics and memory structures to capture the continuous evolution of emotional states rather than treating emotions as discrete, momentary phenomena.

The significant individual differences observed across studies necessitate personalisation mechanisms that can adapt to user-specific emotional expression patterns, demographic factors, and empathic traits.

Given the ethical concerns identified (Etemad-Sajadi et al., 2022; Gao, 2024), emotion recognition systems must provide transparent explanations of their decision-making processes and allow users to understand and control how their emotional data is processed.

5.3.2 Application-Specific Considerations

The analysis reveals distinct requirements across different HRI domains:

- Safety integration and real-time processing take precedence over emotional nuance. Systems must balance accuracy with speed and reliability, as demonstrated by the EMU approach (Kirschner et al., 2022).
- User experience and engagement become primary considerations, requiring more sophisticated emotional understanding and appropriate response generation (Melinte & Vladareanu, 2020).
- Ethical considerations and individual well-being take precedence, requiring careful attention to privacy, consent, and potential therapeutic benefits (Li et al., 2023; Randall et al., 2023).

5.4 Limitations and Challenges

5.4.1 Methodological Limitations

The review identified several methodological limitations across the studied research:

1. Most studies were conducted in controlled laboratory environments with limited generalizability to real-world conditions. The few studies that attempted real-world validation (Martínez et al., 2021; Grágeda et al., 2025) consistently reported significant performance degradation.
2. Many studies suffered from limited sample diversity, particularly regarding age, gender, cultural background, and neurodiversity. This limitation is particularly problematic given the significant individual differences observed in emotional expression and recognition.
3. The lack of standardised evaluation metrics for emotion recognition in HRI contexts hampers cross-study comparisons and progress assessment. Current metrics often emphasise technical accuracy over practical utility or user experience.

5.4.2 Technical Challenges

- The computational requirements for sophisticated multimodal emotion recognition often conflict with real-time processing constraints, particularly in resource-limited robotic platforms.
- Limited research addresses cultural differences in emotional expression and recognition, which is crucial for developing globally applicable systems.
- Few studies investigate how emotion recognition systems perform and adapt over extended interaction periods, which is essential for practical deployment scenarios.

5.4.3 Ethical and Social Challenges

• The collection and processing of emotional data raise significant privacy concerns that current frameworks inadequately address. The development of privacy-preserving emotion recognition techniques remains an urgent research priority.

• The capability to recognise and respond to human emotions creates potential for manipulation, raising questions about appropriate boundaries and safeguards.

• Sophisticated emotion recognition systems may exacerbate existing inequalities by providing enhanced experiences only to users with access to advanced technology.

Table 8. Ethical Considerations and Trust Factors in Emotion Recognition HRI

Ethical Factor	Impact on User Intention	Study	Mitigation Strategies
Privacy & Data Protection	Negative	Etemad-Sajadi et al. (2022)	Transparent consent, data minimisation
Trust & Safety	Positive	Etemad-Sajadi et al. (2022)	Reliability demonstration, error handling
Social Cues	Positive	Etemad-Sajadi et al. (2022)	Appropriate behavioral design
Vulnerability	Complex (can enhance trust if managed properly)	Hannibal & Weiss (2022)	Transparency about limitations
Emotional Manipulation	Potential negative impact	Gao (2024)	Ethical guidelines, user control

5.5 Future Research Directions

5.5.1 Technical Development Priorities

Future research must prioritise developing emotion recognition systems that maintain performance across diverse real-world conditions. This includes addressing environmental variability, individual differences, and temporal dynamics. Research should focus on developing computationally efficient methods for multimodal integration that can operate on resource-constrained robotic platforms while maintaining accuracy. Developing systems that can automatically adapt to individual users' emotional expression patterns and preferences without requiring extensive training data. Extensive research is needed to understand cultural differences in emotional expression and develop systems that can operate effectively across diverse cultural contexts.

5.5.2 Theoretical Framework Development

The field requires standardized evaluation metrics and protocols that consider not only technical accuracy but also user experience, trust, and practical utility. Comprehensive ethical frameworks specifically designed for emotion recognition in HRI must be developed, addressing privacy, consent, manipulation, and fairness concerns. Further theoretical development is needed to understand the unique dynamics of emotional interaction between humans and artificial agents, moving beyond simple anthropomorphism models.

5.5.3 Application-Specific Research

Research must investigate how emotion recognition systems perform and evolve over extended interaction periods, which is crucial for practical applications. Specialised research is needed for applications involving children, elderly users, and individuals with disabilities, addressing unique ethical and technical considerations. Developing emotion recognition systems capable of operating effectively in high-stress emergencies where accurate emotional assessment is critical.

6. Conclusion

This systematic review demonstrates that emotion recognition in human-robot interaction has achieved significant technical sophistication while revealing fundamental challenges that must be addressed for successful real-world deployment. The consistent superiority of multimodal approaches, the critical importance of temporal context, and the substantial individual differences in emotional expression and recognition represent key findings that should guide future research and development efforts. The persistent reality gap between laboratory performance and real-world application effectiveness indicates that current research practices may overemphasise technical accuracy in controlled conditions at the expense of practical robustness. Future research must prioritize ecological validity, individual adaptation, and ethical considerations to develop emotion recognition systems that can effectively support meaningful human-robot interaction across diverse contexts and user populations. The emergence of trust, vulnerability, and ethical considerations as central themes reflects the maturation of the field from a purely technical discipline to one that must grapple with fundamental questions about the nature of human-artificial agent relationships. The development of appropriate theoretical frameworks, evaluation methodologies, and ethical guidelines will be essential for realising the potential of emotion recognition technology to enhance human-robot interaction while respecting human dignity and autonomy.

Acknowledgement
The authors sincerely thank The City School (Chenab Campus) and Beaconhouse International College (Faisalabad Campus) for their valuable collaboration in facilitating this research. Their support in offering a two-month internship provided critical practical exposure and access to authentic educational settings, significantly contributing to the successful execution of the research study. Special thanks are extended to Ms. Kanza and Principal Imran Shahid of The City School, as well as to Ms. Sadia Aziz and Dr. Fawad Naseer from Beaconhouse International College, for their guidance, encouragement, and logistical support throughout the research process.

Conflict of Intest
The authors declare that there is no conflict of interest related to the publication of this research.

REFERENCES
Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in Neural Information Processing Systems, 33, 12449-12460.

Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition, 59-66. https://doi.org/10.1109/FG.2018.00019

Bankins, S., & Formosa, P. (2024). When AI meets PC: Exploring the implications of workplace social robots and a human-robot psychological contract. European Journal of Work and Organizational Psychology, 28(5), 613-625. https://doi.org/10.1080/1359432X.2019.1620328

Baudier, P., De Boissieu, E., & Duchemin, M. (2022). Source credibility and emotions generated by robot and human influencers: The perception of luxury brand representatives. Technological Forecasting and Social Change, 187, 122255. https://doi.org/10.1016/j.techfore.2022.122255

Beck, F., & Kugi, A. (2022). Motion planning for human-robot collaboration. In Automation, robotics and communications for industry 4.0 (pp. 129-140). TU Wien Academic Press. https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_6

Bonarini, A. (2020). Communication in human-robot interaction. Current Robotics Reports, 1(4), 279-285. https://doi.org/10.1007/s43154-020-00026-1

Churamani, N., Kalkan, S., & Gunes, H. (2022). Continual learning for affective robotics: Why, what and how? 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 558-565. https://doi.org/10.1109/RO-MAN53752.2022.9900607

Cucciniello, I., Sangiovanni, S., Maggi, G., & Rossi, S. (2023). Mind perception in HRI: Exploring users' attribution of mental and emotional states to robots with different behavioural styles. International Journal of Social Robotics, 15(5), 867-877. https://doi.org/10.1007/s12369-023-00989-z

De Pagter, J. (2022). Building trust in robots: A narrative approach. In A. Weiss, M. Mara, & M. Vincze (Eds.), Human-robot interaction: Research, development and application (pp. 265-274). TU Wien Academic Press. https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_12

Dwijayanti, S., Iqbal, M., & Suprapto, B. Y. (2022). Real-time implementation of face recognition and emotion recognition in a humanoid robot using a convolutional neural network. IEEE Access, 10, 89876-89886. https://doi.org/10.1109/access.2022.3200762

Etemad-Sajadi, R., Soussan, A., & Schöpfer, T. (2022). How ethical issues raised by human-robot interaction can impact the intention to use the robot? International Journal of Social Robotics, 14(4), 1103-1115. https://doi.org/10.1007/s12369-021-00857-8

Farouk, N. M. (2022). Studying human robot interaction and its characteristics. International Journal of Computations Information and Manufacturing (IJCIM), 2(1). https://doi.org/10.54489/ijcim.v2i1.7

Fartook, M., Patel, D., Bianchi-Berthouze, N., & Baurley, S. (2023). Recognising emotional responses to music with wearable sensors and the sensing garment approach. Sensors, 23(1), 49. https://doi.org/10.3390/s23010049

Fiorini, L., Masala, G. L., Esposito, R., Dario, P., & Cavallo, F. (2024). Investigating older adults' needs and concerns about robot companions for aging in place. Robotics and Autonomous Systems, 174, 104619. https://doi.org/10.1016/j.robot.2024.104619

Frijns, H. A., & Schürer, O. (2022). Design as a practice in human-robot interaction research. In A. Weiss, M. Mara, & M. Vincze (Eds.), Human-robot interaction: Research, development and application (pp. 3-12). TU Wien Academic Press. https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_1

Ganesan, P. (2023). Revolutionising robotics with AI, machine learning, and deep learning: A deep dive into current trends and challenges. Journal of Artificial Intelligence, Machine Learning and Data Science, 1(4), 1124-1128. https://doi.org/10.51219/JAIMLD/premkumar-ganesan/263

Graterol, W., Diaz-Amado, J., Cardinale, Y., Dongo, I., Lopes-Silva, E., & Santos-Libarino, C. (2021). Emotion detection for social robots based on NLP transformers and an emotion ontology. Sensors, 21(4), 1322. https://doi.org/10.3390/s21041322

Hannibal, G., & Weiss, A. (2022). Exploring the situated vulnerabilities of robots for interpersonal trust in human-robot interaction. In A. Weiss, M. Mara, & M. Vincze (Eds.), Human-robot interaction: Research, development and application (pp. 33-42). TU Wien Academic Press. https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_2

Hopko, S., Wang, J., & Mehta, R. (2022). Human factors considerations and metrics in shared space human-robot collaboration: A systematic review. Frontiers in Robotics and AI, 9, 799522. https://doi.org/10.3389/frobt.2022.799522

Jirak, D., Bartos, M., Zelezny, M., & Romportl, J. (2022). Social intelligence in robots: A survey of technical approaches and challenges. Engineering Applications of Artificial Intelligence, 114, 105121. https://doi.org/10.1016/j.engappai.2022.105121

Kewalramani, S., Allen, K., Leif, E., & Ng, A. (2023). A scoping review of the use of robotics technologies for supporting social-emotional learning in children with autism. Journal of Autism and Developmental Disorders. https://doi.org/10.1007/s10803-023-06193-2

Kirschner, R. J., Mayer, H., Burr, L., Mansfeld, N., Abdolshah, S., & Haddadin, S. (2022). Expectable motion unit: Avoiding hazards from human involuntary motions in human-robot interaction. IEEE Robotics and Automation Letters, 7(2), 2470-2477. https://doi.org/10.1109/LRA.2022.3144535

Legler, A., Schmidtke, H. R., & Bock, T. (2023). A review on robot personalities in human-robot interaction. Applied Sciences, 13(5), 2811. https://doi.org/10.3390/app13052811

Li, M., Liang, W., Liu, L., Wang, Y., & Gunes, H. (2023). Multimodal emotion recognition in conversations: Survey, challenges, and opportunities. IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2023.3295750

Lin, Z., Cruz, F., & Sandoval, E. B. (2024). Self context-aware emotion perception on human-robot interaction. arXiv preprint arXiv:2401.10946. https://arxiv.org/abs/2401.10946

Martínez, A., Belmonte, L. M., García, A. S., Fernández-Caballero, A., & Morales, R. (2021). Facial emotion recognition from an unmanned flying social robot for home care of dependent people. Electronics, 10(7), 868. https://doi.org/10.3390/electronics10070868

Melinte, D. O., & Vladareanu, L. (2020). Facial expressions recognition for human-robot interaction using deep convolutional neural networks with rectified Adam optimiser. Sensors, 20(8), 2393. https://doi.org/10.3390/s20082393

Mishra, S., Sturm, B. L., & Dixon, S. (2023). Local interpretable model-agnostic explanations for music content analysis. Proceedings of the 24th International Society for Music Information Retrieval Conference, 719-726.

Naseer, F., Addas, A., Tahir, M., Khan, M. N., & Sattar, N. (2025). Integrating generative adversarial networks with IoT for adaptive AI-powered personalised elderly care in smart homes. Frontiers in Artificial Intelligence, 8. https://doi.org/10.3389/frai.2025.1520592

Naseer, F., Khan, M. N., & Addas, A. (2025). Healthcare transformation through disruptive technologies: The role of telepresence robots. In Advances in science, technology & innovation (pp. 165–180). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-63701-8_14

Naseer, F., Khan, M. N., Addas, A., Awais, Q., & Ayub, N. (2025). Game mechanics and artificial intelligence personalisation: A framework for adaptive learning systems. Education Sciences, 15(3), 301. https://doi.org/10.3390/educsci15030301

Naseer, F., & Khawaja, S. (2025). Mitigating conceptual learning gaps in mixed-ability classrooms: A learning analytics-based evaluation of ai-driven adaptive feedback for struggling learners. Applied Sciences, 15(8), 4473. https://doi.org/10.3390/app15084473

Neerincx, M., van der Waa, J., Kaptein, F., & van Diggelen, J. (2023). Using perceptual and cognitive explanations for enhanced human-AI team performance. Engineering Applications of Artificial Intelligence, 118, 105696. https://doi.org/10.1016/j.engappai.2022.105696

Nigro, M., Akinrintoyo, E., Salomons, N., & Spitale, M. (2025). Social group human-robot interaction: A scoping review of computational challenges. International Journal of Social Robotics. [Note: Updated to reflect anticipated publication]

Obrenovic, B., Gu, X., Wang, G., Godinic, D., & Jakhongirov, I. (2024). Generative AI and human-robot interaction: Implications and future agenda for business, society and ethics. AI & Society. https://doi.org/10.1007/s00146-024-01889-0

Onnasch, L., & Roesler, E. (2020). A taxonomy to structure and analyse human-robot interaction. International Journal of Social Robotics, 13(4), 833-849. https://doi.org/10.1007/s12369-020-00666-5

Rawal, N., Koert, D., Turan, C., Kerzel, M., Meier, M., Chalup, S. K., & Wermter, S. (2022). The role of neural mechanisms in social cognition and behavior: A robotics perspective. Paladyn, Journal of Behavioral Robotics, 13(1), 1-25. https://doi.org/10.1515/pjbr-2022-0001

Safavi, F., Olikkal, P., Pei, D., Kamal, S., Meyerson, H., Penumalee, V., & Vinjamuri, R. (2024). Emerging frontiers in human-robot interaction. Journal of Intelligent & Robotic Systems, 110(2). https://doi.org/10.1007/s10846-024-02074-7

Saxena, A., Khanna, A., & Gupta, D. (2024). Emotion recognition and sentiment analysis in robotics: Current trends and future directions. Expert Systems with Applications, 238, 121847. https://doi.org/10.1016/j.eswa.2023.121847

Seyitoğlu, F., & Ivanov, S. (2024). Service robots as experience co-creators in hospitality: The role of customer readiness to use service robots. International Journal of Contemporary Hospitality Management, 36(1), 300-318. https://doi.org/10.1108/IJCHM-12-2022-1500

Spezialetti, M., Placidi, G., & Rossi, S. (2020). Emotion recognition for human-robot interaction: Recent advances and future perspectives. Frontiers in Robotics and AI, 7, 532279. https://doi.org/10.3389/frobt.2020.532279

Staffa, M., Sgorbissa, A., Ruocco, M., Giglio, L. D., Nunziata, W., & Siciliano, B. (2023). The personality of service robots: How do personality traits affect human-robot interaction? International Journal of Social Robotics, 15(4), 633-651. https://doi.org/10.1007/s12369-022-00868-z

Stock-Homburg, R. (2021). Survey of emotions in human-robot interactions: Perspectives from robotic psychology on 20 years of research. International Journal of Social Robotics, 14(2), 389-411. https://doi.org/10.1007/s12369-021-00778-6

Tariq, R., Aponte Babines, B. M., Ramirez, J., Alvarez-Icaza, I., & Naseer, F. (2025). Computational thinking in STEM education: current state-of-the-art and future research directions. Frontiers in Computer Science, 6. https://doi.org/10.3389/fcomp.2024.1480404

Wachowiak, L., Celiktutan, O., Coles, A., & Canal, G. (2023). A survey of evaluation methods and metrics for explanations in human-robot interaction (HRI). 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 1237-1244. https://doi.org/10.1109/RO-MAN57019.2023.10309540

Wang, Y., Hespanhol, L., & Tomitsch, M. (2021). How can autonomous vehicles convey emotions to pedestrians? A review of emotionally expressive non-humanoid robots. Multimodal Technologies and Interaction, 5(12), 84. https://doi.org/10.3390/mti5120084

Wang, X., Liu, Y., Chen, L., & Zhang, H. (2022). Emotional intelligence in human-robot collaboration: A systematic review. Robotics and Computer-Integrated Manufacturing, 78, 102390. https://doi.org/10.1016/j.rcim.2022.102390

Wang, X., Shen, L., & Lee, L.-H. (2024). Towards massive interaction with generalist robotics: A systematic review of XR-enabled remote human-robot interaction systems. arXiv preprint arXiv:2403.11384. https://arxiv.org/abs/2403.11384

Xie, L., & Park, H. W. (2023). Robot learning of social emotions from multimodal perception based on user feedback. 2023 IEEE International Conference on Development and Learning (ICDL), 298-305. https://doi.org/10.1109/ICDL55364.2023.10364385

Xie, Y., Chen, C., Lantao, Y., Li, H., & Zhu, J. (2024). The role of explainable AI in human-robot trust: A comprehensive framework. IEEE Transactions on Human-Machine Systems, 54(2), 178-189. https://doi.org/10.1109/THMS.2024.3364895

Zafari, S., & Koeszegi, S. T. (2022). Agency in socio-technical systems: How to enact human-robot collaboration. In A. Weiss, M. Mara, & M. Vincze (Eds.), Human-robot interaction: Research, development and application (pp. 229-238). TU Wien Academic Press. https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_10

Zhang, B. J., & Fitter, N. T. (2023). Nonverbal sound in human-robot interaction: A systematic review. ACM Transactions on Human-Robot Interaction, 12(4), 1-46. https://doi.org/10.1145/3583743

Zhang, Y., Ma, Y., & Kragic, D. (2024). Vision beyond boundaries: An initial design space of domain-specific large vision models in human-robot interaction. arXiv preprint arXiv:2404.14965. https://arxiv.org/abs/2404.14965

Zhao, M. (2023). Emotion recognition in psychology of human-robot interaction. Psychomachina, 1(1). https://doi.org/10.59388/pm00331.