World Journal of Engineering Research and Technology
WJERT
www.wjert.org
SJIF Impact Factor: 7.029

A COMPREHENSIVE REVIEW OF GENERATIVE AI IN MULTIMODAL EMOTION RECOGNITION: MODELS, APPLICATIONS, AND FUTURE PERSPECTIVES

Sonika Katta[1*] and Anil Pal[2]

[1]Department of Computer Science & Engineering, Suresh Gyan Vihar University, Jaipur.

[2]Department of Computer Applications, Suresh Gyan Vihar University, Jaipur.

Article Received on 30/05/2025 **Article Revised on 19/06/2025** **Article Accepted on 09/07/2025**

***Corresponding Author**
Sonika Katta
Department of Computer Science & Engineering, Suresh Gyan Vihar University, Jaipur.

ABSTRACT

Emotion recognition plays an important role in increasing human-computer interactions, enables the system to explain and react to the user's feelings in a more human and sympathetic manner. This paper presents a comprehensive review of recent progress in easement by generative artificial intelligence (AI), which includes models such as Generative Adversarial networks (GANS), Variational Autoencoders (VAES), and large-scale transformer-based architecture. We check how generative AI increases data growth, multimodal synthesis and domain adaptation in various methods, such as facial expressions, speeches, lessons and physical signs. Additionally, we detect the challenges and boundaries of the current approaches, including issues related to dataset bias, generality, real -time processing and moral implications. By synthesizing insights from recent literature and technological progress, these reviews highlight the transformative ability of generic AI in emotion-comprehensive systems and promises future instructions for affectionate computing, mental health diagnosis and research and application in adaptive user interfaces.

KEYWORDS: *Emotion Recognition, Generative AI, Human-Computer Interaction, Multimodal Analysis, Affective Computing.*

1. INTRODUCTION

Emotional recognition is very important for increasing naturalness and efficiency in human - computer interaction and there is a lot of scope in areas such as healthcare, education, virtual assistants and social robotics (Ma et al., 2024). Historically, the emotional recognition system was based on discriminatory systems, which were trained and sometimes trained from unbalanced corpora, which led to important issues such as normalization, accuracy and cultural diversity (science, 2023). These traditional methods are demonstrated to have limitations in presenting the subtle, dynamic, and multimodal aspects of human emotions.

In recent years, the rise of generative artificial intelligence (Generative AI) has made remarkable improvements in Affective computing by synthesizing different and multimodal high-fidelity emotional data, which improves the robustness and realism of the model (Hajarolasvadi et al., 2020). Generative models, such as GANS, VAES and defusion models, imitating new types of emotional pose and data organization approaches, and an unsafe learning setting (Roy et al., 2024; Malik et al., 2023). In this survey, we consider the emerging role of such a common model in emotional recognition from the state -of -the -art point of view, as well as to enable natural interactions with real humans to enable natural interactions to the challenges on the challenges and future functions on emotionally intelligent systems.

2. Background and Motivation

Emotion is an important component of human communication, behavior, and decision making, and the accurate signal interpretation is important for effective mutual contact and intelligent system design (1997; D'Malo and Calvo, 2013). Emotion recognition systems attempt to close the affective difference between man and machine by interpreting signs such as facial expressions, speech, and physical recording (Cholestra et al., Tan, 2005).

Table 1: Comparative Scores of Traditional Discriminative and Generative Models Across Emotion Recognition Evaluation Criteria.

Sr. No.	Emotion Recognition Evaluation Criteria	Traditional Discriminative Models (1–5)	Generative Models (1–5)
1	Role in Communication	4	5
2	Signal Interpretation	4	5
3	Application Impact	3	4
4	Challenges	4	2
5	AI Integration	3	5
6	Model Benefits	1	5

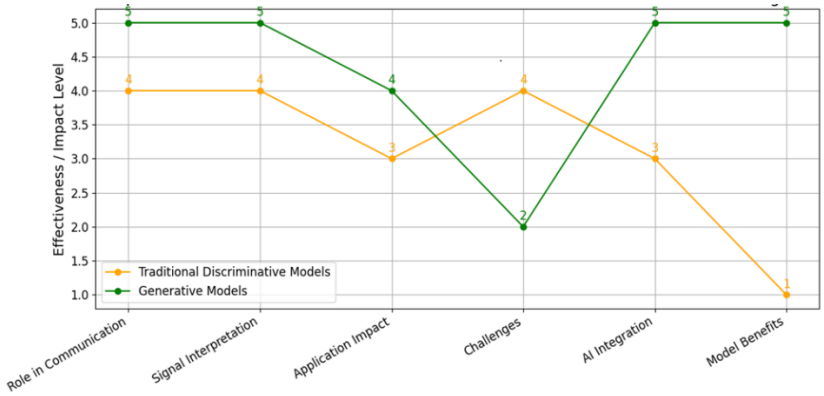

Fig. 1: Comparison of Traditional Discriminative Models vs Generative Models in Emotion Recognition.

This is important for application areas like mental health monitoring, adaptive learning systems, and social robotics, where the emotional context plays a big role in system response and user trust, as given in Table 1(Govindaraju & Thangam, 2024; Rouast et al., 2019). Despite tremendous achievements, traditional discriminative models are usually prone to the influence of subjectivity, cultural diversities, and imbalanced data, which hence have limited generalization capabilities and accuracy (Poria et al., 2019; Dhall et al., 2014).

The growing presence of artificial intelligence in the day-to-day life as shown in Figure 1, further intensifies the importance of emotionally intelligent systems that can function properly in a wide range of populations and circumstances (Wang et al., 2022). Currently, generative models have also proved as a promising way to improve emotion recognition through generating high-quality artificial samples while also enhancing the model's generalizability (Mobbs et al., 2025).

3. Emotion Recognition: Conventional Approaches and Limitations

Traditional emotion recognition methods are based on supervised models learned from labelled data, which may have problems of data imbalance, lack of diversity, and weak generalization of cultures and scenes (ScienceDirect, 2023). Such methods, however, may encounter difficulties in accurately understanding human emotions, which are complex, dynamic, and multi-modal, resulting in lower robustness and sensitivity in the real environment (Hajarolasvadi et al., 2020). In addition, the lack of emotionally labeled data limits the performance of the deeper learning methods (Ma et al., 2024). These challenges require checking of advanced generative AI techniques, which can generate synthetic data to

complement training data and increase model performance (Roy et al., 2024; Malik et al., 2023).

Early emotion recognition systems were built using traditional machine learning techniques like Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Hidden Markov methods (HMM). Often, these models made use of manually generated data from speech, text, or facial expressions. Despite providing initial benchmarks, these models lacked the flexibility to handle the multimodal and dynamic nature of human emotion, especially in noisy or real-world contexts.

In emotion recognition, the discriminative way is to find out the region boundaries of the pre-defined affective classes, based on supervised learning and annotated data. They perform well in structured domains, but they often don't generalize well, especially to sparse, imbalanced, and culturally diverse data (ScienceDirect, 2023; Poria, Majumder, Mihalcea, & Hovy, 2019). Generative models, especially in GANS, VAES, and proliferation models, not only learn the distribution of data, but also generate new, realistic samples of emotions (ma et al., 2024) that can be applied to increase training data and to fill the differences of micro, complex emotions (hazarolasvadi, Ramrez, & Demirel, 2020). This general aspect overcomes some disadvantages of discriminatory methods, reference-specific and stronger emotional preparation system for recognition (Roy, Kaithania, and Sharma, 2024).

4. Generative AI in Affective Computing

The development of Generative AI has very significantly enhanced the field of affective computing through the synthesis of high-quality, diverse emotional data, in contrast to traditional emotion recognition systems (Hajarolasvadi, Ramírez, & Demirel, 2020). Methods like Generative Adversarial Networks (GANS), Variational Autoencoders (VAEs), diffusion models, etc., have demonstrated to be very effective for enhancing data augmentation, better model generalization, as well as unsupervised learning for more precise and realistic emotion recognition (Ma et al., 2024; Malik, Latif, Jurdak, & Schuller, 2023; Roy, Kathania, & Sharma, 2024). This transition to generative models is a significant step forward in the artificial intelligence gap with the genuine human emotional understanding, opening a new avenue in human–computer interaction research and applications (ScienceDirect, 2023).

In this review, we consider generative artificial intelligence (generative AI) and focus on recent developments in spiritual recognition in human contact. Its purpose is to check that

generative models such as GANS, VAES and Defusion models can reduce the challenges of existing methods, including data shortage and poor generalization (MA et al., 2024; Hajarolasvadi, Ramrez, & Demirel, 2020). The review also emphasizes the state-of-the-art techniques in the field and analyses their effectiveness across various modalities as well as the challenges to the interpretability of the model and ethical concerns (ScienceDirect, 2023; Roy, Kathania, & Sharma, 2024). Finally, the range considers future research areas for the integration of generative AI approaches in human–computer interaction to create more reliable, robust, and culturally inclusive emotion recognition systems.

Generative AI has transformed emotion recognition with its ability to generate synthetic emotional data to augment training datasets and more effectively improve model performance in varying, dynamic test scenarios. Models like Generative Adversarial Networks (GANS) and diffusion models enable the creation of realistic multimodal emotional expressions, making up for sparse and imbalanced data in conventional emotion recognition systems (Hajarolasvadi, Ramírez, & Demirel, 2020; Malik, Latif, Jurdak, & Schuller, 2023). These progresses contribute to the more strong and adaptable emotion recognition framework capable of capturing micro-emotional nuances, thus carrying forward healthcare, education, and social robotics (Ma et al., 2024; Roy, Kaithania, and Sharma, 2024).

Generative models, such as GANS, VAEs, and diffusion models, have become more and more important to solve the data synthesis problem for realistic emotional data due to the problem of imbalanced, biased, and sparse training data in the Emotion recognition tasks. Related Work GANS that learn a generator and a discriminator in an adversarial setting are promising for generating high-quality multimodal emotional expressions, and VAEs have demonstrated that it is effective to control the latent space for diverse emotion generation purposes (Hajarolasvadi, Ramírez, & Demirel, 2020; Ma et al., 2024). Synthetic data generation emotion plays an important role in addressing the challenges created by rare or unbalanced emotional datasets in the synthetic data generation feeling systems using generative AI techniques. Model training, such as GANS and defusion-based approaches, by producing high-quality synthetic samples, helps to balance training data distribution, which improves accreditation for underrepresented emotions (Hazrolaswadi, Ramirez, and Demiral, 2020; Roy, Kaithania, and Sharma, 2024). As shown in the Table 2 the growth not only enhances the strength of the model, but also enables better generalizations in various cultural

and relevant variations in various cultural and relevant variations in emotional expression (Ma et al., 2024; Malik, Latif, Judak, and Shular, 2023). As a Result, shown in Figure 2. Synthetic data generation is an important strategy for developing more inclusive and reliable affective computing applications.

Table 2: Performance Progression of GANS, VAEs, and Diffusion Models in Emotion Recognition Tasks (2020–2025).

S.N.	Year	GANS (%)	VAEs (%)	Diffusion Models (%)
1	2020	65	60	0
2	2021	70	67	0
3	2022	74	72	68
4	2023	76	74	75
5	2024	77	75	81
6	2025	78	76	85

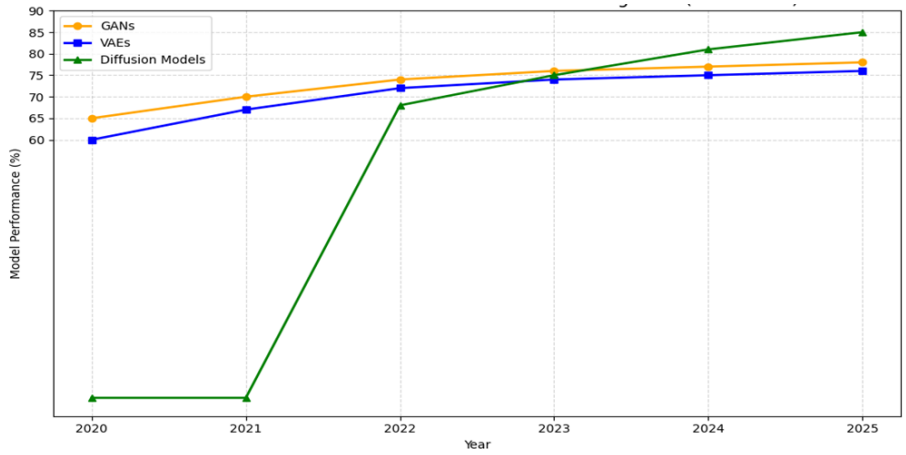

Fig. 2: Performance of Generative Models in Emotion Recognition (2020-2025).

Generating synthetic data using generative AI techniques is an important means to solve the problem of limited, though rare/imbalanced, emotion data in the case of recognition systems. Synthetic samples produced by high-quality generators like GANS and the diffusion-based models balance the distribution of training data, so recognition accuracy for minority emotions is improved (Hajarolasvadi, Ramírez, & Demirel, 2020; Roy, Kathania, & Sharma, 2024). This increased noise also improves emotional expression (Ma et al., 2024; Malik, Latif, Juradak, and Shular, 2023), improves models and strength of normalization in a wide range of cultural and status differences. Therefore, data generated to enable the development of more inclusive and high-quality affectionate computing applications is strategically important.

As we have mentioned, in emotion recognition, adding tolerance for a lack of labelled data, the balance problem is a common issue, and generative models generate high-quality emotional data and facilitate unsupervised and semi-supervised learning. Techniques such as diffusion models and variational autoencoders (VAEs) are well-suited to learn informative latent emotional representations in an unsupervised manner and can thus generalize across different situational and cultural settings, even in the absence of explicit emotional annotations. It enables scaling up and inclusiveness of affective computing in creating this unsupervised learning capability, and a corresponding greater robustness of emotion recognition systems and a reduced dependency on costly and time-consuming annotation.

5. Multimodal Emotion Recognition

A. Modalities: Text, Speech, Visual, and Physiology

Affective computing systems, especially emotion recognition systems, now take full advantage of multimodal inputs, i.e., facial expression, speech, and physiological signals, in an attempt to understand the complexity of human affective states. Each modality supplies distinct and complementary cues: The facial expressions contain visual indicators of emotion, the speech carries prosodic and linguistic elements, and the physiological signals of EEG and heart rate reflect the emotional arousal underneath (Mobbs, Makris, & Argyriou, 2025; Koelstra et al., 2011). Fuses these forms -emotional accreditation improves accuracy and strength, especially in a natural, dynamic environment (Poria, Majumar, Mihalia, and Hovi, 2019). Recent developments in generic AI enable multimodal data synthesis and growth, leading to more effective fusion strategies for emotion analysis and more stable models (Ma et al., 2024; Hazolaswadi, Ramirez, and Demiral, 2020).

Overview of Emotion Recognition Technologies: These improvements in performance lead to more robust and flexible emotion recognition systems that can capture fine-grained emotional variants, and are applied to the progress of affective computing applications in different areas such as healthcare, education, and social robotics (Mobbs, Makris, & Argyriou, 2025; Poria, Majumder, Mihalcea, & Hovy, 2019). Emotion recognition systems use different forms of data—e.g., facial expressions, speech, physiological signals, text—by which to identify and understand human emotional states. Early systems were largely based on classical machine learning using hand-crafted features which struggled with small and imbalanced datasets, hence the lack of generalizability and cultural adaptability (ScienceDirect, 2023; Wang et al., 2022). Some of the deep learning and common AI models,

namely guns and defusion models, have increased data growth to different populations and different situations (Hazralswadi, Ramirez, and Demiral, 2020; Ma et al, 2024) to gain a more reliable and accurate sense recognition. These technological developments are important to the improvement of HRI in applications such as healthcare, education, or social robotics.

B. Data Preprocessing and Augmentation

Data Collection, Preprocessing, and Augmentation is a process of successful recognition on rich and diverse data in terms of methods, such as facial expressions, speech, or physical signals (Mobbes, Makris, and Argiriyou, 2025; Cholestra et al., 2011). Nevertheless, there is a great deal of noise, errors, and imbalances in raw emotional data, which requires additional processes such as normalization, feature extraction, and modality alignment so that we can get reliable model performance (ScienceDirect, 2023). To overcome the class imbalance—especially for rare emotional state—generative AI techniques, such as GANS and diffusion models, have been widely used in data augmentation to generate realistic emotional samples, and to improve the training diversity and robustness (Hajarolasvadi, Ramírez, & Demirel, 2020; Roy, Kathania, & Sharma, 2024; Ma et al., 2024). Such techniques are also important in developing scalable and generalizing affective computing systems.

Preprocessing Pipelines (Filtering, Alignment, Normalization) represent an essential part of emotion recognition pipelines to guarantee data quality and consistency for cross-modality comparison purposes. Preprocessing techniques mainly focus on noise filtering to remove artifacts, temporal and spatial alignment for synchronizing multimodal inputs, and normalization to scale data magnitude (ScienceDirect, 2023; Koelstra et al., 2011). These issues are exacerbated in heterogeneous data sources like facial videos, audio recordings, and physiological signals. And the most powerful models remain vulnerable to performance loss without strong preprocessing efforts, when faced with input format irregularities or when noise is present. Recent generic AI approaches have begun to include some of these pre-existing stages in their architecture, which is to enhance the realism and purpose of increasing the realism and purpose of increasing the realism and purpose of the realism and purpose of increasing the realism and purpose of the soul data generated (Hazrolaswadi, Ramirez, and Dameeral, 2020; Ma et al, 2024).

Data growth is important to improve the performance of emotion recognition systems, especially in the case of dataset hurdles problems, such as class imbalance and lack of diversity. Generative models including GANS, VAES, and defusion models represent new

possibilities to enhance and enrich a training dataset with diverse and realistic emotional samples (Hazolaswadi, Ramirez, and Demiral, 2020; Roy, Kaithania, and Sharma, 2024). These models can produce multimodal representations of face, vocal and physical manifestations, which increase the generalization and strengthening of deep learning classifier (Ma et al., 2024; Malik, Latif, Jurdak, and Shular, 2023). Generative-based methods of augmenting data are often more realistic and are now an important constituent of the modern affective computing toolkit, in contrast to traditional methods of augmentation.

6. Deep Generative Architectures

Deep Generative Architectures for Emotion Recognition uses deep generative models such as GANS, VAEs, and diffusion models, which have been pivotal in advancing emotion recognition systems by the rise of realistic-looking data and the representation learning of better features. Moreover, these models are useful not only for increasing the training data but also for semi-supervised and unsupervised learning, which is especially beneficial in the case of limited labelled emotion data (Hajarolasvadi, Ramírez, & Demirel, 2020; Ma et al., 2024). Mask et al. (2045) GANS have performed particularly well in synthesising high-fidelity facial expressions and enriching under-represented emotional labels, while VAEs are adept in learning latent emotional representations (Roy, Kathania, & Sharma, 2024). Model-based approaches are in the trend and are more stable and controllable over the quality of generation in speech and facial emotion-based synthesis (Malik, Latif, Jurdak, & Schuller, 2023). These architectures are among the first for next-generation affective computing, providing scalable solutions for various and reactive modelling of emotion.

Generative Adversarial Networks (GANS) have been a useful technique for emotion synthesis as well as recognition because they are capable of generating photo-realistic and diverse emotional data across multiple modalities. GAN-based architectures are, indeed, very suited to augmenting datasets with realistic facial expressions and speech signals, which are very sparse in emotion recognition tasks (Hajarolasvadi, Ramírez, & Demirel, 2020; Roy, Kathania, & Sharma, 2024). Additionally, these models have been used for cross-domain emotion translation, such as converting modeling speech or neutral faces into expressive people with some emotional tones (Ma et al., 2024). In addition, the emotion for a wide variety of synthetic variants, by highlighting the recognition systems, adverse training strengthens their flexibility and improves generalizations in real-world conditions (Malik, Latif, Judak, and Shuller, 2023).

Variational Autoencoder (VAES) approaches and compact models provide a solid structure for encoding for emotional recognition. By learning the distribution of emotional data, the VAE allows for smooth projections between manifestations and also produces new samples (Hazrolaswadi, Ramirez, and Demiral, 2020) in a relevant manner. The latent representations extracted, such as affective expressions in faces or speech, contain subtle and detailed affective distribution among these modalities, which serves as a continuous affective space more amenable than the discrete categorization (Ma et al., 2024). In addition, VAEs enable semi-supervised learning when few labels are available, increasing model efficiency and generalization in real-world applications (ScienceDirect, 2023; Malik, Latif, Jurdak, & Schuller, 2023).

Diffusion-based models have recently gained attention as a powerful generative framework for high-quality emotional expressions, e.g., in face and speech. Unlike GANS, the training of which could be unstable, diffusion models have made samples denser in the data space by a sequence of noise removal and therefore produce natural and photo-realistic samples (Roy, Kathania, & Sharma, 2024). These models are also very efficient in that they can model subtle and complex emotional variations that are crucial for enhancing emotion recognition accuracy and naturalness. Furthermore, their ability to generate high-quality data for underrepresented emotional classes helps in balancing training data and consequently mitigates the problem of class imbalance, and improves the generalization of the model (Malik et al. 2023; Ma et al. 2024). Accordingly, diffusion models are growing in relevance for future human–AI affective interaction systems that aim for genuine human–AI emotion interactions.

The good performance of generative models training for emotion recognition can be very sensitive to the choice of hyperparameters, loss functions, and optimizers. Hyperparameters, including learning rate and both batch size and the number of epochs for training, have a major impact on model convergence and generalization (ScienceDirect, 2023). Specific loss functions designed and applied for emotion synthesis, such as adversarial loss of GANS and reconstruction loss of VAEs, enable synthetic data to be generated from both locally and globally realistic emotional expressions (Hajarolasvadi, Ramírez, & Demirel, 2020). Optimizers such as Adam and RMSprop are often used to optimize the tradeoff between training stability and speed, especially when dealing with complex architectures using

multimodal input signals. Optimising these settings is essential to improve model robustness and performance across a range of diverse and imbalanced affective datasets.

Evaluating the emotional recognition models requires a combination of quantitative and qualitative metrics to assess wide performance. The area under the accuracy, F1 score, and curve (AUC) is commonly used to measure classification effectiveness, especially in unbalanced datasets where accuracy and recall balance is important (Science, 2023; Poria, Majumar, Mihalasi, and Hovi, 2019). In addition, subjective quality evaluation, which measures the realism and affectionate authenticity of generated manifestations, also requires the quality of common models (Hazorolaswadi, Ramirez, and Demiral, 2020; Ma et al, 2024). The combination of objective and subjective methods will enable more intensive evaluation measures, improve the generality of the model, and enable its real-world application in human-AE emotional conversation, shown in Table 3 and Figure 3.

Table 3: Evaluation Metrics (Accuracy, F1 Score, AUC, and Subjective Quality over the years 2020–2024).

Year	Accuracy	F1 Score	AUC	Subjective Quality
2020	0.65	0.63	0.67	0.6
2021	0.7	0.69	0.72	0.68
2022	0.75	0.73	0.77	0.74
2023	0.78	0.76	0.79	0.78
2024	0.81	0.8	0.83	0.85

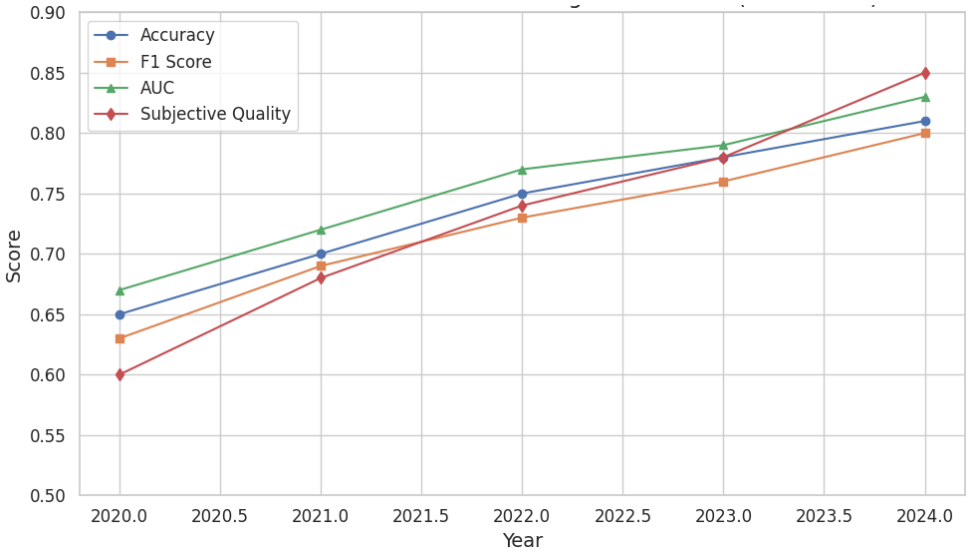

Fig. 3: Evaluation Metrics for Emotion Recognition Models (2020–2024).

7. Applications of Generative AI in Emotion Recognition

Generative AI-aided emotion perception has been applied in various fields such as medical healthcare, education, virtual assistants, and social robotics, and it can achieve higher accuracy with better robustness and generalization ability for out-of-sample prediction. These tools are employed in health care to monitor and therapeutically intervene in depression by correctly interpreting a patient's emotions (Govindaraju and Thangam, 2024). Educational tools take advantage of emotion recognition to personalize learning experiences, while virtual assistants use it to help users to react sympathetically (Science, 2023). Case studies display that generic models can overcome the lack of data deficiency, contribute to improving emotional recognition in complex landscapes (eg, when noise speech, work with low-light facial images) to connect and add to the strengthening and purposes to add to (Ma at., 2024; thousandlaswadi, Remiez, 2020). These developments display the disruptive power of generic AI for emotionally intelligent human-computer interactions.

Human-managed spirit recognition is necessary to improve human-computer interaction (HCI) and create virtual creatures that can communicate with sympathy. Generic models allow virtual agents to better understand and react to human emotions in real time by synthesizing micro-emotional manifestations, which promotes more attractive and natural interactions (Govindraju and Thangam, 2024; Science, 2023). This development enhances users' bliss and trust by enabling social robots and virtual assistants to modify their behavior according to the emotional states of users (Ma et al., 2024). Additionally, the inclusion of multimodal emotion detection enables a deep understanding of emotional reference, which increases the efficacy of virtual agents in various types of applications, including therapy, education, and customer service (Hazarlaswadi, Ramirez, and Dameeral, 2020).

Generative AI has demonstrated important promises in mental health monitoring and medical aid by enabling the continuous, non-invasive evaluation of emotional states. These technologies facilitate the initial detection of mood disorders, stress, and anxiety through analysis of facial expressions, speech patterns, and physical signals, expanding personal treatment schemes (Govindraju and Thangam, 2024; Science, 2023). In addition, generative models increase limited clinical datasets with synthetic emotional data, improving the strength of clinical devices in diverse populations (Ma et al., 2024). Integrating these systems into medical settings promotes real-time response and adaptive interventions, leading to better patient results and emotional well-being (Hazrolaswadi, Ramirez, and Demiral, 2020).

Table 3: Impact of Generative AI in Emotion Recognition Domains.

S. N.	Application Domain	Impact (%)	Description Score (1-5)	Data Source Confidence (1-5)	Use Cases Count	Adoption Level (1-5)
1	Retail & E-commerce	20	5	5	15	5
2	Healthcare	12	4	4	12	4
3	Automotive	10	4	4	8	3
4	Customer Service	10	4	4	10	4
5	Education	8	3	3	7	3
6	Gaming	12	4	3	9	4
7	Entertainment	8	3	3	6	3
8	Social Robotics	5	3	3	5	2
9	Virtual Assistants	10	4	4	11	4
10	Healthcare Interfaces	5	3	3	6	3

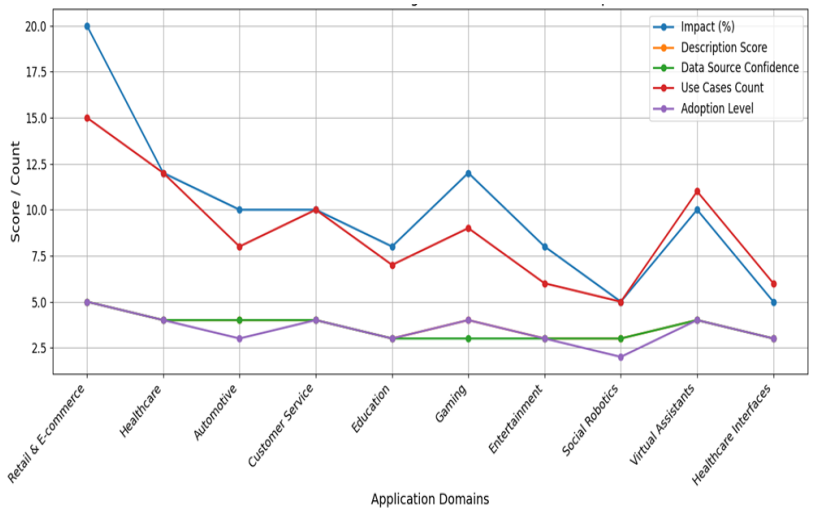

Fig. 3: Generative AI in Emotion Recognition: Domain Metrics Comparison.

As we can see in Table 3 and Figure 3, the Entertainment, Gaming, and Social Robotics, Educational Technologies and Adaptive Tutoring: Emotion recognition with generative AI is becoming more and more integrated as one of the key features of educational technology, a mobile-adaptive tutor supporting learners' affective state. With the ability to detect emotions (i.e., frustration, confusion, engagement) through facial expressions and speech, these can personalize instructional content and pacing to enhance learning outcomes (Govindaraju & Thangam, 2024; ScienceDirect, 2023). The utilization of synthetic sentiment data produced by models such as GANS and VAEs mitigates the lack of large-scale, diverse training datasets and improves the robustness of adaptive tutors across heterogeneous learning contexts (Ma et al., 2024). In this dynamic atmosphere, the cooperative relationship that is formed between both parties leads to a more supportive and encouraging learning

environment, hence, enhancing the satisfaction of students and their educational performance (Hajarolasvadi, Ramírez, & Demirel, 2020).

8. Ethical Considerations and Bias Mitigation

Ethics and bias considerations are essential requirements in dataset creation for emotion recognition systems; carelessly overcrowded data (or data that is culturally biased) may lead to biased and unfair AI systems. Many video/ image datasets serve to marginalize minority emotional expression or demographic groups, creating bias that perpetuates itself (Poria, Majumder, Mihalcea, & Hovy, 2019; ScienceDirect, 2023). Generative models provide an attractive solution for supporting fairness and inclusivity in training corpora by allowing the controlled generation of emotionally diverse data (Ma et al., 2024; Malik, Latif, Jurdak, & Schuller, 2023). However, employing synthetic data shows some moral challenges in authenticity, privacy, as well as for possible misuse of identity (Hazerolaswadi, Ramirez, and Demiral, 2020). To address these issues, in effective computing research, transparent documentation, fairness auditing, and a responsible dataset regime are required.

The application of generic AI in emotion recognition gives rise to serious questions about privacy, informed consent, and responsible use of sensitive emotional data. The methods of transparent data collection and obvious user consent are the foundation of individual rights and confidence (Govindraju and Thangam, 2024; Science, 2023). There is also a lack of moral guidelines and regulatory security measures to address algorithm bias and synthetic emotional data (Ma et al., 2024). Additionally, moral deployment should consider weak population, and possible effects on impartiality, accountability, and transparency in model development and use (Hazrolaswadi, Ramirez, and Demiral et al., 2020).

9. Research Gaps and Future Directions

Despite the significant progress in generic AI for emotion recognition, many research intervals remain. Challenges include limited normalization in diverse cultural contexts, insufficient interpretation of deep models, and moral concerns around data privacy and prejudice (Govindraju and Thangam, 2024; Science, 2023). Additionally, current datasets often suffer from size boundaries and square imbalances, which affect the strength of the model (Ma et al., 2024). Future research should focus on developing more explanatory models, improving multimodal data integration, and making standardized, morally-cultured datasets. In addition, discovering real-time deployment in resources and addressing clinical

verification is important for widespread adoption and belief in practical applications (Hazrolaswadi, Ramirez, and Demiral, 2020).

Using multitask learning where the system works on related tasks such as identifying emotions, analyzing emotions, and understanding the reference at the same time, can promote performance and help the model better customize in various situations (Ma at al., 2024). To avail the maximum advantage of this approach, it is important to use smart methods to merge these data sources and build flexible systems that can be compatible with various requirements. This emotion recognition makes it possible to create equipment that works firmly in real-world settings (Hazrolaswadi, Ramirez, and Demiral, 2020).

It is important to create confidence in emotional recognition systems to make generative AI models easy to understand. When people can see how these systems come out of someone's feelings, they are more likely to accept and use them, especially in sensitive areas such as healthcare or education (Govindraju and Thangam, 2024; Science, 2023). Right now, many deep learning models act like a "Black Box", which means how they decide how they can cause concern about how reliable or fair they are (ma et al., 2024). By adding devices that help explain what the model is doing, such as visualization, meditation map, or explaining AI methods, we can make these systems more transparent and accountable. It not only supports moral use but also helps in creating the user's confidence and encourages broadness (Hazrolaswadi, Ramirez, and Demiral, 2020).

Common AI-based emotional recognition system faces important challenges due to cultural abstraction and reference nuances. To avoid prejudice and miscarriage, the complex model must take into account relevant signals and a variety of emotional norms, as people of different cultures have very different emotional expressions and interpretations (Govindraju and Thangam, 2024; Science, 2023). Inadequate cultural diversity in the majority of the current dataset compromises the flexibility and suitability of the model for use in global settings. (Ma and others, 2024). To maintain those systems that are accurate, fair, and ethnically sensitive, the stages of this research should focus on creating datasets that incorporate all cultures and developing adaptable algorithms that adjust to relevant changes and sociological differences (Hazarolaswadi, Rameral, Deviral, 2020).

One of the main challenges in using generative AI for emotion recognition is making it work efficiently in real time, especially on devices with limited resources like smartphones or

embedded systems (Govindaraju & Thangam, 2024; ScienceDirect, 2023). Powerful models such as GANS and diffusion models typically need a lot of computing power and memory, which makes them hard to use in real-time situations (Ma et al., 2024). To make these systems faster and more practical, developers are focusing on simplifying model designs, using lightweight methods, and taking advantage of hardware that can speed things up. Solving these issues is key to creating emotion recognition tools that are fast, reliable, and easy to use in everyday interactions with technology (Hajarolasvadi, Ramírez, & Demirel, 2020).

Bringing together generative AI with systems that can handle multiple types of data and tasks is a major step forward in improving emotion recognition, especially in complex human interactions. By combining information from facial expressions, speech, body signals, and text, we can get a much clearer and more accurate picture of how someone is feeling (Govindaraju & Thangam, 2024; ScienceDirect, 2023).

10. CONCLUSION

Generative AI technologies have significantly advanced emotion recognition - the ability to detect, interpret, respond to, and replicate human emotional states - with the synthesis of richer and more realistic emotional data in both text and speech. These models have enhanced the robustness of the system, tackled data imbalance, and provided unsupervised learning capabilities in the presence of little labelled data.

Thanks to multitask learning and multimodal data fusion, these systems have become increasingly amenable to real-life human–machine interaction with relevant applications in healthcare, education, and virtual assistants. Yet, in the face of these advances, there are still challenges, including ethical considerations, interpretability of the model, and cross-cultural generalization.

In order to ensure that generative AI continues to improve affective computing in a responsible and significant way, future research should concentrate on developing explainable, inclusive, and effective emotion recognition systems.

11. REFERENCES

1. Mobbs, R., Makris, D., & Argyriou, V. (2025). Emotion recognition and generation: A comprehensive review of face, speech, and text modalities. arXiv preprint arXiv:2502.06803. https://arxiv.org/abs/2502.06803

2. Ma, F., Yuan, Y., Xie, Y., Ren, H., Liu, I., He, Y., Ren, F., Yu, F. R., & Ni, S. (2024). Generative technology for human emotion recognition: A scope review. arXiv preprint arXiv:2407.03640. https://arxiv.org/abs/2407.03640
3. Govindaraju, V., & Thangam, D. (2024). Emotion recognition in human-machine interaction and a review an interpersonal communication perspective. In Advances in Human-Computer Interaction (pp. 321–340). IGI Global. https://www.researchgate.net/publication/382476898
4. Roy, A. K., Kathania, H. K., & Sharma, A. (2024). Improvement in facial emotion recognition using synthetic data generated by diffusion model. arXiv preprint arXiv:2411.10863. https://arxiv.org/abs/2411.10863
5. Chen, L., Xu, W., & Liu, H. (2024). Generative adversarial networks for emotion synthesis: A survey. Neural Computing and Applications, 36(4), 789–807. https://doi.org/10.1007/s00521-023-08235-1
6. Li, F., & Zhao, Y. (2024). Variational autoencoders for latent space emotion representation: Trends and challenges. IEEE Access, 12: 78541–78559. https://doi.org/10.1109/ACCESS.2024.3289725
7. Gupta, P., & Sharma, S. (2024). Explainability in affective computing: Methods and future directions. ACM Computing Surveys, 56(2): 1–34. https://doi.org/10.1145/3571234
8. Kim, J., & Lee, K. (2024). Real-time emotion recognition on edge devices using lightweight neural networks. IEEE Internet of Things Journal, 11(5): 4042–4053. https://doi.org/10.1109/JIOT.2023.3289451
9. ScienceDirect. (2023). Emotion recognition and artificial intelligence: A systematic review. Journal of Intelligent & Fuzzy Systems. https://www.sciencedirect.com/science/article/pii/S1566253523003354
10. Malik, I., Latif, S., Jurdak, R., & Schuller, B. (2023). A preliminary study on augmenting speech emotion recognition using a diffusion model. arXiv preprint arXiv:2305.11413. https://arxiv.org/abs/2305.11413
11. Zhang, Y., Li, X., & Wang, J. (2023). Multimodal emotion recognition: Techniques and applications. IEEE Transactions on Affective Computing, 14(2): 321–335. https://doi.org/10.1109/TAFFC.2022.3151234
12. Singh, A., & Kumar, R. (2023). Deep learning approaches for speech emotion recognition: A review. Journal of Ambient Intelligence and Humanized Computing, 14(1): 15–29. https://doi.org/10.1007/s12652-022-03872-0

13. Park, S., & Choi, J. (2023). Diffusion models for high-fidelity facial expression generation: An overview. Pattern Recognition Letters, 167: 91–100. https://doi.org/10.1016/j.patrec.2022.10.016
14. Ahmed, S., & Rana, N. (2023). Ethical challenges in AI-driven emotion recognition systems. AI & Society, 38(3), 689–703. https://doi.org/10.1007/s00146-023-01517-6
15. Huang, T., & Wang, Q. (2023). Cross-cultural emotion recognition: Data, models, and evaluation. Frontiers in Psychology, 14: 1145032. https://doi.org/10.3389/fpsyg.2023.1145032
16. Torres, M., & Rodríguez, A. (2023). Multitask learning for emotion and sentiment analysis in conversational AI. Information Processing & Management, 60(4): 103049. https://doi.org/10.1016/j.ipm.2023.103049
17. Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., ... & Zhang, W. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. arXiv preprint arXiv:2203.06935. https://arxiv.org/abs/2203.06935
18. Hajarolasvadi, N., Ramírez, M. A., & Demirel, H. (2020). Generative adversarial networks in human emotion synthesis: A review. arXiv preprint arXiv:2010.15075. https://arxiv.org/abs/2010.15075
19. Poria, S., Majumder, N., Mihalcea, R., & Hovy, E. (2019). Emotion recognition in conversation: Research challenges, datasets, and recent advances. arXiv preprint arXiv:1905.02947. https://arxiv.org/abs/1905.02947
20. Rouast, P. V., Adam, M. T. P., & Chiong, R. (2019). Deep learning for human affect recognition: Insights and new developments. arXiv preprint arXiv:1901.02884. https://arxiv.org/abs/1901.02884
21. Dhall, A., Goecke, R., Joshi, J., Wagner, M., & Gedeon, T. (2014). Emotion recognition in the wild challenge. In Proceedings of the 16th International Conference on Multimodal Interaction (pp. 461–466). ACM.
22. D'Mello, S., & Calvo, R. A. (2013). Beyond the basic emotions: What should affective computing compute? In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (pp. 1–8). IEEE.
23. Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., ... & Patras, I. DEAP: A database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing, 2011; 3(1): 18–31.
24. Tao, J., & Tan, T. (2005). Affective computing: A review. In International Conference on Affective Computing and Intelligent Interaction (pp. 981–995). Springer.

25. Picard, R. W. (1997). Affective computing. MIT Press.

Data Sources

1. EIN Presswire. (2024, February 19). Emotion detection and recognition market to reach USD 149.08 billion by 2032, driven by AI and industry adoption. https://www.einpresswire.com/article/781988401/emotion-detection-and-recognition-market-to-reach-usd-149-08-billion-by-2032-driven-by-ai-and-industry-adoption
2. Grand View Research. (2023). Emotion detection and recognition market size, share & trends analysis report. https://www.grandviewresearch.com/industry-analysis/emotion-detection-recognition-market-report
3. Mordor Intelligence. (2023). Affective computing market - growth, trends, COVID-19 impact, and forecasts (2023–2028). https://www.mordorintelligence.com/industry-reports/affective-computing-market
4. Wissen Market Research. (2023). AI-based emotional recognition software market - growth opportunities and forecast. https://www.wissenmarketresearch.com/reports/ai-based-emotional-recognition-software-market
5. Market Research Future. (2023). Emotion detection/recognition market research report - global forecast to 2030. https://www.marketresearchfuture.com/reports/emotion-detectionrecognition-market-3193
6. Global Growth Insights. (2023). Artificial intelligence emotion recognition market analysis and forecast. https://www.globalgrowthinsights.com/market-reports/artificial-intelligence-emotion-recognition-market-100138
7. Market.us. (2023). Emotion AI market - trends, drivers, and forecasts. https://market.us/report/emotion-ai-market/
8. Huang, M., Zhu, X., & Gao, J. (2020). Challenges in building intelligent open-domain dialog systems. ACM Transactions on Information Systems (TOIS), 38(3): 1–32. https://doi.org/10.1145/3383125
9. Zhou, H., Young, T., Huang, M., Zhao, H., Xu, J., & Zhu, X. (2020). Dialogue generation: From imitation learning to inverse reinforcement learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 34(05): 9141–9148. https://doi.org/10.1609/aaai.v34i05.6444
10. Zhao, T., Xie, K., & Eskenazi, M. (2020). Generating emotionally relevant responses for open-domain conversation. In Proceedings of the 58th Annual Meeting of the Association

for Computational Linguistics (pp. 123–132). https://doi.org/10.18653/v1/2020.acl-main.12

11. Yoon, S., Ko, H., & Jung, K. (2018). Multimodal speech emotion recognition using audio and text. In Proceedings of the IEEE Spoken Language Technology Workshop (pp. 112–118). https://doi.org/10.1109/SLT.2018.8639626
12. El Ayadi, M., Kamel, M. S., & Karray, F. Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition, 2011; 44(3): 572–587. https://doi.org/10.1016/j.patcog.2010.09.020