International Journal of Computer Techniques – Volume 12 Issue 4, July - August - 2025

AFFECTIVE COMPUTING AND EMOTIONAL INTERACTION: BRIDGING THE HUMAN-MACHINE EMOTIONAL GAP

Soumya Rishin

Department of Computer Science, Wadihuda Institute of Research and Advanced Studies, Vilayancode-670504, Kannur, Kerala

Corresponding Email: research@wiraskannur.com

Abstract

The goal of the multidisciplinary field of affective computing is to make it possible for machines to perceive, understand, and react to human emotions. Human-computer interaction (HCI) is becoming increasingly dependent on emotional intelligence, and emotionally intelligent systems are changing how intelligent computers and humans interact. This study examines the fundamental ideas of affective computing, examines important enabling technologies like voice modulation, facial expression analysis, and physiological signal processing, and assesses its real-world applications in industries like customer service, healthcare, and education. The study also highlights new issues and potential paths forward, as well as ethical issues pertaining to emotion-aware systems. A paradigm change is marked by the incorporation of affective computing into AI systems, which will enable machines to behave in social situations and comprehend empathy more like humans.

Keywords: Affective Computing, Emotional Intelligence, Human-Computer Interaction

Introduction

The ability of artificial intelligence (AI) to mimic human thought processes has greatly improved in recent decades. However, the ability to understand and respond to human emotions a critical component of effective social interaction has only recently gained attention. In an attempt to bridge this gap by enabling robots to sense and understand emotional states, Rosalind Picard created affective computing in the 1990s. The importance of emotional contact in human-computer systems has grown as intelligent agents, chatbots, and robots have become more common. From therapeutic applications in healthcare to emotion-sensitive teaching systems in schools, affective computing is revolutionizing how people use technology. From rule-based expert systems to data-driven machine learning models, artificial intelligence (AI) has rapidly evolved. But there is still a significant flaw: machines cannot comprehend or react to human emotions.

Emotions are essential for social bonding, communication, and decision-making in human relationships. Researchers have looked for strategies to overcome this emotional gap in recognition of this constraint. Rosalind Picard first used the term "affective computing" in her groundbreaking 1997 study to describe the creation of computers that are able to recognize, understand, and react to human emotions (Picard, 1997). By allowing machines to demonstrate emotional intelligence and improving their social responsiveness, this invention represents a paradigm change in human-computer interaction (HCI). As AI systems become more prevalent in many facets of daily life, the necessity for emotional computing becomes increasingly apparent. Emotionally intelligent systems promise to enhance service quality, customisation, and user pleasure in a variety of fields, including healthcare, education, customer service, and autonomous cars.

Emotion-aware virtual assistants, for example, can recognize signs of impatience or distress in a user's speech and adjust their responses appropriately (McDuff et al., 2016). Real-time teaching strategy adaptation is possible in education using tutoring systems that detect boredom or bewilderment (D'Mello & Graesser, 2012). For groups that need more complex support, including the elderly, people with mental health issues, or students with special needs, these improvements are especially important. Technologically, affective computing draws upon a multitude of disciplines, including psychology, neuroscience, computer science, and cognitive science. It employs methods such as facial expression recognition, speech analysis, physiological monitoring, and natural language processing (NLP) to interpret affective states. Machine learning models are trained on annotated datasets to recognize patterns associated with specific emotions (Zeng et al., 2009).

Furthermore, advances in wearable technology and biosensors have facilitated the non-intrusive collection of emotional data in real-time environments, expanding the scope and applicability of affective systems across different contexts and cultures (Kim & Andre, 2013). However, the creation and application of emotional computing bring up difficult moral and societal issues. It is necessary to address issues with consent, privacy, bias in emotion identification algorithms, and the possible exploitation of emotional data. Building confidence in emotion-aware systems requires ensuring openness and user autonomy (Buolamwini & Gebru, 2018). Its effective incorporation into AI systems is a significant step toward creating socially intelligent, adaptive, and sympathetic machines that can engage in meaningful human contact.

Theoretical Foundations of Affective Computing

Emotion Representation Models

Affective computing is based on an understanding and simulation of human emotions. Emotions are multifaceted psychological states that include expressive acts, physiological reactions, and subjective experiences. Classifying these emotional states and comprehending how they emerge and differ among people and cultures are both necessary for effective emotion detection. As a result, several theoretical models have been put out to direct the development of systems that are sensitive to emotions.

Paul Ekman's notion of universal emotions is among the most important frameworks. Ekman claimed that the six primary emotions such as happiness, sorrow, anger, fear, surprise, and disgust are universally represented through facial expressions and are biologically innate (Ekman & Friesen, 1971). Numerous early affective computing systems that mostly depended on facial expression analysis to identify emotional states were built on this concept. Real-world uses for Ekman's findings include user sentiment analysis, security checks, and lie detection.

The dimensional model of emotions is another popular strategy. This paradigm depicts emotional states within a continuous area delineated by two fundamental axes: valence (pleasant to unpleasant) and arousal (low to high intensity), rather than classifying emotions into distinct groups. One well-known example is Russell's Circumplex Model of Affect, which plots feelings like as enthusiasm and boredom according to where they are located in this two-dimensional space (Russell, 1980). When creating systems that must react to different levels of emotional intensity, this approach is especially helpful.

By focusing on cognitive evaluation processes, appraisal theories of emotion provide a more nuanced viewpoint in addition to these models. Individual evaluations of events based on criteria like novelty, goal relevance, and coping capability lead to emotions, according to Klaus Scherer's Component Process Model (Scherer, 2005). According to this idea, emotion is a subjective perception shaped by experience and context rather than just a response to external stimuli. This emphasizes the significance of contextual awareness in emotion recognition algorithms for affective computing.

Every one of these models makes a distinct contribution to affective computing. Dimensional and appraisal models provide frameworks for comprehending the underlying complexity and dynamism of emotional experience, but Ekman's model makes it easier to classify things discretely using observable clues. Multiple-model hybrid systems are becoming more and more popular because they enable developers to capture both the obvious and subtle facets of emotional expression.

Emotion models also need to be flexible enough to accommodate individual, language, and cultural differences. What one culture views as a sign of happiness or respect may be very different from another. Informed by cross-cultural psychology and trained on a variety of datasets, affective computing systems are better equipped to provide inclusive and accurate emotional assessments. This gives emotion-aware systems an additional level of complexity and richness. To sum up, the development of technology that can identify and react to human affect depends heavily on theoretical models of emotion. These models offer the vocabulary and structure required for machines to meaningfully perceive human emotions, whether through dimensional mapping, cognitive assessment, or discrete categorization. These fundamental notions will continue to be essential to the development of emotional computing.

Human Emotion as a Multimodal Signal

Body language, verbal intonation, facial expressions, and physiological cues are just a few of the channels that interact simultaneously to portray emotions. Because emotional expressions can alter in intensity and clarity depending on the situation, research has shown that depending solely on one modality might be inadequate or deceptive (Calvo & D'Mello, 2010). Consequently, the robustness and dependability of emotion identification systems are improved by combining input from many modalities.

Computer vision techniques are frequently used to capture the rich supply of emotive information found in facial expressions. In the meantime, vocal characteristics such as tone, rhythm, and pitch offer supplementary information about emotional states. A layer of nonverbal communication is added by body posture and gestures, which can support or contradict spoken and facial messages (Pantic & Rothkrantz, 2003). When combined, these techniques provide a more comprehensive understanding of an individual's emotional experience.

Emotional analysis is further enhanced by physiological reactions including electroencephalography (EEG), galvanic skin response, and heart rate variability. These signals can offer objective proof of internal affective states and are frequently involuntary. For instance, alterations in EEG patterns can show cognitive and emotional engagement, whereas increased skin conductance levels may suggest heightened arousal (Kim & André, 2008).

Multimodal data fusion techniques are being used more and more by affective computing systems to process and combine these many emotional inputs. Early fusion, or feature-level integration, and late fusion, or decision-level integration, are components of traditional fusion techniques. More complex fusion designs, such multimodal transformers and attention-based networks, which dynamically assess the significance of each modality, have been made possible by recent developments in deep learning (Zadeh et al., 2018).

The synchronization and contextual alignment between modalities are just as important to the efficacy of multimodal systems as the quality of the data. Recognition performance might be impaired by asynchronous signals or mismatched interpretations. To overcome these obstacles, contextual embeddings and temporal alignment algorithms are used to maintain the semantic coherence of multimodal input (Baltrušaitis, Ahuja, & Morency, 2019). Ultimately, a more precise and sympathetic comprehension of human emotions is made possible by the integration of multimodal data. Affective computing systems are better equipped to function in real-world situations where human emotional expression is complex, dynamic, and context-dependent because to this all-encompassing approach. Emotions can be expressed through a variety of indicators, such as body language, posture, gestures, voice intonation, and facial expressions. Affective computing systems commonly use multimodal data fusion to improve the accuracy of emotion recognition.

Core Technologies and Methodologies

Facial Expression Analysis

One common use of computer vision methods in emotional computing is the analysis of face microexpressions. Even when people try to repress them, these fleeting and involuntary micro-expressions frequently disclose true emotional states. By extracting hierarchical spatial characteristics from facial data, Convolutional Neural Networks (CNNs) have shown a high degree of accuracy in recognizing these subtle facial cues (Ko, 2018). CNN-based models are used for real-time face expression recognition in both the open-source toolkit OpenFace and the for-profit emotion analytics platform Affectiva. These systems are trained on large annotated datasets with thousands of facial photos tagged with emotional expressions, including the Extended Cohn-Kanade Dataset (CK+), FER2013, and AffectNet (Mollahosseini, Hasani, & Mahoor, 2017). Additionally, the Facial Action Coding System

(FACS), created by Ekman and Friesen in 1978, is a key component of many facial expression detection systems.

It breaks down facial movements into action units associated with particular emotions. Deep learning combined with FACS-based analysis allows for more accurate and culturally sensitive emotion detection. Advances in 3D facial modeling and facial landmark detection also help modern systems monitor facial movements more accurately in a variety of head postures and illumination situations. Widespread applications in fields including market research, in-car emotion recognition systems, and mental health monitoring have resulted from these advancements. All things considered, facial expression analysis is still a fundamental component of affective computing and keeps developing with the help of more advanced AI techniques.

Voice Emotion Recognition

Analyzing auditory cues to determine a speaker's emotional state is known as voice emotion recognition, or VER. Prosodic elements including pitch, speed, timbre, intensity, and rhythm are used in speech to convey a wide range of emotions. The physiological alterations linked to emotions such as tension, excitement, or melancholy are reflected in these vocal cues. Pitch contours, spectral energy, and Mel-Frequency Cepstral Coefficients (MFCCs) were among the many acoustic properties that were extracted by early systems. Convolutional neural networks (CNNs) have been used to spectrogram-based representations as deep learning has advanced, enabling more efficient feature extraction straight from unprocessed audio data (Trigeorgis et al., 2016).

The sequential pattern of speech can be well-modeled by recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) models, which capture emotional fluctuations across time. By learning contextual and temporal correlations between speech signals, attention-based processes and transformer models such as wav2vec 2.0 have improved emotion recognition in more recent times (Baevski et al., 2020). Large emotional speech datasets like IEMOCAP, RAVDESS, and CREMA-D are used to train these techniques, which allow for the recognition of a variety of emotional states, including neutrality, fear, anger, and happiness.

Physiological Signal Processing

An individual's internal emotional state can be inferred from physiological cues. Since these signals are frequently involuntary, they are especially helpful for identifying emotions since they might show reactions that might not be communicated through words or facial expressions. Heart rate (ECG), skin conductance (GSR), brain activity (EEG), respiration rate, and body temperature are examples of physiological indicators that are frequently utilized. The autonomic nervous system affects these measures, which show shifts in arousal, stress, or relaxation brought on by emotional experiences. Continuous, real-time data gathering in controlled and naturalistic environments has been made possible by the increasing availability of wearable biosensors and mobile health monitoring equipment, such as smartwatches and EEG headbands (Kim & André, 2008).

Physiological signal processing holds potential, but there are still a number of obstacles to overcome. Individual differences in emotional reactions can be significant, and physiological cues are frequently obscured by movement, surroundings, or health. Machine learning algorithms must therefore be resilient to unpredictability and able to identify minute patterns in a variety of datasets. To increase the accuracy of emotion categorization, methods like hybrid classifiers, deep learning models, and support vector machines (SVMs) have been used (Koelstra et al., 2012). Furthermore, it has been demonstrated that combining physiological data with additional modalities like vocal analysis or facial expressions improves the effectiveness of emotion identification algorithms. Physiological signal processing is being more widely used in domains such as biofeedback therapy, immersive gaming, stress detection, and adaptive user interface design.

Text-Based Sentiment Analysis

A key element of affective computing is text-based sentiment analysis, which focuses on deciphering emotional content in spoken or written language. Techniques from Natural Language Processing (NLP) are used to examine textual sentiment, subjectivity, polarity, and context. To identify affective states, traditional systems used bag-of-words models and rule-based sentiment lexicons, but they frequently had trouble with sarcasm, ambiguity, and context dependency. Word embeddings like Word2Vec and GloVe, which record semantic associations between words and increase the accuracy of sentiment identification across a variety of textual formats like emails, chat transcripts, and social media postings, are now used in modern NLP pipelines (Mikolov et al., 2013).

Text sentiment and emotion identification have been transformed by the creation of deep learning-based transformer models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer). To comprehend the complex emotional tone of words, including idiomatic and implicit expressions, these models make use of contextual embeddings and attention mechanisms (Devlin et al., 2019). Complex affective states like sarcasm, disappointment, and anticipation can now be distinguished in emotion classification tests. Text-based sentiment analysis has applications in a number of fields, such as conversational AI, hate speech identification, customer feedback assessment, and mental health monitoring. Continuous improvements in domain-specific fine-tuning and pre-trained language models have made NLP-based emotion recognition increasingly sensitive, accurate, and flexible.

Applications of Affective Computing

Healthcare

By improving patient monitoring, facilitating early mental health issue detection, and bolstering therapeutic interventions, emotion-aware technology are revolutionizing the healthcare industry. These tools include physiological monitoring, speech pattern recognition, and facial expression analysis to evaluate emotional health. Artificial intelligence (AI)-powered devices are able to monitor mood swings, identify abnormalities in emotional patterns, and give doctors immediate feedback. According to research, robotic companions like PARO can help individuals with depression or dementia feel less alone, less anxious, and more emotionally engaged (Tapus, Mataric, & Scassellati, 2009). Thus, emotionally intelligent systems support more comprehensive and adaptable medical care.

Education

Affective computing makes it possible to create emotionally intelligent learning systems in educational settings that can adjust to each student's unique demands. Using information from body position, speech inflections, and facial expressions, Intelligent Tutoring Systems (ITS) use emotion detection to evaluate learner states including engagement, boredom, or perplexity. In order to re-engage the student, these systems react by changing the speed of instruction, offering encouraging feedback, or displaying other content. Affective systems promote a positive learning environment, enhance retention, and encourage improved academic achievement by identifying and reacting to students' emotional cues (D'Mello & Graesser, 2012).

Customer Service

Through more efficient and sympathetic encounters, emotionally intelligent virtual assistants are transforming customer service. To determine a customer's attitude in real time, these systems examine speech patterns, language sentiment, and tone of voice. Based on this study, when AI agents identify distress, they can modify their dialogue technique to become more proactive, apologetic, or solution-focused. This capacity to tailor interactions increases client pleasure, facilitates more efficient problem-solving, and cultivates enduring brand loyalty. Sentiment-aware interfaces are already being used by businesses like IBM and Amazon to improve customer satisfaction and lower attrition (McDuff et al., 2016).

Automotive Industry

Affective computing is being used by the automotive sector to enhance driver involvement, comfort, and safety. In order to determine the driver's emotional state, in-car emotion recognition systems track physiological information, eye contact, voice tension, and facial expressions. The car can react by sending out notifications, recommending breaks, or changing the cabin's heat, lighting, or music when it detects signs of tension, exhaustion, or distraction. These solutions improve the in-car experience while reducing the risk of accidents. The creation of emotionally intelligent and adaptable autonomous cars depends heavily on emotion-aware technologies (Healey & Picard, 2005).

Ethical and Social Considerations

Deep ethical questions are brought up by robots' increasing emotional intelligence, particularly in relation to consent and privacy. Emotional information, including vocal tones, body language, and facial emotions, is extremely intimate and illuminating. Such data can be misused and violate personal autonomy if it is collected illegally or covertly. Strict data governance procedures are necessary as wearables, smart gadgets, and internet platforms integrate affective computing systems into everyday life. Adherence to regulations such as the General Data Protection Regulation (GDPR) guarantees that users are fully informed and provide their express consent prior to the access, analysis, or storage of their emotional data.

Bias and fairness are also important issues. The absence of ethnic variety in the datasets used to train emotion detection algorithms can lead to misunderstandings or prejudiced results. For instance, facial expression models that have been mostly trained on Western populations might not be able to identify emotional signs in people from other ethnic or cultural backgrounds. Disparities in hiring, law enforcement, and education may result from this. Algorithmic fairness audits, transparent model development, and inclusive datasets are necessary to address these issues and guarantee that systems function fairly for a variety of groups (Buolamwini & Gebru, 2018).

Another major ethical worry is the possibility of emotional manipulation. Emotion-aware systems could be used for manipulative reasons in political campaigns, advertising, or behavioral nudging if they are utilized to identify user vulnerabilities like grief, fear, or exhaustion. Platforms might, for example, modify content to increase user engagement at the expense of mental health. Affective computing systems need to be transparently and responsibly developed in order to reduce these hazards. Users should be informed about the timing and purpose of the use of their emotional data. By demystifying the system's decision-making process, explainable AI (XAI) features can encourage user trust and moral implementation.

Challenges and Future Directions

Affective computing still faces enduring obstacles that limit its widespread applicability, despite tremendous progress. Emotional generalization is one of the most prominent challenges. distinct demographics, such as age, culture, gender, and personality attributes, have quite distinct emotional responses. Emotions in a variety of real-world contexts are frequently difficult for models trained on small or homogeneous datasets to read correctly. A smile, for instance, could signify happiness in one cultural setting but politeness or uneasiness in another. Large-scale, culturally diverse datasets and adaptive learning frameworks which enable systems to gradually adapt to unique behavioral patterns are essential components of affective computing in order to overcome this.

Context awareness is another crucial concern. Situational context is fundamental to emotions; the environment, social interactions, past experiences, and even the time of day influence how emotions are expressed and understood. Affective systems may misinterpret emotional cues in the absence of this context, leading to inappropriate or unproductive reactions. To increase accuracy, contextual data must be integrated using sensors, memory models, and user histories. These complex dependencies may be captured by new methods in temporal sequence modeling and context-aware computing, such as attention-based networks.

Multimodal fusion is a complex yet crucial aspect of affective computing. Human emotions are frequently expressed through a range of indicators, such as language, body language, voice inflections, and facial expressions. However, there are technological hurdles in merging these inputs into a solid and unified emotional inference, particularly when modalities clash or are captured asynchronously. In order to grasp cross-modal links and balance input sources based on consistency and dependability, researchers are currently investigating deep fusion techniques, graph-based models, and transformers. Multimodal integration will need to be mastered in order to create systems that can understand complicated emotions.

Lastly, the effective implementation of emotional systems in interactive contexts like telemedicine, adaptive learning platforms, and virtual assistants depends on real-time processing. The difficulty is in creating algorithms that run on devices with limited resources while being accurate and quick. To address these needs, solutions including edge computing, on-device AI, and model pruning are being investigated. Furthermore, improvements in wearable emotion-tracking technology, closer integration with conversational AI frameworks, and hybrid AI architectures that combine deep learning and symbolic reasoning are likely to be the focus of future study. These approaches seek to develop morally sound, effective, and genuinely human-centered systems that are emotionally aware.

Conclusion

An important turning point in the development of artificial intelligence is represented by affective computing, which moves beyond simply logical computers to ones that can comprehend and react to human emotions. We open the door to more organic, significant, and compassionate human-technology interactions by integrating emotional intelligence into machines. This development has significant potential for use in domains where success depends on an understanding of human affect, such as driverless vehicles, tailored customer service, adaptive education, and mental health. However, there are several difficulties in incorporating affective computing into day-to-day activities. It is imperative that ethical issues pertaining to algorithmic bias, permission, data privacy, and the possibility of emotional manipulation be rigorously and openly addressed. The technical challenges, such as the requirement for correct multimodal fusion, real-time processing, and contextual sensitivity, are as significant. In addition to technology innovation, multidisciplinary cooperation amongst ethicists, psychologists, designers, and legislators is necessary to address these problems. In the future, emotional computing holds promise for creating systems that promote social awareness, empathy, and trust in addition to functionality. These systems need to be able to identify human emotions and react in a morally and appropriately responsible manner.

Acknowledgement

Authors extend the heartfelt gratitude to the Dr. Vineethkumar V., Research Director, Directorate of Research, Wadihuda Institute of Research and Advanced Studies, for the continuous encouragement and support during the preparation of the Manuscript.

References

Baevski, A., Zhou, H., 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. https://arxiv.org/abs/2006.11477

Baltru-aitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423-443.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 81*, 1-15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1*(1), 18-37.

Cho, K., van Merri⎿nboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint* arXiv:1409.1259. https://arxiv.org/abs/1409.1259

Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing, 5*(1), 4-7. https://doi.org/10.1007/s007790170019

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational*

Linguistics: Human Language Technologies, 1, 4171-4186. https://doi.org/10.18653/v1/N19-1423

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145-157. https://doi.org/10.1016/j.learninstruc.2011.10.001

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems, 6*(2), 156-166. https://doi.org/10.1109/TITS.2005.848368

Kim, J., & Andr, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(12), 2067-2083. https://doi.org/10.1109/TPAMI.2008.26

Kim, J., & Andr, E. (2013). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(7), 1325-1335. https://doi.org/10.1109/TPAMI.2011.274

Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors, 18*(2), 401. https://doi.org/10.3390/s18020401

Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., ... & Patras, I. (2012). DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing, 3*(1), 18-31. https://doi.org/10.1109/T-AFFC.2011.15

Lopatovska, I., & Arapakis, I. (2011). Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction. *Information Processing & Management, 47*(4), 575-592. https://doi.org/10.1016/j.ipm.2010.09.001

McDuff, D., El Kaliouby, R., & Picard, R. W. (2016). Applications of affective computing in automotive user interfaces. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-8). https://doi.org/10.1145/3003715.3005461

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint* arXiv:1301.3781. https://arxiv.org/abs/1301.3781

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing, 10*(1), 18-31. https://doi.org/10.1109/TAFFC.2017.2740923

Pantic, M., & Rothkrantz, L. J. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE, 91*(9), 1370-1390.

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

Tapus, A., Mataric, M. J., & Scassellati, B. (2009). Socially assistive robotics [Grand challenges of robotics]. *IEEE Robotics & Automation Magazine, 14*(1), 35-42. https://doi.org/10.1109/MRA.2007.339605

Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. (2016). Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5200-5204). IEEE. https://doi.org/10.1109/ICASSP.2016.7472617

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems, 30*, 5998-6008. https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Zadeh, A., Liang, P. P., Poria, S., Cambria, E., & Morency, L. P. (2018). Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 2236-2246.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39-58. https://doi.org/10.1109/TPAMI.2008.52