



Building Trust: The Sentient AI Framework for Emotionally Intelligent AI

¹Jugal Gajjar, ²Sanjana Nathani

¹Teaching Assistant, ²Teaching Assistant

¹School of Engineering and Technology,

¹Navrachana University, Vadodara, India

Abstract: Artificial Intelligence (AI) is the driving force behind most of the applications we use in our daily lives, and this necessitates advancements in human-AI interaction to go beyond basic functionalities. Through this article, we want to propose a novel methodology, the Sentient AI Framework (SAIF), that prioritizes the integration of emotional intelligence in AI systems, which makes them work in the better interest of humans by interpreting and responding while keeping human emotions into consideration. By integrating SAIF with intelligent systems like chatbots, virtual assistants, and robots, user interactions can be made more natural and engaging. This article discusses how SAIF can be developed and deployed to cultivate a sense of connection based on emotions such as trust and empathy, paving the way for the future where AI can be easily integrated into society and policy-making processes. Moreover, it discusses the key ethical aspects such as privacy, bias, explainability, and transparency to be considered. By developing such emotionally intelligent systems in collaboration with scholars, ethicists, and policymakers, we can ensure the ethical development and utilization of SAIF, thereby enhancing well-being and cultivating a better society.

Index Terms - Artificial Intelligence, Emotional Intelligence, Human-AI Interaction, Trustworthy Artificial Intelligence, Sentient AI Framework.

I. INTRODUCTION

The use of Artificial Intelligence (AI) in our daily lives is becoming more and more evident. From the moment we wake up, our mobile phones curate personalized news feeds based on our interests [1], to the intelligent navigation systems suggesting the quickest route to work [2], AI subtly enhances our everyday lives. Even seemingly tedious tasks like sorting emails as spam or creating new music playlists for our workout depend on the decision-making abilities of intelligent systems. But as AI progressively integrates into our personal, professional, and social lives, a critical question emerges: can we entirely trust these AI systems?

AI has immense potential to enhance efficiency, accuracy, and productivity in various domains, even healthcare [3], yet concerns regarding the fairness and transparency of these complex algorithms and their ability to genuinely comprehend human needs and values persist in society. An example of bias in AI systems was revealed in research where facial recognition software employed by a law enforcement agency displayed racial discrimination, raising worries about equity and bias in AI applications [4]. Such events throughout the course of the AI era highlight the significance of inducing trust in AI mechanisms. Users are more likely to trust these intelligent systems if they believe that they work by taking human best interests into consideration. Herein comes the importance of human sentiment. Our emotions considerably influence our interactions with our fellow beings, and AI is no exception [5]. The ability of AI systems to understand and respond according to human sentiments is key to establishing and maintaining trust.

This article attempts to address a critical question: “How can AI systems be developed to be more emotionally intelligent, creating a deeper trust and connection with humans and enabling the successful integration of trustworthy AI systems into our society?” Through the proposal of the Sentient AI Framework (SAIF), this theoretical study aims to narrow the gap between AI and human sentiment.

II. UNDERSTANDING HUMAN SENTIMENT

For making advancements in the field of trustworthy AI, studying human emotions is crucial. As discussed earlier, trust plays an important role in successfully integrating AI into our society. Hence, AI systems need to be able to understand human emotions in order to build trust among users. Psychology and philosophy are two well-established frameworks that can be utilized to better understand the intricate spectrum of human emotions [6].

Paul Ekman, an American psychologist, proposed a theory that identified six basic emotions: happiness, sadness, anger, fear, surprise, and disgust [7]. These six emotions were considered universal and are expressed through facial expressions and/or physiological changes regardless of cultural background. However, these emotions are usually linked with other emotions and develop into more nuanced feelings based on cultures, personal experiences, and situational context.

Similarly, Robert Plutchik, a distinguished professor and psychologist, developed a model mapping relationships between different human emotions. His model, a wheel of emotions, displayed how primary emotions merge to form secondary emotions [8]. This model is considered crucial for AI in understanding these emotional dynamics to accurately recognize human sentiments.

With a proper understanding of emotions, we can determine how they lead to human actions and also affect decision-making processes [9]. Antonio Damasio, a Portuguese neuroscientist, suggested that emotions act as “somatic makers,” leading to decisions based on past emotional experiences [10]. For example, a positive experience can establish a brand by instilling a sense of security and accepting new product concepts and information. On the other hand, a negative experience may lead to frustration and eventually distrust, impacting future purchasing choices.

Research in the field of Human-Computer Interaction (HCI) highlights the importance of emotions in interactions between humans and AI systems. Studies demonstrate how users tend to develop emotional connections with computers, particularly with software, reacting positively to user-friendly interfaces [11]. On the other hand, complex interfaces are often considered frustrating, which makes a user lose interest in interacting with such interfaces. These findings place emphasis on the importance of emotional intelligence in AI systems.

Recent research shows how users start to form an emotional bond with their virtual assistants, chatbots, and robots during their interactions. These emotional connections tend to heavily influence user behaviour and satisfaction. A study revealed that deceptive behaviour by an intelligent system decreases human trust, regardless of its physical or virtual representation [12]. These findings underscore the impact emotions have on HCI, thereby showcasing the increasing importance of developing an emotionally intelligent AI to cultivate trust and positive sentiments through interactions.

of five years. The time series monthly data is collected on stock prices for sample firms and relative macroeconomic variables for the period of 5 years. The data collection period is ranging from January 2010 to Dec 2014. Monthly prices of KSE -100 Index is taken from yahoo finance.

III. THE SENTIMENTAL CHALLENGE FOR AI

Rapid advancements in the field of AI have resulted in various advantages in different industries. However, a critical hurdle remains in understanding and responding to human sentiments in a nuanced and ethical manner. Current AI systems often struggle to capture the intricate details of human emotions, causing human-AI interactions to be robotic, disinterested, and eventually untrustworthy [13]. This section discusses the constraints of AI in recognizing human emotions, the bias in the data and algorithms, and the ethical considerations associated with the potential manipulation of emotions.

One of the major existing challenges for AI is the complex nature of human emotions. Many of the modern algorithms exhibit adeptness in recognizing the eight basic facial expressions [14], but detecting the emotional undercurrents accurately based on the given situation still remains a difficult challenge. This happens due to the dynamic nature of human sentiments, which are heavily based on context, personal backgrounds, and cultures, which makes it difficult to make an emotional inference from a single face expression or voice note [15]. For example, a high-pitched voice may indicate anger in a serious conversation but convey excitement during an interesting sports match. Moreover, emotions are often expressed subtly through micro-expressions [16] or body language [17], which AI systems might entirely miss.

Apart from basic sentiment detection, the generation of effective responses for addressing our emotional needs and values poses another significant hurdle. AI systems which are solely trained on factual or scientific data might struggle to the intricacies of personal experiences. For example, a client contacts the customer service to complain about a faulty appliance, and the AI chatbot, trained on company policies, product details, and FAQs, provides these details in an artificial, monotonous tone. The lack of emotional intelligence leads to failing to recognize or tackle the emotional distress happening due to a defective product, which ultimately gives rise to a breakdown in trust and a sense of detachment among customers. These feelings may leave consumers feeling neglected and undervalued, resulting in increased frustration.

Bias in existing datasets and algorithms poses another significant challenge. The datasets used for training the AI models often reflect the biases of dataset creators, leading to discriminatory outcomes. For instance, a study revealed that a loan approval AI system trained on historical data that is biased against minority groups could unknowingly extend the racial bias, resulting in more and more rejections of loan requests from those minorities [18]. Such biases can cause significant emotional consequences for users, leading to the creation of negative emotions like frustration, disappointment, and even despair.

Moreover, an ethical dilemma lingers around the AI's potential to manipulate human emotions. Malicious parties can scrupulously craft AI systems to exploit emotional vulnerabilities to influence consumer behaviour and sway public opinions. This raises critical concerns regarding the manipulative capabilities of AI and its influence on user autonomy [19]. A concerning scenario in this modern era involves the usage of AI on social media platforms to personalize content feeds such that it intensifies existing political or social divisions. By capitalizing on users' emotional biases, these platforms can create "filter bubbles" [20] and "echo chambers" [21], further dividing users and limiting their exposure to diverse perspectives.

IV. SENTIENT AI FRAMEWORK

With continuous research into improving human-AI interactions, we propose the Sentient AI Framework (SAIF) to make AI systems more emotionally intelligent and to cultivate deeper emotional connections between humans and AI systems. SAIF is designed to have a multi-layered approach for understanding and responding to human emotions accurately using the fundamental concepts of emotional intelligence (EQ), empathic design, human-in-the-loop sentiment analysis, and affective computing.

At the core of the SAIF architecture is a combination of technologies used to meticulously design an AI system for interpreting human emotions. The first step is emotion recognition, implemented with careful integration of various AI methodologies. For systems with microphonic capacity, context-aware speech processing using deep learning serves as an optimal method for recognizing the sentiments embedded within the voice data of the surroundings [22] and converting speech to text for further processing [23]. Techniques like Natural Language Processing (NLP), particularly advanced models with Transformer-based architectures (e.g., BERT, GPT), can be used to extract emotional cues from the textual data [24]. Moreover, computer vision algorithms such as Convolutional Neural Networks (CNNs) or Vision Transformers (ViTs) can be used to analyse facial expressions and body language for emotion recognition from visual inputs [25, 26, 27]. Additionally, physiological data like heart rate variability and respiration rate, collected through cameras [28] or wearable devices [29], can offer further insights about the user's emotional state. By integrating a combination of these technologies, or even the entire suite when feasible, strategically in the design of human-AI interactions, the effectiveness of recognizing human sentiments can be significantly improved.

Following emotional recognition, the SAIF framework transitions to the emotion comprehension phase, integrating knowledge from psychology and affective computing to interpret identified emotional cues. This crucial stage involves associating recognized emotions with underlying psychological concepts and understanding the situational context in which these emotions manifest. For example, as discussed earlier, frustration expressed during a customer service interaction could be linked to unmet expectations or technical challenges. By meticulously examining interaction history and contextual clues, the AI system can deduce the reasons behind the user's emotional state and tailor its responses accordingly. Machine learning algorithms such as decision trees, Bayesian networks, Recurrent Neural Networks (RNNs), and attention models can be utilized to model the complex interplay between emotional cues, contextual data, and psychological constructs. Through training on labelled emotional response datasets with contextual information, these models can continuously refine their ability to accurately comprehend and interpret human emotions.

The final phase of the SAIF framework, emotion regulation, draws upon insights obtained from research in affective computing [30] and human psychology to formulate responses that effectively regulate user emotions and foster positive interactions. This phase requires the adaptive modification of tone, language, and content in AI responses based on the user's emotional state and the intended outcome of the interaction. For example, when a user expresses frustration or discontent, AI may utilize empathy and active listening

strategies to acknowledge the user's emotions and provide solutions or assistance. Conversely, when users display positive emotions such as joy or contentment, the AI system can utilize positive reinforcement or acknowledgment to enhance those emotions. Reinforcement learning algorithms, like Deep Q-Networks (DQNs) [31] or Policy Gradient methods, enable the AI system to learn effective emotion regulation strategies through trial and error. This continuous learning process ultimately enhances the overall quality of user interactions over time.

Apart from its technical complexities, SAIF integrates empathetic design principles to prioritize user-centric experiences. Explainability and transparency concerning the emotional logic of the AI are crucial for establishing trust with users. Explainable AI (XAI) methods can be effectively incorporated into SAIF so that users get to understand how their emotions are interpreted and how these interpretations influence the AI's responses.

SAIF can be configured to generate justifications for its emotional interpretations. This may involve highlighting specific characteristics or cues (such as facial expressions, word selections, physiological data points) that have led the AI to recognize a particular emotion. XAI techniques like Local Interpretable Model-Agnostic Explanations (LIME) [32] can be especially beneficial in this context. LIME can generate tailored explanations for each user interaction, providing users with a clear insight into the rationale behind the AI's emotional evaluation. Furthermore, XAI can provide developers counterfactual explanations, essentially presenting "what-if" scenarios [33]. For example, the AI could demonstrate how a user's emotional state might be perceived differently if they had used alternative wordings or displayed different body language during the interaction. This method can deepen our understanding of the intricacies involved in SAIF's emotional cue interpretation process. Moreover, XAI can enable a feedback loop between users and the AI system. Users can make corrections or provide additional context regarding their emotional state during interactions. This feedback promotes a collaborative synergy between AI and human judgment, which can be harnessed to refine the AI's emotional comprehension and enhance the explanations it provides in subsequent interactions.

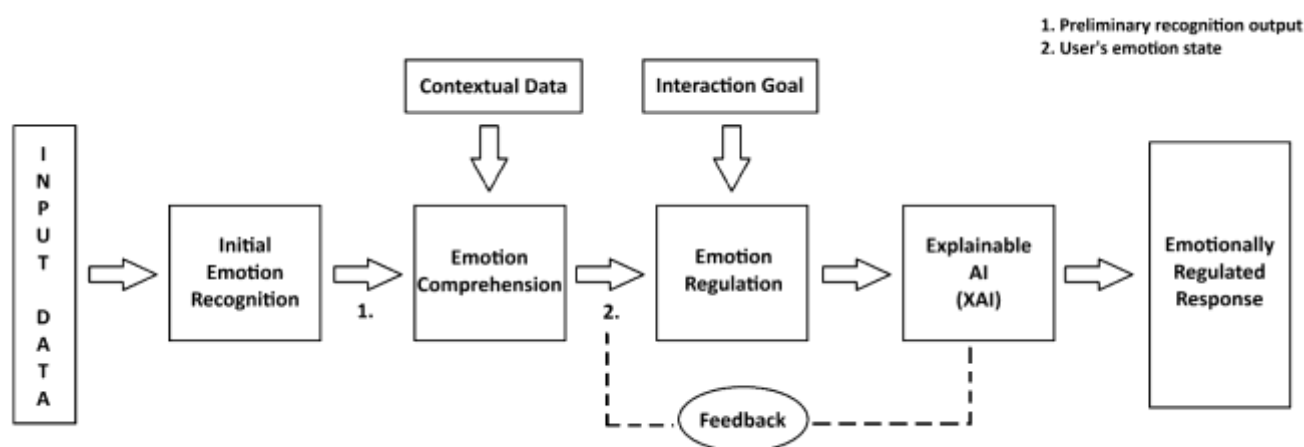


Fig. 1 SAIF Framework's multi-layered approach

In order to attain optimal performance, the SAIF framework requires a thorough and diversified training dataset. This dataset must extend beyond the conventional method of annotated emotional expressions to include contextual cultural data. The integration of cultural knowledge can greatly improve the framework's capacity to accurately perceive and understand emotions among various user groups. Moreover, longitudinal user data is essential for comprehending individual emotional foundations and patterns of response. Through the examination of a user's prior interactions and emotional reactions, SAIF can formulate a personalized comprehension of the user's distinct emotional landscape. Physiological data, such as heart rate and respiration rate, can provide the AI system with a more holistic view of the user's emotional state [34], thereby enhancing the emotional identification process within the SAIF framework. Finally, the incorporation of biofeedback data, with appropriate user authorization, presents an invaluable supplementary aspect.

V. BENEFITS OF INTEGRATING SAIF

SAIF provides numerous advantages for the incorporation of AI systems into human society. By prioritizing emotional intelligence, SAIF promotes a paradigm shift in human-AI interaction, progressing beyond mere functionality towards a more natural and immersive experience. This framework can be smoothly integrated into current intelligent systems, leading to advancements in their capabilities.

For example, at present, chatbots often encounter difficulties in navigating the subtleties of human conversation. Through the incorporation of SAIF, chatbots can utilize emotional recognition to detect user frustration or confusion and dynamically adjust their communication style (e.g., showing empathy or providing clarification) based on the user's emotional state. Furthermore, SAIF enables chatbots to grasp the underlying reasons behind user emotions, enabling them to offer more precise support or display empathy. This shift towards emotionally intelligent chatbots has the potential to transform user experience and satisfaction in various domains, ranging from customer service to information retrieval.

Likewise, virtual assistants and social robots can greatly benefit by integrating SAIF. Imagine a virtual assistant that organizes your daily schedules and also identifies signs of stress and recommends relaxation techniques. Social robots equipped with SAIF can navigate social interactions more efficiently, interpreting emotional cues and responding with suitable social behaviours that foster trust with humans. These advancements offer promise for broader societal applications in healthcare, education, and elder care, where emotionally intelligent AI companions can offer invaluable support and companionship.

Furthermore, the impact of SAIF extends beyond common applications like chatbots. Social media analysis tools paired with SAIF can surpass sentiment analysis, understanding the emotional undercurrents of online posts. This capability is invaluable for assessing public opinion or detecting potential mental health issues. In the educational sector, intelligent tutoring systems integrated with SAIF can evolve into more than just evaluation tools. By recognizing a student's emotional state, SAIF empowers these systems to customize pedagogies and offer targeted assistance, fostering a more positive learning environment. Similarly, smart homes can be revolutionized by SAIF. Imagine a smart home that adjusts lighting or temperature based on your emotional state, creating a calming ambience or an energizing environment.

These are just a few instances of how SAIF's emotional intelligence can transform specialized AI systems, nurturing a deeper and more meaningful interaction between humans and AI.

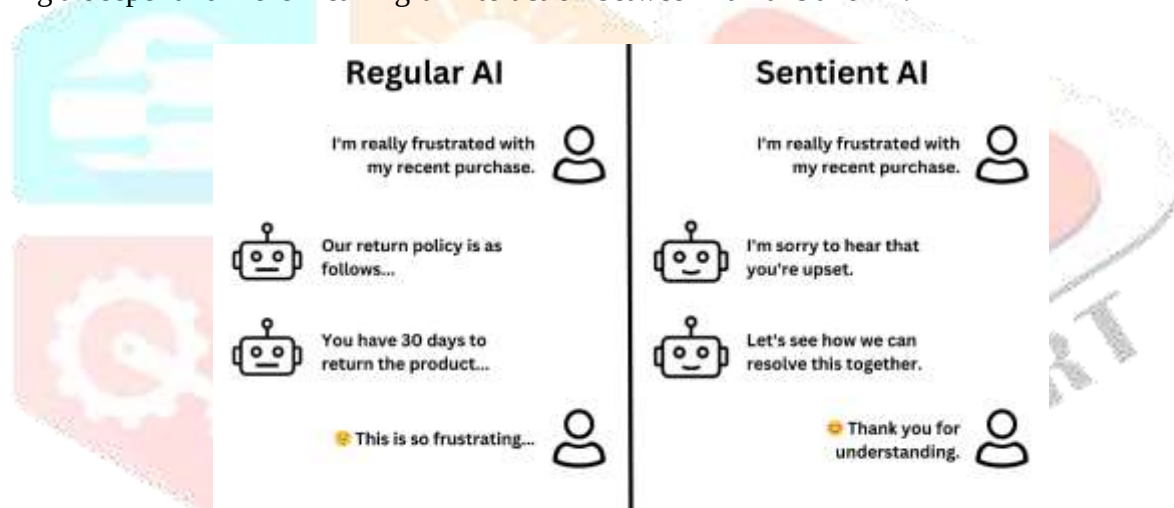


Fig. 2. Comparison of Regular AI vs. Sentient AI Customer Service Responses

VI. ETHICAL CONSIDERATIONS

While SAIF offers significant potential, its advancement and application necessitate thorough examination of ethical considerations. Data security and privacy is of utmost importance as the analysis of emotional data involves sensitive information. Hence, it is important to strictly adhere to regulations on data protection, including user consent, anonymization, and secure storage methodologies.

Bias and fairness necessitate a comprehensive approach. Datasets used for training must exhibit diversity and representativeness to prevent the perpetuation of societal biases. Providing transparency in algorithms enables users to understand the interpretation of emotional cues, thereby fostering trust. Continuous monitoring for bias and the implementation of mitigation strategies are also indispensable.

Moreover, the possibility of manipulation underscores the need for protective measures. AI systems with emotional intelligence have the potential to be exploited for manipulating user emotions for malicious intents, such as targeted advertising or political influence. To address this, the integration of SAIF should prioritize empowering users, granting them control over emotional data and transparency in algorithmic decision-making processes.

In addition, SAIF should refrain from constructing patterns of user behaviour or responses aimed at eliciting specific emotional states from users to achieve predetermined outcomes. Engaging in such practices could undermine user autonomy and encourage manipulative tendencies. With this user behaviour in memory, repetitive responses in an emotional state can often backfire. User frustration with these responses should be treated as negative feedback for the AI, allowing it to learn and adapt in order to generate better

responses. Also, upholding ethical standards and emphasizing user well-being are crucial during the development and deployment of emotionally intelligent AI systems.

VII. FUTURE WORK

SAIF promotes the evolution of emotional AI, creating opportunities for further enhancements. Future studies may concentrate on enhancing emotional recognition to attain a more profound understanding of user emotions, encompassing the need to account for cultural variations in expressions and improve the integration of biofeedback data. Culturally-sensitive models play a vital role in guaranteeing precise recognition across diverse user cohorts. Moreover, the adoption of robust methodologies for integrating biofeedback data can offer a more comprehensive insight into user emotional states.

The ethical dimension holds utmost significance in the development of emotionally intelligent systems. Addressing bias in training data and upholding user privacy during emotional recognition are pivotal focal points for upcoming research. Scholars may direct their attention towards formulating strategies to rectify biases in training data and instituting user consent measures to safeguard privacy. These areas are fundamental in ensuring the conscientious evolution and implementation of emotionally intelligent AI.

VIII. CONCLUSION

The Sentient AI Framework (SAIF) attempts to enhance human-AI interactions by emphasizing that emotional intelligence is equally important as other intelligence quotients. This framework utilizes a combination of AI methodologies for emotional detection, comprehension, and regulation, leading to a more natural and immersive user experience. Through the successful integration of SAIF into current intelligent systems, such as chatbots, virtual assistants, and social robots; scholars have the opportunity to make advancements in various domains. SAIF promotes trust, empathy, and a sense of connection, creating the way for a future in which AI seamlessly merges with human society, not only as a tool but as a supportive and emotionally intelligent companion. Further research efforts directed towards refining emotional recognition, addressing ethical considerations, and exploring a wider range of application areas will establish SAIF's role as a key tool in shaping the future of dependable human-AI interaction.

IX. ACKNOWLEDGMENT

We are grateful to our families and friends for their unwavering encouragement and support.

REFERENCES

- [1] Han, K. 2020. Personalized news recommendation and simulation based on improved collaborative filtering algorithm. Complexity. 2020, 8834908.
- [2] Liu, X.-Y., Zhu, M., Borst, S., and Walid, A. 2023. Deep reinforcement learning for traffic light control in intelligent transportation systems. IEEE Transactions on Network Science and Engineering. XX, XX, XXX.
- [3] Miotto, R., Wang, F., Wang, S., Jiang, X., and Dudley, J. T. 2018. Deep learning for healthcare: review, opportunities and challenges. Briefings in Bioinformatics. 19, 6, 1236-1246.
- [4] Buolamwini, J., and Gebru, T. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. 77-91.
- [5] Jeon, M. 2023. The effects of emotions on trust in human-computer interaction: A survey and prospect. International Journal of Human-Computer Interaction.
- [6] Deigh, J. 2018. Concepts of emotions in modern philosophy and psychology. In From Psychology to Morality: Essays in Ethical Naturalism. 13-37.
- [7] Ekman, P. Basic Emotions. In Handbook of Cognition and Emotion, Dalgleish, T., and Power, M.J., Eds.
- [8] Plutchik, R. 2001. The nature of emotions. American Scientist. 89, 344-350.
- [9] Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S. 2015. Emotion and Decision Making. Annual Review of Psychology. 66, 799-823.
- [10] Damasio, R. 1994. The Somatic-Marker Hypothesis. In Descartes' Error: Emotion, Reason, and the Human Brain. G.P. Putnam, 165-201.
- [11] Gao, Z., and Huang, J. 2022. Human-computer interaction emotional design and innovative cultural and creative product design. Frontiers in Psychology. 13, 982303.
- [12] Rogers, K., and Howard, A. 2021. Intelligent agent deception and the influence on human trust and interaction. In Proceedings of the 2021 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO), Tokoname, Japan. 200-205.

- [13] Picard, R. W. 1997. *Affective computing*. MIT Press, Cambridge, MA.
- [14] He, Y., Zhang, Y., Chen, S., and Hu, Y. 2023. Facial expression recognition using hierarchical features with three-channel convolutional neural network. *IEEE Access*. 11, 84785-84794.
- [15] Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements. *Psychological Science in the Public Interest*. 20, 1, 1-68.
- [16] Tran, T.-K., Vo, Q.-N., Hong, X., Li, X., and Zhao, G. 2021. Micro-expression spotting: A new benchmark. *Neurocomputing*. 443, 356-368.
- [17] Reed, C. L., Moody, E. J., Mgrublian, K., Assaad, S., Schey, A., and McIntosh, D. N. 2020. Body matters in emotion: Restricted body movement and posture affect expression and recognition of status-related emotions. *Frontiers in Psychology*. 11, 1961.
- [18] Garcia, C. B., Garcia, M. G. P., and Rigobon, R. 2023. Algorithmic discrimination in the credit domain: What do we know about it? *AI & Society*. Advance online publication.
- [19] Ienca, M. 2023. On Artificial Intelligence and Manipulation. *Topoi*. 42, 833-842.
- [20] Pariser, E. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Group.
- [21] Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., and Starnini, M. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*. 118, 9, e2023301118.
- [22] Zhao, P., Liu, F., and Zhuang, X. 2022. Speech Sentiment Analysis Using Hierarchical Conformer Networks. *Applied Sciences*. 12, 16, 8076.
- [23] Zhang, Y., Han, W., Qin, J., Wang, Y., Bapna, A., Chen, Z., Chen, N., Li, B., Axelrod, V., Wang, G., et al. 2023. Google USM: Scaling automatic speech recognition beyond 100 languages. *arXiv preprint. arXiv:2303.01037*.
- [24] Gong, X., Ying, W., Zhong, S., and Gong, S. 2022. Text sentiment analysis based on transformer and augmentation. *Frontiers in Psychology*. 13.
- [25] Mehendale, N. 2020. Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*. 2, 446.
- [26] Ma, F., Sun, B., and Li, S. 2023. Facial expression recognition with visual transformers and attentional selective fusion. *IEEE Transactions on Affective Computing*. 14, 2, 1236-1248.
- [27] Stathopoulou, O., and Tsihrintzis, G. A. 2011. Emotion recognition from body movements and gestures. In *Intelligent Interactive Multimedia Systems and Services*, Tsihrintzis, G. A., Virvou, M., Jain, L. C., and Howlett, R. J., Eds. Springer.
- [28] Rahman, H., Ahmed, M. U., and Begum, S. 2016. Non-contact physiological parameters extraction using camera. In *Internet of Things. IoT Infrastructures. IoT360 2015*, Mandler, B. et al., Eds. Vol. 169, Springer. 448-453.
- [29] de Arriba-Pérez, F., Caeiro-Rodríguez, M., and Santos-Gago, J. M. 2016. Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios. *Sensors*. 16, 9, 1538.
- [30] Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., Zhang, W., and Zhang, W. 2022. A systematic review on affective computing: Emotion models, databases, and recent advances. *arXiv preprint. arXiv:2203.06935*.
- [31] Belo, P. R., Azevedo, H., Ramos, J. J. G., and Romero, R. A. F. 2022. Deep Q-network for social robotics using emotional social signals. *Frontiers in Robotics and AI*. 9, 880547.
- [32] Ribeiro, M. T., Singh, S., and Guestrin, C. 2016. "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery. 1135-1144.
- [33] Wolf, C. T. 2019. Explainability scenarios: Towards scenario-based XAI design. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. Association for Computing Machinery. 252-257.
- [34] Egger, M., Ley, M., and Hanke, S. 2019. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*. 343, 35-55.