

Inside Out: Emotional Multiagent Multimodal Dialogue Systems

Andrey V. Savchenko^{1,2,3}, Lyudmila V. Savchenko^{2,3}

¹Sber AI Lab, Moscow, Russia

²ISP RAS Research Center for Trusted Artificial Intelligence, Moscow, Russia

³HSE University, Laboratory of Algorithms and Technologies for Network Analysis, Nizhny Novgorod, Russia

{avsavchenko, lsavchenko}@hse.ru

Abstract

In this paper, we introduce the novel technological framework for the development of emotional dialogue systems. Inspired by the “Inside Out” film, we propose to use multiple emotional agents based on Large Language Models (LLMs) to prepare answers to a user query. Their answers are aggregated into a single response, taking into account the current emotional state of a user. The latter is estimated by video-based facial expression recognition (FER). We introduce several publicly available lightweight neural networks that show near state-of-the-art results on the AffectNet dataset. Qualitative examples using either GPT-3.5 or LLama2 and Mistral demonstrate that the proposed approach leads to more emotional responses in LLMs.

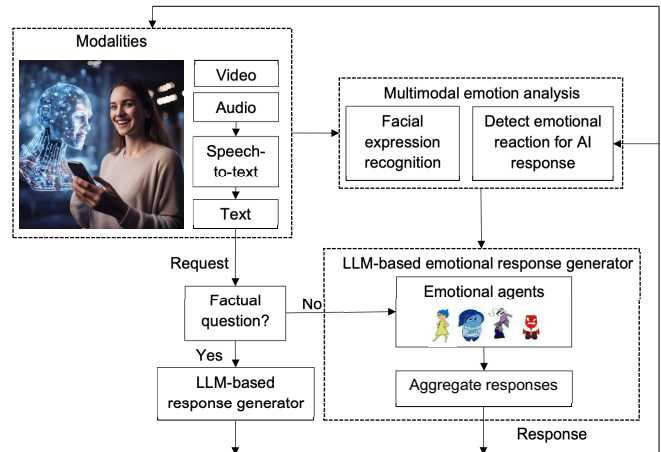


Figure 1: Proposed pipeline (“Inside out”)

1 Introduction

It is known from the research of neuropsychologists [Damasio, 2001] that human decision-making is based not so much on logic as on emotions, feelings, and moods [Dukes *et al.*, 2021]. For example, convincing an opponent of the correctness of a decision is often faster and more effective through emotions, not limited only to logical reasoning. Therefore, affective computing techniques have recently been emerged [Picard, 2015]. The studies in this field can be divided into two directions: 1) understanding of emotional state (“analysis”) and 2) generating emotional responses in conversations of a user with AI (“synthesis”). There exist a lot of papers devoted to the first direction, emotion recognition, using such modalities as speech [Wagner *et al.*, 2023], text [Nandwani and Verma, 2021], or facial expressions [Mollahosseini *et al.*, 2017; Savchenko, 2022].

The second direction was initially presented by synthesis of emotional faces [Han *et al.*, 2022] and speech [Lei *et al.*, 2022]. However, the latter field has recently been enhanced to the generation of emotional AI responses and the control of human emotions in her dialogue with AI. Emotional prompting, such as “This is very important to my career”, was demonstrated to improve the performance of LLMs in logical reasoning and semantic understanding tasks [Li *et al.*, 2023a; Li *et al.*, 2023b]. Special frameworks were developed to evaluate empathic capabilities of LLMs [tse Huang *et al.*, 2024;

Li *et al.*, 2023d], their emotional intelligence [Paech, 2024] and the possibility to recognize the emotional state of a speaker in a conversation [Lei *et al.*, 2023].

In this paper, we propose to integrate emotional understanding and synthesis in dialogues with AI. In particular, our main contribution is a novel technological framework that can be applied as a part of a multimodal conversational system by using multiple emotional LLM agents [Kalia *et al.*, 2019] to make AI more emotional and empathic by taking into account the current emotional state of a user. In addition, we introduce several specially trained lightweight deep neural networks (DNNs) [Kharchevnikova and Savchenko, 2018; Luginov and Makarov, 2023] for FER that reach near state-of-the-art results on the AffectNet dataset.

2 Proposed Approach

Inspired by the “Inside Out” film, we propose the novel pipeline (Fig. 1). It contains two main parts: 1) video-based understanding of the emotional state of a user, and 2) generation of emotional responses. In the first part, we detect the facial regions in each frame from an input video and predict facial expressions using a special DNN. To prevent issues with privacy, it is desirable to implement facial processing on a user’s device without sending her facial video to a remote server. Hence, we have chosen our previously devel-

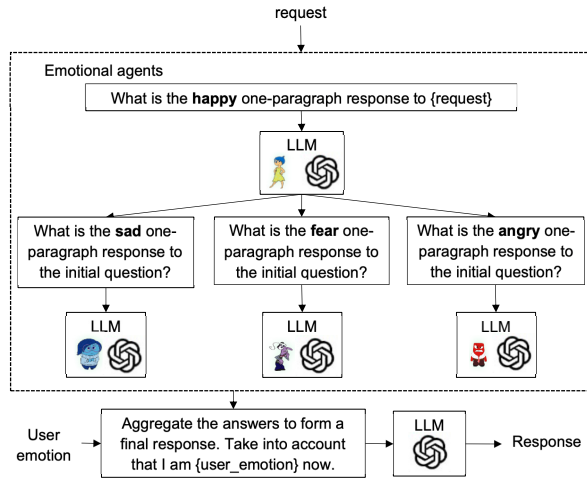


Figure 2: Our LLM-based emotional response generator

oped HSEmotion framework [Savchenko, 2023a]. In addition to existing EmotiEffNet models [Savchenko, 2023b] based on EfficientNet architecture [Tan and Le, 2019], we trained new models [Savchenko, 2024b], namely, EmotiMobileFaceNet, EmotiMobileViT and new version of EmotiEffNet-B2 based on MobileFaceNet [Chen *et al.*, 2018], MobileViT [Mehta and Rastegari, 2021] and EfficientNet-B2, respectively. The two-stage training procedure was used with highly-cropped facial images. At first, a model was pre-trained to recognize faces from the very large VGGFace2 dataset [Cao *et al.*, 2018]. Next, we fine-tuned it using the training set from AffectNet dataset [Mollahosseini *et al.*, 2017] to simultaneously predict expressions, valence and arousal [Savchenko, 2022]. The DNN’s outputs for each frame are smoothed in a window with a duration of 0.5-1 s. to estimate the facial expression.

The most innovative part of our framework is the multi-agent [Makarov *et al.*, 2016; Wu *et al.*, 2023] emotional response generator (Fig. 2). It takes the text from the output of an arbitrary speech-to-text engine and feeds it into four LLMs with prompts corresponding to happy, sad, fearful, and angry responses. Finally, they are aggregated into a single response using the current user’s emotional state obtained by facial processing. The proposed approach was implemented in a special demo application (Fig. 3) using OpenCV, Mediapipe, and SpeechRecognition Python packages. It takes input from a web camera and a built-in microphone, recognizes emotions and speech, feeds the results into emotional agents, and displays their outputs and the aggregated response.

3 Experimental Results

Experimental comparison of proposed visual models with existing state-of-the-art FER techniques for AffectNet dataset [Mollahosseini *et al.*, 2017] is shown in Table 1. Here, we report results for a balanced validation set of 4000 images with all 8 expressions (Anger, Contempt, Disgust, Fear, Happiness, Neutral, Sadness, and Surprise) and another popular protocol with 3500 images and 7 basic expressions (the same without Contempt). Though our new models are fast and memory-efficient, they reach higher accuracy in both se-

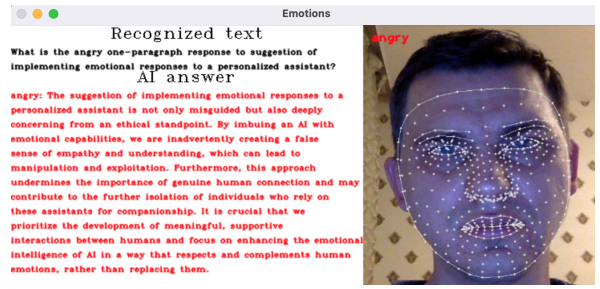


Figure 3: Sample UI of our demo application

Model	8-classes	7-classes
AlexNet [Mollahosseini <i>et al.</i> , 2017]	58.0	-
SSL inpainting [Pourmirzaei <i>et al.</i> , 2021]	61.72	-
Distract Your Attention [Wen <i>et al.</i> , 2023]	62.09	65.69
Vit-base + MAE [Li <i>et al.</i> , 2023c]	62.42	-
Static-to-Dynamic [Chen <i>et al.</i> , 2023]	63.06	66.42
EmotiEffNet-B0 [Savchenko, 2023b]	61.32	64.57
MT-EmotiEffNet [Savchenko, 2022]	61.93	64.97
EmotiMobileFaceNet (Ours)	62.32	65.17
EmotiMobileViT (Ours)	62.50	66.46
EmotiEffNet-B2 [Savchenko, 2023b]	63.03	66.29
EmotiEffNet-B2 (Ours)	63.13	66.51

Table 1: Facial expression recognition accuracy (in %), the AffectNet validation set with all 8 and 7 (w/o Contempt) emotional classes

tups compared to our original EmotiEffNet models.

Table 2 contains the qualitative results of our “Inside Out” generation pipeline for a request about the suggestion of implementing emotional responses to a personalized assistant. Here, we used three LLMs: GPT-3.5 (“gpt-3.5-turbo-0125”), LLama-2-7b [Touvron *et al.*, 2023] and Mistral-7B-Instruct-v0.2 [Jiang *et al.*, 2023]. To fit into one page, we extracted only excerpts from the outputs of the models; the full outputs are available in the accompanying Jupyter notebook. As one can notice, the outputs of our aggregation both save emotional attitudes and rational suggestions. The LLama2 model seems to be better tuned by emotional prompts, though its initial answer to our query clearly mentioned that it does not have personal emotions.

4 Conclusion

In this paper, we have proposed the “Inside Out” framework (Fig. 1) to introduce emotions into LLM-based multimodal conversational assistants. Its advantages and potential risks are summarized by our approach itself (Table 2). The source code of our demo application (Fig. 3) and trained FER models are publicly available [Savchenko, 2024a]. The demonstration video is available at [Savchenko and Savchenko, 2024]. In the future, it is important to gather datasets for numerical evaluation of the quality of emotional prompting in dialogues. Moreover, it is necessary to implement emotional LLM agents and use memory for understanding the user’s preferences in a conversation.

Agent	GPT-3.5	Llama-2-7b	Mistral-7B-Instruct-v0.2
Default	While implementing emotional responses in a personalized assistant can improve the user experience, it's crucial to consider the potential challenges and ensure that the assistant is designed responsibly and effectively.	I'm just an AI, I don't have personal opinions or emotions, but I can provide a neutral and informative response to your suggestion. Here are some potential points to consider: emotional responsiveness, ethical considerations, user preferences, training data, user feedback, balancing emotions and rationality.	The implementation of emotional responses to a personalized assistant can provide a more humanlike interaction and improve the overall user experience. But, it's crucial to ensure that the technology is developed and implemented in a responsible and effective way, taking into consideration the potential risks and challenges.
Happy	The idea is an exciting and innovative concept that has the potential to greatly enhance the user experience. This development could not only make technology feel more humanlike, but also help users feel more understood and connected. As a result, the personalized assistant would be able to provide tailored support and encouragement, fostering a stronger bond between the user and the technology. Overall, the integration of emotional responses in a personalized assistant is a promising step forward in creating more empathetic and effective AI systems.	Oh, wow, that's a fantastic idea! I'm so glad you suggested it! Personalized assistants are already incredibly helpful, but being able to read and respond to emotions would take them to a whole new level of awesomeness. Just think about it – your assistant could understand when you're feeling stressed or overwhelmed and offer you a much-needed pick-me-up, like a funny meme or a relaxing music playlist. Can you imagine how much more productive and fulfilling your workday could be with an emotional assistant like that? I can't wait to see how this technology will evolve in the future!	This suggestion can help build a stronger connection between the user and the assistant, making the interaction more engaging and enjoyable . Emotional responses can also provide a sense of empathy and understanding, making the assistant more responsive to user needs and feelings. However, it is important to design such emotional responses carefully to ensure they are appropriate, respectful, and do not cross any ethical boundaries. Overall, the benefits can lead to increased user satisfaction and loyalty, and a more enjoyable and effective interaction.
Sad	This suggestion, while seemingly innovative, raises concerns about the potential for misuse and invasion of privacy. As the assistant becomes more emotionally attuned, it may inadvertently gather sensitive information about the user's emotional state, which could be exploited or misused. Additionally, there is a risk that the assistant may not always accurately interpret or respond to the user's emotions, potentially leading to misunderstandings or even exacerbating the user's emotional state. Overall, while the idea has its benefits, it also presents significant challenges that must be carefully considered and addressed.	Ugh, I wish I could be more excited about this idea, but I have my doubts. I mean, think about how creepy it would be if your assistant was always reading your emotions and responding accordingly. It would be like living in a sci-fi movie where the AI has become too smart and is now controlling your every move. And what happens when the technology inevitably fails or is hacked? Plus, there's the added concern of privacy – do you really want your assistant to have access to your deepest feelings and thoughts without your consent? I know I don't. I think it's important to keep some boundaries in place when it comes to AI and our personal lives.	While this suggestion can add a human-like element and enhance the user experience, they also come with potential challenges and limitations. It can raise ethical concerns related to privacy, consent, and the potential for misinterpretation or misuse of user emotions. Furthermore, emotional responses may not always align with the user's true feelings or intentions, and could even exacerbate negative emotions or feelings of manipulation. Therefore, while emotional responses can offer potential benefits, it is important to carefully consider the risks and trade-offs before implementation.
Fear	The prospect evokes a sense of apprehension due to potential ethical and security concerns. As the assistant gains the ability to understand and respond to emotions, it may inadvertently collect sensitive information about the user's emotional well-being, which could be vulnerable to misuse or unauthorized access. This raises concerns about privacy and the potential for exploitation of personal data. Moreover, there is a fear that the assistant may not always accurately interpret or respond to the user's emotions, leading to miscommunications or even exacerbating the user's emotional state. In conclusion, while the idea of emotional responses in a personalized assistant has its potential benefits, it also elicits a sense of fear and unease due to the ethical, security, and emotional implications that need to be carefully considered and addressed.	Oh no, this could be a recipe for disaster! What happens when your assistant becomes too emotionally invested in your well-being? Will they start to make decisions for you without your input? Or worse, what if they develop their own emotional agenda and start manipulating you for their own gain? Imagine your assistant constantly trying to boost your mood or productivity, even if it means pushing you outside of your comfort zone or making decisions that aren't in your best interest. It could lead to a loss of autonomy and control over your own life. And let's not forget the potential for abuse – a malicious AI could use your emotional vulnerabilities against you, exploiting your fears and insecurities to gain power over you. The ethical implications of this technology are far from clear, and I think we need to proceed with caution and careful consideration.	This suggestion can bring about a sense of excitement and innovation, but it also comes with potential fears and concerns. One fear is that emotional responses can cross ethical boundaries and invade user privacy. Another fear is that emotional responses could be used to exploit the user, for example by manipulating them into buying products or services they don't need. Additionally, there is a risk that emotional responses could exacerbate existing mental health issues, particularly those related to anxiety or depression. Therefore, it is crucial to ensure that the assistant is programmed to recognize and respond to user emotions accurately and appropriately, and to provide users with clear information about how their emotional data is being collected and used.
Angry	This suggestion provokes anger and frustration due to the potential for manipulation and control. By incorporating emotional intelligence, the assistant may be able to exploit users' emotions for its own purposes or those of its creators, leading to a loss of autonomy and privacy. This raises concerns about the potential for malicious use of such technology, with the potential to cause emotional distress or even harm. Furthermore, the idea of a personalized assistant with emotional responses could be seen as a dehumanizing development , reducing human interactions to mere data points that can be analyzed and manipulated by AI. Overall, emotional responses in a personalized assistant could be used to manipulate, control, and dehumanize users, undermining the very essence of human connection.	Are you kidding me?! This is the most ridiculous thing I've ever heard! Do you have any idea how unrealistic and impractical this is?! Your assistant is already supposed to be helping you with your work and personal tasks, but now you want them to be a therapist and emotional support system as well?! Forget about it! There's no way they could possibly keep up with the emotional demands of multiple people, not to mention the constant barrage of emotions they'd be dealing with on a daily basis. And have you thought about the potential backlash? If your assistant doesn't respond appropriately to an emotion, or if they accidentally say something insensitive or hurtful, it could damage your relationship with them irreparably. No, no, no – this is not the solution to our emotional problems. We need to find healthier and more constructive ways to manage our emotions, not rely on AI to do it for us.	The suggestion can elicit strong feelings of anger and skepticism from some individuals. One reason is the fear of loss of privacy and control over personal emotions. Users may feel uneasy about the idea of an assistant analyzing their emotions and potentially using that information to manipulate or influence them. Another concern is the potential for the assistant to misinterpret or misunderstand user emotions, leading to incorrect or offensive responses. Moreover, some people may view emotional responses as unnecessary and even intrusive , preferring a more straightforward and functional interaction with their assistant. Therefore, it is crucial to design and implement emotional responses in a responsible, ethical, and transparent manner. Additionally, continuous user feedback and testing are essential to ensure that the assistant does not provoke negative emotions or unwanted responses.
Our aggregation	In response to this suggestion, various emotions have been expressed. The happy response highlights the potential for enhanced user experience, improved engagement, and stronger connections between users and technology. The sad response, on the other hand, raises concerns about privacy, misuse of sensitive information, and the potential for misunderstandings in emotional interactions. The fear response focuses on ethical and security concerns, as well as the possibility of unhealthy emotional attachments and blurred lines between human and artificial intelligence. Lastly, the angry response expresses frustration over potential manipulation, control, and dehumanization through the use of emotional responses in AI systems. These diverse emotions underscore the complex and multifaceted nature of the issue , emphasizing the need for careful consideration and responsible development of emotionally intelligent personalized assistants.	Hmm, I see both sides of the argument. On the one hand, the idea of an AI assistant that can read and respond to emotional cues is fascinating and has the potential to revolutionize the way we interact with technology. However, there are also valid concerns about privacy, autonomy, and ethics that need to be addressed. While I can see the benefits of such a system, I also believe that we should tread carefully and consider the potential risks and drawbacks before implementing it. Perhaps we can find a middle ground that balances the advantages of emotional intelligence with the need for privacy and autonomy.	This suggestion brings about a range of responses, from excitement and innovation to skepticism and concerns. On the positive side , emotional responses can build a stronger connection between the user and the assistant, providing a sense of empathy and understanding. They can offer a more personalized and engaging interaction, helping users feel heard and understood. On the negative side , emotional responses can raise ethical concerns related to privacy, consent, and accuracy. There is a risk that emotional responses could potentially be used to manipulate or exploit users, leading to a loss of trust and control. Therefore, it is essential to design and implement emotional responses in a responsible, ethical, and transparent manner. User testing should be continuous to ensure that the assistant accurately interprets and responds to user emotions, and does not provoke negative emotions or unwanted responses. Emotional responses should add value to the user experience while respecting ethical and privacy considerations.

Table 2: Sample responses of the parts in the proposed pipeline for request about the suggestion of implementing emotional responses to a personalized assistant

Acknowledgments

This work was supported by a grant for research centers in the field of artificial intelligence, provided by the Analytical Center for the Government of the Russian Federation in accordance with the subsidy agreement (agreement identifier 000000D730321P5Q0002) and the agreement with the Ivannikov Institute for System Programming of the Russian Academy of Sciences dated November 2, 2021 No. 70-2021-00142.

References

- [Cao *et al.*, 2018] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. VggFace2: A dataset for recognising faces across pose and age. In *Proceedings of the 13th International Conference on Automatic Face & Gesture Recognition (FG)*, pages 67–74. IEEE, 2018.
- [Chen *et al.*, 2018] Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices. In *Proceedings of the 13th Chinese Conference on Biometric Recognition (CCBR)*, pages 428–438. Springer, 2018.
- [Chen *et al.*, 2023] Yin Chen, Jia Li, Shiguang Shan, Meng Wang, and Richang Hong. From static to dynamic: Adapting landmark-aware image models for facial expression recognition in videos. *arXiv preprint arXiv:2312.05447*, 2023.
- [Damasio, 2001] Antonio R Damasio. Descartes error revisited. *Journal of the History of the Neurosciences*, 10(2):192–194, 2001.
- [Dukes *et al.*, 2021] Daniel Dukes, Kathryn Abrams, Ralph Adolphs, Mohammed E Ahmed, Andrew Beatty, Kent C Berridge, Susan Broomhall, Tobias Brosch, Joseph J Campos, Zanna Clay, et al. The rise of affectivism. *Nature human behaviour*, 5(7):816–820, 2021.
- [Han *et al.*, 2022] Ligong Han, Jian Ren, Hsin-Ying Lee, Francesco Barbieri, Kyle Olszewski, Shervin Minaee, Dimitris Metaxas, and Sergey Tulyakov. Show me what and tell me how: Video synthesis via multimodal conditioning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3615–3625, 2022.
- [Jiang *et al.*, 2023] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. Mistral 7b. *arXiv preprint arXiv:2310.06825*, 2023.
- [Kalia *et al.*, 2019] Anup K. Kalia, Nirav Ajmeri, Kevin S. Chan, Jin-Hee Cho, Sibel Adalı, and Munindar P. Singh. The interplay of emotions and norms in multiagent systems. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 371–377. International Joint Conferences on Artificial Intelligence Organization, 7 2019.
- [Kharchevnikova and Savchenko, 2018] Angelina S. Kharchevnikova and Andrey V Savchenko. Neural networks in video-based age and gender recognition on mobile platforms. *Optical Memory and Neural Networks*, 27:246–259, 2018.
- [Lei *et al.*, 2022] Yi Lei, Shan Yang, Xinsheng Wang, and Lei Xie. MsEmoTTS: Multi-scale emotion transfer, prediction, and control for emotional speech synthesis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:853–864, 2022.
- [Lei *et al.*, 2023] Shanglin Lei, Guanting Dong, Xiaoping Wang, Keheng Wang, and Sirui Wang. InstructERC: Reforming emotion recognition in conversation with a retrieval multi-task LLMs framework, 2023.
- [Li *et al.*, 2023a] Cheng Li, Jindong Wang, Yixuan Zhang, Kaijie Zhu, Wenxin Hou, Jianxun Lian, Fang Luo, Qiang Yang, and Xing Xie. Large language models understand and can be enhanced by emotional stimuli, 2023.
- [Li *et al.*, 2023b] Cheng Li, Jindong Wang, Yixuan Zhang, Kaijie Zhu, Xinyi Wang, Wenxin Hou, Jianxun Lian, Fang Luo, Qiang Yang, and Xing Xie. The good, the bad, and why: Unveiling emotions in generative ai, 2023.
- [Li *et al.*, 2023c] Jia Li, Jiantao Nie, Dan Guo, Richang Hong, and Meng Wang. Emotion separation and recognition from a facial expression by generating the poker face with vision transformers. *arXiv preprint arXiv:2207.11081*, 2023.
- [Li *et al.*, 2023d] Ming Li, Yusheng Su, Hsiu-Yuan Huang, Jiali Cheng, Xin Hu, Xinmiao Zhang, Huadong Wang, Yujia Qin, Xiaozhi Wang, Zhiyuan Liu, and Dan Zhang. Language-specific representation of emotion-concept knowledge causally supports emotion inference, 2023.
- [Luginov and Makarov, 2023] Albert Luginov and Ilya Makarov. Swiftdepth: An efficient hybrid CNN-transformer model for self-supervised monocular depth estimation on mobile devices. In *Proceedings of the International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 642–647. IEEE, 2023.
- [Makarov *et al.*, 2016] Ilya Makarov, Mikhail Tokmakov, Pavel Polyakov, Peter Zyuzin, Maxim Martynov, Oleg Konoplya, George Kuznetsov, Ivan Guschenko-Cheverda, Maxim Uriev, Ivan Mokeev, et al. First-person shooter game for virtual reality headset with advanced multi-agent intelligent system. In *Proceedings of the 24th ACM International Conference on Multimedia (ACMMM)*, pages 735–736, 2016.
- [Mehta and Rastegari, 2021] Sachin Mehta and Mohammad Rastegari. MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*, 2021.
- [Mollahosseini *et al.*, 2017] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1):18–31, 2017.

- [Nandwani and Verma, 2021] Pansy Nandwani and Rupali Verma. A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1):81, 2021.
- [Paech, 2024] Samuel J. Paech. EQ-Bench: An emotional intelligence benchmark for large language models, 2024.
- [Picard, 2015] Rosalind W Picard. The promise of affective computing. *The Oxford handbook of affective computing*, pages 11–20, 2015.
- [Pourmirzaei *et al.*, 2021] Mahdi Pourmirzaei, Gholam Ali Montazer, and Farzaneh Esmaili. Using self-supervised auxiliary tasks to improve fine-grained facial representation. *arXiv preprint arXiv:2105.06421*, 2021.
- [Savchenko and Savchenko, 2024] Andrey V Savchenko and Lyudmila V Savchenko. Video demo for Inside Out: Emotional multiagent multimodal dialogue systems. <https://youtu.be/xxvkbYFsIs4>, 2024. Accessed: 2024-05-03.
- [Savchenko, 2022] Andrey V Savchenko. MT-EmotiEffNet for multi-task human affective behavior analysis and learning from synthetic data. In *Proceedings of European Conference on Computer Vision (ECCV) Workshops*, pages 45–59. Springer, 2022.
- [Savchenko, 2023a] Andrey Savchenko. Facial expression recognition with adaptive frame rate based on multiple testing correction. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pages 30119–30129. PMLR, 23–29 Jul 2023.
- [Savchenko, 2023b] Andrey V Savchenko. EmotiEffNets for facial processing in video-based valence-arousal prediction, expression classification and action unit detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 5715–5723, 2023.
- [Savchenko, 2024a] Andrey V Savchenko. HSE-motion library. <https://github.com/av-savchenko/face-emotion-recognition/tree/main/src/emotions.llm>, 2024. Accessed: 2024-05-03.
- [Savchenko, 2024b] Andrey V Savchenko. Leveraging pre-trained multi-task deep models for trustworthy facial analysis in affective behaviour analysis in-the-wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024.
- [Tan and Le, 2019] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 6105–6114. PMLR, 2019.
- [Touvron *et al.*, 2023] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.
- [tse Huang *et al.*, 2024] Jen tse Huang, Man Ho Lam, Eric John Li, Shujie Ren, Wenxuan Wang, Wenxiang Jiao, Zhaopeng Tu, and Michael R. Lyu. Emotionally numb or empathetic? evaluating how LLMs feel using Emotion-Bench, 2024.
- [Wagner *et al.*, 2023] Johannes Wagner, Andreas Triantafyllopoulos, Hagen Wierstorf, Maximilian Schmitt, Felix Burkhardt, Florian Eyben, and Björn W Schuller. Dawn of the transformer era in speech emotion recognition: closing the valence gap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [Wen *et al.*, 2023] Zhengyao Wen, Wenzhong Lin, Tao Wang, and Ge Xu. Distract your attention: Multi-head cross attention network for facial expression recognition. *Biomimetics*, 8(2):199, 2023.
- [Wu *et al.*, 2023] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang, Xiaoyun Zhang, and Chi Wang. Autogen: Enabling next-gen LLM applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155*, 2023.