Assessing the use of Facial Emotion Recognition (FER) for consumer sentiment analysis

Michal Pluta - vsdc48
1609 words

***Abstract*— In this paper, we discuss the ethical considerations necessary for an emotion recognition system used for dynamic ad placement to exist. By conducting an Ethical Impact Assessment informed by Value Sensitive Design, we offer recommendations to ensure the system's implementation is minimally biased and reduces potential harm to stakeholder groups.**

1. Introduction

The first impression a consumer forms upon encountering an advertisement (ad) plays a pivotal role in determining its success in engaging their interest [1], [2]. By understanding how consumers emotionally connect with advertisements, brands can make informed decisions to improve the effectiveness of their marketing. This underscores the importance of analysing consumer sentiment.

The proposed solution is a machine-learning model capable of classifying facial expressions into distinct labels. The model will consist of three key mechanisms:

1. *Object detection & classification* - Identifying and distinguishing human faces within a visual input
2. *Gaze detection* - Detecting which people are currently looking at the advertisement
3. *Emotion classification* - Classifying the facial expression

The model will be trained, validated, and tested using a large, diverse range of human faces each labelled with an emotion. For an in-depth discussion of dataset choice and recommendations for implementation see Section 7.1.

The primary output of the model is an emotion distribution, indicating the model's confidence level in each emotion, similar to that depicted in Figure 1 below.

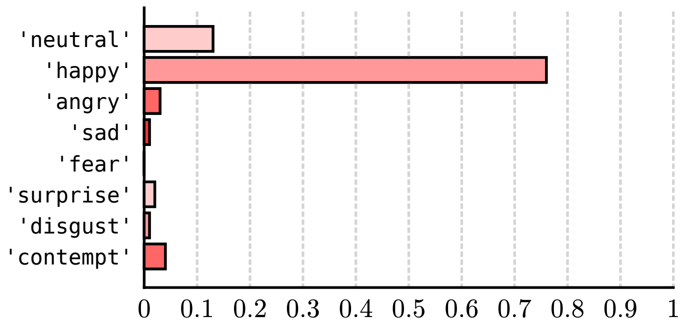

Figure 1: Sample model output distribution [3]

2. Applications of the system

While this system has a variety of potential uses in marketing, for example:

- *Market Research* - Analysing consumer sentiment in a product's development phase, to make small adjustments to the product.
- *A/B testing* - Presenting two versions of a product to a candidate, and measuring which one they react more positively to. This is particularly applicable to websites where most have roughly five seconds to make a lasting impression [2].

, we will focus on *Dynamic ad placement*.

In particular, consider high-traffic pedestrian areas such as bus shelters, train stations, or public squares which are littered with advertising screens.

Figure 2: Example of an advertisement at a bus shelter [4]

By leveraging emotion recognition and continuously surveying the general public through video feeds, advertising agencies can:

- Identify which ads are most well-received and in which areas
- Switch the displayed ad for a passerby if it detects a negative response to the currently displayed ad
- Sell the collected consumer sentiment data back to brands

3. Value Sensitive Design (VSD)

VSD is an approach to system design that systematically integrates human values into the design of technology [5]. It relies on an *"iterative, tripartite methodology consisting of conceptual, empirical, and technical investigations"* [5], [6].

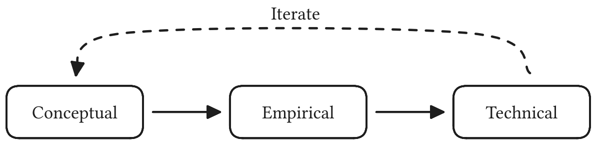

Figure 3: Diagram of the iterative nature of VSD

It consists of three integral investigations:

1. **Conceptual** - Identifying and analysing various stakeholder groups affected by the technology along with their values, rights, and principles to understand which values are important.
2. **Empirical** - Engaging with qualitative and quantitative research methods to gather insights from actual users and stakeholders.
3. **Technical** - Focusing on the design and development of the technology itself and exploring how values identified in the conceptual and informed in the empirical can be embedded in the technology's design.

More broadly, it is a methodology that seeks to identify stakeholders' root values [7], how the values of these stakeholder groups may conflict, and offer compromises or ways to resolve those conflicts in an unbiased way.

Two key benefits of this approach are:

- Human values are deeply rooted within the technology's design from the beginning which helps prevent negative consequences for users and society
- By considering the broader societal and ethical implications of a technology, VSD helps develop solutions that are resilient to changing societal norms. The approach not only addresses pre-existing biases but also proactively guards against the development of emergent bias [8].

By utilising the VSD framework, we conduct an Ethical Impact Assessment (EIA). This crucial step ensures that the implemented AI system aligns with values and principles that prioritise human rights, fairness, and transparency [9].

This proactive approach promotes trust in AI systems by identifying and mitigating potential biases and their potential harms.

4. Immediate Ethical Issues

Informed consent - As this technology is proposed as a passive data collection method, ensuring informed consent is complex, particularly for vulnerable groups (e.g. children) as they may not grasp the associated risks. This highlights the importance of consent mechanisms that respect individual autonomy.

Environmental sustainability - The technology is likely to be implemented city-wide with each location having its own camera and inference model, which is likely to consume a lot of energy. This issue is critical as AI already constitutes a large portion of energy consumption [10].

5. Ethical Impact Assessment

<table>
<tr><th>Stakeholder</th><th>Values</th><th>Potential risks / harms</th></tr>
<tr><td>Advertising Agency (Direct)
Serves as the project initiator and aims to collect data on individuals’ first impressions of advertisements and leverage machine learning algorithms to optimise the location, timing, and content of ad placements.</td><td>• Accuracy
Ensuring the data collected on people’s immediate reactions to advertisements is accurate is vital to understanding the true impact of the ads.
• Accountability
The agency must be held partially or fully accountable for the technology’s use, and potential misuse and consequences.
• Innovation
Advertising is a highly competitive and constantly evolving landscape.
By developing new technologies, the agency can uncover insights that competitors cannot, giving it a competitive edge in the market.
• Consumer Engagement
Engaging consumers effectively is crucial for the success of every advertisement [11].
Higher consumer engagement leads to a boost in consumer loyalty, higher consumer retention [11], and ultimately higher conversion rates [12].</td><td>• Innaccurate data
In recent years, Data-Driven Decision Making (DDDM) has seen increased adoption to help organisations reduce risk and take advantage of opportunities [13]. It does however have its downfalls where faulty data can lead to disaster. This was the case with some companies during the 2008 financial crisis where many data-driven models had hard-coded incorrect data [13] and hence extrapolated incorrectly.
• Inadequate security measures
The agency is responsible for storing consumer sentiment data “including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage” [14].
Failure to adequately protect this data may result in prosecution, civil lawsuits, and irreversible damage to the advertising agency’s reputation.</td></tr>
<tr><td>Brands (Direct)
Companies who approach the advertising agency with a product to advertise to the general public.</td><td>• Consumer Trust
The trust consumers have in a particular brand is directly related to their likelihood of purchasing products from that brand [15], [16]. Moreover, given that the likelihood of converting an existing customer ranges from 60-70%, compared to only 5-20% for new customers [17], it is clear brands will ensure consumers stay loyal.</td><td>• Reputation
If the technology disproportionately impacts/targets certain demographics, brands could face backlash by being associated with this marketing strategy, leading to a loss of consumer trust and reputation.
• Financial Loss
Bad-quality data gathered by the model can lead to incorrect insights, which can be detrimental to the brand’s image and customer engagement. Both of these can impact the brand’s market share and lead to financial loss.</td></tr>
<tr><td>Developers (Direct)
Employed by the advertising agency and tasked with creating the technology.</td><td>• Universal Usability
Developers are likely to prioritise usability by ensuring the technology is accessible to the widest possible audience, including individuals with disabilities, and people from a diverse range of backgrounds.</td><td>• Scapegoating
If the technology is misused by the agency, the developers may be used as scapegoats leading to most of the public backlash being targeted at the developers.</td></tr>
</table>

Stakeholder	Values	Potential risks / harms
	• **Freedom from bias** Valuing freedom from bias directly influences the integrity and fairness of the technology [18]. This responsibility is particularly significant given the increasing regulation around data ethics and AI fairness [19]. • **Quality product** The value placed on delivering a quality product/solution is closely linked to a developer's professional pride and technical reputation. Additionally, for developers, it is clear that if the implemented system matches or goes beyond expectations their career prospects can improve significantly.	
Consumers / Data Subjects (Indirect) Members of the general public who interact with the advertisements.	• **Autonomy** Within consumer choice contexts, autonomy is viewed as "(the) ability to make and enact decisions on (your) own, free from external influences" [20]. By allowing consumers to make independent choices that align with their personal preferences, needs, and values, they may experience higher levels of satisfaction. • **Relevant / Personalised ads** Enables ads to cater to specific interests making the shopping experience more efficient and enjoyable. Reduces the amount of irrelevant advertising noise and makes the experience feel tailored to the consumer.	• **Data Leakage** Personal consumer sentiment data stored in databases could be hacked and leaked. This data may contain sensitive or biometric data which poses a security concern as malefactors could use it for identity crimes. • **Data misuse** There is a risk of the agency misusing the data for its own benefit, such as by selling it to third parties, which is a direct violation of many data protection laws such as GDPR. • **Harmful advertisements** If the model is particularly biased, the ads it suggests may be harmful. Given the unique demographic composition of each location, the model may unintentionally internalise this information leading to harmful, biased model outputs.
Privacy advocates (Indirect) Organisations or activist groups advocating for the general public's privacy rights.	• **Privacy** These organisations consider privacy as a cornerstone of individual freedoms and societal well-being, consequently, they actively confront the development of technologies that encroach upon human privacy [21].	• **Legal and Political Pressure** These groups may feel pressure from the agencies that want to make this technology a reality regardless of ethical considerations. This is a potentially very powerful technology that could be used for national surveillance if not regulated appropriately.

Stakeholder	Values	Potential risks / harms
		• **Loss of trust** If the model is deemed to produce biased, harmful outputs and disregard the public's privacy rights then these groups may be blamed for not protecting the public regardless of whether they made any attempts.
Government, Regulatory bodies, and policy-makers (Indirect) Organisations regulating the development and usage of AI and facial analysis systems	• **Human Welfare** Governments have an ethical duty to ensure the well-being of their citizens, promoting policies that protect the vulnerable and support society as a whole. By prioritising welfare, it ensures social stability, avoiding social unrest. This is most commonly achieved by heavily regulating new technologies [22] and discussing pressing issues with a level of transparency [19]. • **Courtesy** Governments should be courteous in public discourse to promote a cooperative, respectful environment essential for democratic processes and effective implementation of policies. • **Environmental sustainability** Governments have an ethical and moral obligation to ensure companies act in an environmentally sustainable way such that inter-generational equity is not affected.	• **Public backlash** Governments may experience public backlash if the technology is used in invasive ways beyond what was originally specified, or for not interfering and regulating the technology early enough. Furthermore, the proposed technology employs machine learning and is intended to be installed across cities, which is likely to consume significant amounts of energy. This raises environmental concerns which the government may be held accountable for if the technology is not adequately regulated. This is regardless of whether the government is aiding in the creation of this technology.
General Public (Indirect)	• **Informed consent** Companies should not record, store, or analyse the general public's data without explicit informed consent. Failure to obtain such consent constitutes a breach of trust between the agency and the public. • **Privacy** Although there is reduced privacy in public settings compared to personal homes, the general public still retains the right to a certain degree of privacy while in public spaces.	• **Invasion of Privacy** The public may feel like they are constantly being recorded. This may feel like a dystopian society and is likely to affect the general public's welfare. Additionally, a large portion of the general public consists of young people and more specifically children who may be unaware of the risks associated with this sort of technology. This may make it difficult to acquire informed consent.

Table 1: Ethical Impact Assessment using VSD

6. Value Conflict Analysis

Individuals and corporations intrinsically value different things. Individuals often value autonomy, privacy, and security, seeking to protect their personal information and maintain control over their personal lives and decisions. Corporations, on the other hand, typically value profitability and efficiency, seeking to analyse all available data to improve performance.

This divergence in values is identified in Table 1 and cases such as the Cambridge Analytica scandal and the 2017 Equifax breach, underline the importance of resolving this conflict.

In this particular use case, the key conflict lies between the agency wanting consumers' facial and emotional data and the individuals wanting a degree of privacy and assurance that their data is adequately protected.

In order to resolve this conflict it is necessary to conduct empirical investigations to gather qualitative data on how each stakeholder feels towards the technology as well as quantitative data investigating suitable methods for:

- obtaining informed consent
- ensuring transparency through data collection and use
- implementing robust data protection measures

Importantly, there are two more existing conflicts between stakeholders:

1. Developers are likely to prioritise the safety and quality of the model, while the advertising agency may aim for rapid deployment at the expense of thorough testing and safeguarding. Here, prioritising the developer's value for safety is crucial, and suitable protocols should be established to never allow an unsafe model to be deployed for public use.

2. The government has a responsibility to protect public welfare and environmental sustainability, both of which the advertising agency may sacrifice in order to advance development. In this case, the government's values must be prioritised to ensure irreversible damage is not done by the agency and that all actions are heavily regulated.

7. Recommendations

Based on the insights uncovered by the EIA in Section 5 and Section 6, it is clear that many empirical investigations have to be undertaken before designing can begin.

Additionally, in order to create a model that is deemed 'fair' and 'unbiased', it is important to first contextually define these notions. Currently, there are at least 20 admissible definitions for AI fairness [23] including both statistical and causal definitions "with no clear agreement on which definition to apply in each scenario" [23].

We suggest that these notions should first be formally defined and guided by further empirical investigations into stakeholders' perceptions of 'fairness' and 'bias'.

*1. **Choice of Dataset***

As outlined in Section 1, the dataset should consist of human faces labelled with the corresponding emotion.

When selecting such a dataset, there are four key factors to consider:

1. *Accuracy* - Are faces labelled correctly?

2. *Completeness* - Do there exist unlabelled faces? If so, are the remaining faces still representative of the population? Are all demographics equally represented?

3. *Consistency* - Is the labelling consistent?

4. *Relevance* - Are there too many distinct labels? Are all the faces human?

Among the presently available datasets, AffectNet [3] stands out by being the largest collection of images and often being used as a benchmark. Despite this, it is ultimately extremely biased [24] through its over-representation of the white race.

Although diversity considerations were made during its inception, the images lack crucial demographic annotations [25] which makes it difficult to monitor technical bias during the training process. Furthermore, a recent study of AffectNet's underlying bias revealed that a given model's bias is unlikely to be caused by the dataset's bias. This is due to the mitigation strategy of balancing the dataset being unsuccessful as a result of pre-existing bias in the images.

This has led to the creation of face-attributed datasets like FairFace [26] which are labelled by gender, race, and age. Notably, FairFace is comparable in size to AffectNet ($n \approx 100000$), and is also race-balanced, meaning all 7 races identified have an equal representation [27]. However, a potential limitation of this dataset is the distribution is unlikely to align with the population the algorithm is being applied to, leading to lower accuracy.

*2. **Risk and bias mitigation measures***

- **Interpretability and Explainability** - In machine learning, models are often tasked with optimising an objective function, and while the model may be perfectly capable of doing this, it is often difficult for humans to gain insight [28]. This is the key benefit of interpretable models that allows humans to understand what caused the model to make a particular decision [29] and explainable models which involve creating a second model capable of explaining the choices made by the ML system [30].

By incorporating these measures into the design process, it promotes responsible development as the effects of changes made to the model can be verified by examining the decision process. This helps avoid unintentional addition of bias to the system. [31]

- **Continuous monitoring** - This mitigation measure is a long-term strategy that involves continuously monitoring the model's evaluation metrics, which can include accuracy, precision, recall, and f1-score for each demographic indicator (race, gender, age).

  The continuous monitoring of the model's behaviour means that model drift and degradation in performance can be identified and corrected [32]. This allows developers to quickly address the issue (or shut down the entire model) before the model can generate unfair or harmful outputs.

- **Diversity and inclusion** - By encouraging diversity within the development team and involving indirect stakeholders in discussions, the likelihood of identifying and challenging assumptions that could result in biased outcomes increases. This approach not only helps reduce bias and potential harm to end users, but also increases the system's fairness and equity.

3. *Critical Assessment and Limitations*

While these mitigation strategies are theoretically plausible, in practice they offer a multitude of limitations.

Interpretability for example offers a serious trade-off. Often, highly interpretable models may struggle with capturing complex relationships within data [33] (particularly applicable to image data due to its complexity), which can reduce the model's overall accuracy making it more susceptible to producing wrong outputs and harming users.

Moreover, while VSD is a well-studied framework, it is a theoretically grounded approach [5] which does not provide a clear way of embedding values into the design [7].

Bibliography

[1] AIContentfy, "The impact of first impressions in Customer Acquisition," *AIContentfy*, Feb. 2023, [Online]. Available: https://aicontentfy.com/en/blog/impact-of-first-impressions-in-customer-acquisition

[2] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. Wróbel, "Emotion Recognition and Its Applications," *Advances in Intelligent Systems and Computing*, vol. 300, pp. 51–62, 2014, doi: 10.1007/978-3-319-08491-6_5.

[3] "AffectNet Dataset." [Online]. Available: https://paperswithcode.com/dataset/affectnet

[4] [Online]. Available: https://www.outdooradvertisingltd.co.uk/advertising-services/bus-stop-advertising/

[5] B. Friedman, P. Kahn, A. Borning, P. Zhang, and D. Galletta, "Value Sensitive Design and Information Systems," 2006, p. . doi: 10.1007/978-94-007-7844-3_4.

[6] B. Friedman, D. G. Hendry, and A. Borning, "A Survey of Value Sensitive Design Methods," *Foundations and Trends® in Human–Computer Interaction*, vol. 11, no. 2, pp. 63–125, 2017, doi: 10.1561/1100000015.

[7] S. Umbrello, "Beneficial Artificial Intelligence Coordination by Means of a Value Sensitive Design Approach," *Big Data and Cognitive Computing*, vol. 3, no. 1, 2019, doi: 10.3390/bdcc3010005.

[8] B. Friedman and H. Nissenbaum, "Bias in Computer Systems," 2017, pp. 215–232. doi: 10.4324/9781315259697-23.

[9] Unesco, "Ethical impact assessment: A tool of the recommendation on the ethics of Artificial Intelligence." [Online]. Available: https://www.unesco.org/en/articles/ethical-impact-assessment-tool-recommendation-ethics-artificial-intelligence

[10] B. Bailey, "AI power consumption exploding." [Online]. Available: https://semiengineering.com/ai-power-consumption-exploding/

[11] "What is customer engagement? definition, benefits, and strategies for…." [Online]. Available: https://www.outreach.io/resources/blog/what-is-customer-engagement

[12] N. Sharma, "Why is customer engagement important: What are the benefits?." [Online]. Available: https://vwo.com/customer-engagement/why-is-customer-engagement-important/

[13] K. Heires, "Flaws in the data: In an ERA of data-driven decision-making, businesses must avoid the faulty models and algorithms that can poison their results," *Risk Management*, vol. 64, no. 3, p. 38–39, Apr. 2017.

[14] [Online]. Available: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/

[15] J. Dublino, "Returning customers spend 67 percent more than new customers." [Online]. Available: https://www.business.com/articles/returning-customers-spend-67-more-than-new-customers-keep-your-customers-coming-back-with-a-recurring-revenue-sales-model/

[16] [Online]. Available: https://www.salesforce.com/news/stories/customer-engagement-research/

[17] B. May, "Council post: New vs. repeat customers: Where should you focus your marketing?." [Online]. Available: https://www.forbes.com/sites/forbesagencycouncil/2022/07/08/new-vs-repeat-customers-where-should-you-focus-your-marketing/

[18] B. Friedman, "Value-sensitive design," *Interactions*, vol. 3, no. 6, pp. 16–23, Dec. 1996, doi: 10.1145/242485.242493.

[19] [Online]. Available: https://www.gov.uk/government/publications/cdei-publishes-briefing-paper-on-facial-recognition-technology/snapshot-paper-facial-recognition-technology

[20] K. Wertenbroch *et al.*, "Autonomy in consumer choice," *PubMed Central*, Jun. 2020, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7278769/

[21] C. J. Bennett, *The Privacy Advocates: Resisting the Spread of Surveillance*. The MIT Press, 2008. Accessed: Mar. 20, 2024. [Online]. Available: http://www.jstor.org/stable/j.ctt5hhfb6

[22] *ICO*, Oct. 2022, [Online]. Available: https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/10/immature-biometric-technologies-could-be-discriminating-against-people-says-ico-in-warning-to-organisations/

[23] S. Verma and J. Rubin, "Fairness definitions explained," in *Proceedings of the International Workshop on Software Fairness*, in FairWare '18. Gothenburg, Sweden: Association for Computing Machinery, 2018, pp. 1–7. doi: 10.1145/3194770.3194776.

[24] R. Raina *et al.*, "Exploring Biases in Facial Expression Analysis," in *NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research*, 2022. [Online]. Available: https://openreview.net/forum?id=jrddoq9NDP6

[25] I. Dominguez-Catena, D. Paternain, and M. Galar, "Assessing Demographic Bias Transfer from Dataset to Model: A Case Study in Facial Expression Recognition," G. Pedroza, X. C. Chen, J. Hernández-Orallo, X. Huang, H. Espinoza, R. Mallah, J. A. McDermid, and M. Castillo-Effen, Eds., in CEUR Workshop Proceedings, vol. 3215. CEUR-WS.org, 2022. [Online]. Available: https://ceur-ws.org/Vol-3215/12.pdf

[26] J. Joo, "Joojs/fairface: Fairface: Face attribute dataset for balanced race, gender, and age." [Online]. Available: https://github.com/joojs/fairface

[27] K. Karkkainen and J. Joo, "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1548–1558.

[28] A. Huang, J. Li, and N. Shankar, "6 - interpretability." [Online]. Available: https://blog.ml.cmu.edu/2020/08/31/6-interpretability/

[29] M. Schmitt, "Interpretable machine learning." [Online]. Available: https://www.datarevenue.com/en-blog/interpretable-machine-learning

[30] K. N. Vokinger, S. Feuerriegel, and A. S. Kesselheim, "Mitigating bias in machine learning for medicine," *Communications Medicine*, vol. 1, no. 1, p. 25–26, Aug. 2021, doi: 10.1038/s43856-021-00028-w.

[31] C. Molnar, "Interpretable machine learning." [Online]. Available: https://christophm.github.io/interpretable-ml-book/interpretability-importance.html

[32] M. Fayard, "What is AI monitoring and why is it important." [Online]. Available: https://coralogix.com/blog/ai-monitoring/

[33] W. Yang *et al.*, "Survey on Explainable AI: From Approaches, Limitations and Applications Aspects," *Human-Centric Intelligent Systems*, vol. 3, no. 3, pp. 161–188, Sep. 2023, doi: 10.1007/s44230-023-00038-y.