On Legal and Ethical Challenges of Automatic Facial Expression Recognition: An Exploratory Study

Alex Boutin, Lucie Lévêque, Sonia Desmoulin-Canselier

On Legal and Ethical Challenges of Automatic Facial Expression Recognition: An Exploratory Study

ALEX BOUTIN, Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, France
SONIA DESMOULIN-CANSELIER, Nantes Université, CNRS, Droit et Changement Social, France
LUCIE LÉVÊQUE, Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, France

Automatic facial expression recognition (FER) has a lot of potential applications. However, even if it can be beneficial for some areas, e.g. security and healthcare, several legal and ethical challenges arise. In this article, we first present such challenges related to the deployment of FER. Then, we introduce the conduct of a focus group which allowed to highlight interesting points regarding the use of FER in a medical context. Particularly, transparency, data management, diagnoses, liability, best endeavours obligation, and non-discrimination principle are debated. We finally discuss on our study's limitations and directions for future work.

CCS Concepts: • **Applied computing** → *Computers in other domains*; • **Social and professional topics** → *Intellectual property*.

Additional Key Words and Phrases: Emotions, artificial intelligence (AI), facial expression recognition (FER), law, ethics.

ACM Reference Format:
Alex Boutin, Sonia Desmoulin-Canselier, and Lucie Lévêque. 2023. On Legal and Ethical Challenges of Automatic Facial Expression Recognition: An Exploratory Study . In . ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Emotions play a key role in communication, as such that the quality of social interactions depend on how well we can identify others' [14]. To deduce peers feelings, we usually listen to what they actually say. Yet, voice tone, gestures, and facial expressions are the characteristics we analyse [18]. These features have consequently been studied in various domains, including computer science with the advent of affective computing [9].

Facial expression recognition (FER), the most studied modality of non-verbal expression of emotions, can be used in many areas including marketing, education, and security [6]. However, beyond cameras' potential intrusiveness, several works have shown that individuals see more drawbacks than benefits – from creating dependency to technology, to political manipulation [19]. As far as Europe is concerned, entities responsible for interpreting the general data protection regulation (GDPR) issued an opinion regarding uses of artificial intelligence (AI) linked to emotions [2]. They indeed called for a "ban on any use of AI systems designed to infer the emotions of a physical person". If some exceptions seem possible, especially "in medical fields where the recognition of emotions is important", they must be "subject to appropriate safeguards". The Council of Europe also acknowledged that "linking recognition of affect, for instance to the hiring of staff, or access to insurance or education, may pose risks of great concern, both at individual and society levels, and should be prohibited" [1].

Manuscript submitted to ACM

This article aims to provide insights into ethical and legal issues raised by automatic FER. General ethical and legal challenges linked to its deployment are introduced in Section 2. Section 3 presents an exploratory experiment carried out to better understand issues related to FER in the medical field, from the perspective of potential patients. Finally, Section 4 discusses limitations of our experiment and suggests future work.

2 LITERATURE REVIEW

2.1 Ethical challenges

FER systems require access to external states but are a gateway to inner experiences. They consequently provide access to qualia, i.e. "experiential properties of sensations, feelings, perceptions, thoughts and desires" [8]. Furthermore, FER systems, presenting beneficial aspects for health and safety, may also create new risks. For example, devices placed into cars to alert drivers for fatigue could reduce alertness; they could rely on technology and decrease their focus [13]. In an interview with candidates with similar skills but different emotional channelling, the job could be denied to someone unable to hide anxiety. Knowledge of ”vulnerable" consumers' emotions could enable brands to serve their interests. They might shape buyers' behaviour to make them pay for goods [**?**]. Potential risks concerns are for individuals and society; systems designed to target suspicious expressions in a crowd could lead to attribute negative emotions to people without bad intention [19]. Human and financial resources, like security officers or medical aids, could be unnecessarily mobilised, with financial consequences [19]. Another worry is linked to the likely manipulation of public opinion and human behaviour, as political opinions could be reached by detecting emotions on social networks [5].

Malicious uses of FER could violate fundamental rights and freedoms. Manipulating political intentions could infringe on the right to vote freely [19]. Individuals monitored in the workplace, or online, could suffer from privacy invasion [**?**]. To avoid FER restricting freedoms and individual development, prior collection of consent and respect of control possibilities are essential, but are they sufficient [10]? Nowadays, extensive regulations regarding consent collection apply, but terms and conditions can be misleading [**?**]. A study reported criticism of Facebook's manipulation of emotional data [15]. Projecting this to FER would make users concerned, as emotions are intimate and sensitive [**?**]. After consent being obtained, misuses of personal data can still be feared. Using FER in public spaces could be even more undesirable and invasive [23]. In addition, transparency and public debate are major topics [12], which seem crucial as cameras ubiquitous presence lead to uncertainties about data capture and use [24]. Definitions of ”good" and ”bad" FER uses should also be collectively debated [10].

2.2 Legal challenges

Collection and use of personal data fall under the GDPR scope in Europe. A new regulation is currently under discussion, considering FER as a "low-risk practice with transparency obligation".

For the moment, FER is therefore not precisely regulated, but subject to GDPR statements on legality, user information, right of objection, minimisation principle, etc. The technology deployment is also lead by guidelines of the Council of Europe, stating "private companies should not be allowed to use this technique in uncontrolled, freely accessible environments". Such guidelines make some uses presented earlier impossible. Moreover, the use of FER by law enforcement agencies is deemed "acceptable only if strictly necessary and proportionate to prevent imminent and substantial threats to public security". Finally, the idea is to ban FER for the purpose of "emotional recognition" and for identifying "personality traits, inner feelings, mental health or commitment of employees".

Even uses looking beneficial at first sight present potentially negative downsides, which the ban wants to avoid. Yet, mental health applications are different as they could actively contribute to social justice [19]. Abstaining from FER in the health field could even infringe the European Social Charter: "everyone has the right to benefit from all measures enabling them to enjoy the highest attainable standard of health" [4]. In France, the Public Health Code also claims "no person may be discriminated against in access to prevention or care" [3]. Indeed, alexithmic people (unable to feel emotions) would not be guaranteed equal treatment. Reconciling equality and access to FER as new form of care with respect of important legal principles could be a major challenge.

3 EXPLORATORY STUDY

Focus groups, an exploratory research method, facilitate the study of topics not clearly understood or accepted [25]. Amongst the main techniques to conduct focus groups, we used questioning, as it allows to observe individuals' perspective on a well-determined topic [21]. Questions were carefully written beforehand to identify issues that would not have been thought of, without influencing answers.

A group of homogeneous participants was selected [20], composed of ten French students (7 W, 3 M) aged 21 to 47. They are taught AI as well as acceptability of new technologies. Their knowledge of the medical sector was varied; some worked in this domain, some were former medical or psychology students.

Eleven questions were defined prior to the experiment, including general (e.g., "What do you think about FER?", "What would be possible contexts for its use?"), and specific questions ("What would be your conditions for allowing doctors to use FER on you or your children?", "What do you think about using FER to detect neurodevelopmental disorders or neurodegenerative diseases?"), not communicated in advance.

Before the start, participants were given an information sheet and had to give informed consent. The experiment took place in a room specifically designed for focus groups, with an oval table for twelve people, and visual and acoustic insulation. The focus group lasted 110 minutes. The entire session was recorded.

4 RESULTS

4.1 Algorithmic transparency and information

The algorithmic transparency issue was often raised by participants. One subject declared: "I would feel vulnerable if my doctor asked me to automatically recognise my emotions.". The possibility of medical errors - addressed later - accentuates the need to be explicit about risks involved. Participants acknowledged that "there is no such thing as zero risk" and pointed out "bugs can happen".

Informing about possible errors therefore seems essential. However, this is not enough for some participants. One said: "I would trust my doctor more than emotion recognition tools. I do not know how algorithms work.". Another added: "They are not transparent.". While some participants felt the algorithms' complexity is such that they are impossible to understand, others mentioned the doctor's importance; health professionals using FER should be able to summarise how it works. One participant explained that this need to understand functioning would be temporary until accepted. They added: "The term 'AI' scares us.".

Fear can be reflected in the need for transparency regarding data collected: "Not knowing what is collected nor what is done with our data is intrusive.". The first aspect was described as the "main problem" by a participant. They would like to have "direct feedback in real time" to see which parts of the face are captured, and understand how expressions are interpreted. This would be reassuring, allowing to be "aware of what is collected". However, this does

not seem essential to other participants, for whom the minimum requirement is to know "exactly what data is collected". Regarding image processing, participants recognised the need to be "transparent about data" and "purpose" of their capture. We mentioned that certain diseases could be identified using FER; one participant said: "I want diagnosis to stay confidential and. If I have Alzheimer's, I do not want my data to be handed over.".

4.2 Data management

If transparent about risks, functioning, and purposes of data collected, most participants agreed to be subject to FER. However, some declared that obtaining their consent is not sufficient; they mentioned that data belonged to them and are "private rights". The fact that data analysed is a reflection of emotions at a given moment should not be ignored; a psychiatrist could be interested in the evolution of a patient's emotions over time, which requires safe storage.

For one participant, patients should control their emotional data. Another agreed and raised the possibility of using FER regularly to have an emotional dataset that could be given to health professionals. This raises two questions concerning FER in the medical field. First, should the image and associated emotion be considered differently? Second, do they represent medical data? Answering these questions would help decide whether it is ethical to treat emotional data as health data, and make them subject to medical secrecy. Making patients owners of their facial data seems difficult, as "the image of a person is personal data", yet "there is no law that establishes the notion of data ownership" [7].

4.3 Diagnoses establishment

Participants observed that FER systems could be divided into two categories. Doctors would use "Passive" systems to obtain a report of patients' emotions. This way, FER would be a tool to identify emotions, and health professionals in charge of analysing them. "Active" ones would be augmented with decision making for diagnosis.

Regarding passive systems, participants agreed to say that they could be beneficial, as long as used with full awareness of potential errors by doctors. However, opinions were divided about active tools. One participant shared doubts about whether medical experts would need such systems to make a diagnosis; doctors should not depend on FER systems. Another participant agreed: "This technology should only be a support.".

It is therefore necessary to reflect with a global perspective: "individual risk is never acceptable, but, if there is a collective gain, it is perhaps better". Participants suggested a context where one would evaluate levels of pain in an accident. FER could help prioritise the order of care; time saved by task automation could be worth few errors in emotion detection. According to participants, depending on the severity of error consequences, using an active system may be more valuable than a passive one. A participant added that a doctor may not have "seen enough cases to analyse emotional data, while AI is trained with many examples".

4.4 Medical errors liability

Thinking about trusting a FER algorithm to make decisions, participants raised the issue of liability in case of medical errors. In the case of a damaging event, it is necessary to identify to whom fault is attributable. One respondent explained that a doctor who caused a medical error is bound to assume full accountability. However, another one stated that if the doctor uses a tool that could be source of error, "it can inevitably involve a risk of misdiagnosis". Medical malpractice can be defined as "any act, emanating from the caregiver, which resulting in an abnormal damage regarding the predictable evolution of the patient's health condition". Liability for injury would thus lie with the doctor.

Two other entities were suggested as potential responsible for medical malpractice committed through FER. In the case of doctors employed by a hospital, the employer could be implicated. If a FER system leads a victim to seek compensation, the company that conceived the system could be held responsible.

4.5 Best endeavours obligation

The World Health Organisation stated "The enjoyment of the highest attainable standard of health is a fundamental right of every human being" [22]. As discussed above, there may be situations where benefits of using FER in medical settings outweigh risks. We asked participants whether they thought not using it because of possible errors would violate this fundamental principle.

One participant declared: "It is like failing to assist a person in danger.". Another one added: "Considering virtue ethics, we must save the most people, even if there is a risk.". A third one said: "There are different ethical movements: deontological and consequentialist ethics. One cannot say it is unethical not using a technology, as the fact that it could kill people would make one rather choose not to use it.".

Here, participants first gave their opinion saying not using a technology would be malpractice. A nuance was added, specifying that the ethical theory to be adopted is left to the practitioner, who weighs up benefits and risks. A doctor is free to decide on their actions and they must consider advantages, disadvantages, and consequences of the possible choices. However, health professionals are subject to a best endeavours obligation; not using the assistance of a passive FER tool could be a malpractice. Without completely relying on its results, a physician should at least try to use it to diagnose an emotional state. Stakes are therefore two-fold, as they must fulfil their best endeavours obligation while taking responsibility for their choice in the case of an error. This problem could be summarised as: how to provide patients with the care they need using this technology without exposing doctors to sanctions?

Having identified this paradox, and to help doctors cope with such pressure, participants recommended to ensure that tools should be as reliable as possible before being deployed. Another participant suggested that their should be "different algorithms that do the same thing, but trained differently and challenge each other". To limit the biases of FER systems, comparing their results would be an interesting solution [17].

4.6 Non-discrimination principle

The last theme discussed was the non-discrimination challenge; "No person may be discriminated against in access to prevention or care.". However, the lack of representativeness of the data used to train FER systems is source of inequalities [26]. We asked participants: "Do you think it would be fair to use FER in a medical context when some people could not benefit from it?".

Some participants mentioned that FER algorithms working well for only a portion of the population should be used. Their justification is based on the extract of the WHO constitution previously quoted. However, they raised a condition: "Biases should actually be corrected so that others are not neglected.".

Yet one participant said that, before deploying FER in the medical field, "We should wait until biases are corrected; otherwise, it would reinforce societal imbalances.". They added: "We could have built this technology in a much more inclusive way and it would not necessarily have slowed down development.", and that it is necessary to work harder "on sampling, to have a model that is good for training". Non-discrimination is therefore a major issue in the use of FER in the medical field, which is why continuous improvement in its performance and inclusiveness is essential.

5 DISCUSSION

While our results presentation intended to be exhaustive, we acknowledge significant limitations to our work. Individuals selected are not representative of the population and, even though questions were written to be as neutral as possible, they may have guided respondents.

Nevertheless, results reveal interesting preliminary elements. First, there is a clear convergence between issues raised by the ethical literature and our focus group. For instance, algorithmic transparency is one of the main concerns linked to AI in ethical literature [16]. Liability, or non-discrimination principle, is also very present both in the literature and focus group. Second, despite apparent convergence, differences appear. Participants closely linked the need for transparency and for patients to be informed, while literature more clearly distinguishes transparency principle - requiring that any information or communication related to the processing of personal data is easily accessible and understandable - and of patient information - requiring knowing that AI is used in the medical process. Another interesting point is the distinction between "passive" and "active" FER tools, which overlaps the difference between automatic decision making and decision support tool in the literature. Third, unexpected questions or fears appeared during the focus group, like the idea to "feel vulnerable" when physicians use FER. Finally, the use of FER in a medical context leads to consider how to articulate patients rights (i.e., quality care, access to care without discrimination) and personal data rights (i.e., minimise data collection, limit potential misuses).

We therefore suggest extending the analysis beyond this specific group and beyond patients, by looking at opinions of legal experts, ethics specialists, and health professionals. Broadening the research focus of our work could allow some identified themes to be explored in greater depth, and new ones to emerge.

REFERENCES

[1] 2021. Council of Europe, Guidelines on facial recognition, Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data.

[2] 2021. EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).

[3] 2022. Article L1110-3 - Code de la santé publique.

[4] Henriette Roscam Abbing. 2005. The Right to Care for Health: The Contribution of the European Social Charter. *European Journal of Health Law* 12, 3 (2005), 183–191.

[5] Nazanin Andalibi and Justin Buss. 2020. The human in emotion recognition on social media: Attitudes, outcomes, risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.

[6] J Anil and L Padma Suresh. 2016. Literature survey on face and face expression recognition. In *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*. IEEE, 1–6.

[7] Alain Bensoussan. 2022. Vers un droit de propriété des données personnelles. In *Annales des Mines-Réalités industrielles*. Cairn/Softwin, 18–19.

[8] Ned Block. 2004. Qualia. (2004).

[9] Erik Cambria, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco. 2017. Affective computing and sentiment analysis. *A practical guide to sentiment analysis* (2017), 1–10.

[10] Roddy Cowie. 2015. Ethical issues in affective computing. *The Oxford handbook of affective computing* (2015), 334–348.

[11]]ghoroghiprivacy Armin Ghoroghi and Albin Gustafson. [n. d.]. PRIVACY PROTECTION: AI VOICE BIOMETRICS AND EMOTION RECOGNITION. ([n. d.]).

[12] Gabriel Grill and Nazanin Andalibi. 2022. Attitudes and Folk Theories of Data Subjects on Transparency and Accuracy in Emotion Recognition. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–35.

[13] Javier Hernandez, Josh Lovejoy, Daniel McDuff, Jina Suh, Tim O'Brien, Arathi Sethumadhavan, Gretchen Greene, Rosalind Picard, and Mary Czerwinski. 2021. Guidelines for assessing and minimizing risks of emotion recognition applications. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.

[14] Dacher Keltner, Jessica Tracy, Disa A Sauter, Daniel C Cordaro, and Galen McNeil. 2016. Expression of emotion. *Handbook of emotions* 4 (2016), 467–482.

[15] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111, 24 (2014), 8788–8790.

[16] Stefan Larsson and Fredrik Heintz. 2020. Transparency in artificial intelligence. *Internet Policy Review* 9, 2 (2020).

[17] Lucie Lévêque, François Villoteau, Emmanuel VB Sampaio, Matthieu Perreira Da Silva, and Patrick Le Callet. 2022. Comparing the Robustness of Humans and Deep Neural Networks on Facial Expression Recognition. *Electronics* 11, 23 (2022), 4030.

[18] Albert Mehrabian. 1968. Some referents and measures of nonverbal behavior. *Behavior Research Methods & Instrumentation* 1, 6 (1968), 203–207.

[19] Alexandra Prégent. 2021. *Informatique affective: l'utilisation des systèmes de reconnaissance des émotions est-elle en cohérence avec la justice sociale?* Ph. D. Dissertation. Université Laval.

[20] Fatemeh Rabiee. 2004. Focus-group interview and data analysis. *Proceedings of the nutrition society* 63, 4 (2004), 655–660.

[21] Jane Ritchie, Jane Lewis, Carol McNaughton Nicholls, Rachel Ormston, et al. 2013. *Qualitative research practice: A guide for social science students and researchers.* sage.

[22] Michael B Shimkin. 1946. The World Health Organization. *Science* 104, 2700 (1946), 281–283.

[23] Luke Stark and Jesse Hoey. 2021. The ethics of emotion in artificial intelligence systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 782–793.

[24] European Data Protection Supervisor, K Vemou, A Horvath, and T Zerdick. 2021. *EDPS TechDispatch : facial emotion recognition. Issue 1, 2021.* Publications Office. https://doi.org/doi/10.2804/014217

[25] Sue Wilkinson. 1998. Focus group methodology: a review. *International journal of social research methodology* 1, 3 (1998), 181–203.

[26] Tian Xu, Jennifer White, Sinan Kalkan, and Hatice Gunes. 2020. Investigating bias and fairness in facial expression recognition. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 506–523.

Received xxx; revised xxx; accepted xxx