The need for a new "right to refuse" the results of emotion recognition AI*

Roberta Savella [1]

[1] *ISTI-CNR: Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo", Consiglio Nazionale delle Ricerche, via G. Moruzzi 1, 56124 Pisa, Italy*

Abstract

Artificial Intelligence systems are widely used to infer the emotions of people, although this use has been criticized by scholars, associations, and institutions, especially when it is based on the analysis of the facial expression of the individual. In this paper we examine the issues of these technologies, how they are regulated by the new European Artificial Intelligence Act, and the risks they pose from an ethical and legal point of view. From this study, we identify the need for an additional safeguard to protect the individual from decisions taken using facial emotion recognition systems, and we theorize a new "right to refuse" the assessment of these technologies which should allow the person to take back control over the expression of their own emotions.

Keywords

Artificial Intelligence Act, emotion recognition, automated decision-making, right to refuse

1. Introduction and methodology

Artificial Intelligence systems are widely used to detect the emotions of the persons for various purposes, which can range from the medical to the commercial or even the judicial field. However, it is clear now that emotion recognition AI can pose severe risks for the individuals and for society, so in the new European Regulation on Artificial Intelligence (also known as "AI Act") the legislators have included them in the "unacceptable risk" category, for certain uses, and in the "high-risk" category for the other ones. Nevertheless, we argue that this is not enough to effectively protect individuals from the misuse of these technologies or their technical errors.

The purpose of this paper is to highlight the need for an additional safeguard for the rights and freedoms of individuals when an AI system is used to detect their emotions by analyzing their facial expressions, and to propose the creation of a new right which goes beyond the right not to be subject to automated individual decision-making specified by article 22 of the European Regulation on personal data (also known as the "GDPR") [1]. In fact, it is important to point out that the safeguards provided by article 22 of the GDPR [1] fail to protect the individual when an emotion recognition system is involved, because of the intrinsic characteristics of these technologies and for the difficulties in including emotions in the concept of personal data. It is possible to identify a gap of protection in the current legislation on data

*HHAI-WS 2024: Workshops at the Third International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 10—14, 2024, Malmö, Sweden

[1*] Corresponding author.

roberta.savella@isti.cnr.it (R. Savella)

0009-0007-3005-4252 (R. Savella)

protection and the new Artificial Intelligence Act and therefore the need for a new remedy which gives the individual the ability to effectively contest a decision taken by a facial emotion recognition AI. We believe that relying on the individual's consent for the use of these technologies is not enough: first of all, it is not always possible to include information about emotions in the concept of personal data; secondly, article 22 of the GDPR [1] allows automated decision-making in certain cases even without the individual's consent; thirdly, it should be possible for an individual who provided their consent to effectively contest the decision taken by the AI, but this is extremely difficult when the decision involves an immaterial, volatile, and subjective thing such as an emotion. Furthermore, literature on Human-Centered Artificial Intelligence identified the "Golden rules for trustworthy AI" [2] which include the "easy reversal of actions", but when facial emotion recognition AI are involved this is not possible if we don't give absolute prevalence to the affirmation of the individual of their own emotion over the decision taken by the AI: otherwise, it would be the word of the person against the word of the machine, with the risk that the latter would always be perceived as more objective, scientifically sound and accurate and, therefore, right. On the contrary, emotions are subjective and strictly intertwined with the personality and identity of a single person. For these reasons, we argue that there is the need for a new "right to refuse" the output of a facial emotion recognition AI as an additional safeguard which goes beyond the requirements and prohibitions provided by the AI Act.

In this study we will first analyze how the previous versions of the European AI Act regulated emotion recognition AI, and the rules for these technologies which were incorporated in the final version of the text, as well as the debate surrounding the regulation of emotion recognition in the AI Act before its adoption, examining the opinions of relevant Authorities and civil rights associations. Then we will analyze the issues regarding emotion recognition AI from a legal and ethical perspective, we will illustrate the theoretical application to the use of emotion recognition AI of the right provided by article 22 of the GDPR [1] and the scrutiny of the gap of protection, and finally we will propose a possible solution.

2. Context and related work

Emotion recognition technologies can have various kinds of applications and are increasingly used to allow a better interaction between humans and digital systems. For example, uses for this kind of AI can be found in the workplace, to monitor the wellbeing of employees and possible signs of harassment, but also in the retail sector, to provide customized ads and personalization of services. Moreover, emotion recognition AI could be deployed by virtual assistants, by medical systems used to support people with autism, and could be integrated in driver-assistive technologies to improve safety on the roads. The common denominator is the fact that they are systems designed to detect and identify the emotion felt by an individual in a determined moment by analyzing certain subjective features, which can be facial expressions, voice, physiological measurements, body movements, text, and interactions with digital devices and online services [3].

For the purposes of this paper, we will focus on emotion recognition carried out by analyzing facial expressions, as the most problematic one due to the critical issues which have emerged regarding the theory about the detectability of emotion by scrutinizing a person's facial expressions and micro-expressions.

The idea that at least some emotions are deductible by scrutinizing a person's facial expressions and micro-expressions is not new and dates back to Charles Darwin [4], although the most significant voice who theorized the "Basic Emotion Theory" ("BET") was psychologist Paul Ekman [5], who has influenced the debate on this topic for the past 50 years. However, the BET has been heavily criticized by many scholars, so there are now serious concerns regarding the idea that it is possible to detect someone's emotion based on their facial expressions or micro expressions on a universally replicable scale.

As pointed out by Barrett and colleagues [6] the belief that facial movements can justify the inference of a person's emotional state is not sufficiently supported by scientific evidence. In fact, their research found that not only the results of those studies present serious issues regarding their reliability, specificity, and generalizability, but also the validity of this inference was not sufficiently addressed in production and perception studies. To reach this conclusion, Barrett and colleagues analyzed the results of studies regarding how people from different parts of the world and situations (including persons from small-scale and remote cultures, healthy infants and children, and congenitally blind individuals) expressed their emotions, and what they perceived from facial expressions. Their findings led to the affirmation that the expression and interpretation of facial movements regarding emotional states is considerably context dependent and varies across cultures and individuals, so the gestures that the common view (based on the BET) usually associates with the basic emotions are "*best thought of as Western gestures, symbols or stereotypes that fail to capture the rich variety with which people spontaneously move their faces to express emotions in everyday life*". An emblematic example is the fact that the Maori of New Zealand and the Trobriand Islanders in Papua New Guinea interpret the stereotypical expression used in the BET to indicate fear (wide-eyed and gasping face) instead as an indication of intention to harm and anger.

Despite the issues regarding the scientific basis of these technologies, facial emotion recognition systems are increasingly used not only in health-related applications, for which these AI were primarily designed, but also for commercial uses [7] such as, as we have mentioned before, personalized advertising. This raises some serious questions regarding the protection of the *forum internum* and the risks for individuals' fundamental rights and liberties, which were heavily discussed during the drafting of the AI Act, so that there were several and authoritative voices calling for a complete ban of emotion recognition AI.

In particular, in response to the Commission's consultation on the Proposal of the AI Act, non-profit organization Access Now submitted a document [8] in which the treatment of emotion recognition was identified as one of the key issues with that version of the text. At that time, emotion recognition systems were defined as "*an AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data*" and were subject only to transparency rules, identified in article 52 of the first version of the Proposal, except for the cases in which emotion recognition is used for polygraphs, which were already included in the high-risk systems. Access Now pointed out that the definition was potentially flawed and suggested to include the emotions of groups (as well as individuals') and remove the tie between these technologies and biometric data so that it would also cover cases in which applications of emotion recognition use data which does not meet the bar of unique identification (which is required by the strict definition of biometric data given by article 4 of the GDPR [1]). Secondly, the organization considered transparency obligations not enough to effectively address the risks of these technologies, as well as the inclusion of polygraphs in the high-risk category, but they asked instead for a prohibition of all applications of emotion recognition, including in the health sector to protect patients from "*a pseudoscientific product*" [8]. The document quoted Barrett and colleagues' [6] concerns over the possibility of effectively inferring emotions in an automated manner and the risks that the flaws of this technology can pose to fundamental rights, such as the chilling effect on the right to protest, possible discriminations based on individuals' cultures and personal attributes, but also the impact on freedom itself as people would feel pressured to modify their behavior to be positively evaluated by emotion recognition systems, for example showing signs of "happiness" just to get some benefits, or concealing their sadness so they would not be automatically marginalized as "negative" persons.

Among the supporters for a ban of these technologies were also the European Data Protection Board and the European Data Protection Supervisor: in their Joint Opinion on the AI Act [9] they stated that "*use of AI to infer emotions of a natural person is highly undesirable and should be prohibited*", but, unlike Access Now, they accepted the possibility to make some exceptions for certain well-specified use cases, including health and research.

The version of the draft for the AI Act adopted by the European Parliament the 14th of June 2023 [10] contained certain amendments to the original Commission's Proposal that took into account the issues raised. First, in this version, the definition of emotion recognition system was amended to also contain thoughts and states of mind, to include groups as well as individuals, and "*biometric-based data*" in addition to biometric data. Notably and in line with the critics to the theories behind emotion recognition, the Parliament introduced a new recital 26c, which highlighted the "*serious concerns about the scientific basis*" of emotion detection systems also due to the fact that the expression of emotional states varies across cultures, and pointed out that the key shortcomings of these technologies are "*the limited reliability (emotion categories are neither reliably expressed through, nor unequivocally associated with, a common set of physical or physiological movements), the lack of specificity (physical or physiological expressions do not perfectly match emotion categories) and the limited generalisability (the effects of context and culture are not sufficiently considered)*". Recital 26c stated that these risks are particularly significant in real-life situations related to law enforcement, border management, workplace and education institutions, and therefore emotion recognition should be banned for applications in these areas. Consequently, article 5 concerning unacceptable risk systems was amended, including the prohibition of "*the placing on the market, putting into service or use of AI systems to infer emotions of a natural person in the areas of law enforcement, border management, in workplace and education institutions*". Moreover, AI emotion recognition systems not mentioned in article 5 and intended to be used to make inferences about personal characteristics of natural persons on the basis of biometric or biometrics-based data were listed among the high-risk systems of Annex III, as well as polygraphs.

The final text of the AI Act formally adopted by the European Parliament the 13th of March 2024 [11] shows a more lenient approach to emotion recognition systems. The definition was restored to the original one stated in the first Commission's proposal, but some of the prohibitions and obligations provided by the European Parliament in 2023 were maintained. In fact, the placing on the market, putting into service and use of AI to infer emotions of a natural person is now prohibited only in the areas of workplace and education institutions, with the exception of systems used in these sectors but for medical or safety purposes (article 5 par. 1 lect. f). The other AI systems intended for emotion recognition are included in the list of high-risk systems in Annex III and, therefore, subject to the specific requirements and safeguards provided for this category. The transparency obligations are now disciplined by article 50, and paragraph 3 states that "*Deployers of an emotion recognition system or a biometric categorisation system shall inform the natural persons exposed thereto of the operation of the system, and shall process the personal data in accordance with Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, as applicable*". Systems which use emotion recognition to detect, prevent or investigate criminal offenses and which are permitted by law and subject to appropriate safeguards are exempt from the transparency obligation, so in certain cases linked to law enforcement emotion recognition systems can be used even without the person knowing that their emotional states are being detected.

It is also interesting to point out that in this last version of the text [11] the definition of "biometric data" has been slightly changed to remove the requirement of allowing or confirming the unique identification of the natural person. Therefore, there is now a difference between the definition of "biometric data" in the AI Act and in the other legislative text of the European Union (which are all the same as the definition included in article 4 of the GDPR [1]), despite the fact that Recital 14 of the AI Act states that "*The notion of 'biometric data' used in this Regulation should be interpreted in light of the notion of biometric data as defined in Article 4, point (14) of Regulation (EU) 2016/679, Article 3, point (18) of Regulation (EU) 2018/1725 and Article 3, point (13) of Directive (EU) 2016/680*". If the version of the AI Act that will be published in the Official Journal of the European Union will maintain this difference, the notion of "biometric data" for the purposes of the AI Act will be, in fact, broader than the one regulated by all other European laws. This would include in the category of "emotion recognition systems" also the technologies that use biometric information that do not allow or confirm the unique identification of the natural person analyzed by the system.

Looking at these two versions of the text of the AI Act voted by the European Parliament [10] [11], three main differences come to our attention: first of all, the exclusion of the emotion recognition systems used in the areas of law enforcement and border control from the list of prohibited systems and, therefore, the downgrading of the risk they pose from "unacceptable risk" to "high-risk"; secondly, the limitation of the scope of the rules regarding these technologies only to individual emotion recognition, with the deletion of the reference to emotions or intentions of groups from the definition of an emotion recognition system; and thirdly, the return to a strict link between these technologies and biometric data, which could bring to the exclusion of many cases in which the data used to detect the emotional state of a person does not present all the characteristics required by the definition of biometric data given by article 4 point 14 of the GDPR [1]. However, if the definition of "biometric data" as modified in the last version of the AI Act [11] will be maintained in the final text which will be published in the Official Journal of the European Union, it will be possible to consider emotion recognition systems also technologies that use biometric data that do not allow or confirm the unique identification of the data subject, even though there will be a misalignment between the notion of "biometric data" included in the AI Act and the one included in all other European laws.

At the time of writing of this paper we are still waiting for the final steps of the legislative procedure and for the publication in the Official Journal of the European Union, but the text adopted by the Parliament of the 13th of March 2024 [11] is most likely to be the final version of the AI Act. Therefore, we can conclude that the AI Act will probably not stop (maybe not even hinder) the spread of emotion recognition technologies, which will become more and more ubiquitous in individuals' daily lives, and we argue that this elicits the creation of additional safeguards to effectively protect the fundamental rights and freedoms of the people, as we will see further in the next sections of this paper.

3. Problem definition: the need for additional safeguards

As we have seen in the previous section of this paper, eminent scholars and organizations have pointed out serious shortcomings of emotion recognition technologies and risks posed by these systems for individuals and society. In this section we will analyze more deeply the legal and ethical issues posed by facial emotion recognition systems and define the specific problem which our research aims to address.

First of all, Katirai [7] analyzed the literature on emotion recognition technologies and identified three key areas of ethical concern: the risk of biased and unfair outcomes through the use of these systems, also taking into account the issues with the BET highlighted by many scholars [6]; the sensitivity of emotion data, which is often not recognized by the European law on data protection because of the closed definition of sensitive data, in which it is not always possible to include emotion data; and the risk of harm arising from the use of these technologies, especially in delicate sectors such as healthcare and law enforcement. Katirai also points out the danger of the "*automation bias*" as a means to reinforce perceptions of the accuracy of these technologies even when they are fallible and based on faulty premises like the BET.

To examine the practical consequences of the deployment of emotion recognition systems it is interesting to see the report by the international organization Article 19 [12] which shows that the use in China of these technologies poses severe dangers for various human rights. The use cases researched by Article 19 relate to the sectors of public security, driving safety and education. Notably, only uses in one of these areas (education) would be prohibited by the AI Act. The report highlights not only the fact that emotion recognition is not scientifically sound (as we have seen in the previous section) but also that its use imperils human dignity and, in turn, human rights, and in particular the following ones: right to privacy, right to freedom of expression, right to protest, right against self-incrimination, and non-discrimination. A common effect of the use in China of emotion recognition applications, which was identified by Article 19, is mass surveillance, with its chilling effect on freedom, the violation of the personal sphere of the individual, and the possibility of manipulation and influence on people's behavior.

It is important to point out that even the mere possibility of being constantly supervised can impact on how a person interacts with the outside world and can ultimately result in the modification of the individual behavior, self-perception, and even identity. The problematic effects of emotion surveillance have been identified by Steinert and Friedrich [13] who studied the ethical issues of affective Brain Computer Interfaces. They drew attention to the fact that these technologies could infringe on autonomy and authenticity, foster emotion stereotypes, bring to the alienation from one's own emotions, and cause social pressure to self-regulate or enhance control over emotions.

From these considerations it is clear that the right to privacy has a pivotal role in protecting also the other fundamental rights. As highlighted by Alegre [14] "*Privacy, data protection and expression are the gateways to our minds, and, in the digital age, they have so far been serving as the gatekeepers for our rights to freedom of thought and opinion*". Indeed, with the rise of the digital technologies and Artificial Intelligence the right to privacy, especially in its declination as right to data protection, has become an essential safeguard to protect the individual from intrusions in their private sphere, including their thoughts, emotions, and perception of self.

Taking into account this link between data protection and human rights, in the European Union the GDPR [1] provides for specific safeguards to protect the rights and freedoms of the individual regarding their personal data, which is defined as any information regarding an identified or identifiable natural person. Therefore, it is important to ascertain whether emotions can be considered "personal data" according to the European Regulation and what level of protection can be granted to this kind of information.

First of all, emotion data cannot be included in the definition of personal data when it is not connected to an identifier – i.e. the data subject is not identifiable. Moreover, some [15] argue that emotion data provide particularly sensitive information and should afford specific protection with the creation of the new category of "*mental data*" which includes emotions and other information closely related to the *forum internum* (intentions, memories, moods etc.), while others [16] include it in the special categories of personal data regulated by article 9 of the GDPR [1] because of its link with biometric data. Another reason to include emotion data in the special categories of personal data is the fact that it provides information about the mental state of the data subject and therefore could be included in the notion of health data. These instances derive from the fact that the European approach to personal data has always been designed to provide additional safeguards to information linked to the most intimate aspects of one's life, which pose severe dangers when disclosed, such as, for example, those regarding health, sexual preferences and orientation, religion, and political affiliations. From what we have seen above, it is clear that emotion data should in principle afford this level of protection, but the GDPR [1] never explicitly mentions emotions. Consequently, emotion data can afford the special protection provided by article 9 only when the data subject is identifiable, and the information provided by this data is related to one of the areas included in the list of article 9 (for example: when emotions also provide health information, included mental health information).

More specifically, when we talk about emotion recognition AI systems and personal data, now it is necessary to also consider the strict link between emotion recognition technologies and biometric data dictated by the definition of these systems provided by the AI Act [11]. Therefore, it can be argued that the emotion data inferred using the emotion recognition systems regulated by the AI Act should be considered personal data, as it is linked to biometric data which is always an identifier according to its definition provided by article 4 of the GDPR [1] and article 2 of the AI Act [11] states that this Regulation shall not affect the GDPR [1]. However, the difference that we have mentioned above in the definition of "biometric data" between the AI Act [11] and the GDPR [1] opens to the possibility that the biometric data used by the emotion recognition system is not a direct identifier, although it is nonetheless "personal data" as article 3 (33) of the AI Act [11] still defines "biometric data" as "*personal data*", even though it is not required for it to allow or confirm the unique identification of the data subject. Moreover, it is still not clear whether emotion data should be included in the special categories of data listed by article 9, so it will need to be assessed on a case-by-case basis if the emotion data contains information which is granted the highest level of protection.

The GDPR [1] gives the data subjects specific rights to give them back control over what is done with their personal data. Even though we recognize that, from a data protection point of view, to avoid an incorrect assessment of one's own emotions it could be useful to invoke the right to object or the right to rectification, we believe that in the case of facial emotion recognition technologies the data subject would not be granted sufficient protection if we relied only on these rights. In fact, the right to object is only applicable when the processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller (article 6(1)(e) of the GDPR [1]), when it is necessary for purposes of legitimate interest (article 6(1)(f) of the GDPR [1]) or when personal data are processed for direct marketing purposes, and it must be exercised before the processing, so it provides only a preventive remedy. The right to rectification, on the other hand, is applicable to situations in which the data are incorrect or not updated, but to be able to obtain the rectification the data subject should provide the correct information, so in the context of emotion recognition it would force the person to identify the specific emotion they were feeling instead of the one identified by the AI system. However, for all personal data article 22 of the GDPR [1] designs a specific safeguard against the distorted use of new technologies, giving protection to the individual against automated decisions that can significantly affect their personal sphere. Although paragraph 1 of this article literally states that "*The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her*", the Article 29 Working Party in its Guidelines [17] declared that this should be read not as a right of the data subject but as a general prohibition of these practices, which applies "*whether or not the data subject takes an action regarding the processing of their personal data*". This prohibition applies to emotion recognition technologies when are used to automatically take decisions that significantly affect a person, for example: the cancellation of a contract, the entitlement or denial of a social benefit granted by law, the assessment of one's eligibility for credit, decisions that deny someone an employment opportunity or affect access to education [17]. Notably, according to the Guidelines [17], also automated decisions concerning online advertising -such as profiling for personalized ads - could be considered to significantly affect a person when the automated decision system exploits knowledge of the vulnerabilities of the data subjects targeted, which could very well be the case when emotion recognition is employed in this field, for example to profile someone as "sad" and present them ads accordingly.

Paragraph 2 of article 22 lists three exceptions to the prohibition of paragraph 1: when the decision is necessary for entering into, or performance of, a contract between the data subject and a data controller; when the decision is authorized by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or when the data subject gives their explicit consent. However, except for the cases in which the decision is authorized by law, the data controller must always guarantee to the data subject the right to obtain human intervention, to express their point of view and to contest the decision. If the decision is based on special categories of data, the exception to the prohibition applies only when there is the explicit consent of the data subject or the processing is necessary for reasons of substantial public interest, and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests must be in place.

From the analysis of article 22 and the WP29 Guidelines [17] we can conclude that the safeguards against automated decision-making should be applicable also to facial emotion recognition AI systems used to automatically take decisions which significantly affect an individual, and that this would be the case for many uses allowed by the AI Act such as, for example, in the personal advertising field when the profiling exploits the vulnerability of the person due to their emotions.

However, we argue that when these technologies are deployed the data subjects are not sufficiently protected by the mechanisms (prohibition, exceptions and safeguards) introduced by article 22, so additional measures are needed to effectively reach the purpose of this norm, i.e. to protect the people from the potential effects this type of processing may have and to give them back some control over their personal data. To this end it is not sufficient to rely on consent because, even when there is the explicit consent of the data subject, they should be able to contest the decision taken by the automated system.

So how can we effectively protect the individual from the intrusion and influence of emotion recognition systems? Our purpose with this paper is to hypothesize a juridical remedy which goes beyond article 22 of the GDPR [1] to give back to the persons the power to free themselves from external affirmations of something as intimate and as connected to the identity as one's own emotion.

4. A proposed new "right to refuse" the AI decision

For the cases in which it is possible to deploy emotion recognition systems as regulated by the AI Act, we identified the need for a right to effectively contest the assessment of a facial emotion recognition system, even if it is not practically possible to prove that the AI made a mistake.

Otherwise, given the volatility and subjectivity of emotions, the burden of proof on the person to correct a faulty assessment of these technologies would be impossible to fulfill in most cases, especially when the system analyzes only the facial movements to infer the emotional state. We have seen that there are serious doubts on the accuracy of facial emotion recognition systems, and their mistakes can lead to discrimination, breach of human rights and intrusions into the most intimate part of oneself, with possible repercussions not only on one's external behavior but ultimately even on their own identity.

We argue that the automation bias [7] undermines the ability of a person to contest the decision of a facial emotion recognition system, thus nullifying the last safeguard given by article 22 of the GDPR [1] against the faults of automated decision-making. Therefore, when these systems are deployed, there is the need to guarantee to the individual that their assessment of their own emotions, which are strictly connected to their own personality and identity, will prevail over an automated recognition.

For these reasons, we propose to integrate the existing regulations on data protection and Artificial Intelligence with a new "right to refuse" the decision taken by a facial emotion recognition system.

This new right should integrate and go beyond article 22 of the GDPR [1]: as the latter should be interpreted as providing not a right but a prohibition of automated decision-making that significantly affects the individual, it would apply by default, without the need for the person to make a positive demand - indeed, this is the reason why the WP29 [17] interprets article 22 as a prohibition, going against the letter of the law which mentions a "*right not to be subject to*" these decisions.

However, in the cases where this prohibition would not apply, the "right to refuse" the decision would ensure the possibility for the individual to reject the automatic assessment of their emotion when they perceive it as faulty.

We propose to identify the material scope of application of this new right starting with the same limitation as article 22 of the GDPR [1] which applies only to decisions that produce legal effects on the data subject or similarly significantly affect them. The WP29 clarifies [17] that decisions that produce legal effects are the ones that affect someone's legal rights, legal status or rights under a contract, while for a decision to "*similarly significantly affect*" the individual it must have the potential to significantly affect the circumstances, behavior or choices of the individual concerned, have a prolonged or permanent impact on the data subject or, at its most extreme, lead to the exclusion or discrimination of the person. In extreme cases, even profiling for online advertising could significantly affect the individual, and we have already mentioned that we believe that this is the case when emotion recognition is used to exploit a person's vulnerability for marketing purposes.

Applying this limitation of scope also to the "right to refuse" should ensure a balance between the need to grant the individual a high level of protection of their personal sphere and the possible benefits of the use of these technologies in certain fields when there are no significant risks for the person.

Yet, we recognize that it is not possible to apply the material scope of article 22 in its entirety, because article 22 applies to decisions "*based* ***solely*** *on automated processing*" (emphasis added), but this would exclude every case in which there is a human intervention in the decision-making process, including when the assessment of the AI is used by a natural person to take a decision. This outcome would nullify the purpose of the "right to refuse" in many cases, for the same reason that raised the need to go beyond article 22 in the first place: even when a human being is involved (in the decision-making process or in the re-evaluation of the decision), the automation bias [7] would likely affect their choice to follow the assessment of the facial emotion recognition AI even against the objections of the data subject. Therefore, even when the human intervention in the decision-making process is sufficient to exclude the application of article 22, the individual should still have the "right to refuse" if the other conditions for the application of this right are met.

Furthermore, we are proposing the "right to refuse" in the context of facial emotion recognition systems, because we have narrowed down our research on the technology we recognized as most problematic for the issues with the BET identified by many scholars [6]. It could be interesting to assess whether a right such as this would be necessary also for other kinds of emotion recognition AI, but it goes beyond the scope of this paper, and we hope that this question will be considered in future research.

As a preliminary conclusion, in order to assess whether the "right to refuse" is applicable to a specific situation, we suggest the following checklist which includes the key questions that summarize the elements of the material scope of this right that we have seen so far:

1. Is the person identifiable from the emotion data?
2. Is the emotion assessed by a facial emotion recognition system?
3. Does the assessment determine a decision which produces legal effects on the data subject or similarly significantly affects them?

If the answer to all these questions is affirmative, the "right to refuse" applies.

We summarized this assessment in Figure 1:

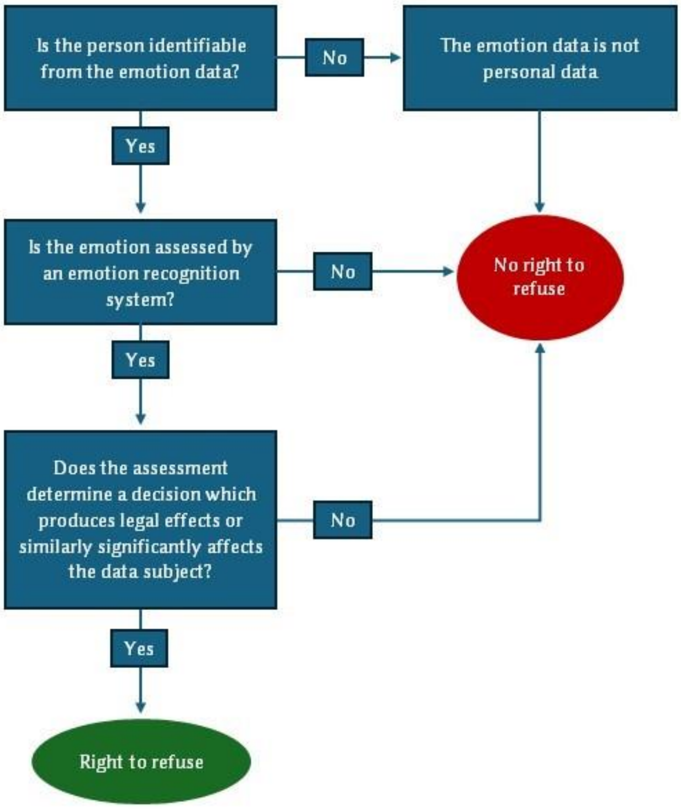
Is the person identifiable from the emotion data?
No
The emotion data is not personal data
Yes
Is the emotion assessed by an emotion recognition system?
No
No right to refuse
Yes
Does the assessment determine a decision which produces legal effects or similarly significantly affects the data subject?
No
Yes
Right to refuse

Figure 1: flowchart to assess whether the "right to refuse" applies to a specific situation.

As for the contents of the new "right to refuse", it should entail the possibility of rejecting the decision taken by a facial emotion recognition AI, even when the individual gave their consent to the use of the system, and the absolute prevalence of the affirmation of the individual about their own emotion, without the need to prove it (as it would be extremely difficult to do so when emotions are involved).

On the contrary, we believe that this right should not include the need for the individual to affirm which emotion they were feeling instead of the one assessed by the AI, as in most cases it is extremely difficult for a person to identify a specific emotion felt. In fact, in real life the emotive states tend to be more complicated than the schematic representations which are normally given by the BET and facial emotion recognition systems, as can be inferred by the many conceptualizations of emotions [18]. Therefore, to effectively contest the decision, it should be sufficient for the individual to say that the assessment of the AI was wrong.

Recognizing the "right to refuse" as we have theorized it would give back control to the individual over the affirmation of their own emotions, minimizing the risks of violation of human rights, manipulation, and infringement of autonomy and identity.

Moreover, this approach would promote a human-centric design of AI and enhance the trustworthiness of these systems by giving a way to ensure the easy reversal of actions as mandated by the "Golden rules for trustworthy AI" [2]. It would also be possible to create a synergic interaction between the individual and the facial emotion recognition AI, which would improve the accuracy of the latter by integrating the contestation as feedback.

In our analysis we have taken for granted that the individual is informed that they are exposed to an emotion recognition system, as it is mandated by article 50 paragraph 3 of the latest version of the AI Act [11]. We recognize that this is a prerequisite for the exercise of the "right to refuse", because it would be impossible for the individual to ask for this remedy without the knowledge that a facial emotion recognition system is used to assess their emotions, and without having access to the results of this assessment. However, we have seen that in certain cases the transparency obligation laid down in article 50 does not apply because the system is used for biometric categorization and emotion recognition which are permitted by law to detect, prevent, and investigate criminal offences. In these situations, the "right to refuse" would not be exercisable, so a different remedy would be necessary to effectively protect the individual.

5. Conclusions and proposed future research

We have seen that there are serious concerns over emotion recognition Artificial Intelligence in general, but the use of these technologies is particularly alarming when they are employed to detect emotional states on the basis of the facial expressions of an individual, because in this case there are also various doubts about the scientific validity of the assessment itself.

Looking at how these systems are regulated by the European AI Act [11], also considering the legislative path during which the initial proposal was changed to take into account instances posed by associations and institutions in the consultation process, we argue that it is reasonable to conclude that the new Regulation does not provide sufficient safeguards for the rights and freedoms of natural persons when these technologies are deployed. In fact, except for the areas of workplace and education institutions, it is still possible to use facial emotion recognition systems for any (lawful) purpose, provided that the obligations for high-risk systems listed in the AI Act [11] and other applicable laws (including the GDPR [1]) are respected, including the transparency requirements stated in article 50 of the AI Act.

Therefore, we identified the need for an additional layer of protection for individuals subject to facial emotion recognition systems. We started from article 22 of the GDPR [1], which provides a safeguard against automated decision-making in the context of personal data protection, and we imagined how it would be applicable to these specific technologies. We concluded that the characteristics of facial emotion recognition do not allow the purpose of article 22 to be fulfilled, and that the automation bias makes it extremely difficult (if not impossible) for the individual to effectively access this kind of protection. Consequently, we hypothesized a remedy which we designed as a new "right to refuse" the decision taken by a facial emotion recognition system. This should give the data subject the means to reject the assessment of these technologies when it causes legal effects concerning the data subject or similarly significantly affects them, without the need to prove the validity of their own affirmation or to state the different emotion they were feeling in that particular moment.

For the purpose of this research, we selected a specific kind of application – emotion recognition systems that analyze the facial expressions and micro-expressions of a person to infer their emotion – but future research could assess whether the conclusions we have reached in this paper could be applicable also to other emotion recognition systems (for example: technologies that analyze the vocal tones). Additionally, we designed the "right to refuse" from an individualistic point of view, as we started our theory from the principles enshrined in the prohibition of automated decision-making processes stated by article 22 of the GDPR [1], but further research could analyze the possible implications of a similar right in respect of anonymized data or even group's emotions, as some scholars are already identifying the importance of privacy also for Big Data [16]. Also, we selected article 22 of the GDPR [1] as the starting point for our theory for the "right to refuse" because we intended to find an equivalent remedy applicable irrespective of the legal basis for the processing of the personal data and that would not be invalidated by the automation bias and by the difficulties in identifying one's own emotions. We recognize that the GDPR [1] provides for other mechanisms which could be used by the data subject such as, for example, the right to object and the right to rectification, and further research could investigate how the "right to refuse" could interact with these other remedies.

Acknowledgements

This work is supported by the European Union – Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities", Grant Agreement n.871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu), NextGenerationEU – National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) – Project: "SoBigData.it – Strengthening the Italian RI for Social Mining and Big Data Analytics" – Prot. IR0000013 – Avviso n. 3264 del 28/12/2021, and TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No 952215.

The author gratefully acknowledges the Coordinator of SoBigData++ Roberto Trasarti for his inputs and valuable feedback on this paper.

References

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[2] Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., and Elmqvist, N., Designing the User Interface: Strategies for Effective Human-Computer Interaction, 6th. ed., Pearson, (2017).

[3] Monteith, S., Glenn, T., Geddes, J. et al. Commercial Use of Emotion Artificial Intelligence (AI): Implications for Psychiatry. Curr Psychiatry Rep 24 (2022), 203–211, URL: https://doi.org/10.1007/s11920-022-01330-7.

[4] Darwin, C., The Expression of the Emotions in Man and Animals, 3rd. ed., Oxford University Press, New York, (1872).

[5] Ekman, P., Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.), Nebraska Symposium on Motivation, 1971 (pp. 207–283). Lincoln: University of Nebraska Press, (1972).

[6] Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D., Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. Psychological Science in the Public Interest 20 (2019), 1–68.

[7] Katirai, A. Ethical considerations in emotion recognition technologies: a review of the literature. *AI Ethics* (2023). URL: https://doi.org/10.1007/s43681-023-00307-3.

[8] Access Now's submission to the European Commission's adoption consultation on the Artificial Intelligence Act, (2021). URL: https://www.accessnow.org/wp-content/uploads/2021/08/Submission-to-the-European-Commissions-Consultation-on-the-Artificial-Intelligence-Act.pdf.

[9] 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), (2021). URL: https://www.edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf.

[10] Artificial Intelligence Act, Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), P9_TA(2023)0236, (2023). URL: https://artificialintelligenceact.eu/wp-content/uploads/2023/06/AIA-%E2%80%93-IMCO-LIBE-Draft-Compromise-Amendments-14-June-2023.pdf.

[11] Artificial Intelligence Act, European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), P9_TA(2024)0138, (2024). URL: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf.

[12] Article 19, Emotional Entanglement: China's emotion recognition market and its implications for human rights, A19/DIG/2021/001, (2021). URL: https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf.

[13] Steinert, S., Friedrich, O. Wired Emotions: Ethical Issues of Affective Brain–Computer Interfaces. *Sci Eng Ethics* **26**, (2020), 351–367. URL: https://doi.org/10.1007/s11948-019-00087-2.

[14] Alegre, S. Regulating around freedom in the "forum internum". *ERA Forum* **21**, (2021), 591–604. URL: https://doi.org/10.1007/s12027-020-00633-7.

[15] Ienca, M., Malgieri, G., Mental data protection and the GDPR, *Journal of Law and the Biosciences*, Volume 9, Issue 1, January-June 2022, lsac006. URL: https://doi.org/10.1093/jlb/lsac006.

[16] McStay, A., Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society*, *7*(1), (2020). URL: https://doi.org/10.1177/2053951720904386.

[17] Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01), (2018).

[18] Stark, L., Hoey, J., The Ethics of Emotion in Artificial Intelligence Systems. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, (2021), 782–793. URL: https://doi.org/10.1145/3442188.3445939.