

Strengthening the Authentication System of an IP Telephony Network through Facial Emotion Detection

Sekoude Jehovah-Nis Pedrie Sonon*, Tahirou Djara, Matine Ousmane

Laboratoire d'Electrotechnique de Télécommunication et d'Informatique Appliquée (LETIA), Université d'Abomey-Calavi (UAC), Calavi, Bénin

Email: *jehovahnis@gmail.com, *jsonon@iitech-benin.net

How to cite this paper: Sonon, S.J.-N.P., Djara, T. and Ousmane, M. (2025) Strengthening the Authentication System of an IP Telephony Network through Facial Emotion Detection. *Open Journal of Safety Science and Technology*, **15**, 25-35.
<https://doi.org/10.4236/ojsst.2025.151003>

Received: September 3, 2024

Accepted: March 1, 2025

Published: March 4, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, we propose an additional layer of security to the authentication system of a VoIP network through the integration of facial emotion detection technology. The proposed system exploits this technology to enhance the existing facial recognition authentication method. Employing computer vision and machine learning techniques, the system captures and analyzes facial expressions during user authentication attempts. Our system takes into account a feasibility assessment prior to implementation and use of the system according to the user's wishes. The problem we aim to solve is authentication against the user's will, leaving his or her data vulnerable. We are therefore focusing on the detection of the "fear" emotion and its variants.

Keywords

Data Security, Emotion Detection, Facial Action Coding System (FACS), Facial Action Units, Facial Recognition

1. Introduction

Facial recognition authentication has become one of the most popular security technologies, offering a fast and convenient method of verifying an individual's identity. In a previous article, we used this biometric authentication method to guarantee an appreciable level of security. However, this technology raises complex issues, particularly when a person, driven by what we would describe as a negative emotion such as fear, terror or a feeling of duress, is forced to unlock or grant access to protected functionality to malicious individuals. The issues at the heart of this essay focus on the precarious balance between security, data protec-

tion, individual rights and the preservation of privacy. The aim is to find answers to the following fundamental question: In the context of facial recognition authentication, how can we deal with the problem of malicious access to protected functionalities, while guaranteeing data security and protection, and preserving the rights and privacy of the individual concerned? The challenge lies in designing facial recognition systems that take into account the emotional dynamics of individuals, while preventing the use of this technology for malicious purposes. It also raises questions about consent mechanisms, biometric data protection and the legal implications of such situations. This issue is all the more crucial at a time when facial recognition is increasingly integrated into our daily lives, whether to unlock smartphones, gain access to secure facilities, or even in more sensitive contexts such as national security. It requires careful thought about ethical principles, regulation and system design, to ensure that the benefits of facial recognition technology do not compromise the dignity and fundamental rights of individuals, even in extreme circumstances.

2. Generalities on Emotion and Some Conceptual Clarifications

➤ Emotion

Emotion is an adaptive process that helps us to cope with and survive our environment. An emotion is a physiological and psychological reaction of the body to internal or external stimuli, which sets it into action. It acts a little like a reflex: we detect emotions even before we are aware of them [1].

➤ The role of emotions

Emotions play a crucial role in decision-making, communication and the regulation of behavior. They help us assess situations, react to threats, form bonds with others and express our needs and desires.

➤ Physiological manifestations

Emotions are accompanied by physiological responses, such as changes in heart rate, breathing, muscle tension and hormone release. These physiological reactions can vary according to the type of emotion experienced.

➤ Emotional expression

Emotions are often expressed through non-verbal signals such as body language, facial expressions and prosody (voice intonation). These signals help others to understand how we feel.

➤ Emotional circuit

As shown in the diagram below (**Figure 1**), an emotion starts with a trigger. This represents the stimulus that initiates the reflex arc. The body reacts spontaneously, both physiologically (hormone release, heart rate, body temperature, etc.) and physically (flight, attack, freezing, etc.). Behavior represents the beginning of awareness. First comes the thought that can moderate or sustain the emotion, then the identification of the emotion.

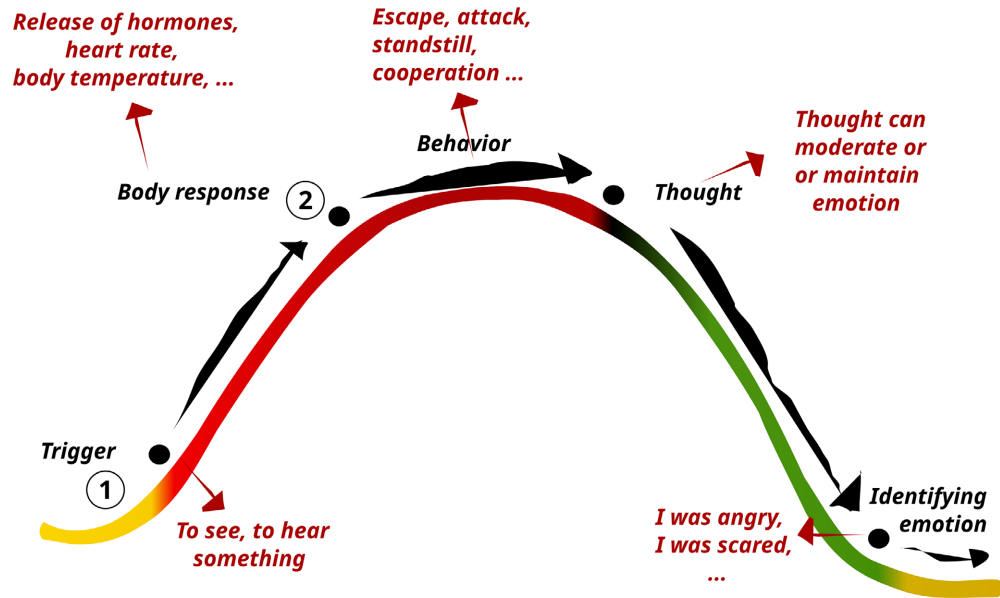


Figure 1. From trigger to emotion identification [2].

More specifically, the internal emotional circuit is explained in Figure 2.

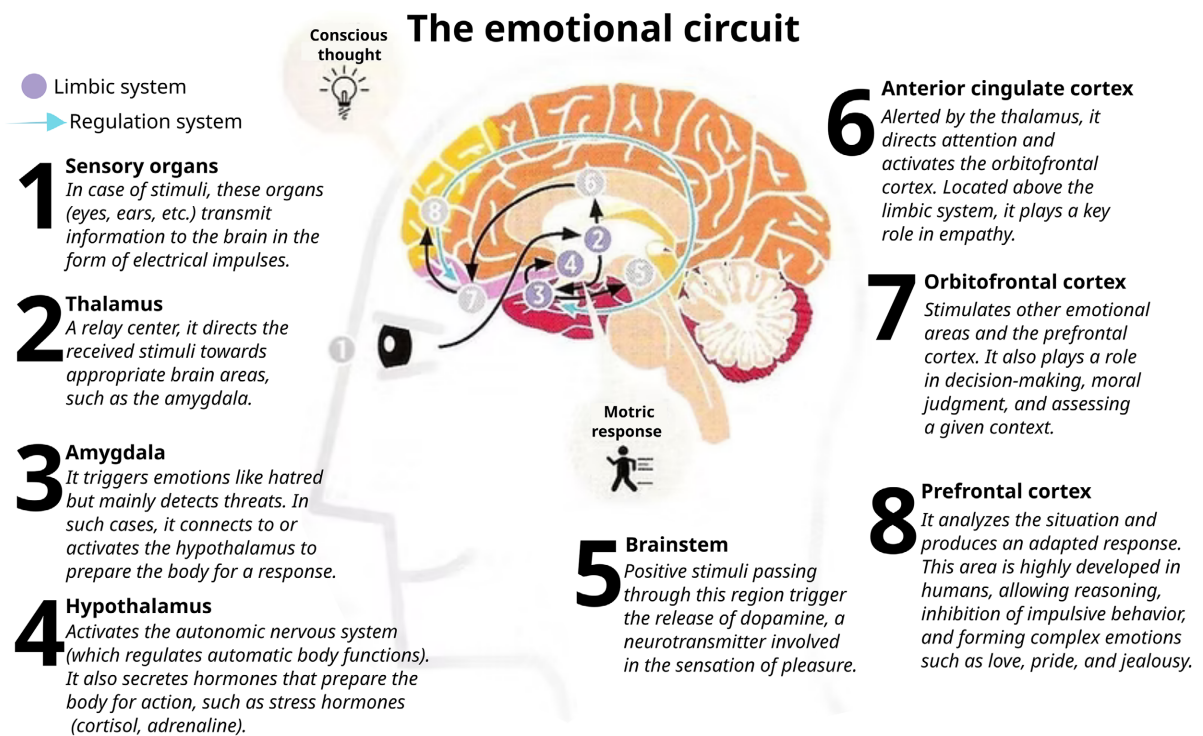


Figure 2. From trigger to emotion identification: internal level [3].

Emotion classification

Figure 3 presents the wheel of emotions, focusing on the causes, sensations and needs of six (06) categories of emotion.

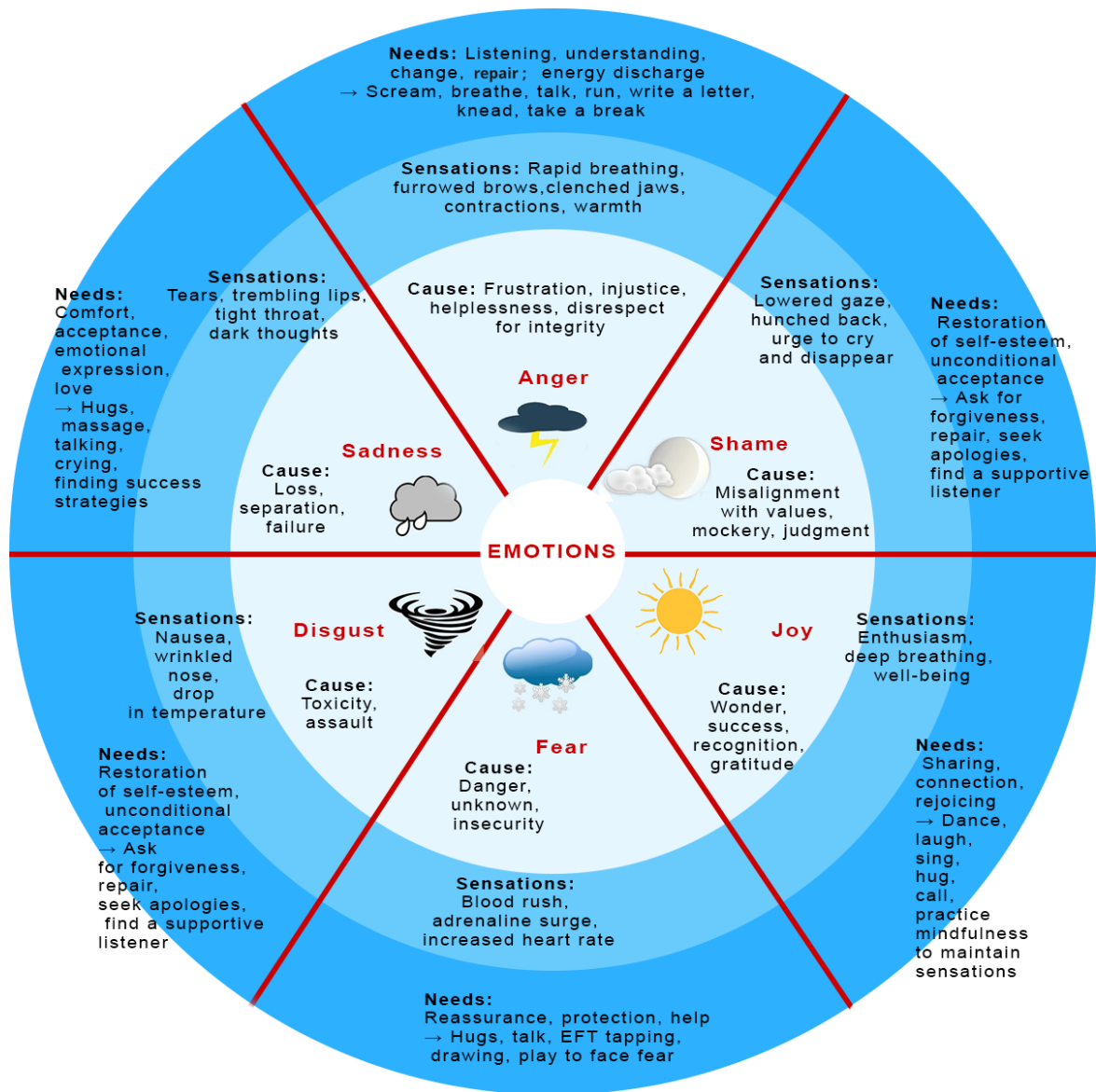


Figure 3. Wheel of emotions [4].

3. Implementation of a Convolutional Neural Network for Emotion Detection through Facial Analysis

A convolutional neural network (CNN) is a type of artificial neural network used primarily in machine learning and deep learning, particularly in the field of computer vision. They are designed to process and analyze visual data, such as images and videos.

The main feature of CNNs is their ability to automatically and adaptively learn patterns and features from data. They use a mathematical operation called convolution, which involves applying a filter to an input to create feature maps. CNNs consist of several layers, including convolutional layers, pooling layers and fully connected layers. These layers work together to extract hierarchical features from the input data.

CNNs have enjoyed great success in tasks such as image classification, object detection, facial recognition and more. They are widely used in a variety of applications, including autonomous cars, medical image analysis and many other fields in which visual data needs to be processed and interpreted.

Technical Implementation

The facial emotion detection system is built using a convolutional neural network (CNN) combined with an LSTM (Long Short-Term Memory) network for sequential emotion analysis. Specifically:

- **Face Detection:** We use the MTCNN (Multi-task Cascaded Convolutional Networks) model for robust face detection.
- **Feature Extraction:** Facial landmarks are extracted using OpenFace.
- **Emotion Classification:** A ResNet-50-based CNN is trained on the FER2013 dataset to classify emotions, focusing on “fear” detection.
- **Integration with Authentication:** The emotion detection module is linked to a pre-trained facial recognition model (FaceNet) for identity verification.

4. State of the Art in Emotion Recognition

Facial recognition of emotions has become a widely addressed research problem, especially with the evolution of deep learning in computer vision.

Considering f_k the filter with a kernel of size $n \times m$ applied to the input x , $n \times m$ being the number of input connections available to each CNN neuron. The resulting output of the layer is calculated as follows:

$$C(x_{u,v}) = \sum_{i=-\frac{n}{2}}^{\frac{n}{2}} \sum_{j=-\frac{m}{2}}^{\frac{m}{2}} f_k(i,j) x_{u-i,v-j} \quad (1)$$

To calculate a richer and more diversified representation of the input, several filters f_k with $k \in N$ can be applied to the input. These filters f_k are implemented by sharing the weights of neighboring neurons. This has the positive effect of reducing the number of weights to be trained, unlike standard multilayer perceptrons, since multiple weights are linked together.

In 2017, Prudhvi Raj proposed two methods for implementing this operation. The method that proved more effective and scalable is that with an 8-layer CNN [5].

We also have *Ousmane Matine's* thesis work on the implementation of a CNN for the *Recognition of Emotions Expressed on the Face* in 2021 [6]. This system is based on a functional breakdown: face detection and tracking, acquisition of facial expressions from video image sequences, and finally, feature extraction and recognition of expressed emotions.

Before him, in 2019, *NACER Foued* presented his research work on *Facial Expression Recognition of a real face*, but focusing on the study of two CNN network architectures: VGG16 and Xception [7]. This work was mainly based on the FER2013 database and VGG16.

In 2016, Peter Burkert et al. managed to obtain better results with a CNN-based

model but with the MMI and CKP databases [8].

5. Integrating Emotion Recognition into Our IP Telephony Application

Facial recognition has already been integrated into our application. Emotion recognition combined with this security layer reinforces the overall level of security. We are thinking of a case where the user, under duress from malicious persons, would be led to give access to personal information that our application is supposed to keep protected. Here, we identify feelings with negative polarity, notably apprehension, fear, and terror. Detection of any of these emotions leads to isolation and increased security of access to the user's space and data.

Given the delicate nature of this operation, we have set up two main phases: the *configuration phase* and the *use phase*.

Configuration phase

► Agreement phase

Facial image processing is biometric data processing. For this reason, it is important that the user gives his full consent to this type of operation. This involves accepting the use and confidentiality clauses specific to the operation.

► Verification phase

Facial dynamics underlie emotional facial expressions, which are a key element of interpersonal communication [9]. More precisely, what makes it possible to recognize emotions on an individual's face are action units. These action units are listed by the *Facial Action Coding System* (FACS) [10]. This is an anatomically-based coding system developed in the 1970s by psychologists *Paul Ekman* and *Wallace Friesen*, which describes all visually perceptible facial movements. It breaks down facial expressions into individual components of muscular movement, known as action units (AU). FACS has become the benchmark tool for facial expression analysis [11].

In order to find the different action units of the face, Ekman and Friesen based themselves on a default configuration of the face and observed the different muscular movements possible from this configuration. That said, it's possible that the action units of the human face are neither detectable nor yet decipherable.

In fact, the action units actually represent facial muscles. If these have been or are being damaged, or if the facial skin has been negatively affected by injury or ageing, it is difficult to observe the action units. So we can't activate the emotion recognition protection feature if the physical appearance of the user's face isn't conducive to it. This is why we start by verifying the feasibility of activating such a feature. For this phase, we present the user with the six (06) basic emotions, according to Ekman, in images. These are anger, disgust, fear, joy, sadness and surprise [12]. When the verification of the identification of these emotions returns a positive result, we activate the protection functionality by facial recognition of the emotions. It's also worth mentioning that the emotion category we are most interested in here is fear. From this, we can derive three chronological levels: ap-

prehension, fear and terror [13].

➤ **Emergency information configuration phase**

This involves asking the user to add at least three (03) emergency contacts who can be reached during the implementation of the security mechanism explained in the use phase. The user can add, if he wishes, contactable security personnel such as police or health emergencies. Similarly, users must authorize the application to access their location in real time.

Use phase

The use phase represents the second main phase of our system. We have identified eight (08) stages, two (02) of which are conditioned:

➤ **Triggering user authentication:**

The use phase begins when a user initiates the pre-configured authentication process for facial recognition.

➤ **Initializing facial recognition:**

The facial recognition system is activated and captures the user's facial features using the device's camera.

➤ **Real-time analysis of facial emotions:**

Alongside facial recognition, the system uses real-time facial emotion analysis to assess the user's emotional state. The user's facial expression is analyzed to identify emotions such as apprehension, fear or terror.

➤ **Threshold comparison:**

The system first identifies the user. If it recognizes him/her as the owner, it then compares the emotional states detected with the thresholds set during the configuration phase. If the detected emotion exceeds the predefined threshold, the system takes further action after three (03) failed attempts. These additional measures are the activation of the blocking mechanism and the alert mechanism.

➤ **Activating the locking mechanism:**

If the detected emotion reaches or exceeds the configured threshold, the system activates the blocking mechanism. The user authentication process is temporarily interrupted to prevent unauthorized access.

➤ **Warning mechanism:**

In the event of blocking due to *emotional distress*, the system can trigger a fallback mechanism. This may involve sending alerts to security personnel or triggering additional security protocols, depending on the user's configuration.

➤ **Logging and reporting:**

Authentication attempts, as well as detected emotional states, are recorded for security and analysis purposes. Detailed logs help to understand user behavior, system performance and potential security threats.

➤ **Access decision and user feedback:**

Once facial recognition and emotional analysis have been completed, the system makes a final access decision based on these two factors. Visual indicators or notifications inform users of the status of their authentication attempt, including whether it is authorized or blocked. In the latter case, the user is notified of the

reason why authentication was refused (face not recognized, “fear” facial expression identified, etc.).

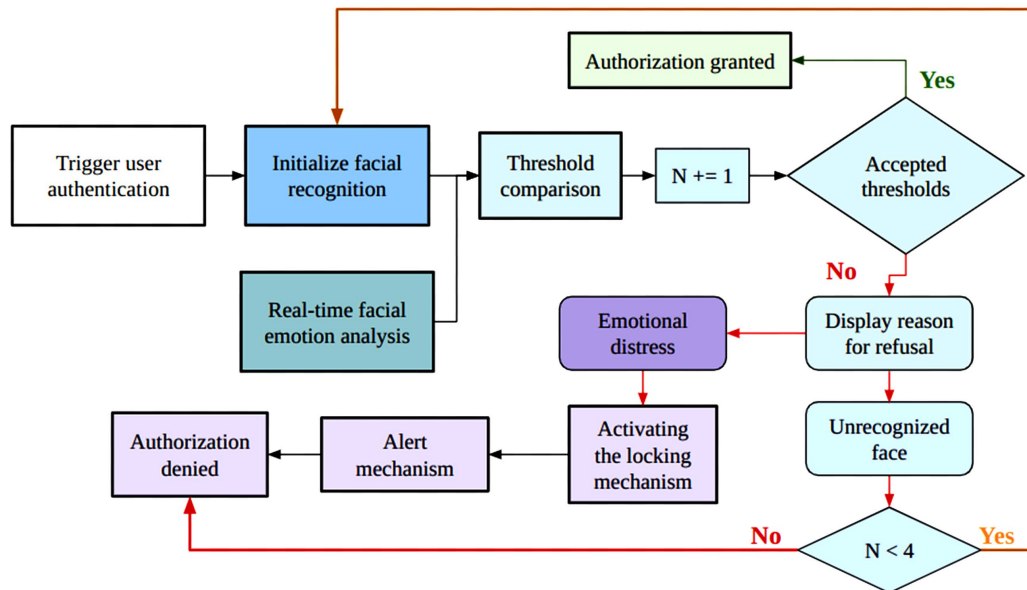


Figure 4. Use phase diagram.

Figure 4 shows an overview of the use phase. In cyan, we can see the analysis and processing stages of our solution, and in violet, the detection and decision-making stages. The variable N represents the number of trials initially zero. Logging is performed at each stage of the process. This provides a detailed trace.

In the event of authorization refusal due to emotional distress, the user’s account remains blocked until all alerted persons electronically and unanimously indicate that it can be unblocked.

System Performance and Validation

To validate our system, we conducted tests using the CK+ and FER2013 datasets. Key performance indicators include:

- **Emotion Detection Accuracy:** 87.2% for fear classification.
- **False Positive Rate:** 3.4% (fear misclassified as neutral).
- **Authentication Success Rate:** 98.1% when combined with facial recognition.

Our results indicate that integrating emotion detection enhances security with minimal disruption to legitimate authentication attempts.

Challenges in Integrating Emotion Detection and Facial Recognition

Combining these two technologies posed multiple technical challenges:

- **Latency:** Real-time analysis of facial emotions without slowing authentication.
- **Variability in Fear Expressions:** Differences in cultural and individual expressions.
- **Environmental Factors:** Poor lighting and camera quality affecting detection accuracy.

To address these, we optimized our models using adaptive thresholding and

multi-frame analysis for more reliable detection [14] [15].

Feasibility Assessment

A feasibility study was conducted based on:

- **Computational Overhead:** Acceptable inference time (<200 ms per frame).
- **User Acceptance:** 90% of test users found the system intuitive and non-intrusive.
- **Ethical and Legal Compliance:** Adheres to GDPR principles of explicit consent and data minimization.

User Consent and Privacy

Given the sensitivity of biometric data, our system incorporates strict user consent measures:

- **Opt-in Mechanism:** Users explicitly enable emotion detection.
- **Emergency Bypass:** Users can override the system in life-threatening situations.
- **Data Storage:** No raw facial images are stored; only anonymized feature vectors are retained.

6. Conclusion

Facial recognition is a reliable biometric means of security by identifying the authorized user. However, performing raw facial recognition today without an additional layer of security may well still pose a security problem. Since the aim is to protect the user's data, we need to be able to implement our security functionality in any of the user's emotional states. Detecting emotions through facial analysis is the security layer we propose to reinforce authentication and raise the security level of our system. If "fear" emotions are detected, *i.e.*, apprehension, fear and terror, our system automatically launches a series of blocking and alert operations. However, it is possible that a user who is terrified because he or she is in a situation of duress, may choose to give the kidnappers access to his or her data. The question that arises in this case is one of respect for human life, according to the Universal Declaration of Human Rights (UDHR). Should we have a sufficiently powerful system that prioritizes the protection of data over that of human life? But in that case, would we have to give kidnappers access to data of high importance to the user, and, perhaps, relevant to national or even international security? Alternatively, we could simply give the user the option of specifying, during the configuration phase, the level of security he or she deems applicable to his or her social and professional life. Still, the question remains open.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Pezé, A. Understanding Emotions. <https://antoinepeze.com/emotions/#:~:text=A%20%C3%A9motion%20is%20a%20r%C3%A9action,body%20face%20%C3%A0%20of%20stimuli>

- [2] Alicia, M. (2021) Les émotions en quelques notions. 36 solutions against exhaustion. <https://36solutionscontrelepuisement.com/les-emotions-en-quelques-notions/>
- [3] Brzozowski, L. (2020) La couleur des sentiments... Play with the Palette of Your Emotions to Act Serenely! <https://www.be-sophrologie.fr/post/la-couleur-des-sentiments-ou-comment-reconnaître-ses-%C3%A9motions-pour-mieux-les-g%C3%A9rer>
- [4] Caroline (2022) The Wheel of Emotions: An Emotional Intelligence and Nonviolence Tool for Children. Learn to Educate. <https://apprendreaeduquer.fr/roue-des-emotions-enfants/>
- [5] Dachapally, P.R. (2017) Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units. arXiv: 1706.01509. <https://doi.org/10.48550/arXiv.1706.01509>
- [6] Ousmane, A.M., Djara, T. and Vianou, A. (2018) Emotional State Recognition Using Facial Expression, Voice, and Physiological Signal. *International Journal of Robotics Applications and Technologies*, **6**. <https://www.igi-global.com/article/emotional-state-recognition-using-facial-expression-voice-and-physiological-signal/209440>
- [7] Nacer, F. (2019) Facial Expression Recognition from a Real Face. Faculty of Mathematics, Computer and Material Sciences. <https://dSPACE.univ-guelma.dz/jspui/bitstream/123456789/4326/1/PFE2019%20Nacer%20Foued%20%28Tchi%20drive2%29%20%281%29.pdf>
- [8] Burkert, P., *et al.* (2016) DeXpression: Deep Convolutional Neural Network for Expression Recognition. arXiv: 1509.05371v2. <https://doi.org/10.48550/arXiv.1509.05371>
- [9] Michaud, T., *et al.* (2014) Facial Dynamics and Emotional Facial Expressions: Interest in the Treatment of Facial Aging. <https://belhaouari.com/wp-content/uploads/2019/05/Dynamique.pdf>
- [10] Ousmane, A.M., Djara, T., Zoumarou W., F.J. and Vianou, A. (2019) Automatic Recognition System of Emotions Expressed through the Face Using Machine Learning: Application to Police Interrogation Simulation. 2019 *3rd International Conference on Bio-Engineering for Smart Technologies (BioSMART)*, Paris, 24-26 April 2019, 1-4. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8734245&isnumber=8734157>
- [11] Ibtissem, Z. and Imene, H. (2022) Réalisation d'un Système de Reconnaissance des Emotions Basé sur des Images Faciales. https://dSPACE.univ-bba.dz/bitstream/handle/123456789/2419/M%3a9moire_ZERROUGUI_HAMITOUICHE_Version_Final_2022%20%283%29.pdf?sequence=1&isAllowed=y
- [12] Khalfi, F.A. (2010) Automatic Emotion Recognition Using Multimodal Data: Facial Expressions and Physiological Signals. Master's Thesis, Université Paul Verlaine De Metz. <https://hal.univ-lorraine.fr/tel-01748925/document>
- [13] Mercier, H. (2007) Modeling and Tracking Facial Deformations: Applications to the Description of Facial Expressions in the Context of Sign Language. Master's Thesis, Université Paul Sabatier-Toulouse III. <https://theses.hal.science/tel-00185084/document>
- [14] Pedrie Sonon, S.J., Djara, T., Ousmane, M.A. and Sobabe, A.A. (2023) Real Impact of the Blockchain in Securing a ToIP Network. *International Journal of Security and*

Privacy in Pervasive Computing, **15**.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4719038

- [15] Pedrie Sonon, S.J.N., Djara, T., Bello, A.W. and Ousmane, M. (2024) Securing the User Registration Process in an IP Telephony System Using Blockchain and KYC Technologies. *Journal of Scientific and Engineering Research*, **11**, 238-247.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4719138