2024 18th International Conference on Automatic Face and Gesture Recognition (FG)

A Framework for Assessing Proportionate Intervention with Face Recognition Systems in Real-Life Scenarios

Pablo Negri[1,2] and Isabelle Hupont[3] and Emilia Gomez[3]
[1] Instituto de Investigacion en Ciencias de la Computacion (ICC), UBA-CONICET.
[2] Computer Department, FCEyN, Universidad de Buenos Aires, Buenos Aires, Argentine
[3] European Commission, Joint Research Centre, Sevilla,Spain

***Abstract*— Face recognition (FR) has reached a high technical maturity. However, its use needs to be carefully assessed from an ethical perspective, especially in sensitive scenarios. This is precisely the focus of this paper: the use of FR for the identification of specific subjects in moderately to densely crowded spaces (e.g. public spaces, sports stadiums, train stations) and law enforcement scenarios. In particular, there is a need to consider the trade-off between the need to protect privacy and fundamental rights of citizens as well as their safety. Recent Artificial Intelligence (AI) policies, notably the European AI Act, propose that such FR interventions should be proportionate and deployed only when strictly necessary. Nevertheless, concrete guidelines on how to address the concept of proportional FR intervention are lacking to date. This paper proposes a framework to contribute to assessing whether an FR intervention is proportionate or not for a given context of use in the above mentioned scenarios. It also identifies the main quantitative and qualitative variables relevant to the FR intervention decision (e.g. number of people in the scene, level of harm that the person(s) in search could perpetrate, consequences to individual rights and freedoms) and propose a 2D graphical model making it possible to balance these variables in terms of ethical cost vs security gain. Finally, different FR scenarios inspired by real-world deployments validate the proposed model. The framework is conceived as a simple support tool for decision makers when confronted with the deployment of an FR system.**

I. Introduction

Face recognition (FR) is a flexible biometric technology capable of identifying people at a distance, even without the active cooperation of the captured subjects. In the last decade, FR systems have been used for many different purposes, such as access control [18], border control [7], device/machine unlocking [37], control of attendance [33], missing people identification [25] and face tagging [3].

This paper focuses on the most technically advanced, albeit ethically controversial, FR application: its real-time use to identify specific subjects in moderately to densely crowded spaces (e.g. public open spaces, sports stadiums, train stations, airports, malls) and for law enforcement purposes. Such FR scenarios typically make use of a multi-camera system to identify people who represent a potential threat (e.g., thieves, criminals, terrorists on police records) or are being searched (e.g. missing people or kidnapped people) over multiple video streams [4]. Software products conceived for this specific purpose are widespread on the market [14], and they are deployed by polices and law enforcement agencies worldwide.

From a technical perspective, FR nowadays performs successfully even in highly uncontrolled situations with tens -to- hundreds of individuals in the scene, changing lighting conditions and low face resolutions. State-of-the-art face identification algorithms achieve accuracy metrics above 95% with a false acceptance rate of 10^{-4} in these contexts [20], [6]. Although demographic fairness (e.g. race and gender biases) in FR is still an open research area [13], [30], some mitigation measures are being developed, and the FR community is raising awareness of this matter and encouraging its research [12].

While algorithmic robustness and fairness are undoubtedly key requirements for the development of FR systems, critical ethical aspects related to deployment phases have been widely under-considered. Even assuming that an FR system is almost perfectly accurate, fair and deployed by authorities for the exclusive purpose of improving public security, its use inevitably involves an invasion of privacy as the faces of all the subjects passing by a designated area are processed to search for a potential match with a person on a watchlist. In this scenario, the captured subjects might not wish to be under FR surveillance and might not be aware of the system operation. Other rights may also be affected when FR is used in this context, such as *the right to freedom of expression, peaceful assembly, and association, as well as freedom of movement*, according to [36]. The authority in charge of the system deployment should therefore establish the most strictest privacy-preserving mechanisms and carefully assess the use of these technologies considering the trace-off between *security and privacy (or, more broadly, fundamental rights)*.

Recent Artificial Intelligence (AI) policies addressing FR have acknowledged the importance of this trade-off. The European AI Act proposal [8]mandates a *proportionate* and *strictly necessary* use of *real-time remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement* and requires that their deployment shall be subject to prior authorisation by a competent authority. The World Economic Forum has also called for *responsible limits on facial recognition* [21] in *law enforcement investigations*, highlighting its *necessary and proportional use*.

Fig. 1 illustrates four intervention alternatives that could

The views expressed in this scientific publication are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

979-8-3503-9494-8/24/$31.00 ©2024 IEEE

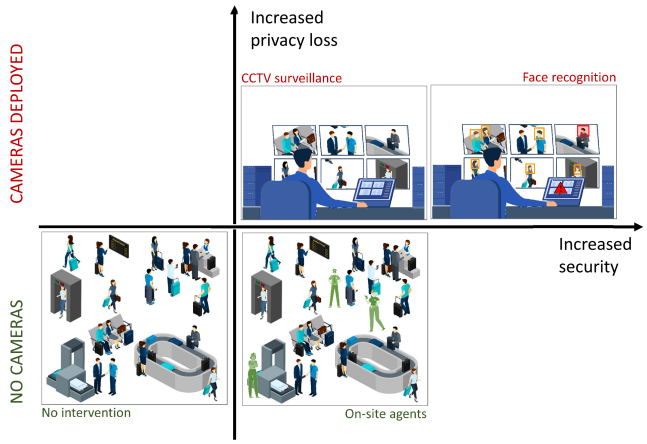


Fig. 1. Different types of interventions that can be considered for a law enforcement scenario (Source: pictures modified from [9], [10]).

be considered by authorities when confronted with a *security vs privacy* trade-off. *No intervention* might be the best suited decision in contexts with no or very limited security needs. *On-site agent intervention*, i.e., placing police agents to patrol the site, could be an alternative when security needs are higher and it is not possible to deploying cameras. *CCTV surveillance* can be a suitable solution if cameras are available on site and real-time human supervision of streaming videos is deemed sufficient to ensure the required level of security. Finally, *face recognition intervention* would additionally make use of an FR system to automatically analyze videos in search of faces on a watchlist and send identification alarms to security bodies. This is the most privacy-invasive solution, although it might be needed in case of severe security threats.

To the best of our knowledge, thus for no concrete guidelines on how to address the concept of *proportional use* in FR deployments have been developed. Authorities would benefit from them to formalize, visualize and guide their decision on whether the deployment of an FR system is proportionate or not in a given situation. This paper proposes a 2D framework for this assessment. First, the main quantitative and qualitative variables relevant to the FR deployment decision (e.g., number of people in the scene, scale of the threat, consequences on individual rights and freedoms) are identified. Then, a 2D model making it possible to weight these variables in terms of ethical cost (including privacy and related fundamental rights) vs security gain is proposed. The framework is designed to support decision makers confronted with the choice of deploying FR or not. Finally, the model is applied different face recognition scenarios inspired by real-world deployments, for the purposes of simulation and validation of the proposed framework.

II. Background

A. The citizen perspective

People make use of face recognition in their everyday life. For instance, FR is commonly used to unlock devices such as smartphones, access e-bank accounts, pass controls at airports or tag friend in social networks. However, when it comes to scenarios that are not so well-known or used on a daily basis, including the large-scale use of FR by law enforcement authorities for security purposes, which is the focus of this paper, studies reveal deeper reluctance to use and mistrust.

Seng et al. [32] recently analyzed people's perception of FR in 35 different scenarios, ranging from device unlocking to financial transactions, personalized marketing, control of attendance and surveillance at public events. They showed 35 FR scenarios to 314 participants in the form of vignettes, and asked questions related to usefulness, comfort level and privacy concerns. Their results confirm that perceptions of FR are strongly dependent on the specific context in which it is applied. Participants feel more comfortable in scenarios where they trust the entities collecting their facial information and where this information is stored in their personal devices, which gives them a sense of control over their sensitive data. Another key finding of this study, also raised in [14], is that users who do not find a clear benefit in the use of FR in a given scenario tend to consider the technology as an invasion of privacy. Indeed, from the 35 scenarios, only two of them relate to the large-scale use of FR at public events. They differ in one aspect: while the objective of the first one is left open ("*FR is used to track people attending a public event*"), the second one specifies that the purpose is "*public safety and law enforcement*". Participants found the second scenario to be more useful and reported feeling more comfortable compared to the one that did not state the purpose behind FR surveillance. This is in accordance with *the social contract theory* [24], which states that individual privacy often needs to be sacrificed for the greater good such as national security.

In addition to the benefit perceived in the use of FR for public safety purposes, recent studies have analyzed citizen trust on law enforcement agencies as the entities behind FR deployments. A survey with 4,109 adults run by the Ada Lovelace Institute [15], and another one with 2,291 participants by the Monash University [2] showed that, although people have certain fears and there is no unconditional support for police use, they are open to the use of the technology for law enforcement purposes as long as there is a demonstrable public benefit, as well as regulations and privacy safeguards in the management of biometric data. Nevertheless, the public perception of the use of FR for law enforcement purposes is found to be closely related to cultural background. A study on public attitudes towards face identification in criminal justice in the USA, China, United Kingdom and Australia [29] found that US respondents are more accepting of citizen tracking, even though they are less trusting of the police than people in the UK and Australia. This illustrates the need to take into account the cultural perspective in decisions related to FR deployment by public bodies. The intertwining of culture and education is also important. As highlighted in [14] relevant aspects such as the lack of knowledge of these systems by citizens (e.g., their working principles, limitations and applications) are closely

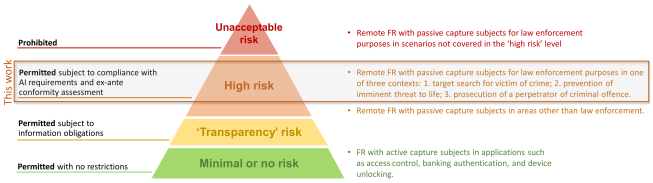


Fig. 2. Illustration of the four risk levels proposed by the AI Act with the corresponding applications of face recognition. This paper focuses on the high-risk use of FR for law enforcement purposes (highlighted with an orange box).

connected to acceptance of this technology.

B. The policy perspective

Recent global policies refer to the *proportional* use of FR technologies. This section considers two relevant examples. The first one is the European AI Act [8] proposal, aiming at trustworthy and safe development, implementation and use of AI systems. The AI Act adopts a risk based approach where AI systems are subject to different requirements according to their risk level, which is linked to their context of use and depends on how the system may impact fundamental rights. When this paper was written, the proposal was being refined by the European co-legislators, therefore there could be modifications to the following summary. We refer here to the European Commission's proposal published in 2021, which defines four risk levels: (1) Prohibited or unacceptable risk; (2) High-risk, where AI systems are subject to a set of requirements including, for example, the implementation of risk mitigation measures, appropriate levels of accuracy, robustness, cybersecurity, data governance, technical documentation and human oversight strategies; (3) Transparency risk, implying only information obligations; (4) Minimal risk, where AI systems are permitted with no restrictions. Hupont et al. [14] analyze the landscape of facial processing applications, linking them to different risk levels like in the AI Act proposal. In terms of FR (Fig. 2), the study identifies as low risk applications those intended to verify a person's identity provided that the subject has an *active* role. This includes applications for access control, banking authentication or device unlocking. At the other side of the risk dimension, the study considers FR scenarios where subjects have a *passive* role (referred to as *remote* scenarios), which are linked to high or unacceptable risk based on the context.

The proposal pays special attention to *real-time remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement*, which would include the real-time use of FR for law enforcement purposes studied in this paper. Article 5 explicitly regulates the use of FR for law enforcement purposes in points 1 to 3. Article 5(1d) states that this use is only allowed as far as it is *strictly necessary* for one of the following three objectives: *1. the targeted search for potential victims of crime; 2. the prevention of a specific, substantial and imminent threat to life or terrorist attack; and 3. the localisation, identification or prosecution of a perpetrator or suspect of a criminal offence*. The AI Act introduces security as an aspect to consider for the proportional use of FR, as Article 5(2) further specifies that this deployment shall take into account *the seriousness, probability and scale of the harm caused in the absence of the use of the system* and *the consequences of the use of the system for the rights and freedoms of all people concerned*. Finally, Article 5(3) states that in any case this type of deployment shall be subject to a prior authorisation by a judicial or other relevant authority. Although as mentioned before, there could be modifications regarding FR in the final text, the proposal mentions its proportional use by authorities in law enforcement scenarios.

Another relevant initiative at the international level is led by the World Economy Forum (WEF), which has recently developed a policy framework made of a set of principles for the use of FR in law enforcement [21]. The proposal identifies *necessary and proportional use* as one of the principles to be followed, which is related to the trade-off between security threats and fundamental rights. It states that *the decision to use facial recognition technology should always be guided by the objective of striking a fair balance between allowing law enforcement agencies to deploy the latest technologies, which are demonstrated to be accurate and safe, to safeguard individuals and society against security threats, and the necessity to protect the human rights of individuals. As a general principle, FR is considered to be linked to a cause and need as otherwise it would undermine human and fundamental rights.* This principle also refers to the need to document and justify the deployment of FR, specifying the classes of crimes or investigations for which its use is acceptable and/or lawful, and limiting the collection of images from public and publicly accessible spaces in terms of area and time period. In particular, it calls to consider alternatives to the use of FR and to ensure that its use is appropriate, limited and exclusively related to investigative purposes.

Even though they are not policy initiatives, this section includes the efforts made by some private companies, research institutions and public sector organizations around the world to build ethical principles and guidelines for AI. There is no consensus yet about the actual constituent elements of AI ethics, but the exhaustive analysis of 84 AI ethical principles/guidelines carried out by Jobin et al. [16] finds that a global agreement is emerging around the following key principles: transparency, fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy and trust. These principles therefore apply to face recognition systems and are, indeed, aligned with both the aforementioned specific FR policies and citizens' concerns.

III. Intervention Models from Other Fields

Two dimensional (2D) frameworks have been widely used as a simple but robust tools for resource allocation and policy intervention decisions in different fields as varied as Meteorology [35], [38], Economy [23] or Medicine [1], [26]. In general, these frameworks compare costs vs benefits to assess the desirability of a project, a decision, or any other

type of intervention. Although the words *cost* and *benefit* might sound purely economic, it should be noted that the trade-off between these two terms does not necessarily have to be monetary. For example, the cost of implementing an intervention or not can also be ethical (e.g., losing a fundamental right such as privacy) or medical (e.g., contracting a disease). Some 2D frameworks from other fields that have inspired this paper are described below.

Wilks proposes an economic cost/loss framework for weather forecast conceived for decision-makers [38]. On the one hand, this framework considers cost C of implementing measures (i.e., to intervene) to protect against the effects of a potential severe weather condition and probability forecast p that such event occurs. On the other hand, the occurrence of adverse weather events without this intervention would result in damage loss L. The intervention is considered to be economically viable when the cost/loss ratio is below the probability of occurrence of the adverse weather event, i.e., $\frac{C}{L} < p$. Thus, this framework transforms a weather forecast into a GO/NO-GO decision. Also related to weather, Keith [17] proposes a flight deviation intervention model in the case of severe climate threats. The intervention based on an adverse forecast involves loading additional fuel to reach an alternative airport. If no protection measures are taken and the event occurs, the flight should return to the airport of departure paying a higher cost in terms of fuel and delays.

The field of Medicine has been using Cost-Effectiveness Analysis (CEA) for a long time to decide whether intervene when confronted with a health threat. Decisions such as the allocation of extra health care resources [31] or population vaccination (e.g., for COVID-19 [19]) are assessed. Black [5] proposes a visual approach to CEA in Medicine by using a 2D plane, where the x-axis represents effectiveness (E), and the y-axis is cost (C). It defines a linear function with slope $K > 0$, representing the maximum acceptable cost/effectiveness ratio, which splits the space into two regions. An intervention strategy is considered cost-effective if it provides more effectiveness than costs. Geometrically speaking, this implies that the point in the 2D plane representing the intervention is located at the bottom of the space where $E > \frac{C}{K}$. Two alternative intervention strategies, I_1 and I_2, can be evaluated in terms of their distance to the line with slope K, to decide which one is more cost-effective. It is important to highlight that, in the case of Medicine, the cost is economic, but the benefits are purely health-related (e.g., cost per COVID-19 contagion averted).

This paper defines *intervention* as the decision whether to deploy an FR system for law enforcement purposes as a protective action in the case of imminent public threat. While most papers focus on analyzing and improving the accuracy and performance of FR models, the assessment of real-world interventions has been widely ignored. As seen above, many factors might affect this decision. Just like a severe weather threat case, FR watchlist suspects can cause varying levels of loss, damage, and harm affecting society from an economic and human life perspective. However, while severe weather cannot be stopped, a watchlist suspect can. Another factor

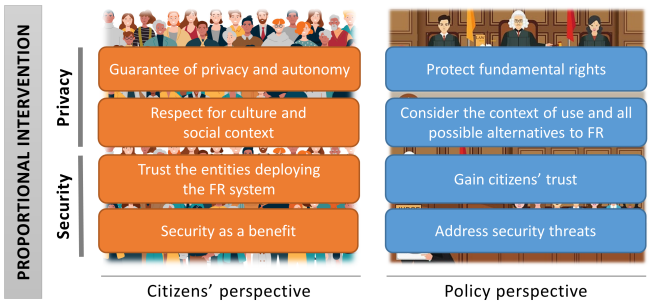


Fig. 3. Key elements that a face recognition intervention decision must weigh, according to citizen and policy needs.

to consider is the specific context in which the intervention would take place (e.g., in an indoor/open, more or less crowded space). In addition, FR deployment pays an ethical cost in terms of privacy-related fundamental rights [36]. This represents a trade-off between ethical concerns and security needs, as represented in Fig. 3.

IV. Proposed Framework

The cost/loss policy paradigm is used as a basis to propose a 2D graphical framework to assess the proportional and adequate use of FR in relevant contexts. It considers key factors, such as type of surveillance scenario, security risk, citizens' privacy concerns, and intends to assist authorities to arrive at an intervention decision.

The decision framework consists of two elements: a static 2D plane with *Privacy Loss* vs *Security Harm* variables, and a dynamic function s_i driven by the implementation details.

A. Proportional 2D-plane

Our framework is based on a 2D cartesian plane modeling the proportional use of an FR intervention. The y-axis considers the ethical cost of deployment, related to privacy and potential breaches of fundamental rights and associated with a loss of citizen trust in authorities. The x-axis represents the level of harm of the security threat, i.e., the harm potentially caused by not finding the individual on the watchlist (e.g., threat of not finding a criminal or a missing person). It should be noted that the economic dimension is intentionally not considered in our framework, as the focus is exclusively on the ethical aspects of FR interventions.

As the ethical cost is strongly driven by invasion of privacy, the y-axis dimension has been named *Privacy Loss*. Its p value is formalized as mainly dependent on two variables:

- d – the density of people (e.g. people/hour) circulating in the deployment site and thus subject to FR. The higher the density of people under FR surveillance, the higher the overall loss of privacy.
- c – the ethical cost linked to the site of deployment, which might be considered differently depending on its characteristics (e.g., public open space, indoor space, critical infrastructure), the intensity of surveillance (e.g., number of cameras in place, area covered) and the cultural context (e.g. benefit perceived by society in the country of deployment).

The x-axis dimension has been named *Security Harm*, representing the value of harm h, which could be mitigated by an FR deployment in site i. It covers both potential material and human harm with varying levels, from physical harm to human lives, and it depends on d and variable l, representing the level of harm that the individual(s) being searched could potentially imply or cause.

Table I provides some examples of scenarios and how they may be linked to different p and h values respectively. It should be noted that, even though the framework's dimensions are conceived as continuous, conceptual values are provided (p_n for *Privacy Loss* and h_m for *Security Harm*) as their concrete numerical value might need to be adapted to the particular context of deployment including for instance cultural considerations.

Privacy Loss		
Privacy	Var	Description
p_1	d+ c+	FR deployed in a public open space with moderate people flow density (tens to hundreds of people per hour), i.e. streets, squares, neighbourhoods, etc.
p_2	d++ c++	FR deployed in an indoor space with a moderate people flow density (hundreds of people passing by per hour) whith restricted access. Examples: airports or stadiums where people may enter with a ticket, such as a football match or musical concert.
p_3	d+++ c+++	FR deployed in a critical infrastructure with a high people flow density (circulation of hundreds to thousands of people per hour). This scenario could be, for instance, a mall, train, bus or metro station.
Security Harm		
Harm	Var	Description
h_1	l++	Security issues involving human life such as murder, kidnapping, or missing people.
h_2	l+++	Security issues concerning terrorists attacks linked to many human lives.

TABLE I

EXAMPLES FOR DIFFERENT LEVELS OF *Privacy Loss* p AND *Security Harm* h. THE '+' SIGN INDICATES THE VARIABLE'S LEVEL OF CONCERN.

In the case of *Privacy Loss* values p_n, index n increases with the level of invasion of privacy. Its lowest value p_1 represents a scenario involving a small-to-medium-sized group of people captured by the FR system in outdoor spaces. In the second privacy level, p_2, we consider scenarios involving a larger flow of people but this time in indoor spaces such as stadiums, airports or concerts. In this kind of spots, it is common for authorities to implement security measures at the entrance such as asking for IDs or tickets. Moreover, severe security incidents have recently occurred in these scenarios, which has raised awareness and fear in the population [27], consequently making them more open to FR intervention for the sake of security. The highest privacy level considered, p_3, is directly associated with FR intervention at so-called *critical infrastructures* which include bus, metro or train stations [11] with very high circulation of people. Serious security incidents have also recently occurred in these scenarios. In these contexts, hundreds or even millions of people might be walking in front of the FR system everyday, unaware of the fact that their faces are being matched against those on a watchlist. As for *Privacy Loss*, two levels are defined for *Security Harm* h_m, where index m increases with the severity of the harm that an individual on the watchlist might perpetrate. Table I describes the two proposed levels of *Security Harm* h_1 and h_2, which are linked to human lives, distinguishing between murders/kidnapping/missing people and terrorist attacks, respectively. Note that in the case of kidnapping, the individual(s) being searched may be either the kidnapper(s) or the kidnapped person (or both). The rationale behind these harm levels is that the search for suspects of kidnapping, murders and terrorist attacks may be deemed sufficient to justify an FR intervention. Similarly, preventing any of these events, especially when there is a high probability of appearance w of the searched person(s), may also be considered a justification for deployment as a protective action. Article 5, point d of the European AI-Act, considers harm levels in Table I within the permitted *"use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement"*[8]. Note that the security issues involving material damage, such as robbery or property damage, are considered in this study as a non-proportional use of FR.

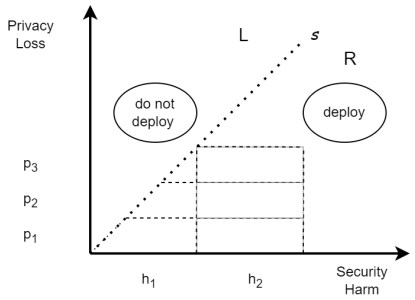


Fig. 4. Proportional 2D-plane proposed for FR intervention assessment and dynamic function s spliting the plane into Deploy and Not-Deploy areas.

Just like the cost/effectiveness 2D visualization proposed in [5] for the assessment of medical interventions, Fig. 4 depicts our 2D plane which is divided into two regions by identity function s. In region R the value of *Privacy Loss* p is below that of *Security Harm* h, and thus the use and deployment of FR may be deemed proportional. In region L, *Privacy Loss* is above *Security Harm* and FR deployment should be considered, in principle, non-proportional.

Fig. 4 also divides the discretized 2D space into rectangular regions or blocks based on the defined values of *Privacy Loss* p_n and *Security Harm* h_m in Table I. The height/width (H/W) ratio of the blocks drives the graphical analysis of the *Privacy Loss* vs *Security Harm* trade-off proposed in this paper, which is further illustrated in the following sections.

Authorities facing this myriad of scenarios and having the responsibility to authorize an FR intervention would benefit from a decision framework helping them to weigh all these variables for citizen security. The proportional 2D-plane

considers all those highly complex variables and provides a graphical and intuitive 2D representation to address the intervention decision.

B. The dynamic implementation function

Fig. 4 depicts identity line function s dividing the 2D-plane into "deploy" and "not-deploy" regions. In practice, the proposed framework uses a new dynamic implementation function, s_i, which will depend on the following variables:

- w – probability that the individual(s) on the watchlist may appear in scenario i. This information might be provided, for example, by authorities or intelligence agencies based on previous investigations.
- r – FR system's reliability, for instance, in terms of false positives/negatives, false positive identification rate (FPIR) and demographic bias issues. For example, a false positive could result in the arrest of a wrong person and the consequent mistrust in the authorities.
- t – period of time when the system is deployed (e.g., 24/7 in a venue, for a limited length of time during an event).

Thus, the dynamic function relates *a priori* knowledge about the watchlist individual(s) in variable w, specification about the deployment in variable t, and details of the FR system in variable r. This function is defined with the following equation:

$$s_i(h) = w \cdot h^r - t \tag{1}$$

where w is a probability variable defined in the range of $(0, 1)$, and consisting on the following events: 0.0 (Do not occur), 0.1 (Very unlikely to occur), 0.3 (Unlikely to occur), 0.5 (May occur), 0.75 (Likely to occur), and 1.0 (Very likely to occur). Variable r is also defined in range $(0, 1)$ and can be associated with the F1-score of the FR system. Finally, variable t takes discretized values $[0, 0.25, 0.5]$ representing an FR deployment for a period of less than one week, a couple of weeks, and more than one month, respectively.

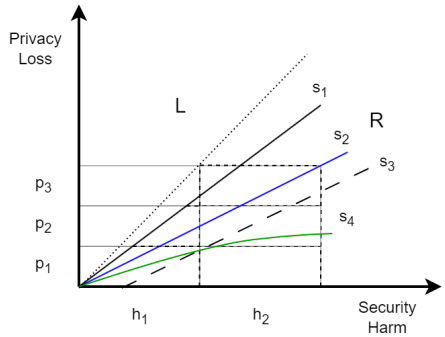


Fig. 5. Examples of dynamic functions for different variables.

Fig. 5 shows of different dynamic functions with different variable values. Dynamic functions s_1 and s_2 have the following values $r = 1$, $t = 0$, and $w = 0.75$ and $w = 0.5$ respectively. Function s_3 has these values $r = 1$, $w = 0.5$, and $t = 0.25$, showing the same slope as s_2 with a rightward displacement, and the *Privacy Loss* concern increases because of the longer deployment of the FR system. Finally, s_4 is defined by $r = 0.75$, $w = 0.5$ and $t = 0$.

For the sake of simplicity, in the following examples provided in this paper, variables $r = 1$ and $t = 0$ will be used, to work with straights lines without displacement.

C. To deploy or not to deploy

Both, the proportional 2D-plane (section IV-A), and the dynamic function (section IV-B) determine the framework to address the intervention decision for a given law enforcement case. Thus, the site of intervention and the watchlist individual(s) indicate the coordinates (h_m, p_n) of the corresponding block in the 2D plane grid, as depicted in Fig. 4. The specific variables of this case and the FR system shape the dynamic function, which will split the plane into a deployment region and a non-deployment region. But instead of focusing on the entire R plane, the analysis is made at the block level.

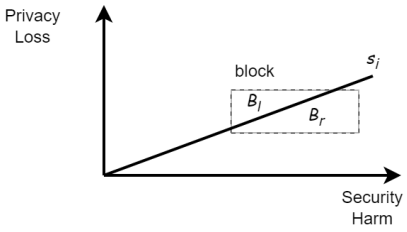


Fig. 6. Block analysis based on surfaces B_l and B_r at position (h_m, p_n).

Fig. 6 shows the graphical decision-making procedure. Dynamic function s_i splits block B in two areas: B_l and B_r. Then, the 2D framework rule defines that the FR intervention on site i is proportional if and only if $B_r > B_l$.

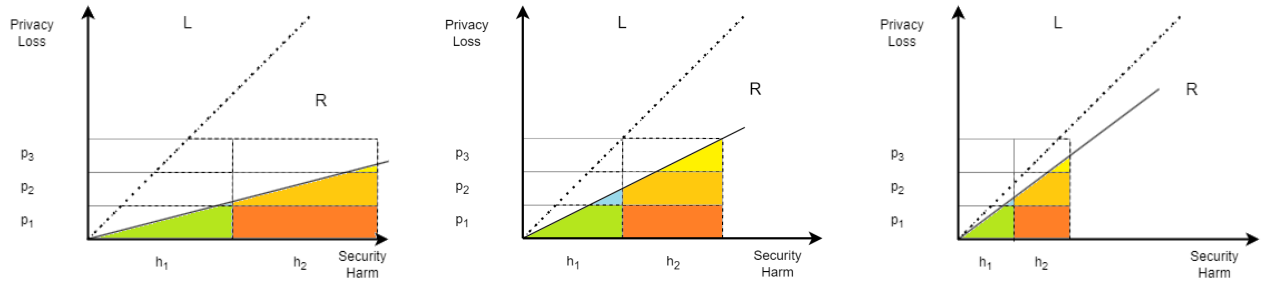
Now, we can return to Fig. 5 to evaluate the proportional decision based on the different dynamic implementation functions. Let us take block (p_3, h_2). FR deployment with function s_1 determines an intervention decision, i.e., $B_r > B_l$, while function s_2 does not. As we know, the difference between both functions is appearance probability w. A low w at this high Privacy Loss level rules out the deployment decision.

Taking now block (p_1, h_1), dynamic implementation function s_2 determines an intervention decision, but functions s_3 and s_4 do not. This time, the difference for s_3 is a longer deployment time, and for s_4 it is a lower F1-score performance of the FR algorithm.

D. Cultural Contexts

As mentioned before, public perception of facial processing applications on a wide range of scenarios concerning social good highly relates to cultural background [29], [14].

Fig. 7 illustrates how the 2D plane would be used to drive an FR intervention decision in different cultural contexts. From fig. 6, the intervention or non-intervention decision is based on the analysis of areas B_l and B_r, showing the relevance of the height/width (H/W) ratio of the block. Graphically, the FR deployment would be allowed when the



$s_i@w = 0.25$ in tolerant context $s_i@w = 0.50$ in moderate context $s_i@w = 0.75$ in conservative context

Fig. 7. Graphic illustration of how the proposed 2D plane would be used for the assessment of a FR intervention in sites i with different associated s_i dynamic function and H/W ratio.

colored filled area in a block is greater than the uncolored area, i.e., when $B_r > B_l$. Fig. 7 depicts examples of different H/W ratios modeling different society perspectives about FR systems and Privacy Loss. In the three examples, blocks (p_1, h_1), (p_1, h_2), (p_2, h_2) fulfill the $B_r > B_l$ condition, and blocks (p_2, h_1), (p_3, h_2) do not follow the rule. Values of w leading to a similar intervention behavior are chosen, while the H/W ratio changes, with $r = 1$ (meaning perfect FR performance), and $t = 0$ (implying that the FR system will be deployed for a short time).

The first example (Fig. 7-left), with $H/W = \frac{3}{13}$, representing a context with high *Security Harm* concerns, permits an FR intervention in block (h_2, p_2) even with a low probability of subject appearance $w = 0.25$, as shown by the corresponding dynamic function s_i. This context accounts for a tolerant society towards FR systems and a moderate *Privacy Loss* concern. The second moderate context (Fig. 7-center) has a ratio of $H/W = \frac{3}{9}$. FR intervention in block (h_2, p_2) would be deemed proportional when the probability of appearance is $w = 0.5$ representing a reasonable value to deploy an FR system. The last example (Fig. 7-left), where $H/W = \frac{3}{5}$, accounts for a more conservative policy context in terms of *Privacy Loss* preservation. In this case, FR deployment in block (h_2, p_2) would only be worth with a very high probability of appearance of watchlist individual. It means a dynamic function s_i with $w = 0.75$.

These examples demonstrate how cultural differences can drive the decision whether to intervene with FR or not, as well as the importance for policy makers to adequately address the *Privacy Loss* vs *Security Harm* trade-off. Some countries and their citizens might be more willing than others to sacrifice part of their privacy in return for increased security. However, thus far no concrete studies and figures on this matter have been developed.

V. The Framework in Practice

In the following section, the graphical 2D framework is applied to assess different types of FR interventions inspired by three real-world law enforcement scenarios.

1) Metropolitan Police Service Live Facial Recognition Trials: In 2020, London's Metropolitan Police Service presented a report about the deployment of Facial Recognition between August 2016 and February 2019[22]. The report details ten deployments in public spaces and a set of metrics of interest for a complete evaluation, including: duration, average of recognition opportunities, watchlist size, number of false alarms, number of people engaged by a police officer, and the number of actions/arrests. Our focus will be on two of these trials, which used the same software version of the FR algorithm and similar equipment (surveillance camera). Firstly, the "Stratford Westfield 28 June 2018" trial concerns a deployment on the street furniture for 6 hours. The watchlist had 486 'Wanted Missing' individuals chosen by geographic area (proximity to Westfield Stratford). The deployment produced 5 alerts over 10,000 detected and evaluated people. It is worth mentioning here that the FR alerts (a match over a threshold) follow an operator adjudication process. An operator is a qualified officer who has received advanced training in the facial recognition system and its features. Thus, only one of these alerts resulted in engagement by an agent, but no action/arrest was performed. Secondly, the "Romford February 2019" trial street deployment consisted of a 6:45 hour surveillance, detecting 10,100 pedestrians, with a larger watchlist of 1996 people including individuals wanted for violent offenses and filtered by geographic area. The deployment returned 13 positive matches, which on 3 occasions led to an arrest. In our framework, this *Privacy Loss* scenario can be considered as p_1, given that it is deployed in a public open space with a moderate people flow density as defined in Table I. The variables to define dynamic implementation function s_{met} are: $t = 0$ (limited time), $r = 0.85$ (obtained from detailed performance statistics on the report, such as: False Alarms, and Positive Identifications at each trial), and probability $w = 0.3$ (unlikely to occur), which is a moderate value, because the watchlist is filtered by geographic area (individuals living in the watched neighborhood). The information about the level of harm associated with the individuals on the watchlist is missing. This would allow authorities to determine the security harm value and, therefore, its proportionality, according to our model. In fig. 8 this scenario is depicted, with (p_1, h_1) and (p_1, h_2) blocks colored under the dynamic function $s_{met}(h)$. At this value of w, both areas below $s_{met}(h)$ indicate an *Intervention* recommendation, but only if the *Security Harm* caused by individuals on the watchlist corresponds to h_1 and h_2 levels.

2) Arrest of Terrorist Suspect in London: A 21-year-old member of the British Army turned suspected terrorist and spy in January 2023 was sent to HMP Wandsworth prison. He escaped from the prison on the morning of Wednesday 6 September and was recaptured on Saturday 9 September. The four-day search was coordinated at the Counter Terrorism Operations Centre (CTOC) in West Brompton, central London [34]. The situation room at the center had access to "cutting edge" spy technology including facial recognition, a CCTV camera network, and phone tracking data. This case represents a real scenario with one individual on the watchlist of the FR system. Technical information about the deployment of the FR system is not available. However, this kind of search involves the use of FR in scenarios with different privacy losses: p_1, p_2, and p_3. While the level of harm cannot be exactly defined, the charges against the individual involve national security, and, thus, it can be assigned h_2. Other variables of the 2D framework can also be estimated to draw the dynamic implementation function s_{run}. The time parameter was less than one week, i.e., $t = 0$. Variable r can be considered as $r = 0.9$. Finally, if it is considered that the FR deployment is performed in places where the public reported sighting of the suspect, appearance probability is $w = 0.75$ (likely to occur). Also, scenarios involving p_3 typically correspond to those places where a fugitive in the run can show up and escape using subways, trains, or plains. In this case, deployment could be considered within the area of proportional use, if it is associated with a high potential harm, which should be regarded with respect to a maximum privacy threshold. Fig. 8 shows blocks (p_1, h_2), (p_2, h_2) and (p_3, h_2) colored under the dynamic function validating the *Intervention* recommendation.

3) Brøndby IF's STADIUM: Brøndby IF is a professional football club in the Danish Superliga [28]. In the summer of 2019 at Brøndby IF's stadium, Panasonic installed "FacePRO," a facial recognition system. Individuals who have been caught breaking stadium rules are banned from coming back to games and are registered on a watchlist. Brøndby IF has an average home game attendance of roughly 14,000 people, and approximately up to 100 people are registered on the watchlist on average. For the graphical 2D framework, this scenario would correspond to a p_2 level of privacy, given that it is deployed in an indoor space as specified in Table I. The time variable is $t = 0$ because the deployment corresponds to a short period (the match time). The appearance probability has a relatively high value, i.e., $w = 0.5$, because the fans are likely to be present at the match. The technology on the FR model is based on [39], and the evaluation point from the official evaluation report of the National Institute of Standards and Technology (NIST) states $r = 0.95$. However, when the level of security harm caused by individuals on the watchlist, cannot be determined from Table I, as the subjects are banned for violent behavior not for severe crimes. Our 2D graphical framework does not place the scenario in the *Intervention-Non Intervention* plane, which means that the FR deployment is not recommended. Thus, other types of interventions that can be envisaged from Fig. 1.

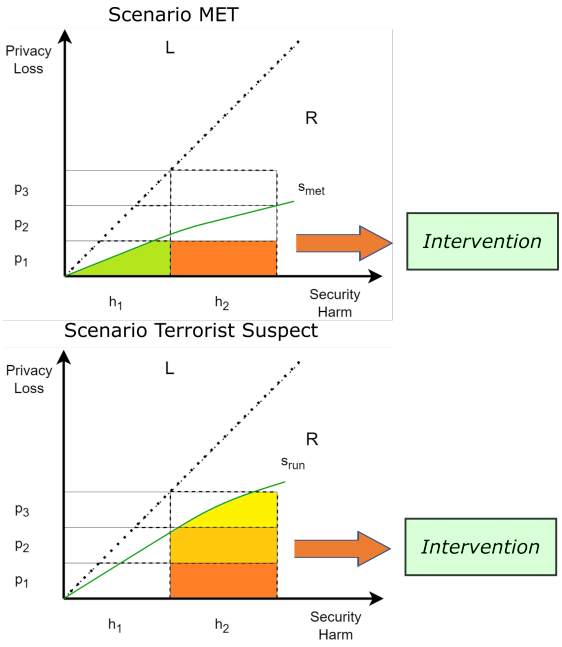


Fig. 8. Framework in practice.

VI. Conclusions and Future Work

A 2D graphical framework was proposed to assess the proportional use of FR systems in real-world scenarios, grounded on an ethical cost vs security gain model. The two dimensions consider variables from recent studies and policies on face recognition and related citizen privacy concerns. To the best of our knowledge, this is the first framework addressing the problem of FR intervention, which might have a high impact for decision makers and lead to new research considering the principle of proportionality in FR. It will also hopefully contribute to open discussion, in line with worldwide regulations such as the European AI Act [8], on the proportional and strictly necessary use of FR technology.

Our framework has, however, some limitations. In its practical implementation, a simple linear approach has been used with a broad discretization of the 2D plane into large intervention blocks. This model can be improved by incorporating in its design stakeholders directly involved in FR deployment (e.g., citizens, decision makers, etc.). To address this, future work may include developing simulations of different FR scenarios and conducting a large-scale user survey to understand which of them are deemed proportional as well as culture-related information. This will allow us to come up with a more fine-grained mathematical model taking advantage of the continuous nature of the variables, such as the H/W ratio. Indeed, the framework needs to identify different cultural and ethical preferences by countries or world regions. Future work should also test the usability and usefulness of the framework with policy-makers and authorities.

REFERENCES

[1] J. P. Anderson, J. Bush, M. Chen, and D. Dolenc. Policy space areas and properties of benefit-cost/utility analysis. *Jama*, 255(6):794–795, 1986.

[2] M. Andrejevic, R. Fordyce, L. Li, and V. Trott. *Australian attitudes to facial recognition: a national survey*. Clayton Victoria Australia: Monash University, 2020.

[3] S. Balakrishnan, S. Chaudhuri, and V. Narasayya. Autotag'n search my photos: leveraging the social graph for photo tagging. In *Proceedings of the 24th International Conference on World Wide Web*, pages 163–166, 2015.

[4] G. Barquero, C. Fernández, and I. Hupont. Long-term face tracking for crowded video-surveillance scenarios. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2020.

[5] W. C. Black. The CE plane: a graphic representation of cost-effectiveness. *Medical decision making*, 10(3):212–214, 1990.

[6] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1578–1587, 2022.

[7] L. R. Carlos-Roca, I. Hupont, and C. Fernández. Facial recognition application for border control. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2018.

[8] European Commission. Proposal for a Regulation on Artificial Intelligence. Available: `https://bit.ly/3PGxraY`, 2021. [Online; accessed May 19, 2024].

[9] Freepic. Monitoring screens. `https://bit.ly/3LsxdSd`. [Online; accessed May 19, 2024].

[10] Frepic. Airport passengers. `https://bit.ly/3EHDsOn`. [Online; accessed May 19, 2024].

[11] D. Gritzalis, M. Theocharidou, and G. Stergiopoulos. Critical infrastructure security and resilience. *Springer International Publishing*, 10:978–3, 2019.

[12] P. Grother, M. Ngan, and K. Hanaoka. *Face recognition vendor test (FRVT): Part 3, demographic effects*. National Institute of Standards and Technology (NIST), 2019.

[13] I. Hupont and C. Fernández. Demogpairs: Quantifying the impact of demographic imbalance in deep face recognition. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–7. IEEE, 2019.

[14] I. Hupont, S. Tolan, H. Gunes, and E. Gómez. The landscape of facial processing applications in the context of the european ai act and the development of trustworthy systems. *Scientific Reports*, 12(1):10688, 2022.

[15] A. L. Institute. Beyond face value: Public attitudes to facial recognition technology. `https://bit.ly/44ZLRr4`, 2020-07-24 2019. [Online; accessed May 19, 2024].

[16] A. Jobin, M. Ienca, and E. Vayena. The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1(9):389–399, 2019.

[17] R. Keith. Optimization of value of aerodrome forecasts. *Weather and Forecasting*, 18(5):808–824, 2003.

[18] H. Lee, S.-H. Park, J.-H. Yoo, S.-H. Jung, and J.-H. Huh. Face recognition at a distance for a stand-alone access control system. *Sensors*, 20(3):785, 2020.

[19] R. Li, H. Liu, C. K. Fairley, Z. Zou, L. Xie, X. Li, M. Shen, Y. Li, and L. Zhang. Cost-effectiveness analysis of bnt162b2 covid-19 booster vaccination in the united states. *International Journal of Infectious Diseases*, 119:87–94, 2022.

[20] F. Liu, M. Kim, A. Jain, and X. Liu. Controllable and guided face synthesis for unconstrained face recognition. In *European Conference on Computer Vision*, pages 701–719. Springer, 2022.

[21] S. Louradour and L. Madzou. A policy framework for responsible limits on facial recognition, use case: Law enforcement investigations. In *World Economic Forum*, 2021.

[22] Metropolitan Police Service. Metropolitan Police Service Live Facial Recognition Trials. Available: `https://bit.ly/3vjXe0N`, 2020. online.

[23] E. J. Mishan and E. Quah. *Cost-benefit analysis*. Routledge, 2020.

[24] A. D. Moore. *Privacy, Security and accountability: ethics, law and policy*. Rowman & Littlefield, 2015.

[25] P. Negri, S. Cumani, and A. Bottino. Tackling age-invariant face recognition with non-linear plda and pairwise svm. *IEEE Access*, 9:40649–40664, 2021.

[26] P. J. Neumann, G. D. Sanders, L. B. Russell, J. E. Siegel, and T. G. Ganiats. *Cost-effectiveness in health and medicine*. Oxford University Press, 2016.

[27] A. Oksanen, M. Kaakinen, J. Minkkinen, P. Räsänen, B. Enjolras, and K. Steen-Johnsen. Perceived societal fear and cyberhate after the november 2015 paris terrorist attacks. *Terrorism and Political Violence*, 32(5):1047–1066, 2020.

[28] Panasonic. Peace of mind on match day: facial recognition solution at the football stadium. Available: `https://bit.ly/47lWnKr`, 2019. online.

[29] K. L. Ritchie, C. Cartledge, B. Growns, A. Yan, Y. Wang, K. Guo, R. S. Kramer, G. Edmond, K. A. Martire, M. San Roque, et al. Public attitudes towards the use of automatic facial recognition technology in criminal justice systems around the world. *PloS one*, 16(10):e0258241, 2021.

[30] J. P. Robinson, G. Livitz, Y. Henon, C. Qin, Y. Fu, and S. Timoner. Face recognition: too bias, or not too bias? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–1, 2020.

[31] L. B. Russell, M. R. Gold, J. E. Siegel, N. Daniels, and M. C. Weinstein. The role of cost-effectiveness analysis in health and medicine. *Jama*, 276(14):1172–1177, 1996.

[32] S. Seng, M. N. Al-Ameen, and M. Wright. A first look into users' perceptions of facial recognition in the physical world. *Computers & Security*, 105:102227, 2021.

[33] T. Sutabri, A. K. Pamungkur, and R. E. Saragih. Automatic attendance system for university student using face recognition based on deep learning. *International Journal of Machine Learning and Computing*, 9(5):668–674, 2019.

[34] The Guardian. Surveillance centre hailed as critical in capture of escaped terror suspect. Available: `https://bit.ly/3TNZuaR`, 2023. online.

[35] J. C. Thompson. On the operational deficiences in categorical weather forecasts. *Bulletin of the American Meteorological Society*, 33(6):223 – 226, 1952.

[36] United Nations High Commissioner for Human Rights. The right to privacy in the digital age. Available: `https://bit.ly/48bNzbP`, 2021. online.

[37] Z. Wang, Z. Cheng, H. Huang, X. Zhou, and Y. Liu. Design and implementation of vehicle unlocking system based on face recognition. In *34rd youth academic annual conference of chinese association of automation (YAC)*, pages 121–126. IEEE, 2019.

[38] D. Wilks. A skill score based on economic value for probability forecasts. *Meteorological Applications*, 8(2):209–219, 2001.

[39] L. Xiong, J. Karlekar, J. Zhao, Y. Cheng, Y. Xu, J. Feng, S. Pranata, and S. Shen. A Good Practice Towards Top Performance of Face Recognition: Transferred Deep Feature Fusion, 2018.