Operational Fairness for Facial Authentication Systems

Mélanie Gornet, Claude Kirchner, Catherine Tessier

▶ **To cite this version:**

Mélanie Gornet, Claude Kirchner, Catherine Tessier. Operational Fairness for Facial Authentication Systems. ERCIM News, 2022. hal-03837716

HAL Id: hal-03837716
https://hal.science/hal-03837716v1

Submitted on 14 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Operational Fairness for Facial Authentication Systems

by Mélanie Gornet (Télécom Paris), Claude Kirchner (CCNE, CNPEN, Inria), and Catherine Tessier (ONERA/DTIS, Université de Toulouse)

How to design a facial authentication system, taking into account both performance and fairness? We consider the choices that a developer makes when coding such a system, such as the training parameters, the architecture of the neural network, or the authentication threshold. We evaluate their impact on the global fairness of the system, showing that fairness is not only affected by the training data but also by the multiple choices that are made when coding the model.

Numerous international recommendations have been issued over the past five years, listing values, principles and criteria to be considered during the development, and more generally the life cycle, of a machine learning system. These recommendations, although paving the way for standardised methods to design algorithms, do not explain how to actually implement these criteria. For example, what should researchers and engineers do to design "fair" machine learning based systems?

We focus on fairness through the eyes of a developer who has to design a facial authentication system. This study was conducted at the French National Committee for Digital Ethics [L1] and is going on as part of a doctoral research. The code is available on GitHub [L2].

Facial Authentication, Performance and Fairness

Automated facial recognition has been particularly criticised for reinforcing overall discrimination that exists in societies. For instance, it was shown that face analysis systems from big tech companies were misclassifying dark-skinned women much more often than light-skinned males [1]. This was later confirmed by the US National Institute of Standards and Technology (NIST) that conducted a study to quantify the accuracy of face recognition algorithms for demographic groups defined by sex, age, and race or country of birth, revealing significant discrepancies [2]. Yet, researchers have long reduced the problem of fairness to a data issue: "Garbage in, garbage out", if the data is unbalanced, the system is quite likely to be biased. But this mindset overlooks other parameters or coding choices that are also likely to affect fairness.

We define digital facial authentication as the comparison of recorded biometric data with those presented by a person. It is a one-to-one matching system, and its output is binary: if the output is "yes", authentication is validated, otherwise it is rejected. We have developed a system using a convolutional neural network (CNN), trained by triplet loss for facial authentication [3]. This process requires many technical choices that are usually made by the developer according to what yields the best performance. We have investigated seven of these choices (see Figure 1) through several metrics for both performance and fairness.

For model selection, a high performance corresponds to a low validation loss at the end of the training phase. For model validation and evaluation, it also corresponds to a high accuracy, a high triplet learned rate (TLR, a metric measuring how well the system has learned), and low error rates.

Fairness is considered here as having the same probability of being recognised by the system in similar conditions, whoever you are. This implies checking, as the NIST did, that for different subgroups of population the system has the same accuracy, TLR and error rates (group fairness). A discrepancy between two groups is significant if the 90% confidence intervals on a given metric do not overlap.

Study Results

Data processing:

- Surprisingly, the data sampling method that yields the best results for fairness measures is the random one, compared to the model prioritising certain underrepresented individuals.

* *All face images come from the dataset Labeled Faces in The Wild [L3]*

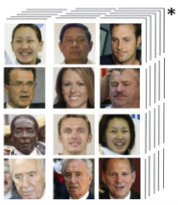
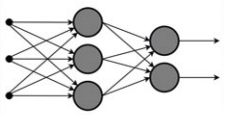
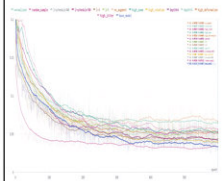
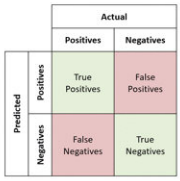
Data Acquisition	Data Processing	Neural Network	Training	Evaluation
				
- Image resolution - Number of images - Number of images per person - Labels - Consent - Acquisition technique - Balance - ...	- Normalization - Sampling technique - Augmentation - Number of batches - Train/Test separation - Framing - ...	- CNN - Triplet Loss - Initialization - Layers arrangement - Network depth - ...	- Loss function - Margin - Number of epochs - Hard mining - Regularization - Optimizer - Dropout - Earlystopping - Learning rate and scheduler - Mitigation techniques - ...	- Threshold - Pairs - Metrics - Groups - "Acceptable" value - ...

Figure 1: List of design choices for a facial authentication system and investigated choices (in green).

- Data normalisation seems to degrade measures on majority groups but does not affect minority groups.
- Data augmentation improves performance including for minority groups but widens gaps between groups; if data is not augmented, there are fewer gaps but to the detriment of performance.

Neural network:
- The depth of the network does not seem to affect fairness very much but still affects performance.

Training:
- Changing the margin of the loss function can improve fairness but results in a small reduction in performance.
- The choice of the learning rate and its scheduler can affect the local optimum the network will reach and thus yield very different results; here, the model that has the best performance is also the best for fairness.

Evaluation:
- The authentication threshold that separates positive and negative pairs strongly affects the error rates: a high threshold increases the number of matches but generates more false matches, whereas a low threshold prevents some people from being correctly identified. The value of the threshold should thus depend on the use case and on what type of error is the less harmful to the people involved.

Trade-offs

International recommendations about "the ethics of AI" hardly mention that all the proposed criteria cannot be met at the same time and that trade-offs are often necessary. Moreover, fairness is not only a data issue but involves the coding of the model itself. Therefore, ethical thoughts involving all the stakeholders should come with the design of machine learning systems, making the conflicts explicit and guiding the decisions concerning the code implementation as well as the main decision of whether or not to deploy such digital processes.

Links:

[L1] https://www.ccne-ethique.fr/
[L2] https://github.com/mgornet/CNPEN
[L3] http://vis-www.cs.umass.edu/lfw/

References:

[1] J. Buolamwini, T. Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", in Conf. on Fairness, Accountability and Transparency, p. 77–91, PMLR, January 2018.
[2] P. J. Grother, M. L. Ngan, K. K. Hanaoka, "Face Recognition Vendor Test Part 3: Demographic Effects", Technical report, December 2019.
[3] F. Schroff, D. Kalenichenko, J. Philbin, "Face Net: A Unified Embedding for Face Recognition and Clustering", in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), p. 815–823, June 2015.

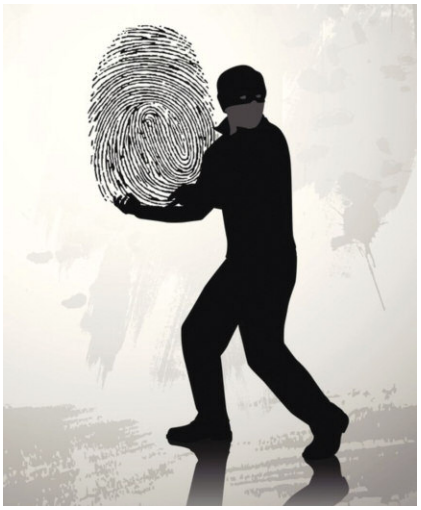
Please contact:

Mélanie Gornet
Télécom Paris, France
melanie.gornet@telecom-paris.fr

Enhancing Biometric Data Security by Design

by Bernhard Strobl (Austrian Institute of Technology, AIT) and Margherita Natali (United Nations)

This article will give an insight into some key problems and related solutions concerning the implementations of a privacy-preserving biometric matching system. We propose three by-design possibilities, strictly in compliance with human rights and data protection regulations, to improve the security of authentication systems: contactless fingerprint scanning, use of a distributed ledger system for biometric matching, and homomorphic encryption. These technical solutions would potentially constitute a step forward for governmental use of authentication procedures under the international security agenda while supporting ethically aligned design principles.



Identity management represents one of the key items on national and international security agendas. In the private domain the use of own identity is predominantly used for granting access in basic and common transactions or actions, whereas governments, in the public domain, more often implement such systems to manage social phenomena such as migration or the illicit activities of organised criminal groups. One of the most common uses of identity management on the global scale is the authentication of official identification documents (e.g., identity cards, passports, driver's licences, and other civil-registry-issued certifications) to monitor and facilitate the legitimate movement of individuals.

Authentication processes can be built upon three basic and very distinct pillars:
- What is known (password, passphrase, PIN, etc.)
- What is available (key, card, stick, document, QR Code, sign, etc.)
- Who the person is (biometrics: DNA, face, fingerprint, iris, veins, etc.)

Sometimes a combination of these pillars is chosen to perform a secure authentication. Depending on the application, different interests may shape the technological choice. For instance, in the case of a commercial service, the need for a speedy and