EU INNOVATION HUB FOR INTERNAL SECURITY

POLICY BRIEF

INDEPENDENT EVALUATION OF BIOMETRIC TECHNOLOGY IN THE EU:

STATE OF PLAY AND FUTURE OPTIONS

Contents

EXECUTIVE SUMMARY

Considering the widespread use of biometric recognition technologies, their potential and actual impact on society should not be underestimated. Decisions based on the output produced by a biometric recognition algorithm may have wide ranging repercussions for individuals. Therefore, it is essential that organisations deploying such systems ensure that they are accurate, reliable, implemented ethically respecting fundamental rights, and in compliance with the applicable legal framework.

Independent evaluation is one of the most effective means to ensure that biometric systems deployed in the EU meet the necessary requirements for accuracy, robustness and transparency, according to standard performance metrics. However, today, the evaluation of biometric technology in the EU is fragmented and inconsistent. Public authorities primarily rely on the results of work carried out by the U.S. National Institute for Standards and Technology (NIST).

This Policy Brief argues that **establishing an independent capability for the evaluation of biometric technology in the EU** will bear a number of benefits, including:

- support EU and national authorities in the acquisition of biometric recognition systems,
- support policy makers in regulating biometric technologies,
- support organisations deploying biometric technologies in compliance with the EU Artificial Intelligence Act and other applicable legislation,
- improved societal trust in biometric technology,
- and ultimately contribute to the EU's technological sovereignty.

The Policy Brief therefore recommends the **creation of two complementary initiatives** to support EU-level capability for biometric evaluation:

1. **common biometric data repository** as the first key step towards establishing an independent EU capability. The datasets contained in the repository should reflect the demographics and use-cases for the deployment of biometric technologies relevant for EU authorities and should be compliant with the relevant data protection regulations, e.g., EUDPR and GDPR.
2. **centralised biometric evaluation and testing platform** as the second step, to supplement the proposed data repository. This platform would be linked with the common biometric data repository and would also provide standardised protocols for the evaluation of biometric technologies.

Considering the sensitive nature of the datasets and the testing platform, **a secure data centre environment** should be set up and **managed by a trusted authority.**

1. INTRODUCTION

The **use of biometric recognition technologies** today **is widespread across public and private sector** activities. Smartphone users increasingly rely on either face or fingerprint recognition to access their device. When opening a new bank account online, the applicant may be asked to use their smartphone to provide a face image and a picture of the identity document. Such remote enrolment of new customers in financial services is also enabled by biometric recognition technology. Law enforcement authorities have been relying on biometric recognition technologies for over fifty years. And, of course, biometric recognition technologies are widely used in border control and migration management, including systems based on face, fingerprint or iris recognition.

Why is the use of biometric recognition technologies so common? In essence, biometric recognition technology has become foundational to **modern identity management systems** thanks to the balancing of speed, convenience, accuracy and security, and acceptance by users and society. The use of biometrics has introduced a new paradigm in identity management technologies: you are not recognised by what you know (e.g., password, PIN) or by what you have (e.g., card, key), but by who you are.

Considering the ubiquitous use of biometric recognition technologies, their impact on society should not be underestimated. Decisions based on the output produced by a biometric recognition algorithm may have **wide-ranging repercussions for individuals**. For example, as a result of human-made decisions based on outputs of biometric recognition systems, a person may be denied access to financial and healthcare services, investigated by law enforcement authorities, or denied entry into a country. Therefore, it is essential that the organisations deploying such systems ensure that these systems are accurate, reliable, and are implemented ethically, ensuring respect for fundamental rights, and in compliance with the applicable legal framework.

Independent evaluation is one of the most effective means to ensure that biometric systems deployed in the EU meet the defined requirements in terms of accuracy, robustness and transparency, according to standard performance metrics. However, today, there is limited capacity to perform independent vendor-agnostic evaluation of biometric recognition systems in the EU, and organisations deploying biometric recognition systems rely primarily on the evaluations carried out by the U.S. NIST[1].

This Policy Brief explains the benefits of independent evaluation of biometric technology and what it entails. It also provides an overview of independent biometric technology evaluation capabilities available internationally, and within the EU. It then presents some of the key **challenges and risks stemming from over-reliance on evaluations carried out in non-EU countries.** In place of conclusion, some options are proposed for the development of an EU capability for the independent evaluation of biometric recognition technologies.

The core argument of this Policy Brief is that the development of an independent evaluation capability for biometric recognition systems will not only help to ensure compliance with applicable EU law but will also reinforce the EU's technological sovereignty.

1 U.S. National Institute for Standards and Technology (NIST).

2. SCOPE OF THE POLICY BRIEF

2.1. Biometric evaluation versus biometric certification

It is important to differentiate between 'biometric evaluation' and 'biometric certification'. Although both terms and tasks are related and often sequential, there is a clear difference between them in terms of purpose, formality, authority and possible legal effect.

Biometric evaluation

In essence, a 'biometric evaluation' is a technical and scientific assessment of a biometric system or algorithm to measure its performance. The main objectives of biometric evaluations are to:

1. understand how well a biometric system performs;
2. measure accuracy, error rates, bias, robustness, and limitations;
3. compare systems or configurations;
4. support R&D, procurement, or policy decisions.

The parameters assessed in a biometric evaluation are accuracy metrics such as the 'false match rate' (FMR) and 'false non-match rate' (FNMR), performance bias across demographic groups, robustness to attacks or to low-quality samples, or data quality. Normally, the outcome of an evaluation is a report that describes the evaluation protocols and provides the results of performance measurements based on which it identifies strengths and weaknesses of the assessed algorithms or systems. In summary, biometric evaluation answers the question: ***'How well does this biometric system/algorithm perform under defined conditions?'***

Biometric certification

On the other hand, a 'biometric certification' is a formal, third-party conformity assessment that attests that a biometric system or algorithm meets predefined standards or regulatory requirements. Its main purpose is to:

1. demonstrate compliance with specific standards, regulations, or procurement requirements;
2. to support legal acceptance, market access, or operational deployment;
3. to provide assurance to authorities, users, and stakeholders.

A biometric certification can only be conducted by an **accredited certification body**, and it is based on fixed standards or schemes, so that it results in an official certificate which is in most cases time-limited, version- and data-specific. Typically, a certification assesses characteristics such as compliance with technical standards, minimum performance thresholds, security and robustness requirements, data protection and privacy safeguards, or process and quality management aspects. The outcome of the certification is a **formal/official certificate of conformity** with a clearly defined scope, validity period, and conditions. The certification process and its outcome may be the result of a legal obligation. In summary, biometric certification answers the question: ***'Does this biometric system comply with the required rules and/or standards?'***

The **present Policy Brief focuses solely on *'biometric evaluation',*** which is a task, for instance, carried out by NIST in the United States.

2.2. Types of biometric evaluations

The ISO/IEC 19795-1:2021[2] standard on biometric performance testing defines three types of evaluations:

1. **Technology evaluation:** only algorithms are evaluated under controlled laboratory conditions using a reference biometric database.

2. **Scenario evaluation:** end-to-end system performance is evaluated in standardised environmental conditions and with a standard population.

3. **Operational evaluation:** the biometric system is tested in a specific application environment using a specific target population.

Technology evaluations should be carried out on a regular basis, as a way to keep track of the evolution of the state of the art of a given technology, and to compare performance between different vendors. Therefore, technology evaluations are a valuable source of information for authorities procuring biometric recognition systems, as they help define realistic and specific performance requirements, and provide an objective ranking of vendors based on a comparison elaborated on the same benchmark (i.e., database, evaluation protocol and evaluation metrics).

Scenario evaluations measure end-to-end system performance in an environment that models a real-world target application. In operational evaluations, the goal is to measure the performance of a biometric system in a specific application environment using a specific target population. Both scenario and operational evaluations are more complex when it comes to execution and more demanding in terms of administrative resources. Scenario evaluations may also require significant upfront investment in testing infrastructure.

Although each of the evaluation types is valuable in its own right, **this Policy Brief focuses exclusively on technology evaluations**, which provide a clear view on the capabilities of a given biometric technology (e.g., face recognition or fingerprint recognition), and can be carried out by an independent institution over previously acquired data (under conditions as close as possible to the operational setting). Scenario and operational evaluation modalities are application-dependent, and therefore need to be conducted in an operational setting specifically designed for that evaluation. This fact limits the possibility for an external independent organisation to organise this type of evaluations, on a regular basis for technology monitoring purposes.

In particular, the document focuses on the evaluation of the **performance of biometric recognition workflow**, with special attention to the **comparison algorithm** and the evaluation of **biometric accuracy**.

This way, as mentioned above, the document does not discuss the evaluation of “end-to-end complete biometric systems” which are the focus of “biometric scenario evaluations” and “biometric operational evaluations”. These two latter types of evaluations (i.e., scenario and operational) may be the topic of a second phase of the building process of biometric evaluation capabilities within the EU.

2 ISO/IEC 19795-1:2021 IT— Biometric performance testing and reporting. Part 1: Principles and framework, edition 2, 2021.

3. WHY IS THE EVALUATION OF BIOMETRIC TECHNOLOGY NECESSARY AND WHAT IS THE MAIN CHALLENGE?

Products and services provided within the EU, particularly those in critical sectors, are frequently subject to rigorous regulatory oversight. For example, how do we know that a motorcycle helmet will protect the rider in case of a crash? Because the ECE 22.06 standard mandates manufacturers to test protective characteristics of motorcycle helmets under conditions that resemble typical motorcycle accidents.[3] This gives consumers assurance of the protection the helmet provides in case of an accident.

Like the protective characteristics of motorcycle helmets, the performance[4] of biometric recognition systems can only be determined through evaluation, using data as representative as possible of the data the systems will process in real-world operation. The evaluation of biometric recognition systems is therefore **essential to ascertain how a system will perform when exposed to the target population** (e.g., travellers crossing Schengen borders) in operational conditions (e.g., capture of biometric characteristics at a border-crossing point).

If the aim of such evaluations is to produce results that are representative of the **system's expected performance in real-world conditions**, then the biometric data used must be representative of the operational data to which the system will be exposed to in terms of the following aspects:

- **quality of biometric data** (which is a predictor of a system's accuracy),
- **demographics** (parameters such as age or gender),
- **capturing setup,** and
- **environmental conditions.**

In the context of evaluating biometric recognition systems, access to **representative datasets** at appropriate scale has been and continues to be a key issue for the biometric community. Some of the challenges related to data availability for testing purposes, in particular compliance with the legal framework for protecting privacy (e.g., GDPR) and demographic biases, can be addressed by generating **synthetic datasets.**[5] Some early studies have shown promising initial results in the use of synthetic biometric datasets to complement real data for training or evaluation of facial recognition algorithms. However, for the time being, the use of **real operational data** is still strongly recommended for the evaluation of biometric accuracy. Synthetic data may be of use for measuring other performance properties such as throughput.

3 United Nations Economic Commission for Europe (UNECE / ECE), UN Regulation No. 22, Uniform provisions concerning the approval of protective helmets and of their visors for drivers and passengers of motor cycles and mopeds, August 2021.

4 Performance of biometric recognition technologies includes not only throughput or accuracy, but also other characteristics, such as vulnerability to attacks, efficiency of template protection schemes, fairness or bias with regard to different demographic groups, robustness to data quality variability, etc.

5 *Synthetic biometrics:* artificially generated biometric data, which exhibits meaningful biological characteristics as measured by an existing biometric system, as defined in Johnson P. et al. (2023) Texture Modelling for Synthetic Fingerprint Generation.

4. WHY IS IT IMPORTANT TO ENSURE THE INDEPENDENT EVALUATION OF BIOMETRIC TECHNOLOGY?

Evaluations can be performed by:

1. the **developer/provider/vendor** of the system,

2. the **deployer** of the system or

3. an **independent third party**.

Each of these approaches has its own advantages and drawbacks.

Vendors of biometric recognition algorithms perform bespoke evaluations of the technologies they provide, however, there are some limitations:

- **limited transparency:** Vendors may not be willing to provide full transparency regarding the evaluation protocol or methodology applied, and in particular regarding the origin and nature of the dataset used. As a result, the meaningful interpretation of the results of such evaluations is not possible;

- **limited comparability:** Bespoke evaluations carried out by vendors of biometric technology normally evaluate the performance of algorithm(s) provided by the vendor that carries out the evaluation, and are in general carried out on non-public datasets.

These two facts limit the comparability of results with technologies provided by other vendors. The previous two points have a significant risk of **undermining trust** in biometric recognition technology which, in turn, may lead to high-profile controversies and fuel public scepticism regarding the accuracy and reliability of biometric recognition systems.

In addition to evaluations carried out by themselves, vendors can also engage **private certification laboratories** that offer evaluation and certification services for providers of biometric recognition technologies following well-established standards such as the ISO/IEC 19795[6] or the Common Criteria.[7]

There are several challenges related to the reliance on private (often for-profit) entities for the performance and safety testing of technologies, including biometric recognition systems. For example, some certification bodies may be industry-led, which may raise concerns regarding their independence, potential conflicts of interest, as well as reliability of quality assurance in such entities. In addition, evaluation protocols, datasets and the results of the evaluations carried out on behalf of vendors are not disclosed, thus significantly limiting the transparency and reliability of such evaluations.

Furthermore, as was the case with the self-assessments carried out by vendors, evaluations performed by private for-profit laboratories in most cases include only one system from one specific vendor, therefore not enabling comparability with the performance of other systems.

Several other limitations should be considered as well. For example, proprietary evaluation or certification frameworks often lack mechanisms for evaluating fairness or demographic bias of biometric technologies. Consideration of legal or ethical compliance, including the compliance of biometric technologies with such principles as transparency and accountability, is often also excluded from proprietary evaluation frameworks.

6 ISO/IEC 19795-1:2021 IT— Biometric performance testing and reporting. Part 1: Principles and framework, edition 2, 2021.

7 Common Criteria for Information Technology Security Evaluation, Part 1: Introduction and general model, November 2022.

Considering the limitations outlined above, evaluations carried out by commercial certification laboratories should be viewed as complementary to independent performance evaluations, especially for biometric technologies to be deployed in settings where decisions may have legal consequences.

In summary, evaluations carried out by technology vendors or by third-party private for-profit organisations limit transparency, reliability and trust in biometric technology.

One of the main recommendations to overcome these limitations, **is to entrust such evaluations to independent testing laboratories** (e.g., public or non-profit), that can deliver comparable and trustworthy results by being transparent with regard to their evaluation methodologies (incl. using standardised performance metrics and benchmarks) and most importantly, the datasets used.

5. STATE OF PLAY OF INDEPENDENT BIOMETRIC TECHNOLOGY EVALUATION INTERNATIONALLY COMPARED TO THE EU

International outlook

Outside of the EU, the evaluations performed by the **U.S. National Institute for Standards and Technology (NIST)** are widely considered as the benchmark for high-quality, independent, reliable and transparent biometric technology evaluations.[8] Over the past 30 years, NIST has become the world's most trusted, independent evaluator of biometric technology, shaping the development of fingerprint, face, iris, and voice recognition systems. Its large-scale tests, demographic fairness evaluations, and work on biometric data quality, are used as benchmarks for accuracy, security and responsible use of biometric technology around the world.

NIST began creating common datasets and evaluation protocols for face and voice recognition technologies in the 1990s with two key programmes:

- Face Recognition Technology (**FERET**), 1993-1997,[9]
- Speaker Recognition Evaluations (**SRE**), 1996-2024.[10]

Based on the success of these two evaluation programmes, NIST expanded its biometric evaluations portfolio in 2000 with two systems that drive its biometric evaluations to date:

- Face Recognition Vendor Test (**FRVT**),[11]
- Fingerprint Vendor Technology Evaluation (**FpVTE**).[12]

All these programmes have helped NIST establish itself as the **global leader in large-scale, industry-neutral, transparent comparison of biometric technologies**, supporting their adoption by the U.S. government in areas such as border control, law enforcement, and defence. Today, NIST is one of the main organisations driving the development of ISO/IEC standards on biometric recognition technologies, thus contributing to interoperability and enhancing the quality of biometric recognition systems deployed internationally.

Capabilities at EU level

According to an expert assessment and based on publicly available information, the **EU currently lacks an independent capability** for testing biometric technologies comparable to the work carried out by NIST in the U.S. Several limited evaluation campaigns have been carried out in the past; however, they were mostly one-off initiatives linked to EU-funded projects or initiatives funded and carried out by academic institutions for the purpose of advancing scientific research.[13]

At the **Member State level** we are starting to see some initiatives in the field of independent biometric evaluation especially by national forensic laboratories that tend to perform their own biometric evaluations. For example:

- **German Biometrics Evaluation Centre** (BEZ) can be considered as possibly the most advanced initiative carried out by a public

8 NIST is responsible for the testing and assessment of materials and technologies, ranging from nanotechnology, cybersecurity, to artificial intelligence, and construction materials, among others.

9 Face Recognition Technology (FERET).

10 NIST 2024 Speaker Recognition Evaluation (SRE24).

11 Face Recognition Vendor Test (FRVT).

12 Fingerprint Vendor Technology Evaluation (FpVTE).

13 See Annex 1 for a list of examples.

institution in the EU, focusing specifically on the evaluation of biometric technologies.[14]

- **French National Institute for Research in Digital Science and Technology** (Inria) includes assessment of biometric technologies as one of its activities.[15]
- **Netherlands Forensic Institute** (NFI) is also involved in the evaluation of biometric recognition technology, including through operational evaluations, pilots and proof-of-concept projects.[16]

However, the work of these national institutions in the EU, while valuable, is in general not made public and is very limited compared to the evaluations carried out by NIST.

To sum up, although expertise in the evaluation of biometric systems is available in the EU, and some capabilities to evaluate biometric systems exist at the national level in some Member States, they are **not comparable to the scale and scope** of biometric technology evaluations carried out by NIST in the U.S., and do not, at this moment, meet the requirements of EU or Member State authorities interested in the acquisition of biometric recognition technologies.

14 Biometrics Evaluation Centre **(BEZ).**

15 National Institute for Research in Digital Science and Technology **(Inria).**

16 Netherlands Forensic Institute **(NFI).**

6. REGULATORY LANDSCAPE FOR THE EVALUATION OF BIOMETRIC TECHNOLOGY IN THE EU

The European regulatory framework **sets clear boundaries on the processing of personal data**, in particular:

- EU Data Protection Regulation (**EUDPR**),[17]
- General Data Protection Regulation (**GDPR**),[18]
- Law Enforcement Directive (**LED**).[19]

These regulations lay down significant limitations on how such data may be processed in the context of development, training, testing and validation of biometric tools and systems. As a result, all personal data needed for the evaluation of biometric recognition technologies is subject to **strict purpose limitation rules:** it may only be processed for the specific purpose for which it was originally collected, unless a clearly compatible or legally authorised secondary use is identified. At the same time, the **principles of data quality and fair processing of personal data** that are enshrined in the EU's core data protection regulations mentioned above can be guaranteed only by **rigorous and accurate evaluation** of biometric technologies.

This need for robust and representative datasets is further reaffirmed by the obligations put in place under the **EU Artificial Intelligence Act**,[20] which will require that all high-risk AI systems placed on the EU market, or whose outputs are used in the EU, comply with strict requirements. These include obligations around data quality, representativeness, and bias mitigation. If a biometric recognition system uses artificial intelligence, it must also comply with the AI Act. According to Annex IV and Article 10 of the AI Act, developers must demonstrate that their systems have been trained and tested on datasets that are relevant, free from errors, and representative of the people or environments in which they will operate. As already presented in this document, without access to real-world, context-specific datasets — including those originally collected for border control—developers of biometric recognition technologies and systems may struggle to meet these requirements. Additionally, deployers of such systems must complete a fundamental rights impact assessment prior to the first use of such systems to evaluate all possible risks emanating from AI-based systems and take concrete action on their mitigation.

Biometric recognition systems in the EU

At the EU level, biometric recognition technologies are an **integral part of large-scale IT systems deployed in the EU's justice and home affairs (JHA)** domain and used mainly by law enforcement, immigration and judicial authorities. All these JHA information systems are **managed by eu-LISA,** which is mandated to carry out measurement and testing of biometric

17 Regulation (EU) 2018/1725 on the protection of natural persons with regard to the processing of personal data by EU institutions, bodies, offices and agencies (EUDPR).

18 Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation, GDPR).

19 Directive (EU) 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data (Law Enforcement Directive, LED).

20 Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).

recognition systems, including on performance and potential bias.[21]

In fact, the **Entry/Exit System (EES)** Regulation (EU) 2017/2226 explicitly mandates eu-LISA and participating national authorities in the Member States (e.g., border and law enforcement agencies) to use actual, real and representative data of subjects for testing these systems in the development phase.[22]

Additionally, the **European Data Protection Supervisor** (EDPS) has recommended in its opinion on the shared Biometric Matching Service (sBMS) that eu-LISA should use real, representative sampled production data to ensure that the system meets accuracy requirements, as the reassurance of bias minimization would outweigh other risks of using personal data.[23] Specifically, regarding the monitoring of comparative performance of biometric systems in the Interoperability Regulation and on the sBMS, the EDPS argued that, in general, the use of production data for testing purposes should be avoided in the absence of a **clear legal basis**, but that at the same time the **use of synthetic data** would not provide strong enough reassurances against bias, neither would only trusting the results of proprietary product accuracy results provided by vendors.

Therefore, the legal certainty regarding the **use of operational data for evaluation purposes** is established for the evaluation of the sBMS in the context of the Entry/Exit System (EES). However, there is no clear legal basis for the use of operational data for evaluation purposes for other EU JHA systems managed by eu-LISA and relying on biometrics, e.g., Visa Information System (VIS), Schengen Information System (SIS). Even if the necessary legal basis would exist within the scope of other EU large-scale IT systems in the JHA policy area, this would only serve the needs of eu-LISA and Member State authorities relying on these systems. Using this data for the testing of any other biometric technologies procured by other EU agencies or Member State authorities that are not covered by this legal basis would likely not be possible under the existing legal framework.

21 Regulation (EU) 2018/1726 on the EU Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice (eu-LISA). For more, please refer to eu-LISA's website on the EU's large-scale IT systems in the JHA domain.

22 Regulation (EU) 2017/2226 establishing an Entry/Exit System (EES) to register entry and exit data and refusal of entry data of third-country nationals crossing the external borders of the Member States and determining the conditions for access to the EES for law enforcement purposes (EES Regulation). The Commission Implementing Decision (EU) 2019/329 laying down the specifications for the quality, resolution and use of fingerprints and facial image for biometric verification and identification in the Entry/Exit System (EES) specifies how monitoring of biometric accuracy performance shall be carried out.

23 European Data Protection Supervisor (EDPS) Second EDPS Supervisory Opinion on eu-LISA's shared Biometric Matching Service (sBMS) DPIA, case 2021-0757, August 2022.

7. WHY IS AN EU-LEVEL CAPABILITY FOR THE EVALUATION OF BIOMETRIC TECHNOLOGY NEEDED AND WHAT ARE THE RISKS OF NOT HAVING ONE?

From a technical perspective, evaluations of biometric recognition technologies carried out by NIST represent the state of the art; however, we cannot be sure that the results of such evaluations will continue to be made public outside the U.S. Furthermore, these evaluations do not necessarily meet **EU-specific requirements.** For example, NIST performs evaluations according to the needs and specifications of the U.S. government entities, such as the U.S. Customs and Border Protection (CBP), the federal law enforcement agency under Department of Homeland Security, or the Federal Bureau of Investigation (FBI), under the Department of Justice.

Considering this wider context, relying solely on independent evaluations carried out by organisations outside of the EU (e.g., NIST) does not provide any possibility to control what and how is being evaluated, and therefore will not address EU-specific needs, especially in terms of alignment with:

- **the regulatory framework** applicable in the EU, e.g. data protection regulations, and the EU AI Act,
- **the needs of European authorities,** e.g., need to test performance of biometric recognition systems using other biometric characteristics; new systems specific to Schengen border crossing processes, such as biometric corridors.

Finally, establishing an independent EU-level capability for the evaluation of biometric recognition systems will contribute to strengthening the **EU's technological sovereignty, resiliency and economic competitiveness,** all of which have been emphasised in strategic policy documents recently published by the EU institutions.[24]

As previously mentioned, biometric technologies underpin the functioning of some of the EU's key large-scale IT systems in the areas of internal security, border management and justice, that are being used by authorities at both EU and national levels. Therefore, reducing the EU's reliance on NIST and private certification service providers in favour of an independent EU-level evaluation framework (or organisation mandated to carry out such work) is not merely a technical choice but a **strategic investment** in autonomy, trust, and the **sovereignty of the EU's digital infrastructure.**

Not strengthening this capability within the EU may result in a number of risks, including:

- **Lack of neutral, trusted benchmarks.** As argued in this policy brief, in the absence of an independent evaluator, testing will have no other choice but to rely on vendors or private certification service providers or national labs, which may have conflicts of interest or inconsistent methodologies. As a result, it may be more difficult for policymakers, procurement agencies, and the public to trust reported performance that may support vendor-driven narratives overstating accuracy and understating the limitations of the technologies they provide.

24 For example, see The Draghi report on EU competitiveness (2024), EU Competitiveness Compass (2025), 2030 Digital Compass (2021), or the European Parliament Report on European technological sovereignty and digital infrastructure (2025).

- **Fragmentation across Member States.** Although evaluation could be left to each individual Member State, this would likely lead to fragmentation due to lack of harmonisation and potential inefficiencies across the EU, as each individual MS may decide to rely on its own datasets, evaluation protocols and metrics. Furthermore, it would be difficult to compare results across evaluations.

- **Weaker position in global standardisation.** NIST evaluations (e.g., FRTE, FpVTE or SRE) serve as a global reference point used by governments and industry worldwide and this puts the EU at a significant disadvantage, having less influence in international standardisation bodies (e.g. ISO/IEC, ICAO) as it lacks independent large-scale results to justify proposals.

- **Limited transparency on demographic bias and attack vulnerability.** Aside from NIST, evaluation of biometric systems with regard to such performance characteristics as demographics or robustness to presentation or morphing attacks is rarely performed by other parties systematically. Therefore, without an independent EU-based capability, authorities deploying biometric technology and civil society will not have the possibility to independently verify to what extent a technology is biased, or how it performs in real-world scenarios, thus increasing the likelihood of public distrust and legal challenges to the deployment of biometric systems.

- **Slower and riskier deployment of biometric recognition technologies at the EU and Member State level.** As already mentioned, large-scale IT systems used in the EU's JHA domain (e.g., VIS, EES or SIS), rely on biometric recognition technologies. At Member State level, a wide range of national authorities also deploy biometric recognition technologies. In these circumstances, the lack of independent testing capability may lead to a number of suboptimal outcomes such as: authorities have no other choice but to rely on vendor declarations of performance; slow rollout of new systems or biometric tools because they need to be assessed on a case-by-case basis; in more extreme cases it may even lead to the deployment of systems with unverified or insufficiently verified performance.

- **Policy and regulatory gaps.** Without the independently verified evidence of biometric performance, the EU authorities responsible for regulating biometric technologies must rely on information provided by the U.S. NIST, vendors, academic studies or national laboratories, which, inter alia, increases dependence on political and strategic decisions of the U.S. government. This may lead to suboptimal decisions regarding the regulation of technical aspects concerning the use of biometric technologies.

- **Weaker global position of EU biometric industry.** In the long term, the overreliance on biometric evaluations carried out by organisations based in the U.S. could undermine the global competitive position of EU-based industry, especially if their access to take part in these evaluations is limited or even discontinued.

In conclusion, without a strong and clear plan for the independent evaluation of biometric technology at EU level, **Europe risks being a rule-taker rather than a rule-maker in biometrics**, both technically and politically. At the same time, establishing an EU-level independent capability for the evaluation of biometric technology will ensure a level playing field for the industry, incentivise industry to meet EU-specific regulatory requirements and standards, as well as stimulate innovation in biometric technologies across the EU.

8. RECOMMENDATIONS

As explained in previous sections, comprehensive and systematic evaluation of biometric recognition technologies conducted regularly over time and not as a one-time stand-alone exercise, requires stable datasets that are appropriately large and representative of the operational data to which European biometric systems will be exposed.

The Policy Brief therefore recommends the **creation of two complementary initiatives** to support EU-level capability for biometric evaluation:

1. **common biometric data repository** as the first key step towards establishing an independent EU capability;
2. **centralised biometric evaluation and testing platform** as the second step, to supplement the proposed data repository.

8.1. Common biometric data repository

As the first key step towards **establishing an independent EU capability for the evaluation of biometric technologies**, a common data repository should be created. This repository should meet at least the following requirements:

- **legal compliance:** the datasets stored in the common data repository must be compliant with the relevant data protection regulations, such as EUDPR, GDPR and LED. Specifically, this also means that such datasets can be used solely for the purposes of evaluation of biometric recognition technologies;
- **representativeness:** the data stored in the repository should reflect the demographics and use-cases for the deployment of biometric technologies relevant for EU authorities and those based in the Member States and Schengen associated countries;
- **security:** the data repository should be kept in a secure data centre that meets the requirements for storage of sensitive personal data;
- **minimisation:** the datasets stored in the repository should not contain any other personal data that may be linked to the biometric samples, except for characteristics that are necessary for the effective performance of biometric technology evaluation (e.g., demographic features such as the subject's age, or biological sex);
- **authority:** the maintenance of the common data repository should be delegated to a trusted authority.

Biometric datasets for the repository

In conjunction with the proposed data repository, we also need to **create biometric datasets** contained in the repository that should reflect the demographics and use-cases for the deployment of biometric technologies relevant for EU authorities and should be compliant with the relevant data protection regulations.

Considering the sensitive nature of biometric data, it must meet the following two requirements:

1. the dataset must be **representative** of the data that may be used in operational biometric systems used by relevant authorities, and also
2. the dataset must be **sufficiently large** to be used for evaluation purposes so that the statistical significance of the results obtained is as high as possible.

From the point of view of the legal framework for the protection of personal data this is especially challenging, as the dataset needs to be available for multiple evaluations over a longer time period (e.g., 10 years).

Below we provide a non-exhaustive list of approaches to the creation of such dataset:

- **Option 1:** Make use of **existing operational data** stored centrally at EU level. For example, the biometric data necessary for evaluation purposes is already stored in the EU's large-scale IT systems managed by eu-LISA, i.e., VIS, SIS, EES, Eurodac and will be stored in the near future in the ECRIS-TCN, when this system will be operational. These data are already managed by a reliable and trustworthy EU agency (eu-LISA) and stored in a highly secured environment with an on-premises data centre.

- **Option 2:** Engage with Member State authorities responsible for managing large-scale biometric datasets and operating in the areas of border management, and law enforcement, and **set up a shared dataset** in collaboration with them. The main constraint that may effectively prohibit this particular approach is national legislation, which may not allow for cross-border sharing of operational data or the use of operational data for evaluation purposes.

- **Option 3:** Gather biometric data for evaluation purposes using the **consent of third-country nationals entering the Schengen Area.** This approach presents several challenges with respect to Option 2. It may yield a high-quality dataset, which will be representative of a particular use-case, namely border-crossing scenario, but not other use-cases, which may be more specific to law enforcement. It is also very difficult to assess: 1) the size of the database that could be collected in this way (as it largely depends on the willingness of international travellers to provide their biometric data for this specific purpose); 2) the time it would require to collect a large enough database.

All options presented above have certain constraints:

- biometric data stored in the national and EU large-scale IT systems mentioned throughout this document will likely not cover **all relevant use-cases,** as these systems are mainly focused on border management scenarios. Therefore, additional data will need to be sourced in order to cover the entire spectrum of scenarios and applications of biometric technologies as is relevant for EU authorities;

- the **legal basis** for the use of biometric data for evaluation purposes is unclear, and such legal basis may need to be created by amending the relevant regulations; Option 3 introduces an additional operational challenge, as any extra steps taken during the enrolment procedure will likely affect the speed of processing of international travellers at border crossing points, thus further exacerbating the issues that Member State authorities have voiced their concerns about.

The **proposal for amending EU data protection rules**, put forward by the European Commission in November 2025 as part of the Digital Omnibus Regulation Proposal,[25] provides for the possibility to process special categories of personal data for the development and operation of AI, in case of **legitimate interest.** The independent testing of biometric technologies to ascertain whether those perform in line with the defined specifications, in particular non-discrimination, may constitute such legitimate interest, thus potentially providing the legal basis for the use of existing or collection of new data for this purpose, i.e., all three options listed above.

Based on the practical considerations outlined above, the **most feasible option** for the creation of a common data repository is to **utilise the biometric data already available** in the EU's JHA systems managed by eu-LISA (at least for the use cases where the stored data is representative of the operational scenario), provided that the legal constraints to the use of such data for evaluation purposes are addressed.

There is already at least one precedent demonstrating the feasibility of the proposed Option 1: the one-time evaluation of the shared Biometric Matching Service (sBMS) carried out jointly by

25 Digital Omnibus Regulation Proposal, European Commission press release, 19 November 2025.

eu-LISA and the European Commission's Joint Research Centre (JRC), ahead of the entry into operation of the EES. With the approval of the European Data Protection Supervisor (EDPS), this evaluation was carried out using real operational data (close to 500K records) extracted from the VIS.

8.2. Centralised biometric evaluation and testing platform

As a second step, a centralised biometric evaluation and testing platform should also be set up to **complement the common data repository.** This centralised evaluation and testing platform would be linked with the common data repository and provide **standardised protocols for the evaluation of biometric technologies.** The platform can be set up in different configurations and can gradually evolve to provide a more extensive catalogue of services:

- **Stage 1:** Set up an evaluation and testing platform that allows for the **automated evaluation** of biometric technologies submitted by vendors. This stage will require a certain level of upfront investment to set it up; however, the long-term operational costs will be limited. The main benefits offered by this kind of setup are the continuous tracking of the performance of biometric technologies offered by vendors and the availability of comparative data for benchmarking these technologies.
- **Stage 2:** Complement the automated technology evaluations described above, with ***ad-hoc* tailored scenario and operational evaluations**. The objective is to assess the performance of not only biometric algorithms, but also of end-to-end systems based on specific operational scenarios or in actual operational environments (i.e. including biometric data acquisition devices). For this configuration, the upfront investment required for setting up the evaluation platform, would have to be supplemented by the coverage of additional operational costs. These operational costs would include expert support for conducting evaluations, as well as additional physical infrastructure to support scenario and operational evaluations.

As with the common data repository, considering the sensitive nature of the testing platform, it needs to be set up in a **secure data centre environment** and managed by a trusted authority.

ANNEX I - BIOMETRIC EVALUATION ORGANISATIONS, PROJECTS AND INITIATIVES IN THE EU

National organisations in EU Member States

Germany: Biometrics Evaluation Centre (BEZ)[26]

Operational since 2021, the BEZ is a joint effort between the German Federal Office for Information Security (BSI) and Hochschule Bonn-Rhein-Sieg's Institute for Safety and Security Research (ISF). This is likely the **most advanced** public organisation at the Member State level with regards to the evaluation of biometric industrial products ready for deployment. The laboratory focuses on the continuous and regular examination of biometric systems in the form of **long-term test series** and also performs eventual **one-off assessments** on demand. The test criteria are system performance (recognition accuracy), usability (ease of use, human-machine interaction) and transmission security. The tested solutions range from full electronic passport control systems (eGates) for border crossing points, and fingerprint scanners, to 3D and thermal facial sensors.

France: National Institute for Research in Digital Science and Technology (Inria)[27]

The mission of Inria is to accelerate, through digital research and innovation, the construction of France's scientific, technological and industrial leadership in Europe. As a public research and innovation infrastructure, Inria contributes to the construction of **France's digital sovereignty** in the European ecosystem. In this capacity, it contributes to the implementation of public policies with far-reaching implications, such as the National Artificial Intelligence Research Programme or the development of a French ecosystem for digital security (in partnership with ANSSI). As one of the major independent in-house consultants for the French Government, its activities and projects also cover the assessment of new IT technologies, including biometrics.

The Netherlands: Netherlands Forensic Institute (NFI)[28]

The NFI's mission is to provide forensic services using state-of-the-art science and technology, contributing to the investigation and prosecution of suspects, and the exoneration of innocent parties. Although the institute is **fully focused on forensic science,** the NFI can be a valuable example of a testing institution offering services to multiple stakeholders and continuously moni-

26 Biometrics Evaluation Centre **(BEZ).**

27 National Institute for Research in Digital Science and Technology **(Inria).**

28 Netherlands Forensic Institute **(NFI).**

toring and evaluating the latest technological developments in a given field. Furthermore, one of the main areas of NFI's expertise is biometric technology, where they have produced over the years very relevant contributions in the form of pilot projects, operational evaluations and proofs of concept.

EU funded projects

Biometrics Evaluation and Testing (BEAT) project[29]

The Biometrics Evaluation and Testing project (BEAT) was an EU project funded under the FP7 programme led by the Swiss research institute IDIAP. BEAT was created specifically to address the challenge of developing a **standardised European framework** to compare reliability, robustness and privacy aspects across biometric systems.

The project's main outcome was an open-source online platform — **BEAT platform** — for evaluating biometric algorithms together with metrics, benchmarks, algorithms (to provide baseline performance), and perhaps most importantly, datasets. The project provided, for the first time in the EU, a means for **evaluating the reliability of biometric systems** in a comparable and measurable way, enabling to benchmark performance, robustness against attacks and privacy of biometric technologies.

The platform's key contribution is that it facilitates reproducibility and comparison of results, thereby increasing transparency and aiding open access and scientific rigour in biometric evaluation. This was possible because the platform's (trusted) host had full control of the datasets used for evaluation, while registered users could only submit their algorithms to be tested on the data, without ever having access to it. While BEAT was a **pioneering project** in terms of biometric evaluation in the EU, it was a one-off effort. The open BEAT platform can still be downloaded and run by any institution that wants to provide biometric evaluation services and to make its own datasets exploitable by others (without being directly accessed). However, there is no support provided for the software, which has probably become outdated in the meantime.

BioSecure project[30]

The BioSecure project was a prominent Network of Excellence (NoE) project funded under the EU's FP6 programme, running from June 2004 to September 2007. Among its key objectives was the development of a **common evaluation infrastructure,** including a standardised biometric multimodal database, baseline reference systems (open-source algorithms) and benchmarking protocols across multiple characteristics. In 2007, they ran a one-time evaluation campaign of the systems developed within the project. Today, the **BioSecure Association** established post-project, continues distributing the datasets, evaluation tools, and benchmarks for potential future evaluations, under licensing agreements and with data protection oversight.

29 Biometric Evaluation and Testing (BEAT).

30 BioSecure.

Projects: D4FLY / iMARS / METICOS / TReSPAsS

This section presents a selection of recent EU projects, funded under the Horizon 2020 programme, that have integrated some type of **biometric evaluation components** within their work, for example:

- assessment of vulnerabilities to presentation and morphing attacks (**iMARS**);[31]
- operational and field testing in border crossing environments (**METICOS**); [32]
- assessment of usability and accuracy of multiple biometric characteristics such as 2D and 3D face, fingerprints, iris or thermal facial imaging (**TReSPAsS**); [33]
- specific scenarios such as biometrics on-the-move (**D4FLY**).[34]

While these efforts highlight ongoing interest and innovation in the area of biometric evaluation, they are not a viable substitute for a sustainable, EU-level evaluation framework anchored in public accountability and repeatability.

Competition series

Fingerprint Verification Competition (FVC) series[35]

The FVC series is, arguably, the most consistent initiative organised at EU level with regards to biometric technology evaluations. This series was launched in 2000 by the Biometric Systems Laboratory from the University of Bologna, which ran four different evaluations on a biennial schedule between 2000 and 2006: FVC2000, FVC2002, FVC2004 and FVC2006. After FVC2006, the organisers shifted from fixed-schedule competition campaigns closed in time, to an **open-ended, continuously available evaluation platform — FVC-onGoing** — where developers can submit their fingerprint recognition algorithms at any given point in time and get evaluated. FVC focuses specifically on the **technical accuracy of fingerprint comparison** and remains a widely respected benchmark in both academic and industrial contexts, providing a scientifically grounded framework that supports the advancement of fingerprint technologies.

31 Image Manipulation Attack Resolving Solutions (iMARS).

32 Platform for Monitoring and Prediction of Social Impact and Acceptability of Modern Border Control Technology (METICOS).

33 RobusT Risk basEd Screening and alert System for PASSengers and luggage (TReSPAsS).

34 Detecting Document frauD and iDentity on the fly (D4FLY).

35 Fingerprint Verification Competition (FVC) series, University of Bologna Biometric Systems Laboratory.

Liveness Detection (LiveDet) Competition series[36]

The Liveness Detection Competition series (LivDet) are international benchmarks designed to evaluate the state-of-the-art in **biometric liveness detection** across multiple characteristics, principally **fingerprint, iris, and face.** Organised approximately every two years since 2009, with participants from academia and industry, these competitions provide standardised datasets, protocols, and performance evaluations for both algorithm-only (software) and full-system (hardware + software) **presentation attack detection (PAD) solutions.** While this is not strictly a purely EU initiative, the promoter of the first edition (LivDet-fingerprint 2009) was the University of Cagliari in Italy, which remains one of the main organisers of the series, together with Clarkson University from the U.S., and a number of other partners from academia.

36 Liveness Detection (LiveDet) Competition series.

ACKNOWLEDGEMENT:

This policy brief was produced by the following members of the **Biometrics Cluster of the EU Innovation Hub for Internal Security**: European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice (eu-LISA - lead), European Commission's Directorate-General for Migration and Home Affairs (DG HOME), European Commission's Joint Research Centre (JRC), European Union Agency for Law Enforcement Cooperation (Europol), and the European Border and Coast Guard Agency (Frontex).

Manuscript completed in March 2026

ISBN 978-92-95237-12-4 ISSN: 2443-8103 doi:10.2857/8468491 Catalogue number: EL-01-26-002-EN-N