Artificial Intelligence Advisory Council

Advice to Government: FRT Use by An Garda Síochána

Advice Paper No. 1/2024 June 2024

The role of the Artificial Intelligence Advisory Council is to provide independent expert advice to the whole of Government on artificial intelligence policy, with a specific focus on building public trust and promoting the development of trustworthy, person-centred AI.

This advice to Government is an output of the Sub-Group of the AI Advisory Council on Biometrics in the Public Service in Ireland which has been approved by the AI Advisory Council.

Membership of Sub-Group of AI Advisory Council on Biometrics in the Public Service in Ireland

Professor Deirdre Ahern (Chair)

Dr Patricia Scanlon

Dr Abeba Birhane

Dr Susan Leavy

Mr Ronan Murphy

Mr Barry Scannell

Membership of AI Advisory Council

Dr Patricia Scanlon (Chair)

Professor Deirdre Ahern

Dr Abeba Birhane

Mr Bernard Harbor

Professor Stephen Kinsella

Dr Susan Leavy

Mr Seán Mullaney

Mr Ronan Murphy

Professor Barry O'Sullivan

Ms Emma Redmond

Ms Bronagh Riordan

Ms Sasha Rubel

Mr Barry Scannell

Professor Alan Smeaton

Executive Summary

This report of the Artificial Intelligence Advisory Council ('AIAC') providing advice to Government addresses societal, legal, procurement, and deployment considerations for Facial Recognition Technology ('FRT') in law enforcement. FRT has the potential for significant efficiency benefits but poses substantial risks to fundamental rights, particularly through potential misidentifications or false matches. The report emphasises the need for public trust, transparency, and accountability in FRT evaluation, procurement, and usage. The AIAC highlights some significant factors in relation to deciding to legislate for, procure and operationalise FRT and the importance of them being comprehensively addressed.

Critical concerns include accuracy and bias in FRT systems, which perform variably under different conditions and across demographics. False positives and false negatives can have severe consequences and relying on a 'human in the loop' has been shown in studies to not mitigate these risks. The report illustrates that while FRT algorithms have been previously cited at 99% accuracy from US based National Institute of Standards and Technology (NIST), deeper investigation reveals that false match rates vary dramatically between demographics, highlighting a need for more robust evaluations. Furthermore, the NIST evaluations do not include any assessments on real-world data and scenarios; evaluations results were from data under ideal conditions with optimal lighting and cooperation, such as at border control, passports, or mugshots. The report therefore recommends comprehensive independent evaluations by an AI expert group to assess the performance of FRT across all demographics and in real-world conditions.

Legislating for FRT use by An Garda Síochána, in whatever form it may take, would represent a watershed development in policing and its interface with society in Ireland. The report highlights the importance of public trust around an AI-powered step change in policing power. In choosing to legislate, the report highlights the need to ground FRT in a robust legislative framework that takes due account of the EU AI Act, fundamental rights, data protection, privacy, and equality laws. It also advocates for robust operational protocols and ongoing monitoring to safeguard data privacy and ensure accountability in FRT usage. Ultimately, the report offers a detailed guide for policymakers to ensure that FRT evaluation, procurement, and deployment is accurate, fair, transparent, and respects individual rights. The AIAC advises against deploying FRT in Ireland without thorough independent evaluation and resolution of identified risks.

A. INTRODUCTION

FRT brings both opportunities and challenges. FRT software can automate existing human-driven workflows and perform analyses previously impossible with human review. It provides the potential to speed up the processes of investigating and apprehending offenders and finding missing persons while using less police time and personnel. When used in law enforcement, potential efficiency gains must be balanced against the impact on rights. A complex range of harms may potentially occur in deploying FRT including misidentifying crime suspects. Therefore, as recognised in the AI Act, used in a law enforcement context, FRT is a high-risk technology given the potential consequences of its use for individuals.

B. PUBLIC TRUST AND HUMAN-CENTRED VALUES

In line with the National AI Strategy, *AI – Here for Good*, maintaining strong public trust in AI as a force for societal good is critical. Not only must AI be safe and appropriately governed, it must also be ethical and serve our values as a democratic society with legally and constitutionally protected rights to privacy, equality, freedom of expression and assembly and to a fair trial. Introducing FRT in a policing context, in whatever form it may take, would represent a step change in policing power in Ireland whose potential impacts on society and upon trust in An Garda Síochána require careful consideration. The privacy of many could be impinged on by surveilling or searching images of the population at large. Furthermore, the potential for a gradual mission creep towards an untargeted mass surveillance State is a legitimate concern.

Advice:

Public trust must be a cornerstone in the use of AI by An Garda Síochána. Public transparency, engagement and accountability is crucial to building public trust around any contemplated use of FRT in policing and its operational parameters.

C. EVALUATING THE ACCURACY OF FRT SYSTEMS

FRT demonstrates impressive capabilities in controlled environments, such as one-to-one matching in passport kiosks. While people may experience these systems working effectively in such settings, this use case is very different from real-world law enforcement FRT applications 'in the wild', such as one-to-one or many-to-one matching in real-world scenarios. These complex conditions in law enforcement raise concerns about accuracy, bias, and human rights, and thus require thorough evaluation by experts.

In controlled settings, consistent lighting, cooperative subjects, and high-quality images allow for accurate one-to-one matching. In contrast, law enforcement scenarios often involve many-to-one matching using images from varied conditions—such as poor lighting, different angles, and non-cooperative subjects captured by lower-quality CCTV or bodycam footage. These factors significantly degrade FRT performance.

FRT errors fall into two main categories:

- False Negatives: When the system fails to recognise two photos of the same person. This is often due to poor image quality, such as inadequate lighting for dark-skinned individuals or incorrect camera angles.
- False Positives: When the system incorrectly matches photos of different people. This typically results from insufficient training data for certain demographic groups.

Accuracy in FRT is crucial, especially in law enforcement where errors can have severe consequences. False positives can lead to wrongful suspicion, arrest, and lasting damage, while false negatives allow criminals to evade justice, undermining public safety and trust. Even small error rates can lead to significant misidentifications in large populations, highlighting the potential for substantial harm.

Incorporating a "human in the loop" is often proposed to mitigate inaccuracies in facial recognition technology (FRT). While this approach aims to improve decision-making, it may introduce challenges related to cognitive bias. Studies have shown that FRT algorithm outcomes can bias human decision-making; for instance, officers presented with an FRT match may be unduly influenced by the system's suggestion, potentially amplifying errors. Therefore, it is crucial not to simply rely on the "human in the loop" approach as a solution to address the inaccuracies and biases of FRT algorithms.

When evaluating the performance of biometric systems such as FRT, several key factors must be considered:

- Real-World Conditions: Accuracy metrics derived from ideal datasets may not reflect real-world conditions, which are often more complex and challenging.
- Matched and Unmatched Domains: Reported accuracy is often based on matched domains, as in the NIST results, such as mugshot-to-mugshot comparisons, rather than more difficult real-world scenarios like mugshot-to-CCTV footage.
- Demographic Disparities: Presenting evaluation results as a single, averaged accuracy figure can obscure significant disparities in performance across different demographics, potentially masking poorer outcomes for specific groups.

Evaluating FRT systems is complex. An Garda Síochána has referenced NIST results for the cloudwalk_mt_007 algorithm, highlighting its 99% accuracy. To illustrate this statistic, in a full Croke Park stadium of 82,300 people, European males aged 20-35 would have a false match likelihood rate (FMR) of 0.00012, potentially resulting in around 10 false matches. In contrast, West African females would have an FMR of 0.00710, potentially resulting in around 584 false matches. This deeper analysis (and its demographic distribution), corroborated by NIST researchers, show that there is still significant bias in this algorithm. In addition, it is important to note that the NIST results were obtained using 'ideal' datasets with respect to subject cooperation, good lighting, and high-quality images such as at border control, passports, visa applications and mugshots. In real-world law enforcement, images from bodycams and CCTV are of significantly lower resolution (due to distance from subjects), involve non-cooperative subjects, and suffer from poor lighting, occlusions, and obscured facial features, resulting in worse performance than in the illustration above.

The FMR statistics above from NIST were obtained by specifically comparing mugshot images to other mugshot-style images (matched domains). In real-world scenarios, it is more common to match mugshot or passport-style images to 'wild' data, such as images from bodycams or CCTV footage. In these less ideal real world circumstances, an algorithm using data from mismatched domains would likely perform more poorly than the NIST results above.

Advice:

NIST researchers have acknowledged significant issues with the accuracy and bias of FRT algorithms and challenges in evaluating real-world FRT using wild image data from sources like CCTV and bodycams.

Given the limitations of current evaluations, the AIAC advises against procuring or deploying FRT until satisfactory independent evaluations are conducted under real-world conditions relevant to Irish law enforcement. It is recommended that an independent Irish AI expert group be established to assess existing and emerging FRT evaluation methods from NIST and other international studies.

Their goal would be to determine whether FRT algorithms are fit for intended law enforcement purposes in Ireland.

D. LEGAL AND REGULATORY CONSIDERATIONS

Legislating for FRT has complex operational implications for data protection, data privacy and fundamental rights that must be taken on board. As emphasised by the European Data Protection Board, the legality principle and rule of law dictates that the legal basis and the public interest justification for using FRT (see UK *Bridges* decision, Court of Appeal, 2020) must be proportionate and clearly articulated in legislation.

We also highlight the importance of complying with the public sector equality duty during both procurement and deployment phases to avoid potential indirect race or gender discrimination (see *Bridges* decision).

Advice:

If a decision is made to proceed to legislate, primary legislation must clearly establish the legal basis and use cases for FRT. The AIAC recommends close consultation with the Data Protection Commission, the Human Rights and Equality Commission and the EU AI Office to appropriately navigate the rights and regulatory considerations which FRT in law enforcement give rise to.

In providing for and operationalising FRT, compliance with the EU AI Act framework would be necessary including fundamental rights impact assessments taking place before procurement and deployment. Work would need to be undertaken to pinpoint the complex risks to the rights of individuals (e.g. rights to respect for private life, human dignity, respect for personal data, freedom of thought, conscience and religion, assembly and association, presumption of innocence and the right to a fair trial) and to identify credible measures that could be taken to address these risks. Given the challenging and uncharted nature of this territory, it may be prudent to await the EU AI Office's deliverables around conducting Fundamental Rights Impact Assessments and designing risk mitigation measures.

Advice:

***In providing for and operationalising FRT, compliance with the EU AI Act framework should be built in. A Fundamental Rights Impact Assessment in alignment with the EU AI Act should be provided for before procurement and deployment of an FRT system by An Garda Síochána*.**

Currently there is an unfortunate level of opacity around the intended sourcing and use of facial image data by An Garda Síochána. We underline the imperative of complying with data protection and data privacy law in relation to compilation of facial image reference databases for searching against a captured image. These concerns matter hugely for public trust and fairness as well as to complying with legal expectations (*Glukhin v Russia*, European Court of Human Rights, 2023).

Advice:

Reference databases of images used in any matching exercise must have defined parameters and an express legislative basis provided for the data collected there.

Applications for approval for deployment and applications for judicial redress should be from suitably trained members of the judiciary.

To be suitably robust, the operational parameters should ideally be expanded upon in the form of secondary legislation rather than a Code of Practice.

Access to an FRT system should be restricted to protect data privacy and to minimise the risk of errors as a result of deployment by personnel who are not appropriately trained.

Robust complaints and judicial redress provisions should be expressly included in any legislation.

Periodic independent auditing of the use of FRT should be provided for.

E. PROCUREMENT CONSIDERATIONS

Were FRT to be rolled out, it would be inadequate to rely on generic public sector procurement frameworks. A bespoke procurement framework adapted for this context would help to safeguard the effectiveness of an FRT system used in law enforcement and to ensure that at procurement stage and during its lifecycle it meets strict standards for accuracy, transparency, fairness, and privacy.

Advice:

We recommend the adoption of a bespoke procurement framework for FRT systems, in consultation with AI experts, to ensure their reliability in meeting best practice.

During procurement, prospective vendors should be expected to deal effectively with relevant matters such as those detailed below.

<u>Data Sourcing and Training Transparency</u>

- Verification that no data privacy violations have occurred and that data collection complies with applicable data protection and privacy laws.
- Transparency regarding the data used to train models.

<u>Bias and Fairness Evaluation</u>

- Independent bias assessment reports to check that the system performs equitably across different races, ethnicities, skin colours, genders, and age groups (with post-selection selected, periodic re-evaluation).

<u>Accuracy in Diverse Real-World Use cases and Conditions</u>

- To ensure real-world effectiveness, the system's accuracy needs to be evaluated across a wide range of conditions. This includes variations in camera angles, distances, positions, and lighting. The system should be able to handle images with multiple faces, blurry footage, and low resolutions typical of body cameras and CCTV. Importantly, evaluations should encompass both controlled settings like mugshots and border control and uncontrolled environments with "wild" images from bodycams and CCTV as well as FRT performance when matching between these domains. The system's ability to identify individuals in one-to-one as well as

many-to-one scenarios, regardless of image capture style, is crucial for real-world deployment.

Ongoing Monitoring and Vendor Support

- Vendor support should be contracted for. This will ensure system updates, patches and addressing security vulnerabilities.
- The contact should mandate regular performance reviews, system audits and third party independent bias assessment and transparency in updating algorithms.

F. OPERATIONALISATION

Individuals have well-defined rights around their personal data and use of it. Any legislation must clearly mesh into data protection frameworks in relation to its vision for operationalisation.

Advice:

If FRT is operationalised, ensuring legitimacy of data use and processing would be crucial.

Accountability demands establishing clear audit trails to track the origin and usage of facial recognition data and to address potential misuse. Restricting access through An Garda Síochána implementing the principle of least privilege (in alignment with best practice around data protection) would help to ensure responsible access and use.

Advice:

Accountability and securely storing and managing facial recognition data demand the development of extremely robust protocols.

The EU AI Act emphasises the importance of competence, training and support around deployment of AI systems in law enforcement.

Advice:

Adequate training, support and proper oversight around interpreting AI outputs is essential.

G. CONCLUSION

While AI has the potential to offer significant opportunities, the State must approach the adoption of AI that impacts fundamental rights with caution and rigour. Furthermore public trust in AI is crucial. FRT in law enforcement is high-risk and cannot be responsibly implemented without addressing issues of accuracy, discriminatory effects, data privacy, data security, and fundamental rights. This necessitates robust legislation, procurement, operational, and accountability frameworks. NIST evaluations of FRT algorithms have revealed significant biases, as discussed earlier.

More fundamentally, no independent studies, including those by NIST, have evaluated FRT in real-world conditions and with use cases such as body cams and CCTV. Therefore, it is the advice of the AIAC that FRT systems should not be deployed in Ireland without independent review and evaluation by AI experts, considering unresolved risk factors and their potential impacts.