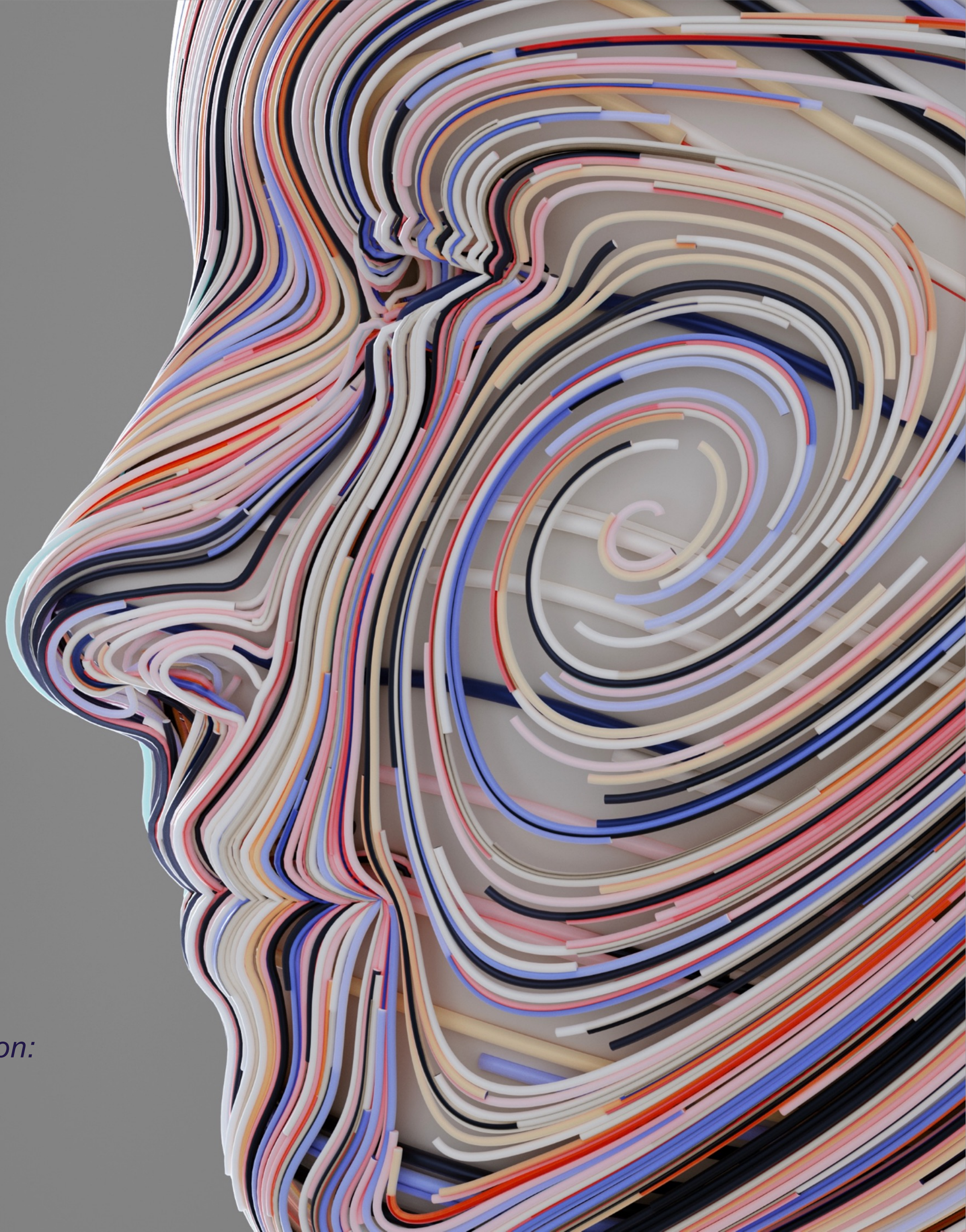
NATIONAL ACADEMIES *Sciences Engineering Medicine*

Facial Recognition Technology

Current Capabilities, Future Prospects, and Governance



EDWARD W. FELTEN, Robert E. Kahn Professor of Computer Science and Public Affairs (emeritus), Princeton University

JENNIFER L. MNOOKIN, chancellor, University of Wisconsin-Madison

Co-chairs, National Academies Committee on Facial Recognition: Current Capabilities, Future Prospects, and Governance

Study Committee and Staff

EDWARD W. FELTEN (NAE), Princeton University, Co-chair

JENNIFER L. MNOOKIN, University of Wisconsin, Madison, Co-chair

THOMAS D. ALBRIGHT (NAS), Salk Institute for Biological Studies

RICARDO BAEZA-YATES, Northeastern University

BOB BLAKLEY, Team8

PATRICK GROTHER, National Institute of Standards and Technology

MARVIN B. HAIMAN, Metropolitan Police Department, Washington, D.C.

AZIZ Z. HUQ, University of Chicago

ANIL K. JAIN (NAE), Michigan State University

ELIZABETH JOH, University of California, Davis

MICHAEL C. KING, Florida Institute of Technology

NICOL TURNER LEE, The Brookings Institution

IRA S. REESE, Global Security and Innovative Solutions

CYNTHIA RUDIN, Duke University

Staff

JON K. EISENBERG, Senior Board Director, CSTB

BRENDAN ROACH, Program Officer, CSTB

STEVEN KENDALL, Senior Program Officer, CSTL

GABRIELLE M. RISICA, Program Officer, CSTB

EMILY BACKES, Deputy Board Director, CLAJ

SHENAE A. BRADLEY, Administrative Assistant, CSTB

Statement of Task (summarized)

- Provide a broadly accessible explanation of FRT.
- Assess the strengths, capabilities, risks, and limitations of FRT.
- Consider current approaches to governing the use of FRT and describe implications of the use of FRT and requirements for adequate safeguards.
- Consider concerns about the impacts of FRT in public and private settings on privacy, civil liberties, and human rights.
- Develop recommendations to govern the use and performance of FRT.

Study sponsored by the U.S. Department of Homeland Security through contract number 70RSAT21G00000003/70RDAD21FR0000159 with the National Academy of Sciences and by the Federal Bureau of Investigation with DHS's assistance.

About the Report

Summary—Overview of conclusions and recommendations

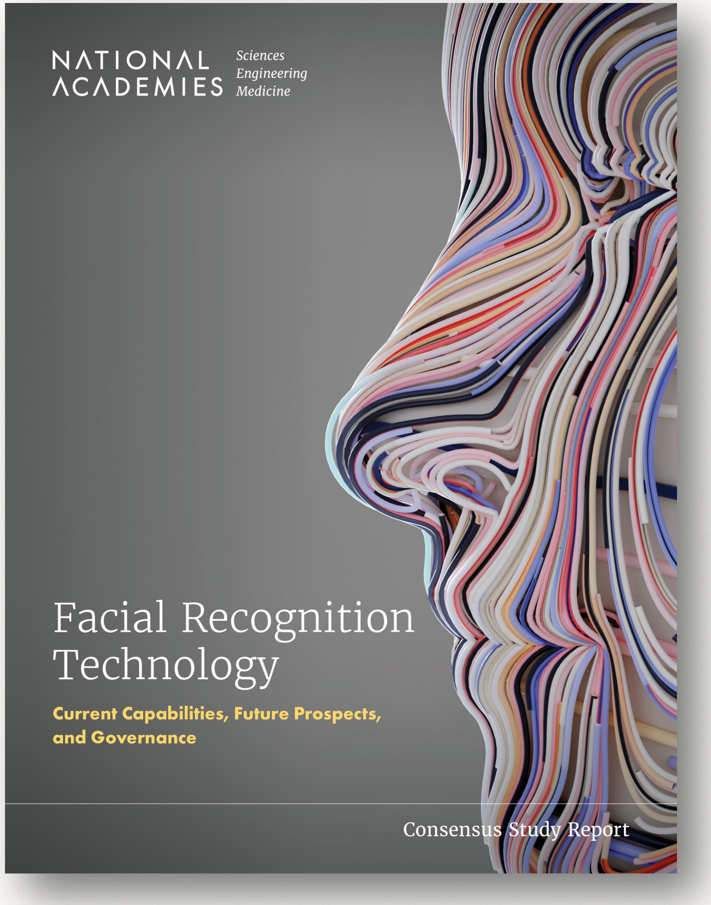
Chapter 1—Introduction

Chapter 2—Current state of technology, accuracy, error rates, and demographic disparities in performance

Chapter 3—Use cases with brief vignettes to illustrate potential benefits and concerns presented by current and potential uses

Chapter 4—Equity, privacy, and civil rights and associated governance issues and options

Chapter 5—Study conclusions and recommendations along with an initial sketch of a risk management framework and its application to several illustrative cases



NATIONAL ACADEMIES *Sciences Engineering Medicine*

Facial Recognition Technology

Current Capabilities, Future Prospects, and Governance

Consensus Study Report

Characteristics of FRT

Highly personal—the face is a uniquely individualizing part of the body and is much more visible than fingerprints or iris patterns

Pervasive—large and growing number of images available from cameras operated by governments, businesses, and individuals

Ubiquitous—many if not most people can be recognized

Stealthy—difficult to detect use and purpose

Inexpensive—automation makes marginal cost of use very low

Expanding Scope and Scale

Technical development has accelerated in the past decade

- Adoption of deep convolutional neural network techniques
- Test and evaluation—NIST Face Recognition Vendor Test
- Experience gained from industrial adoption and deployment

Collection and curation of increasingly comprehensive “reference galleries”

- Government: driver’s license, passport, and arrest photos
- Private sector: Internet or on-premises

Numerous and growing sources of “probe images”

- Security cameras
- Street cameras
- Doorbell cameras (sometimes incentivized by local government)

Deployment in many different applications

- Unlocking smartphones and other personal devices
- Law enforcement investigations
- Airports and international borders
- A variety of other government and commercial applications

Complex Governance Challenges

- Policy questions arise from the development of the technology to deployment and use
- Numerous unsettled legal and policy questions
- Potential for governance at national, state, and local levels
- Implicates a core set of interests related to freedom from state and private surveillance, privacy, civil liberties, and equity

Utility for Identity Verification and Identification

- Process large numbers of individuals quickly
- Identify high-risk individuals among large numbers of people entering a location without delaying others
- Aid for law enforcement in criminal and missing person investigations
- Convenient identity verification

Two Sets of Concerns

- **Concerns about poor performance of the technology**—e.g., unacceptable false positive or false negative rates or unacceptable variation of these rates across demographic groups.

- **Concerns about problematic use or misuse of the technology**—technology with acceptable technical performance sometimes produces societally undesirable outcomes because of either inadequate procedures or training for operating, evaluating, or making decisions using FRT or the deliberate use of FRT to achieve a societally undesirable outcome (including uses not foreseen by FRT developers or vendors).

Equity, Privacy, and Civil Liberties Concerns

- Powerful tool for pervasive surveillance
- Potential adverse equity and privacy impacts in the largely unregulated commercial sphere
- Abuse by private individuals
- Implicated in at least 6 high-profile wrongful arrests of Black individuals

FRT has been implicated in at least 6 high-profile wrongful arrests of Black individuals

- Incidents likely represent a small percentage of known arrests involving FRT but no comprehensive data on the prevalence of FRT use, how often FRT is implicated in arrests and convictions, or the total number of wrongful arrests that have occurred on the basis of FRT.
- Backdrop of deep and pervasive distrust by historically disadvantaged and other vulnerable populations of policing methods that have often included a variety of forensic, surveillance, and predictive technologies.
- Distrust of FRT exacerbated because all the reported wrongful arrests associated with the use of FRT have involved Black defendants.
- Testing has demonstrated that false positive match rates for Black individuals and members of some other demographic groups are relatively higher (albeit low in absolute terms) in FRT systems that are widely used in the United States.

Demographic disparities

Much progress has been made in recent years to characterize, understand, and mitigate phenotypical disparities in the accuracy of FRT results. However, these performance differentials have not been entirely eliminated, even in the most accurate existing algorithms.

- With the most accurate algorithms, and if both the probe and reference images are of high quality, **false negative rate differentials** are extremely small. However, false negatives can become significant with low quality images.
- For identify verification (1-to-1 comparison) algorithms the **false positive match rates** for certain demographic groups when using even the best performing facial recognition algorithms designed in western countries and trained mostly on White faces, are relatively higher (albeit quite low in absolute terms) even if both the probe and reference images are of high quality.

Demographic differentials present in verification algorithms are usually but not always present in identification (1-to-many comparison) algorithms.

Conclusions

With a few exceptions the U.S. does not have authoritative guidance, regulations, or laws that adequately address concerns broadly.

Study committee members all agreed:

- Some use cases of FRT should be permissible.
- Some use cases should be allowed only with significant limits or regulation.
- Other uses cases likely should be altogether prohibited.

Study committee members did not reach a fully shared consensus on:

- Precisely which use cases should be permitted
- How permitted uses should be regulated or otherwise governed

This reflects the complexity of the issues raised; individual assessments of the risks, benefits, and tradeoffs; and individual perspectives on the underlying values.

However, the committee is in **full agreement with the recommendations** that follow.

Overview of Recommendations

Mitigating potential harms and laying the groundwork for more comprehensive action

Rec 1: The federal government should take prompt action…to mitigate against potential harms of facial recognition technology and lay the groundwork for more comprehensive action.

Fostering trust and mitigating bias and other risks

Rec 2: Developers and deployers of FRT should employ a risk management framework and take steps to identify and mitigate bias and cultivate greater community trust.

Enacting more comprehensive safeguards

Rec 3: The Executive Office of the President should consider issuing an executive order on the development of guidelines for the appropriate use of FRT by federal departments and agencies and addressing equity concerns and the protection of privacy and civil liberties.

Rec 4: New legislation should be considered to address equity, privacy, and civil liberties concerns raised by FRT, to limit harms to individual rights by both private and public actors, and to protect against its misuse.

Detailed Recommendations

The detailed recommendations in this briefing are summarized from the report. Key points are in bold.

Recommendation 1: The federal government should take prompt action…to mitigate against potential harms of facial recognition technology and lay the groundwork for more comprehensive action.

Recommendation 1-1: The National Institute of Standards and Technology should sustain a vigorous program of FRT testing and evaluation to drive continued improvements in accuracy and reduction in demographic biases.

Testing and standards are a valuable tool for driving performance improvements and establishing appropriate testing protocols and performance benchmarks, providing a firmer basis for justified public confidence, for example, by establishing an agreed-on baseline of performance that a technology must meet before it is deployed

Rec 1-2: The federal government, together with national and international standards organizations (or an industry consortium with robust government oversight), should establish:

a. **Industry-wide standards** for evaluating accuracy and demographic variation

b. **A tiered set of profiles defining standards for uses with different levels of sensitivity**

   - Minimum quality for probe and reference images
   - Acceptable overall false positive and false negative rates
   - Acceptable thresholds for accuracy variation across different phenotypes

c. **Methods for evaluating false positive match rates** for probe images captured by closed-circuit television or other **low-resolution cameras** (which have been implicated in erroneous arrests of several Black individuals).

d. **Process standards** in such areas as data security and quality control.

Rec 1-3: DOJ and DHS should establish a multidisciplinary and multi-stakeholder working group on FRT to develop and periodically review standards for reasonable and equitable use, as well as other needed guidelines and requirements for the responsible use of FRT by federal, state, and local law enforcement.

That body, which should include members from law enforcement, law enforcement associations, advocacy and other civil society groups, technical experts, and legal scholars, should be charged with developing:

a. Standards for appropriate, equitable, and fair use.

b. Minimum technical requirements for FRT.

c. Minimum image quality standards for probe images.

d. Guidance for whether FRT systems should **(1) provide additional information about confidence levels for candidates or (2) present only an unranked list of candidates above an established minimum similarity score.**

e. Requirements for the **training and certification** of law enforcement officers and staff.

f. Policies and procedures to address failures to adhere to procedures or attain appropriate certification.

g. **Mechanisms for redress** by individuals harmed by FRT misuse or abuse.

h. Policies for the use of FRT for real-time police surveillance of public.

i. Retention and auditing requirements for search queries.

j. Guidelines for **public consultation and community** oversight of law enforcement FRT.

k. Guidelines and best practices for assessing public perceptions of legitimacy and trust.

l. Policies and standardized procedures for reporting of **statistics on the use of FRT** in law enforcement.

Rec 1-4: Federal grants and other types of support for state and local law enforcement use of FRT should require that recipients adhere to the following technical, procedural, and disclosure requirements:

a. Provide verified results with respect to accuracy and performance across demographics.

b. **Comply with the industry standards** called for in Rec 1-2—or comply with future certification requirements.

c. Use FRT systems that present only candidates who meet a minimum similarity threshold.

d. Adopt minimum standards for the quality of both probe and reference gallery images.

e. Use FRT systems only with a **human-in-the-loop** and not for automated detection of offenses.

f. Limit the use of FRT to being one component of developing investigative leads. FRT should be **only part of a multi-factor basis for an arrest** or investigation, in line with current fact-sensitive determinations of probable cause and reasonable suspicion.

g. Restrict operation of FRT systems to law enforcement organizations that have sufficient resources to properly deploy, operate, manage, and oversee them.

h. Adopt policies to **disclose** to criminal suspects, their lawyers, and judges the role played by FRT in law enforcement procedural actions.

i. **Publicly report** on a regular basis de-identified data about **arrests** that involve the use of matches reported by FRT.

j. **Publicly report** on any instances where **erroneous arrests** have been made on the basis of FRT.

k. Conduct periodic independent audits of the technical optimality of an FRT system and the skills of its users.

NATIONAL ACADEMIES *Sciences Engineering Medicine*

Rec 1-5: The federal government should establish a program to develop and refine a risk management framework to help organizations identify and mitigate the risks of proposed facial recognition technology applications with regard to performance, equity, privacy, civil liberties, and effective governance.

Valuable tool for:

- Identifying and managing sociotechnical risks,
- Defining appropriate measures to protect privacy,
- Ensuring transparency and effective human oversight, Identifying and mitigating concerns around bias and equity
- Forming the basis for future mandatory disclosure laws or regulations.

NIST would be a logical organization to be charged with developing this framework given its prominent role in FRT testing and evaluation as well as in developing risk management frameworks for other technologies.

Some issues that might be addressed by a risk management framework (1)

Technical performance—including accuracy and differential performance across standardized demographic groups, quality standards for probe and reference images, and adequate indication of the confidence of reported matches.

Equity—including the extent to which there are statistically and materially significantly different probabilities of error for different demographic groups, the extent to which these are attributable to technical characteristics or other, and the parity of use among different populations.

Privacy—including privacy protection for faces used in training the template extraction model and whether use of FRT significantly increases the scope or scale of the identification being performed

Data collection, disclosure, use, and retention policies for both subject and reference images and templates—including data retention policies to limit, for example, inappropriate use of probe images for searches beyond pre-defined operational needs.

Data security and integrity—including adequately protecting information in training data sets and reference databases from exfiltration and misuse.

Civil liberties—including whether FRT is being used to control access to a public benefit or service and whether the use of FRT will have a reasonably foreseeable negative impact on the exercise of civil rights, such as free speech or assembly, whether by individuals or groups.

Governance—including whether there is an important public interest or legitimate business purpose; who decides whether and how to deploy FRT and who assumes the risks/benefits of its use; consultation with the public at large or with affected groups and meaningful consideration of results; and appropriate safeguards, oversight, and quality assurance.

Some issues that might be addressed by a risk management framework (2)

Disclosure—including where, when, and for what purpose the system is used.

Consent—including whether consent is opt-in or opt-out and whether consent is meaningful and uncoerced, and in the case of mandatory use, whether the justification is clear and compelling.

Training—including what sort of capabilities or competencies the operator of an FRT system, and those using its output, need to demonstrate and whether the training or certification regimes meet the needs of the system usage.

Human in the loop—including whether there is an individual responsible for all significant decisions made on the basis of an FRT match.

Accountability—including who is responsible for addressing systematic technical issues with an FRT system, the manner in which it is used, and ethical and societal concerns that arise from the social environment in which it is used, and whether and how frequently audits are conducted.

Adverse impacts and their distribution—including the potential adverse impacts of a false positive or false negative match in the proposed use, identifying who bears the consequences of those impacts, and indicating whether costs are borne primarily by the individual subject or the operator of the technology.

Recourse—including whether recourse mechanisms provide redress proportional to potential consequences, whether they are available to individuals who will experience adverse outcomes, and whether the organization has a mechanism for receiving complaints.

Rec 1-6: The federal government should support research to improve the accuracy and minimize demographic biases and to further explore the sociotechnical dimensions of current and potential FRT uses.

<u>NIST, DHS Maryland Test Facility, or other institution</u>

- The accuracy of FRT systems in a variety of non-optimal settings, including non-optimal facial angle, focus, illumination, and image resolution.
- The development of representative training datasets for template extraction and other methods that developers can safely apply to existing datasets and models to adjust for demographic mismatches between a given dataset and the public.
- The performance of FRT with very large galleries (i.e., tens or hundreds of millions of entries), to better understand the impacts of false positive and false negative match rates as the size of galleries used continues to grow.

<u>NSF or similar research sponsor</u>

- Developing privacy-preserving methods to prevent malicious actors from reverse-engineering face images from stored templates.
- Mitigating false positive match rate variance across diverse populations and building better understanding of the levels at which residual disparities will not significantly affect real-world performance.
- Developing approaches that can reduce demographic and phenotypical disparities in accuracy.
- Developing accurate and fast methods for directly matching an encrypted probe image template to an encrypted template or gallery, e.g., using fully homomorphic encryption.
- Developing robust methods to detect face images that have been deliberately altered by either physical means such as masks, makeup, and other types of alteration or by digital means such as computer-generated images.
- Determining whether FRT use deters people from using public services, particularly members of marginalized communities.
- Determining how FRT is deployed in non-cooperative settings, public reaction to this deployment, and its impact on privacy.
- Determining how FRT may be used in the near future by individuals for abusive purposes, including domestic violence, harassment, political opposition research, etc.
- Determining how private actors might use FRT in ways that mimic government uses, such as homeowners who deploy FRT for private security reasons.
- Researching future uses of FRT, and their potential impacts on various subgroups of individuals.

Recommendation 2: Developers and deployers of FRT should employ a risk management framework and take steps to identify and mitigate bias and cultivate greater community trust.

Rec 2-1: Organizations deploying facial recognition technology should adopt and implement a risk management framework addressing performance, equity, privacy, civil liberties, and effective governance to assist with decision making about appropriate use of FRT.

Until the recommended risk management framework is developed, the issues in the framework (Rec. 1-5) may serve as a useful point of departure. Future standards documents (Rec. 1-2) may also provide relevant guidance.

Rec 2-2: Institutions developing or deploying FRT should take steps to identify and mitigate bias and cultivate greater community trust—with particular attention to minority and other historically disadvantaged communities.

a. Adopting more inclusive design, research, and development practices.

b. Creating decision-making processes and governance structures that ensure greater community involvement.

c. Engaging with communities to help individuals understand the technology’s capabilities, limitations, and risks.

d. Collecting data on false positive and false negative match rates in order to detect and mitigate higher rates found to be associated with particular demographic groups.

Imperative to help address mistrust about bias in FRT’s technological underpinnings and to respond to broader mistrust, especially in communities of color, about the role of technology in law enforcement and similar contexts

Enacting More Comprehensive Safeguards

Recommendation 3: The Executive Office of the President should consider issuing an executive order on the development of guidelines for the appropriate use of facial recognition technology by federal departments and agencies and addressing equity concerns and the protection of privacy and civil liberties.

Recommendation 4: New legislation should be considered to address equity, privacy, and civil liberties concerns raised by facial recognition technology, to limit harms to individual rights by both private and public actors, and to protect against its misuse.

a. Limitations on the storing of face images and templates

b. Specific uses of concern

c. User training

d. Certification

Rec 4 (a): Limitations on the storing of face images and templates

- Consider **prohibiting storing of face images or templates in a gallery unless** the gallery will be used for a **specifically allowed purpose**. Possible allowed uses include:
  - For prescribed government functions—e.g., at international arrival/departure points
  - Where there is explicit consent for a specific purpose
  - Where there are threats to life and physical safety

Precisely which uses are and are not allowed merits careful consideration by legislators and the public at large. The risk management framework (Rec 1.5) may provide a useful tool for considering these questions.

- Consider whether it is appropriate to use **images collected from the Internet without consent** or knowledge and the implications of including low-quality or synthetic images collected in this manner.

Rec 4 (b): Specific uses of concern

- **Commercial practices that implicate privacy** (through either broader privacy legislation addressing FRT risks or an FRT-specific federal privacy law),
- Harassment or blackmail,
- Unwarranted exclusion from public or quasi-public places,
- **Especially sensitive government FRT uses** (e.g., pertaining to law enforcement or access to public benefits or federally subsidized housing),
- Public and private uses that tend to **chill the exercise of political and civil liberties**—both intentional or from the emergent properties of use at scale—and
- **Mass surveillance or individual surveillance** other than that properly authorized for law enforcement or national security purposes.

Rec 4 (c): User training

- Consider requiring **training for operators and decision makers**—in applications where the operator must apply judgment or discretion in when/how to use FRT systems or in interpreting their results and where a false match may result in significant consequences for an individual
- Training less critical—in applications where the fallback in case of a failure is simply to inspect a government-issued ID

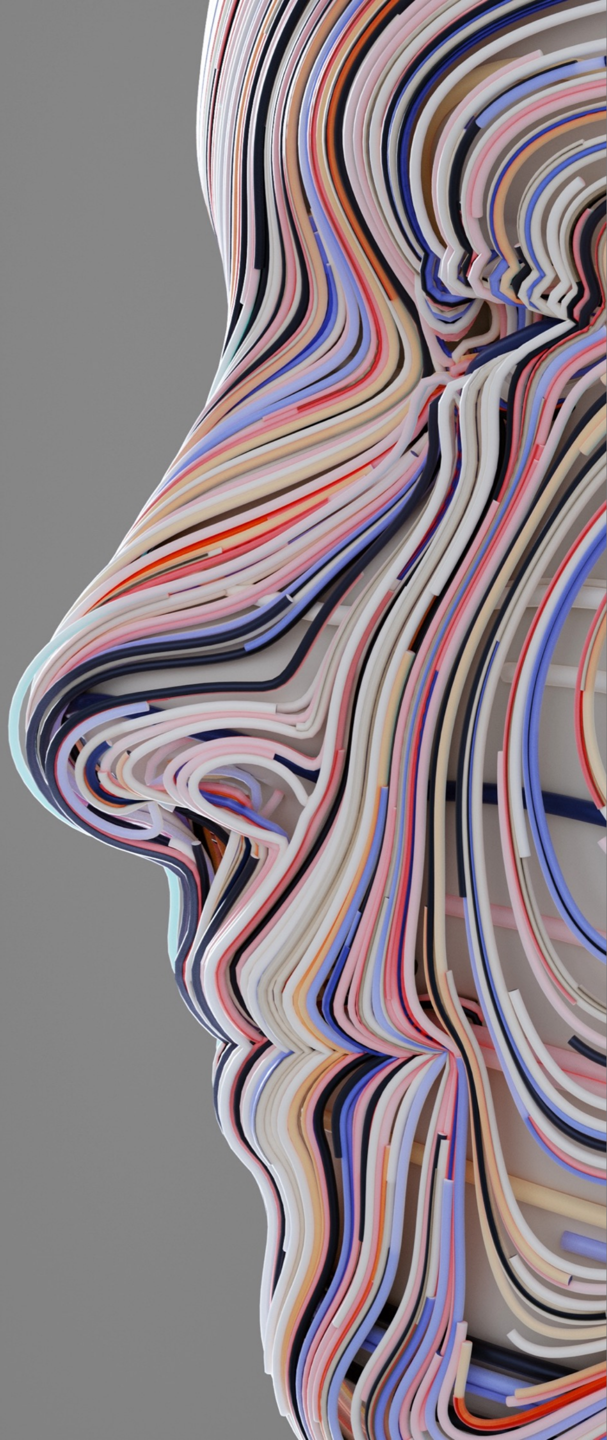
Rec 4 (d): Certification

- Consider requiring **certification of operators** and other users or **certification of organizations** that operate FRT systems for applications where **technical or procedural errors can significantly harm subjects**, notably in law enforcement.

FRT is a powerful tool with profound societal implications. Its attributes make it especially salient for privacy and civil liberties.

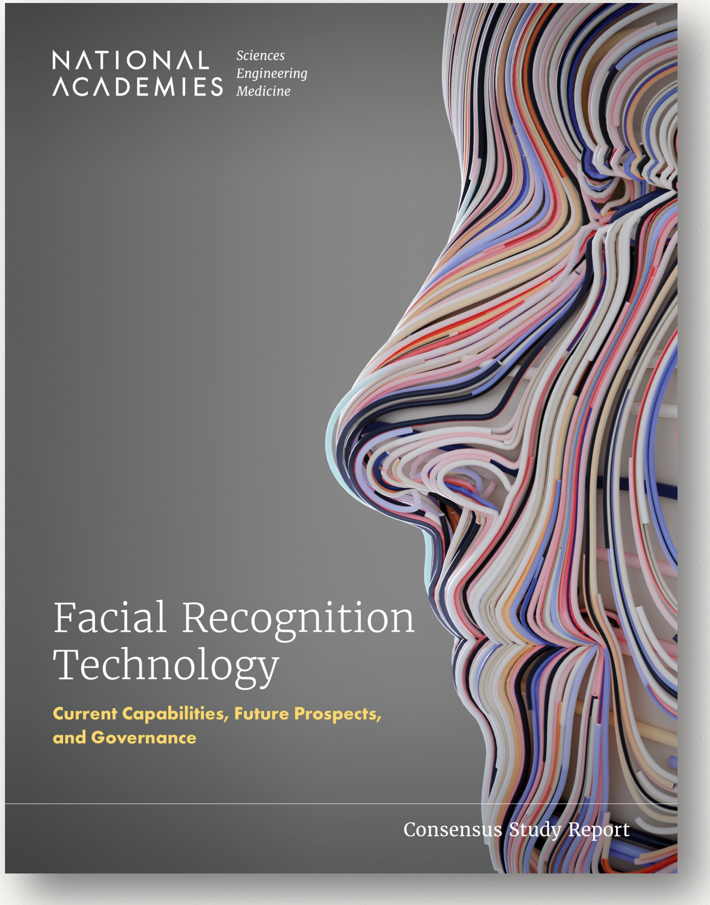
FRT is developing quickly, and governments have fallen behind in addressing its implications.

In developing stronger safeguards, governments and other institutions will need to consider the views of a wide range of constituencies, especially the communities most affected by the technology.



Thank you!

Report is available for download from National Academies Press at: https://nap.edu/27397



NATIONAL ACADEMIES Sciences Engineering Medicine

Facial Recognition Technology

Current Capabilities, Future Prospects, and Governance

Consensus Study Report