

Science & Technology Innovation Program



W | **Wilson
Center**

W Science and Technology
Innovation Program



Author
Anne Bowser

Beyond Bans: Policy Options for Facial Recognition and the Need for a Grand Strategy on AI

September 2020





In [an open letter](#) to Congress published on June 8, 2020, IBM CEO Arvind Krishna announced the sunset of IBM’s general-purpose facial analysis program. Calling to reform a biased criminal justice system, Krishna expressed support for the [George Floyd Justice in Policing Act of 2020](#) (H.R.7120). Among other provisions, this bill would prohibit the use of facial recognition technology to analyze images from in-car or body-worn video cameras. Krishna also called for new policies around bias testing and auditing in artificial intelligence (AI) systems, noting “now is the time to begin a national dialogue on whether and how facial recognition technology should be employed by domestic law enforcement agencies.”

Other technology companies quickly chimed in. [Amazon implemented a one-year moratorium](#) on police use of their facial recognition technologies, hoping this “might give Congress enough time to implement appropriate rules.” Microsoft [also announced new limits](#). In an industry that rarely self-regulates, these decisions reveal the magnitude of issues relating to bias in facial recognition, and its impact on civil liberties in policing and criminal justice.

But what’s more—all three companies not only announced self-regulation, but actively called for policy reform. While policy reform is needed, simply banning facial recognition is not enough.

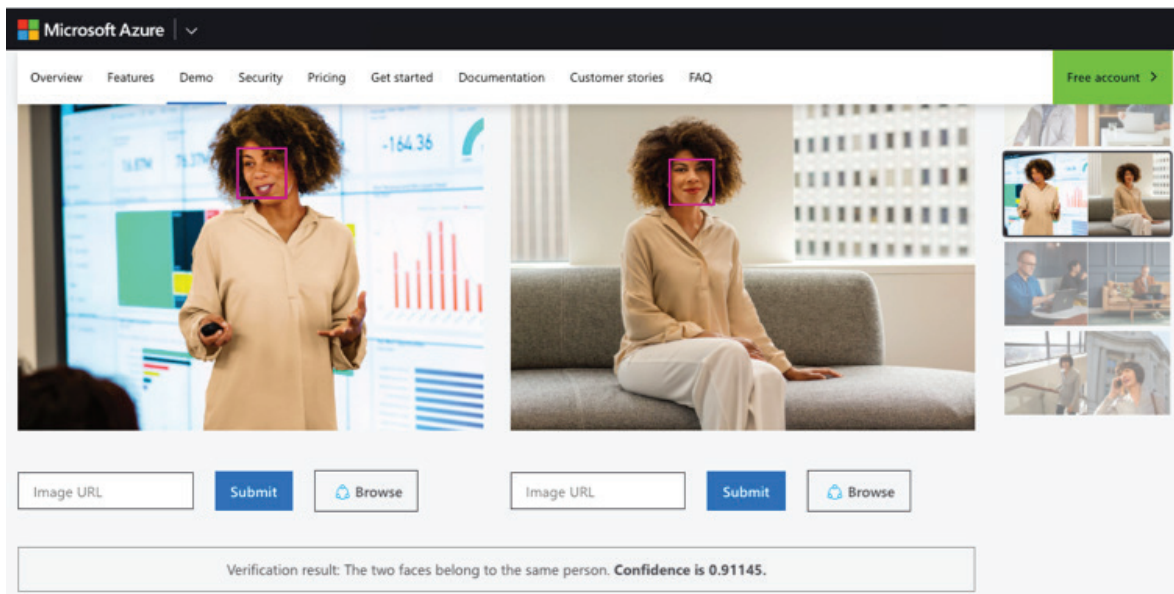
Bias is also an issue in criminal justice when AI is used to support parole decisions. When considering facial recognition more generally, ethical considerations extend beyond bias and also encompass privacy concerns. Planning to support ethical artificial intelligence requires moving beyond short-term, targeted efforts like bans to encompass bigger picture policy thinking. The U.S. needs a strategy for artificial intelligence that includes a framework for considering key ethical issues in all research, development, and use. The issue of bias in facial recognition can become an important and timely case study to help explore what effective policy options might include.

Understanding the Immediate Issue: Facial Recognition and Bias

Facial recognition is an important application of AI and a subset technology called machine learning (ML). In computational facial recognition, an algorithm analyzes a two- or three-dimensional image to first identify a face, and then extract identifying features. This “faceprint” is compared to one or more images retrieved from a database to determine a similarity score, or the likelihood that a match between two images exists.

From a policy perspective, facial recognition is classified as one type of [biometric](#) technology, or as a technology where the measurement of physiological characteristics including—but not limited to—fingerprints, iris patterns, or facial features are used to identify an individual. From a technological perspective, facial recognition is one application of facial analysis, a cluster of techniques that analyze and extract information from faces. Other, non-biometric applications of facial analysis include [automated gender recognition](#) (AGR) technologies, and tools that estimate the age of an individual.

Facial recognition can be useful for verification through “one-to-one” matching. In these cases, a new image is compared to one or more reference images to confirm the identity of an individual. One-to-one matching is helpful in applications ranging from unlocking a smartphone, to participating in an automated program for airport security screening or passport control. One-to-one matching can also be used in criminal justice, for example, to determine whether the mugshot of a suspect matches a surveillance camera’s video still.



Microsoft's [Cognitive Service Demo](#) is openly available for potential users to explore. This screenshot shows Microsoft's one-to-one facial recognition capabilities, which are used for verification. In this case, the likelihood that both photos depict the same individual is offered with 0.91145 confidence (approximately 91%).

In "one-to-many" matching, a new image is compared to a larger database of images taken from many individuals with the intent to identify a match. Delhi police recently [made headlines](#) for using one-to-many facial recognition in Operation Smile, a campaign to tackle child labor and missing children. The app, described as a game changer by a human rights NGO, helped police reunite over 1,500 missing children with their families in less than one month.

One-to-many matching is also used to analyze images from body cameras. Within the U.S., images from body cams are typically compared to images in databases of criminal mugshots, though many states also use databases [including DMV records](#). Due to advances in cloud storage, the feasibility of real-time one-to-many analysis is already being investigated by groups [like Customs and Border Protection \(CBP\)](#).

Different types of facial recognition algorithms fail by making one of two errors. A false positive identifies a match between two images that do not depict the same individual. A false negative fails to detect a match when one exists.

While both types of errors are problematic, false positives are particularly problematic when considered through the lens of civil rights concerns. For example, the 2018 [Gender Shades](#) study evaluated the performance of different facial recognition algorithms. The authors reported strong demographic effects: darker-skinned women were misclassified by commercial algorithms with a rate as high as 34.7%, while the maximum error rate for white males was reported as 0.8%. While this research has been [criticized](#) for methodological issues,



Gender Shades is still important for elevating the conversation around bias in facial recognition. The study also demonstrated the relevance of intersectionality, a theoretical framework that explores how different aspects of an individual's identity, such as race and gender, can interact to create unique types of discrimination or privilege.

These findings have also been validated in other studies. In 2019, the National Institute of Standards and Technology (NIST) conducted [a large-scale assessment](#) of the performance of facial recognition related to demographic effects. In regard to racial differences, NIST found that false positives were highest in West African, East African, and East Asian Individuals, with the exception of algorithms developed in China. False positives were slightly higher for South Asian and Central American people, and lowest for Eastern Europeans. NIST also reported demographic effects related to gender and age.

Higher rates of false positives for darker-skinned individuals means that Black and brown communities are more likely to be falsely identified by facial recognition technologies. These findings are particularly concerning given that there is no federal standard, or agreed-upon threshold, that determines how accurate a facial recognition algorithm must be before it is used in law enforcement.

The Gender Shades study isolated one important variable: training data. Researchers found that many benchmark training data sets were overwhelmingly biased towards lighter-skinned individuals. While NIST did not attempt to identify causation, the agency did suggest that attempts to mitigate demographic differentials could benefit from research on aspects including "threshold elevation, refined training, more diverse training data, and discovery of features with greater discriminative power."

Higher rates of false positives for darker-skinned individuals means that Black and brown communities are more likely to be falsely identified by facial recognition technologies.

Bias in Criminal Justice Is More Pervasive Than Facial Recognition

Thanks to advances in AI and ML, long-term issues of bias in criminal justice have been brought to the fore. Many recent conversations are centered around facial recognition, a ML tool designed to augment limitations to human perception. Other applications of ML in criminal justice are used to support human decision-making, for example through the use of risk assessment instruments (RAIs) such as the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) in parole decisions.

These tools are meant to help mitigate human bias by providing data-driven, and therefore "objective," assessments of the likelihood that a paroled criminal will commit another crime. But RAIs are only as effective as the data that they are trained on. While data related biases can be an issue, these applications also elucidate the importance of understanding systemic bias in AI.

In criminal justice, systematic bias is a paramount concern. Empirical research demonstrates that [Blacks and non-white Hispanics are incarcerated at much higher rates than whites](#), are more likely to be incarcerated for the



same behaviors, and are incarcerated at younger ages. When RAIs like COMPAS are trained on such data, they are likely to make predictions on recidivism that lead to outcomes that negatively (and unfairly) impact individuals. In this case, training data are biased not because of poor selection, but because the trends reflected in training data are indicative of, and ultimately re-enforce, larger systemic social inequalities. Without proper guardrails, these systemic biases will be perpetually codified in AI applications, driving outcomes that exacerbate inequalities and unfair treatment in a vicious feedback loop.

Approaches like standards and testing—as exemplified by NIST’s Facial Recognition Vendor Test program—can effectively assess performance limitations that are rooted in data-related biases. Contending with systemic bias is a much more complex undertaking. For example, if a RAI was evaluated to assess whether its predictions matched outcomes from a historical benchmark data set, the RAI could demonstrate technical accuracy while still



*The Black Lives Matter movement shows a growing awareness of racial bias in criminal justice.
Photo by Chris Henry on Unsplash*



producing biased recommendations. In other words, the RAI could be accurate in reflection of real-world trends, but might not be considered “fair.”

Standards and testing help bring transparency to the performance of a ML application. Requirements for transparency can also be met through other processes, like making training data or algorithmic code available as open source. For example, one method for facilitating transparency is the [model cards](#) approach, which offers a structure for sharing information including details about the model and its intended use, relevant factors (including those related to demographics), performance metrics, and information on the data used in training or evaluation.

Theoretically, approaches to transparency like model cards can help crack the “black box” that makes systems seem opaque. However, many real-world machine learning applications are often classified as trade secrets. U.S Customs and Border Protection was [unable to determine the causes of failure in a tool for iris recognition](#) using proprietary code. In criminal justice, this [prevents a judge](#) or other official from evaluating the likelihood of bias and determining whether an algorithm could be considered fair. Approaches to transparency are also limited when there is no party able and empowered to understand and act on the information shared.

Notably, while no federal statute on facial recognition exists, a handful of states and municipalities have banned the use of facial recognition in criminal justice.

Notably, while no federal statute on facial recognition exists, [a handful of states and municipalities](#) have banned the use of facial recognition in criminal justice. In June 2019, the Somerville, MA city council unanimously voted to ban the use of facial recognition by agencies including the police department. In October 2019, California became the first state to place a three-year moratorium on the use of facial recognition by law enforcement. Legislation is now pending in dozens of other states. But, as Georgetown University Law Center’s [The Perpetual Line-Up](#) project illustrates, broader use of facial recognition in law enforcement prevails.

Bias in Facial Recognition Is Not the Only Concern

While bias is one ethical concern in facial recognition, other ethical issues also need attention. Chief among these is privacy, and the potential for facial recognition technologies to violate legal or ethical rights.

In 2015, the U.S Government Accountability Office (GAO) published a report to the U.S. Senate’s Subcommittee on Privacy, Technology, and the Law, titled “[Facial Recognition Technology: Commercial Uses, Privacy Issues, and Applicable Federal Law](#).” Based on an extensive multi-stakeholder consultation, GAO identified privacy concerns related to “the ability to identify and track individuals in public.” These included threats to reasonable expectations of privacy and questions about whether individuals subject to facial recognition analysis had appropriate knowledge and consent. Additional concerns were raised regarding the sensitive nature of biometric data, including the lack of meaningful control over one’s personally identifiable information (PII), the potential for



disparate treatment or discrimination, and cybersecurity threats. Many of these same issues are highlighted in [The Perpetual Line-Up](#) risk framework, which also identifies variables including whether facial recognition analysis is conducted in real time or after the fact, and whether there is precedent for conducting similar activities with different biometric technologies (like fingerprints).

Among stakeholders interviewed by GAO, there was general consensus that while some federal regulations apply, more work is needed to address the full range of privacy concerns. First, while statutes like the Privacy Act of 1974 protect personal information collected by federal agencies, no similar, cross-cutting legislation applies to the private sector. Instead, laws typically cover a limited application area—such as the Health Insurance Portability and Accountability Act (HIPPA)’s protections for health and medicine—or address vulnerable populations, such as the Children’s Online Privacy Protection Act (COPPA). These findings are supported by [a recent graduate thesis](#) on regulatory gaps in AI that explored a case study of facial recognition technologies, concluding that “no unified set of rules governs their use; instead, multiple laws and regulations create a disjointed policy environment, limiting the extent to which privacy and bias concerns can be mitigated for these implementations.”



Citing privacy concerns, a handful of states have enacted legislation limiting commercial activity in biometric identification including facial recognition. California’s laws [are some of the most restrictive](#), while Illinois’ laws have held up to judicial scrutiny. In [Patel vs. Facebook](#), which [ultimately settled out of court for \\$550 million](#), a class action lawsuit found Facebook’s use of facial recognition in violation of privacy rights protected by the Illinois Biometric Information Privacy Act. This ruling was significant because it did not identify specific grievances or harms related to privacy violations, but rather found that privacy violations are inherent to facial recognition technologies. In other words, all uses of facial recognition will raise privacy concerns.

As explored earlier, there is some evidence of private sector self-regulation around facial recognition in criminal justice.

*Critics worry that facial recognition may violate privacy concerns, especially when combined with surveillance technologies.
Photo by Ennio Dybeli on Unsplash*



Non-profit and other NGO stakeholders have also helped develop codes of conduct to shape norms. Beginning in 2013, the National Telecommunications and Information Administration (NTIA) led a multi-stakeholder process to address the commercial use of facial recognition. The process culminated in 2016, with a set of [Privacy Best Practice Recommendations for Facial Recognition Technology](#).

Key principles include transparency, developing good data management practices, use limitations, security safeguards, data quality, and problem resolution and redress. However, these principles do not apply to areas outside of consumer privacy (including in national security or to law enforcement). But there is still disagreement on whether the principles should be voluntary or enforceable.

In addition to filling data gaps directly, research conducted through citizen science can help advance new methodologies.

The Need for a Bigger Picture Perspective

Addressing the use of facial recognition in law enforcement might seem like a powerful first step. But, as demonstrated by the prevalence of bias in other criminal justice processes, addressing facial recognition in isolation will not help us contend with broader systemic issues around justice and race. Similarly, while bias is an important ethical concern at the forefront of the facial recognition policy debate, concerns such as privacy also need attention.

Despite these considerations, the potential for AI and ML to bring economic and social benefits is unmatched by any technology since the Internet. Programs including the Trump Administration's initiative on [Artificial Intelligence for the American People](#) are accelerating AI through direct investment, reduction of regulatory barriers, workforce training, and strategic international cooperation. Recognizing the accelerated pace of innovation, there is an urgent and compelling opportunity to move beyond bans to proactively explore a grand strategy or national action plan for AI. Such a strategy should include an ethical framework for AI that begins with high-level principles and becomes progressively more concrete.

Building blocks: Principles. In the earliest stages, work on ethics often begins with principles. The United States has already contributed to and endorses the OECD's [five values-based principles for responsible stewardship of trustworthy AI](#). The Department of Defense (DoD) issued [five AI ethical principles](#) to cover the unique challenges faced by defense and security communities. Most recently, the White House issued a draft memorandum proposing [ten legally binding principles](#) to help agencies formulate regulatory and non-regulatory approaches.

Ethical principles identify guiding priorities. For example, one of the OECD's values-based principles states:

"AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards—for example, enabling human intervention where necessary—to ensure a fair and just society." (OECD)



The concept of a fair society is echoed in the draft White House AI principles “Fairness and Non-Discrimination.” This is perhaps unsurprising. [A meta-review of 84 sets of ethical principles](#) found that, at least among Western democracies, principles are unlikely to contradict one another, and converged around themes including transparency, justice and fairness, responsibility, and privacy.

Principles can provide context for understanding why various ethical concerns matter. For example, bias in facial recognition could impede progress towards a “fair and just society.” But while principles codify expectations, they do not suggest how expectations can be met.

Clarify principles through definitions. One step towards making generic principles actionable is writing definitions, which may happen while crafting policy or through standards development processes. Artificial intelligence was formally defined in the [John S. McCain National Defense Authorization Act for Fiscal Year 2019](#). Defining artificial intelligence was a critical first step towards cementing the importance of this concept and policy discussions, and clarifying what is in and out of scope during discussions about AI.

More definitions are required. A second OECD principle states, “There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.” However, NIST notes, “without clear standards defining what algorithmic transparency actually is and how to measure it, it can be prohibitively difficult to objectively evaluate whether a particular AI system...meets expectations.”

The European Commission [describes trustworthy AI](#) as encompassing human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; environmental and societal well-being; and accountability. An evaluation of trustworthy facial recognition technology should consider all aspects of this definition, including privacy and non-discrimination. In this way, cohesive definitions of important concepts like transparency can be helpful for providing an initial checklist of specific requirements or concerns.

Codify definitions in standards. Technical standards and other guidance suggest exactly how systems might be evaluated as (for example) “trustworthy.” Domestically, NIST is leading government development of standards in close coordination with other federal agencies. Internationally, U.S. participation in standards-setting bodies by NIST and others should accommodate different levels of activity including monitoring and participating, but especially influencing and leading.

[According to NIST](#), domestic standards should leverage international developments to the greatest degree possible. Developing standards for trust, along with related work on non-technical standards addressing ethical, governance, and privacy concerns, will provide a litmus test for determining whether ethical requirements are satisfied. Work on related tools, such as data sets and benchmarks—challenge statements created to measure performance around strategically selected scenarios—is also needed.

For example, NIST might develop training data sets that represent full demographic diversity to help lead to less biased facial recognition technologies. And work on benchmarks could investigate whether clever assessments could be designed to identify systemic bias in decision-making algorithms used to support parole decisions.



Create general and domain-specific requirements. Building on principles, definitions, and standards, requirements are fully prescriptive, and can be technical or non-technical.

One technical opportunity is the development of architectures for trustworthy AI. Requirements could encompass mandatory and/or prohibited behaviors. For example, a technical requirement for facial recognition used in police body cams could be that the system does not display matches below a certain confidence threshold. Technical requirements can also codify the role of humans “in the loop” by preventing certain decisions from being made without human review.

Non-technical requirements can look across the AI lifecycle to create provisions for development, procurement, and use. Building on the success of privacy impact assessments, algorithmic impact assessments [are being used by U.S. allies like Canada](#) to help government agencies consider and mitigate risk before a system is developed or deployed. Provisions for continued auditing, including internally and by third parties, can help ensure that systems are functioning as they should regarding ethical, safety, and other performance concerns. Notably, technical and non-technical requirements can be applied to systems that are developed in house, and also, as [the World Economic Forum points out](#), through procurement.

Using tools like algorithmic impact assessments for facial recognition would provide scaffolding for thinking through issues like bias and privacy. Making the outcomes of algorithmic impact assessments open and available for stakeholders like the public could help unpack the black box to support greater transparency and trust.

Cross-cutting efforts to support and advance AI. In addition to an ethics framework, broader efforts are needed to support and advance American AI. First, a formal assessment of regulatory gaps is needed. The 2015 GAO study offers a helpful starting point, but it is limited by the focus on privacy and commercial use. The White House Draft Memorandum suggests that agencies should describe statutory restrictions on collecting and sharing information, and report on regulatory barriers. Building on both of these, additional work on technologies like facial recognition is needed to identify gaps in protections that could be addressed by clarifying existing statute, through new law, by delegating authority to other levels of government, or through voluntary approaches.

Second, AI needs ongoing coordination and support from a strong federal authority. The White House Office of Science and Technology Policy (OSTP)’s coordination of federal agencies is a valuable starting point, but is insufficient to ensure ongoing alignment with the private sector or as well as domestic and international NGOs. Relatedly, efforts to engage civil society should look beyond the federal register to build public understanding and foster trust through mechanisms like roundtables and two-way discussion forums.

Provisions for continued auditing, including internally and by third parties, can help ensure that systems are functioning as they should regarding ethical, safety, and other performance concerns.



Lastly, investments in education and workforce development are urgently needed. For students, new technical training and education opportunities could help meet growing workforce demands, and should include mandatory content on ethics. Training for agency employees should target leadership, procurement, and technical roles. Innovations are also needed to help civilians move into government roles, and could include leveraging direct hiring authorities, relying on subject matter experts over human resource (HR) professionals to make selections, and offering referral bonuses for highly qualified individuals.

Now is the time to begin a national dialogue—but not just about facial recognition. As the pace of innovation continues to accelerate, so will the need for a set of checks and balances to promote the development of safe and ethical AI. A Grand Strategy for AI that includes an ethical framework will enable us to take advantages of AI's social and economic advances while aligning with American values, and demonstrate leadership on the global stage.

Acknowledgements

Thanks to Bob Greenberg, Meg King, and Morgan Livingston for a comprehensive first draft review. Elizabeth Newbury and Metis Meloche provided helpful comments while editing the final version.








WOODROW WILSON INTERNATIONAL CENTER FOR SCHOLARS

The Woodrow Wilson International Center for Scholars, established by Congress in 1968 and headquartered in Washington, D.C., is a living national memorial to President Wilson. The Center's mission is to commemorate the ideals and concerns of Woodrow Wilson by providing a link between the worlds of ideas and policy, while fostering research, study, discussion, and collaboration among a broad spectrum of individuals concerned with policy and scholarship in national and international affairs. Supported by public and private funds, the Center is a nonpartisan institution engaged in the study of national and world affairs. It establishes and maintains a neutral forum for free, open, and informed dialogue. Conclusions or opinions expressed in Center publications and programs are those of the authors and speakers and do not necessarily reflect the views of the Center staff, fellows, trustees, advisory groups, or any individuals or organizations that provide financial support to the Center.





Woodrow Wilson International Center for Scholars
One Woodrow Wilson Plaza
1300 Pennsylvania Avenue NW
Washington, DC 20004-3027

The Wilson Center

 www.wilsoncenter.org
 wwics@wilsoncenter.org
 facebook.com/woodrowwilsoncenter
 [@thewilsoncenter](https://twitter.com/thewilsoncenter)
 202.691.4000



STIP

 www.wilsoncenter.org/program/science-and-technology-innovation-program
 stip@wilsoncenter.org
 [@WilsonSTIP](https://twitter.com/WilsonSTIP)
 202.691.4321

