Bias and AI
ENSURING INCLUSIVE TECH

The increasing use of artificial intelligence (AI) for decision-making has highlighted its potential for bias and discrimination, resulting in unintended societal consequences and liability for organizations. Organizations that develop or leverage AI-enabled technology must identify AI blind spots, keep current with AI ethics trends, and monitor recommendations for inclusive technology (inclusive tech) to ensure fair, equitable, and transparent AI use.

ELIZABETH M. ADAMS
CEO
EMA ADVISORY SERVICES

Elizabeth is an AI ethics and organizational culture advisor with substantial expertise in technology leadership, civic technology policy, social advocacy, inclusive tech design, and diversity and inclusion. She has led large-scale technology initiatives for Fortune 500 companies and government organizations. Elizabeth also serves as the Chief AI Ethics Advisor for Paravison and as the Global Chief Culture & Ethics Officer for Women in AI.

Organizations continue to develop and leverage AI systems at a rapid pace to enhance workflows, streamline operations, and improve efficiency. With the increasing adoption of AI, there has been pressure at the federal, state, and local levels to regulate AI's use and ensure consumers and other individuals are protected from discriminatory impacts, among other potential abuses.

Awareness is growing about the potential for bias and discrimination in using AI for decision-making, especially when used in areas such as criminal prosecution and hiring decisions. Unfortunately, leaders at many organizations are removed from the concept and design phase of AI-enabled technology and often miss key opportunities to direct the fair, equitable, and transparent use of AI. Failing to include diverse perspectives when engaging in technology development projects often results in AI blind spots that can yield unintended societal consequences and increase potential liability for organizations.

This article:

- Provides a high-level overview of AI, the typical AI technology development life cycle, and digital surveillance technologies, such as facial recognition technology (FRT).
- Highlights recent AI ethics trends.
- Discusses AI bias, including biometric privacy and FRT bans.
- Makes the case for inclusive tech and provides recommendations organizations can adopt to ensure technology is inclusive.

Search Trends in AI Regulation: 2020 for information on key developments in US federal and state regulation of AI.

Search Artificial Intelligence (AI) in the Workplace for information on the legal issues raised when employers use AI tools to perform human resources and employee management functions.

AI BACKGROUND

AI is the ability of a machine to think, learn, and perform tasks like humans, but at a much faster pace. AI is all around us and impacts various aspect of our lives. For example, AI is used to:

- Predict traffic patterns.
- Unlock a mobile phone through FRT.
- Offer preferred user content in streaming services, such as Netflix.
- Power home security systems and home automation.
- Diagnose diseases.
- Preserve environmental resources, from combating the effects of climate change to improving recycling systems.
- Improve education.

Recent AI developments continue to show how AI affects our daily personal and professional lives, such as:

- Tesla Bot, a humanoid robot that uses Tesla's vehicle AI.
- Singapore's patrol robots, which patrol public areas and deter poor social behavior.
- Meta's Horizon Workrooms, a virtual reality application that allows co-workers to interact in virtual offices.

Given the proliferation of AI, it is important to analyze and explore:

- The AI development process through each stage of the technology development life cycle to determine the origin, capabilities, and programming decisions made.
- The issues and challenges raised by FRT and other digital surveillance technologies.

TECHNOLOGY DEVELOPMENT LIFE CYCLE

There are specific technical stages of technology development (for more information, search Software Development: Understanding Agile and Scrum Methodology on Practical Law). Additional stages of developing and launching a new technology product are relevant from an organizational perspective. Traditionally, adjacent teams become involved towards the end of the life cycle, as set out in *Figure 1*.

***Figure 1*: Technology Development Life Cycle (Traditional Approach)**

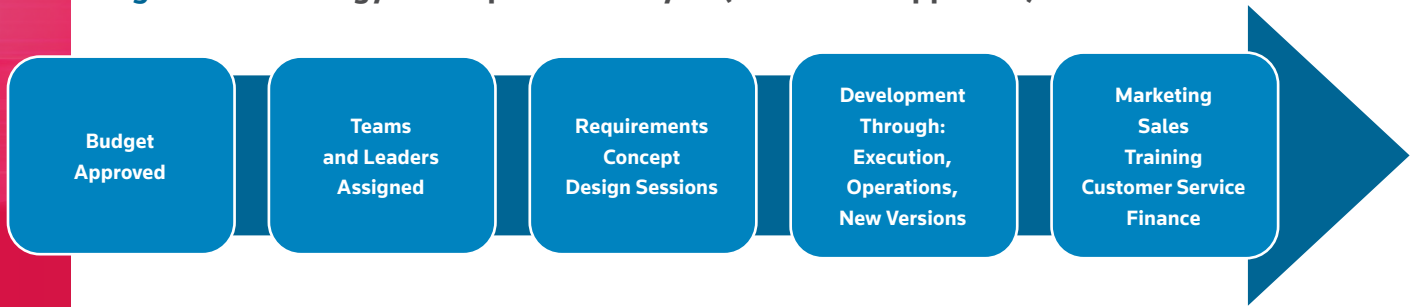
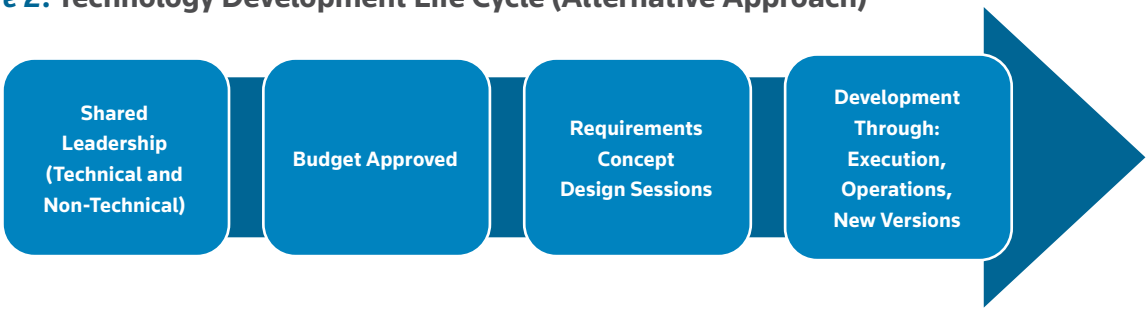


***Figure 2*: Technology Development Life Cycle (Alternative Approach)**



Other approaches include adjacent teams in the stages of technology development throughout the entire life cycle, as set out in *Figure 2*.

Technology development projects require collaboration and face constraints on price, time, functionality, and resources. Organizations typically approach the technology development process using a certain methodology, such as traditional waterfall methodology or agile methodology, each of which addresses these constraints in different ways. Considering these constraints, many approvals are typically needed to progress from one stage of the technology development life cycle to the next.

Collaboration within the technology development life cycle involves, among other things:

- System testing.
- User testing.
- Deciding whether to move forward at each stage.
- Discussing trade-offs concerning certain requirements, such as minimally viable product capabilities.

The technology development life cycle therefore presents various opportunities to:

- Evaluate whether efforts have been made to mitigate bias when developing AI.
- Incorporate accountability at an organization to prevent discriminatory outcomes.

An organization's leaders can also direct fair and equitable product development by being aware of and involving themselves in the technology development life cycle during each stage.

DIGITAL SURVEILLANCE TECHNOLOGIES

Digital surveillance technologies are creating challenges in cities across the globe. Policymakers continually face digital surveillance issues, such as:

- How to address data privacy, digital justice, and digital inclusion.
- How societies are impacted by digital surveillance technologies.
- Whether governments and organizations can be certain that digital surveillance technologies are trustworthy.

To begin to address these issues, it is important to understand:

- FRT, a widely used digital surveillance technology.
- Other types of digital surveillance technologies.

FRT

Facial recognition is a form of computer vision. A computer program is trained by reviewing large datasets of images to find patterns, classify images, and detect faces, features, and objects. A dataset can contain millions of images to help train systems to spot specific attributes, such as eye and skin color and facial hair.

The first step of FRT is enrolling an image, which involves:

- Capturing the image of a face (taking a picture).
- Converting the image.
- Storing the image for possible future comparison and match.

The technology converts images by using nodal points on a face to gather numeric information (the distance from one point to another or the angle between points) and then create a numeric code that is a faceprint. The image of a face is mapped using an algorithm that embeds recognition points into the image. To detect someone, the technology uses algorithms to compare that face (the probe) to the other previously stored images. The matches are usually rated by percentage, for example, a 58% match.

Use of FRT is being regulated and monitored (see below *FRT Bans*). In February 2017, the National Institute of Standards and Technology (NIST) released an Ongoing Face Recognition Vendor Test (FRVT) that started a new evaluation of FRT (see NIST, Ongoing Face Recognition Vendor Test (FRVT): Part 1: Verification (NISTIR Draft) (Nov. 22, 2021), available at *nist.gov*). The FRVT is considered the gold standard for assessing the use of FRT for identity verification. It measures the performance of automated FRTs applied to a wide range of civil, law enforcement, and homeland security applications, such as:

- Verification of visa images.
- De-duplication of passports.
- Recognition across photojournalism images.
- Identification of child exploitation victims.

In addition, NIST released the Face Recognition Vendor Test (FRVT): Part 7: Identification for Paperless Travel and Immigration (NISTIR 8381) in July 2021, available at *nist.gov*, to test FRT's accuracy for airplane boarding.

Because FRTs are based on algorithms, the concern is not necessarily about technology that has already been developed, but rather what types of thinking and design impact FRT outcomes. The life of the algorithm (that is, the decisions that influenced where the algorithm started versus where the algorithm ended up) is part of many AI-powered systems used by governments to help govern society and by law enforcement to assess an individual's potential threat to a community. The risk, however, is that the algorithms comprising FRTs may create flawed predictions.

Other Digital Surveillance Technologies

Other types of digital surveillance technologies include:

- **Computer vision.** This includes gait recognition, which analyzes structural properties of the human body to characterize an individual's gait pattern, and license plate readers.
- **Location tracking.** This includes GPS, cell site simulators, drones and aerial surveillance, and Amazon's Ring video surveillance doorbell.
- **Biometric surveillance.** This includes DNA, thermal cameras, and infrared lasers, such as the US military's Jetson device that can identify individuals by heartbeat.

Many of these tools are available to law enforcement, who use them to augment their work.

Search Biometrics and Local Government Issues for more on the types of biometric identifiers.

Search Tracking Technologies: Privacy and Data Security Issues for more on the privacy issues surrounding common consumer tracking techniques.

AI ETHICS TRENDS

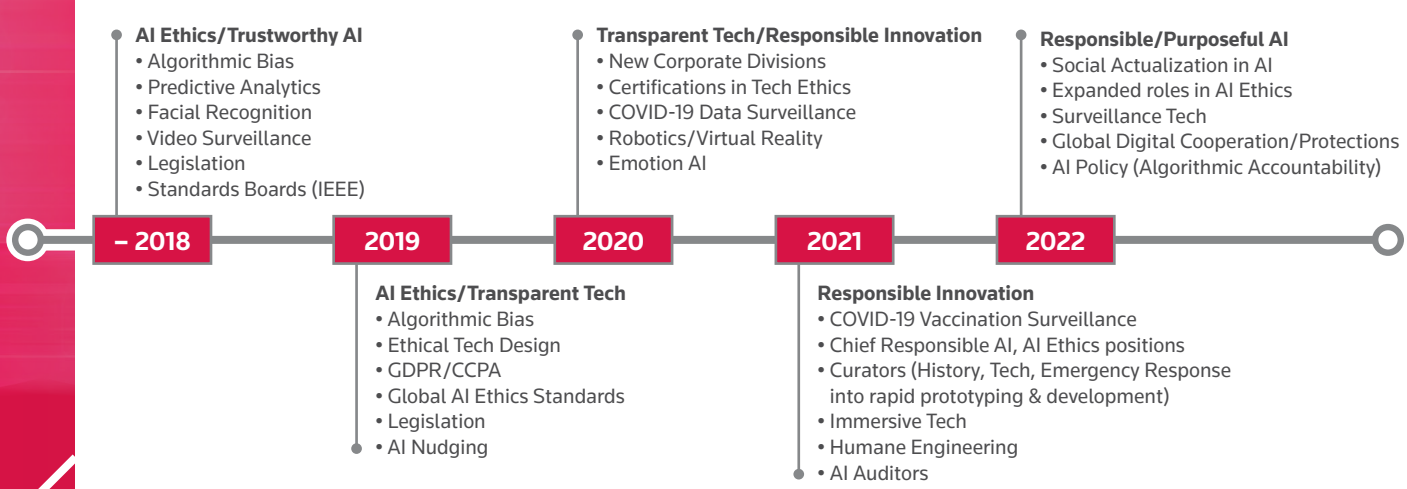
AI ethics involves a system of principles governing the use of AI. In a recent study commissioned by Infosys, Amplifying Human Potential: Towards Purposeful Artificial Intelligence (Infosys Report), available at *infosys.com*, 1,600 IT and business decision-makers were interviewed from organizations with more than 1,000 employees and $500 million or more in annual revenue across ten sectors and seven countries. Of those interviewed, 53% believe ethical concerns prevent AI from being as effective as it can be and only 36% believe that their organization completely considered the ethical issues related to AI use.

Organizations must properly balance maximizing AI's potential with addressing ethical concerns. Several trends have emerged over the past few years to address the ethical use of AI, including developments related to:

- Responsible innovation.
- Transparent technology.
- Trustworthy AI.

(See *Figure 3*.)

Figure 3: Embracing AI Trends



To avoid potential AI pitfalls, organizations should staff individuals in technical and non-technical AI roles to ensure ethics by design.

One AI ethics trend in 2022 is moving towards the use of purposeful AI, which aims to ensure a purposeful approach when leveraging AI to transform businesses without losing sight of the values and ethics involved (see Infosys Report). Aspects of purposeful AI include:

- **Social actualization in AI.** Social actualization is akin to self-actualization at a community or social level. It aims to use AI to help society reach its highest potential by helping society thrive and bridging gaps to decrease division.
- **Expanded AI ethics roles.** To avoid potential AI pitfalls, organizations should staff individuals in technical and non-technical AI roles to ensure ethics by design. For example, Salesforce, a cloud-based software company, created a responsible AI program that includes technical and non-technical AI roles, as well as other procedures and frameworks to address potential harms, such as algorithmic bias.
- **Oversight of surveillance tech.** It is important to ensure responsible uses of technology related to surveillance tools and data privacy, and address domestic and international policies (see Federation of American Scientists, A More Responsible Digital Surveillance Future: Multi-Stakeholder Perspectives and Cohesive State & Local, Federal, and International Actions (Feb. 2021), available at *fas.org*).
- **Global digital cooperation.** In 2020, the United Nations issued the Secretary-General's Roadmap for Digital Cooperation, available at *un.org*, to address how the international community could work together to optimize the use of digital technologies and mitigate risks.
- **AI policy (algorithmic accountability).** AI policy on issues such as algorithmic accountability is another aspect of purposeful AI. The Algorithmic Accountability Act of 2019 (S. 1108 and H.R. 2231, both introduced on April 10, 2019) remains relevant currently as the first federal legislative effort to regulate AI across industries. These bills would direct the Federal Trade Commission (FTC) to require entities that use, store, or share personal information to conduct data protection impact assessments and impact assessments on automated decision systems, including AI-based systems. (For more information, search Federal Privacy-Related Legislation Tracker on Practical Law.)

An opportunity and need for diverse voices, perspectives, and interests is vital when considering the future of AI. Organizations should prioritize AI ethics and continually evaluate whether their products or development process creates an additional discriminatory burden on historically excluded communities and vulnerable populations. There are unlimited possibilities for innovation and technological advancement that can occur while also embracing inclusive tech and an inclusive culture across an organization.

AI BIAS

AI bias may result when there is a lack of diversity in the data used to train an AI tool. For example, AI bias occurs in FRT when the training data does not include an equal mix of ethnicities, genders, ages, and skin tones.

When evaluating AI bias, it is important to consider:

- Real-world examples.
- State biometric privacy statutes.
- Bans on FRT.

EXAMPLES OF AI BIAS

Bias occurs in AI generally, such as in:

- **Using FRT to identify alleged criminals.** In January 2020, New Jersey's Attorney General ordered all state police and county prosecutors to stop using FRT by Clearview AI, a company that uses FRT to identify alleged criminals, and sent a cease and desist letter to Clearview directing the company to stop using the office and its investigations to promote its products (Blake Nelson, New Jersey Cops Told to Halt All Use of Controversial Facial-Recognition Technology, NJ.com (Jan. 24, 2020), available at *nj.com*).
- **Profiling to predict recidivism.** The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a risk assessment tool that predicts recidivism of defendants. The tool is used in sentencing and assesses placement in the appropriate prison facility. However, the algorithm used in the tool incorrectly labeled Black defendants as "high

risk" to commit a future crime twice as often as their white counterparts (Julia Angwin et al., Machine Bias, ProPublica (May 23, 2016), available at *propublica.org*; see also, Anne L. Washington, How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate, 17 Colo. Tech. L.J. 131 (2018)).

- **Determining health care risk.** Researchers found that a widely used health care risk prediction algorithm, which determines which patients would likely need extra medical care and heavily favored white patients over Black patients, demonstrated racial bias because the metric used to define need was faulty (Ziad Obermeyer et al., Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations, Science 366 (6464), 447-453 (2019), available at *science.org*).
- **Using HR recruiting tools.** The Amazon hiring tool used AI to give job candidates scores ranging from one to five stars. However, the tool's ratings for candidates for software developer jobs and other technical posts were not gender neutral. (Jeffrey Dastin, Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women, Reuters (Oct. 10, 2018), available at *reuters.com*.)

BIOMETRIC PRIVACY

Some states have passed laws specifically governing the collection, use, disclosure, and destruction of biometric information. Other states have enacted more limited restrictions on using certain biometric information or considered similar legislation. (For more information, search Biometrics in the Workplace on Practical Law.)

For example, the Illinois Biometric Information Privacy Act (BIPA) (740 ILCS 14/1 to 14/99) requires companies that collect or possess biometric identifiers or biometric information to obtain written consent and disclose how they collect, retain, disclose, and destroy this information. The biometric identifiers or biometric information include:

- Retina or iris scans.
- Fingerprints.
- Voiceprints.
- Scans of hand or face geometry.
- Other biometric information from the public.

(740 ILCS 14/10.)

BIPA provides individuals a private right of action to sue for liability up to $1,000 per negligent violation and $5,000 per reckless violation, as well as attorneys' fees, costs, and injunctive relief (740 ILCS 14/20). (For more information, search US Privacy Litigation: Overview and Trends in Privacy and Data Security: 2020 on Practical Law.)

FRT BANS

FRT can have adverse effects on human and civil rights. Many communities experience additional discrimination as a result of innovation such as FRT and AI in general. These communities experience a reinforcement of racial biases due to lack of transparency and lack of regulation of AI, which has led to false arrests and harassment, among other issues.

Many state and local jurisdictions ban the use of FRT by agencies, governments, and law enforcement, such as:

- **California.** The Body Camera Accountability Act (AB 1215), effective October 8, 2019 and expiring January 1, 2023, temporarily stops California law enforcement from installing, activating, or using any biometric surveillance system in connection with an officer camera or data collected by an officer camera (2019 Cal. Legis. Serv. ch. 579 (2019)).
- **Washington.** SB 6280, enacted in March 2020 and effective July 1, 2021, places restrictions on the government's use of facial recognition software to establish safeguards to prohibit uses of facial recognition software that threaten democratic freedoms.
- **Baltimore.** On August 9, 2021, the Baltimore Mayor's office approved Council Bill 21-0001, an ordinance prohibiting most uses of certain facial surveillance technology in Baltimore through the end of 2022 and requiring the city to prepare and publicly post an annual surveillance report. Each day of noncompliance constitutes a separate offense against the ordinance, and each offense is a misdemeanor punishable by both a fine up to $1,000 and imprisonment up to 12 months. The ordinance does not apply to federal or state government entities or their contractors. (For more information, search Baltimore Enacts Facial Recognition Moratorium on Practical Law.)

Many communities experience additional discrimination as a result of innovation such as FRT and AI in general.

- **Minneapolis.** On February 10, 2021, the Minneapolis City Council unanimously passed an ordinance that bans the city from buying or using data derived from FRT (Libor Jany, Minneapolis Passes Restrictive Ban on Facial Recognition Use by Police, Others, Star Tribune (Feb. 12, 2021), available at *startribune.com*).
- **Portland.** On September 9, 2020, the City of Portland, Oregon issued an announcement that Mayor Ted Wheeler and the Portland City Council unanimously passed a leading-edge ordinance to prevent private entities from using defined FRTs in places of public accommodation within the city's boundaries. The ban reflects the city's framework to prevent discrimination in these spaces. (For more information, search Portland, Or. Bans Private Entity Use of Face Recognition Technologies in Public Spaces on Practical Law.)

Regulation is also being considered at the federal level.

Search Trends in AI Regulation for more on state and local regulation of facial recognition software.

INCLUSIVE TECH

Until a technology priority is placed around the needs of underrepresented communities, and an inclusive strategy is embraced and acted on, underrepresented communities will continue to find themselves on the outskirts of lifesaving innovations and protections. A lack of interest, attention, and action will keep systemic oppression alive.

Inclusive tech explores innovative solutions to technology diversity and inclusion. Now, more than ever, it is important for executives, hiring managers, human resources personnel, data scientists, educators, entrepreneurs, investors, policymakers, and diversity and inclusion advocates to come together to drive solutions and ensure technology is inclusive.

Inclusive tech:

- Minimizes algorithmic bias.
- Considers societal issues.

Organizations must take action to ensure inclusive tech.

MINIMIZING BIAS IN ALGORITHMS

In technology, bias and inequalities are quite literally products of design. The design choices of a product reflect the people who make it, and for whom they chose to design it. To fix this, organizations need to redesign the design process itself by incorporating inclusive tech principles, such as:

- Designing with excluded and diverse communities, not for them.
- Fostering belonging through representation.
- Strengthening culture, training, and processes.
- Promoting accountability.

Artificial Intelligence Toolkit

The Artificial Intelligence Toolkit available on Practical Law offers a collection of resources to assist companies and their counsel in identifying potential legal issues concerning AI. The Toolkit features a range of continuously maintained resources, including:

- Artificial Intelligence Key Legal Issues: Overview
- Trends in AI Regulation: 2020
- Expert Q&A: Developing and Implementing AI Ethics Principles
- Artificial Intelligence and Legal Ethics
- US Privacy and Data Security Law: Overview
- Use of Bots in Consumer Transactions Checklist
- Artificial Intelligence (AI) in the Workplace
- Using Artificial Intelligence in Law Departments

- Normalizing inclusion at a systemic level.
- Ensuring data represents all equally.
- Enforcing data governance to ensure ethical practices are being met.
- Checking for bias after training a system and, if bias is found, determining the cause and mitigating.
- Monitoring deployed systems over time.

TECHNOLOGY AND SOCIETY

There is a quantifiable tension between the promise of AI and what society experiences from AI use. Vast investments of time and money are allocated to AI projects that promise to be exciting technological initiatives. However, society often does not experience the desired outcome. Once society experiences challenges, the alarms sound to advocate for bans, transparency, oversight, and accountability to lessen the impact on what many members of historically excluded communities of color and vulnerable populations experience. Organizations should therefore take preventative action to ensure that racism and other forms of discrimination do not take root in our AI systems.

ENSURING INCLUSIVE TECH

Organizations are not bound to maximize profits only. Societal impact is embraced by large organizations around the globe. Some steps organizations can take to ensure inclusive tech and make a societal impact regarding AI include:

- Implementing responsible AI leadership.
- Complying with FTC guidance.
- Monitoring digital surveillance to limit its effect.
- Engaging personnel at an organization to implement and monitor compliance with AI ethics principles.

Implement Responsible AI Leadership

To implement responsible AI leadership, organizations can:

- Use a playbook for guidance to get started or evolve, such as UC Berkeley's Equity Fluent Leadership Playbook, available at *haas.berkeley.edu*,

which explains that bias exists in AI systems and organizations should address bias and execute strategies to mitigate bias, such as enabling diverse and multi-disciplinary teams to work on algorithms and AI systems, and establishing responsible AI governance and internal policies to mitigate bias.

- Adopt a shared leadership function, such as Leadership of Responsible AI™ (available at *eadams.tech*), between technical and non-technical leaders who adopt processes and procedures to support responsible AI.
- Detail how datasets can harbor bias and outline strategies to mitigate bias to drive explainability and accountability.
- Offer a course on the ethics of AI to management.
- Create an AI ethics council.
- Train leaders to operationalize AI and data governance and measure engagement.
- Engage diverse communities and organizations.
- Support inclusive collaboration.

Comply with FTC Guidance

The FTC recently warned that apparently neutral algorithms can create legal exposure for companies if those algorithms are biased (Elisa Jillson, Aiming for Truth, Fairness, and Equity in Your Company's Use of AI, FTC: Business Blog (Apr. 19, 2021), available at *ftc.gov*). The FTC noted, for example, that the sale or use of racially biased algorithms is prohibited by Section 5 of the FTC Act.

To use AI truthfully, fairly, and equitably, the FTC recommends that organizations:

- Start with the right foundation by evaluating datasets.
- Watch out for discriminatory outcomes by testing their algorithm before using it and periodically thereafter to ensure it does not discriminate on the basis of race, gender, or other protected classes.
- Embrace transparency and independence, for example, by:
  - using transparency frameworks and independent standards;
  - conducting and publishing the results of independent audits; and
  - opening data or source code to outside inspection.
- Refrain from exaggerating the algorithm's capabilities or whether it can deliver fair or unbiased results.
- Tell the truth about how an AI tool uses data (for more information, search FTC Announces Settlement with Photo Storage App Company Over Improper Facial Recognition Use Allegations on Practical Law).
- Reduce the FTC's ability to challenge the use of an AI tool as unfair by doing more good than harm.
- Hold the organization accountable or be ready for the FTC to do so.

Monitor Digital Surveillance

Digital surveillance is changing the fabric of society because its use and disparate outcomes often cause loss of trust in the ability of individuals to live their lives safely and autonomously. Often, communities and governments are pitted against technology companies that are in favor of digital surveillance, or communities are pitted against governments and technology companies that are in favor of digital surveillance, hindering innovation and the overall societal benefits of embracing technology.

Ensuring oversight of digital surveillance technologies is one way of reforming abusive and discriminatory practices, which provides a solid foundation for larger social change and transparency. It is important to embrace technology proactively so that when AI tools are thought of, procured, and even implemented, communities have shared decision-making in the use of these AI tools to ensure and monitor that they are being used responsibility and not used to target certain populations.

Engage the Organization

It is important for an organization to:

- Promote engagement in the technology life cycle through questions, short checklists, or learning events. This creates opportunities to help executives prepare for the future of work and society.
- Encourage innovation and advancements for all by interacting with governments and society in ways that encourage change.
- Guide its executives to direct projects, teams, and efforts to be more inclusive and human-centered.
- Create opportunities for employees to become involved in corporate social responsibility. As many employees now desire hybrid work experiences, their relationships within their own community are strengthened and they are becoming stronger corporate social responsibility advocates. PL