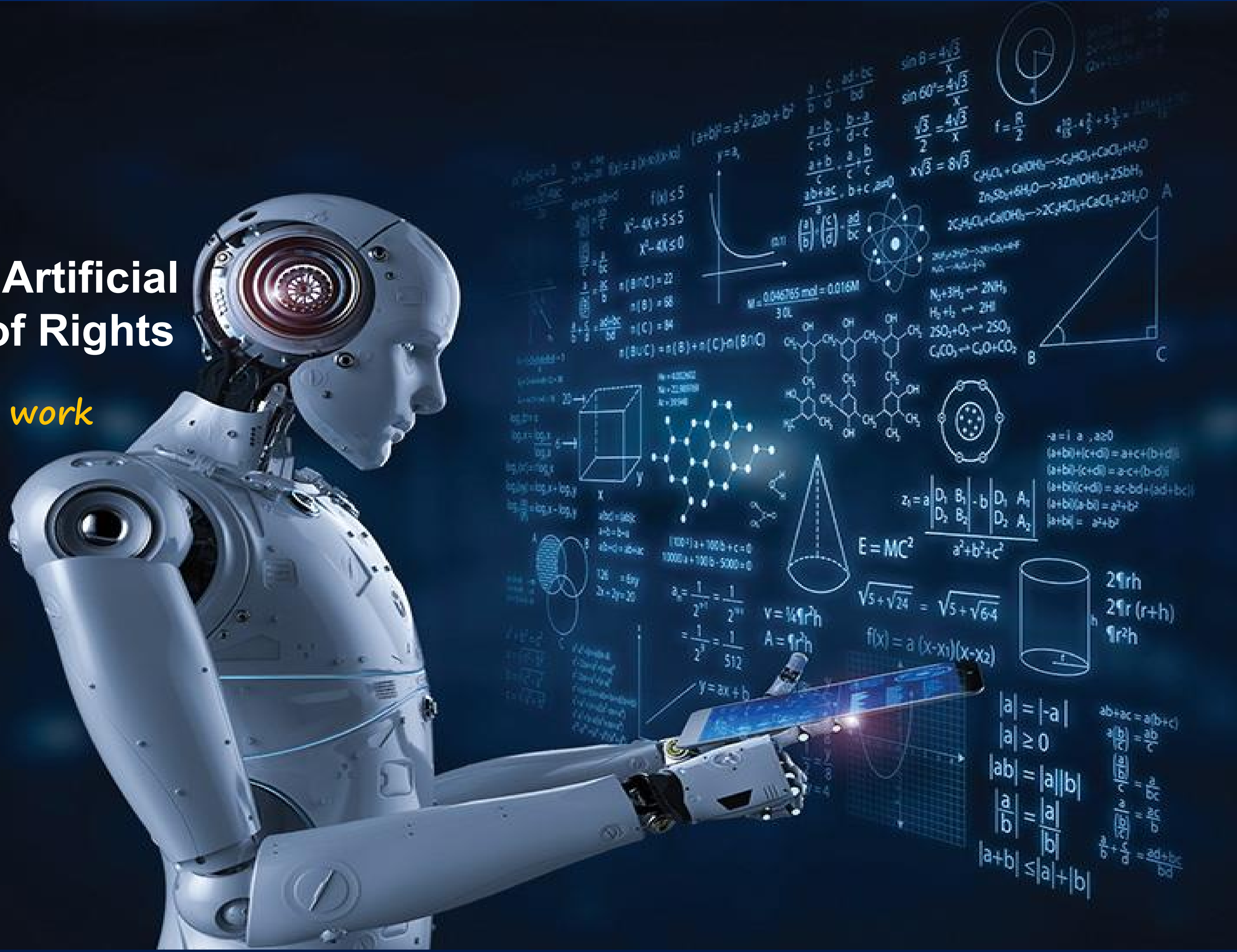
Management Solutions
Making things happen

WH- Blueprint for an Artificial Intelligence (AI) Bill of Rights

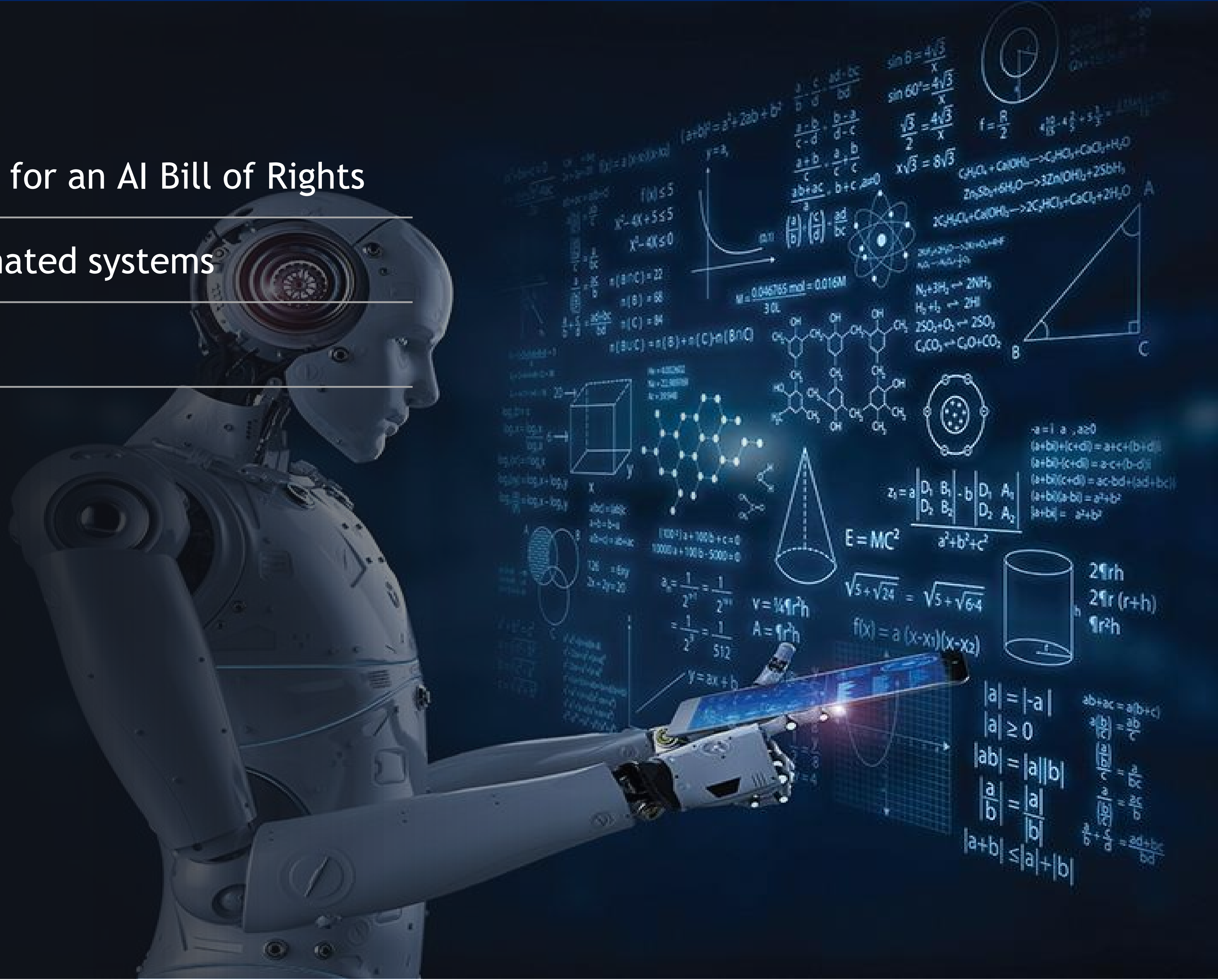
Making automates system work



Overview of the Blueprint for an AI Bill of Rights

Expectations about automated systems

Annex



1 General overview of the Blueprint for an AI Bill of Rights

The Blueprint for an AI Bill of Rights is a set of five principles and associated practices (expectations about automated systems) to help guide the design, use, and deployment of automated systems to protect the rights of the American public in the age of AI

Context

The White House Office of Science and Technology Policy published the Blueprint for an AI Bill of Rights in October 2022 which is an exercise in envisioning a future where the American public is protected from the potential harms, and can fully enjoy the benefits, of automated systems. It describes principles that can help ensure these protections. Some of these protections are already required by the US Constitution or implemented under existing US laws.

Principles

1	Safe and effective systems	Automated systems should be developed with **consultation** from **diverse communities**, **stakeholders**, and **domain experts** to identify concerns, risks, and potential impacts of the system.
2	Algorithmic discrimination protections	Designers, developers, and deployers of automated systems should take **proactive and continuous measures** to **protect individuals** and **communities from algorithmic discrimination** and to use and design systems in an equitable way.
3	Data privacy	Designers, developers, and deployers of automated systems should seek for permission and **respect people's decisions regarding** collection, use, access, transfer, and deletion of **their data** in appropriate ways.
4	Notice and explanation	Designers, developers, and deployers should provide a clear description of: i) the overall system functioning and the role automation plays; ii) notice that such systems are in use; iii) the individual or organization responsible for the system.
5	Human alternatives, consideration and fallback	Opting for automated systems in favor of a human alternative, **where appropriate**. Appropriateness should be determined based on reasonable expectations in a **given context** and with a focus on ensuring **broad accessibilit**y and protecting the public from especially harmful impacts.

*For further information regarding real life examples of these principles see Annex 1 and Annex 2

Access the entire Document

2 Expectations about automated systems

Safe and effective systems

1

Automated systems should be developed with consultation from diverse communities, stakeholders, and domain experts to identify concerns, risks, and potential impacts of the system

1 *Protect the public from harm in a proactive and ongoing manner*

Consultation	Public should be consulted in the **design**, **implementation**, **deployment**, acquisition, and maintenance phases of **automated system** development.
Testing	Undergo **extensive testing before deployment**. This testing should follow **domain-specific best practices.**
Rik identification and mitigation	Before deployment, and in a proactive and ongoing manner, **potential risks should be identified and mitigated**.
Ongoing monitoring	Ongoing monitoring procedures to **ensure that performance does not fall below an acceptable level over time**, based on changing real-world conditions or deployment contexts, post-deployment modification, or unexpected conditions.
Clear organizational oversight	Include **clearly-stated governance procedures** before deploying the system, as well as **responsibility** of specific individuals or entities to oversee ongoing assessment and mitigation.

2 *Avoid inappropriate, low-quality, or irrelevant data use and the compound harm of its reuse*

Relevant and high-quality data	**Data used as part of any automated system's** creation, evaluation, or deployment should be relevant, of high quality, and tailored to the task at hand.
Carefully track and review derived data sources	**Data that is derived from other data** though the use of algorithms, such as data derived or inferred from prior model outputs, should be identified and tracked.
Data reuse limits in sensitive domains	**Data reuse**, and especially data reuse in a new context, can result in the **spreading and scaling of harms**. Accordingly, such data should be subject to extra oversight to ensure safety and efficacy.

3 *Demonstrate the safety and effectiveness of the system*

Independent evaluation	Independent evaluators, should be given **access to the system and samples** of associated data, in a manner consistent with privacy, security, law, or regulation in order to perform such evaluations.
Reporting	Provide **regularly-updated reports**, including: i) an overview f the system; ii) system goals; iii) any human-run procedures.

2 Expectations about automated systems

Algorithmic discrimination protections

Algorithms should not be discriminatory, and systems should be used and designed in an equitable way

1 *Protect the public from algorithmic discrimination in a proactive and ongoing manner*

Proactive assessment of equity in design

Review potential input data, associated historical context, accessibility for people with disabilities, and societal goals to identify potential discrimination and effects on equity resulting from the introduction of the technology.

Representative and robust data

Any data used should be **representative of local communities**, reviewed for bias based on the historical and societal context of the data, and sufficiently robust to identify and help to mitigate biases and potential harms.

Guarding against proxies

Identify proxies by testing for correlation between demographic information and attributes in any data used.

Ensuring accessibility during design, development & deployment

Consideration of a **variety of disabilities**, adherence to relevant accessibility standards, and user experience research to identify and address any accessibility barriers to the use or effectiveness of the automated system.

Disparity assessment

Test systems by using **demographic performance measures**, overall and subgroup parity assessment, and calibration measures to assess whether the system components produce disparities.

Disparity mitigation

Evaluate multiple models and select the one that has the **least adverse impact**, modify data input choices, or identify a system with fewer disparities. If this is not possible, then the use of the automated system should be reconsidered.

Ongoing monitoring and mitigation

Regularly monitor automated systems to assess algorithmic discrimination that might arise from unforeseen interactions of the system with inequities not accounted.

2 *Demonstrate that the system protects against algorithmic discrimination*

Independent evaluation

Allow independent evaluation of potential algorithmic discrimination caused by automated systems they use or oversee.

Reporting

Provide reporting of an appropriately designed algorithmic impact assessment, with clear **specification** of who performs the assessment, who evaluates the system, and how **corrective actions** are taken in response to the assessment.

ManagementSolutions
Making things happen

2 Expectations about automated systems

Data privacy (1/2)

Users should be protected from abusive data practices via built-in protections and have agency over how data about the user is used

1 *Protect the privacy by design and by default*

Privacy by design and by default	Automated systems should be **designed** and built with privacy protected by default.
Data collection and use-case scope limits	**Data collection** should be **limited in scope**, with specific, narrow identified goals.
Risk identification and mitigation.	**Attempt** to proactively **identify harms** and seek to manage them when collecting, using or storing sensitive data.
Privacy-preserving security	Entities creating, using, or governing automated systems should **follow privacy** and security best practices designed to ensure data and metadata do not leak beyond the specific consented use case.

2 *Protect the public from unchecked surveillance*

Heightened oversight of surveillance	Surveillance or monitoring systems should be subject to **heightened oversight** that includes at a minimum assessment of potential harms during design.
Limited and proportionate surveillance	**Surveillance should be avoided** unless it's necessary to achieve a legitimate purpose and it's proportionated to the need.
Scope limits on surveillance to protect rights and democratic values	**Civil liberties** and **civil rights** must **not** be **limited** by the **threat of surveillance** or harassment facilitated or aided by an automated system.

2 Expectations about automated systems

Data privacy (2/2)

Users should be protected from abusive data practices via built-in protections and have agency over how data about the user is used

3 *Provide the public with mechanisms for appropriate and meaningful consent, access, and control over their data*

Use-specific consent.

Consent practices should **not allow for abusive surveillance** practices.

Brief and direct consent requests.

Short, plain language consent requests should be used so that users understand for what **use contexts, time span**, and entities they are providing data and metadata consent.

Data access and correction.

People whose data is collected, used, shared, or stored by automated systems should be able to **access data and metadata** about themselves.

Consent withdrawal and data deletion.

Entities should **allow withdrawal** of data access consent.

Automated system support.

Entities designing, developing, and deploying automated systems should **establish and maintain** the **capabilities** that will allow individuals to use their own automated systems.

4 *Demonstrate that data privacy and user control are protected*

Independent evaluation.

Entities should allow **independent evaluation** of the claims made regarding data policies.

Reporting

When members of the public wish to know what data about them is being used in a system, the entity responsible for the development of the system should **respond quickly** with a report on the data it has collected or stored about them.

2 Expectations about automated systems

Notice and explanation

Users should be notices of the use and understand how and why the automated system contributes to outcomes that impact them

1 *Provide clear, timely, understandable, and accessible notice of use and explanations*

Generally accessible plain language documentation

The entity responsible for using the automated system should ensure that **documentation** describing the overall system.

Accountable

Notices should clearly identify the **entity responsible** for designing each component of the system and the entity using it.

Timely and up-to-date

Users should **receive notice of the use of automated systems** in **advance** of using or while being impacted by the technology.

Brief and clear

Notices and **explanations** should **be assessed**, such as by research on users' experiences, to ensure that the people using or impacted are able to easily find notices and explanations, read them quickly, and understand and act on them.

2 *Provide explanations as to how and why a decision was made or an action was taken by an automated system*

Tailored to the purpose

Explanations should be **tailored to the specific purpose** for which the user is expected to use the explanation, and should clearly state that purpose.

Tailored to the target of the explanation

Explanations should be targeted to **specific audiences** and clearly state that audience. An explanation provided to the subject of a decision might differ from one provided to an advocate, or to a domain expert or decision maker.

Tailored to the level of risk

An **assessment** should be done to determine the level of risk of the automated system.

Valid

The explanation provided by a system should **accurately reflect the factors** and the influences that led to a **particular decision**, and should be meaningful for the particular customization based on purpose, target, and level of risk.

3 *Demonstrate protections for notice and explanation*

Reporting

Document the determinations made based on the above considerations.

2 Expectations about automated systems

Human alternatives, consideration and fallback (1/2)

5

Users should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems they encounter

1 *Provide a mechanism to opt out from automates systems in favor of human alternative*

Brief, clear, accessible notice and instructions.	Those impacted by an automated system **should be given a brief**, clear notice that they are entitled to opt-out, along with clear instructions for how to opt-out.
Human alternatives provided when appropriate	When automated systems make up part of the attainment process, **alternative timely human-driven** processes should **be provided**.
Timely and not burdensome human alternative	**Opting out** should be **timely** and not unreasonably burdensome.

2 *Provide timely human consideration and remedy by a fallback and escalation system if an automated system fails*

Proportionate	The availability of **human consideration** and fallback should be **proportionate** to the potential of the automated system.
Accessible	Mechanisms for human consideration and fallback should be **easy to find**.
Convenient	Mechanisms for human consideration and fallback **should not be unreasonably burdensome** as compared to the automated system's equivalent.
Equitable	Consideration should be given to **ensuring outcomes of the fallback** and escalation system are equitable.
Timely	Human consideration and fallback are only useful if they are conducted and concluded in a **timely manner**.
Effective	Organizational structure surrounding processes for consideration and fallback should be designed so that if the human **decision-maker** determines that it should be overruled, the new decision **will be effectively enacted**.
Mantained	Human consideration and fallback process and any associated automated processes should be **maintained and supported** as long as the relevant automated system continues to be in use.

2 Expectations about automated systems

Human alternatives, consideration and fallback (2/2)

Users should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems they encounter

3 Institute training, assessment, and oversight to combat automation bias and ensure any human-based components of a system are effective

Training and assessment

Anyone administering, interacting with, or interpreting the outputs of an automated system should receive **training** in that system.

Oversight

Human-based systems have the potential for bias. The results of assessments of the efficacy and potential bias should be **overseen** by governance structures to update the operation of the human-based system in order to mitigate these.

4 Implement additional human oversight and safeguards for automated systems related to sensitive domains

Narrowly scoped data and inferences

Human oversight should ensure that automated systems in sensitive domains are **narrowly scoped** to **address a defined goal**.

Tailored to the situation

Human oversight should ensure that automated systems in sensitive domains are **tailored to the specific use case** and real-world deployment scenario.

Human consideration before any high-risk decision

Automated systems, where they are used in sensitive domains, may **play a role** in directly **providing information**.

Meaningful access to examine the system

Designers, developers, and deployers of automated systems should **consider limited waivers of confidentiality** here necessary in order to provide meaningful oversight of systems used in sensitive domains.

5 Demonstrate access to human alternatives, consideration, and fallback

Reporting

Reporting on the accessibility, **timeliness**, and **effectiveness** of human consideration and fallback should be **made public** at regular intervals for as long as the system is in use.

A Annex

Real life examples (1/2)

1 Safe and effective systems

Executive Order 13960 on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government requires that certain federal agencies adhere to nine principles when designing, developing, acquiring, or using AI for purposes other than national security or defense.

The law and policy landscape for motor vehicles shows that strong safety regulations—and measures to address harms when they occur—can enhance innovation in the context of complex technologies.

From large companies to start-ups, industry is providing innovative solutions that allow organizations to mitigate risks to the safety and efficacy of AI systems, both before deployment and through monitoring over time.

The Office of Management and Budget (OMB) has called for an expansion of opportunities for meaningful stakeholder engagement in the design of programs and services.

The National Institute of Standards and Technology (NIST) is developing a risk management framework to better manage risks posed to individuals, organizations, and society by AI.

Some U.S government agencies have developed specific frameworks for ethical use of AI systems.

The National Science Foundation (NSF) funds extensive research to help foster the development of automated systems that adhere to and advance their safety, security and effectiveness.

Some state legislatures have placed strong transparency and validity requirements on the use of pretrial risk assessments

2 Algorithmic discrimination protections

The federal government is working to combat discrimination in mortgage lending

The Equal Employment Opportunity Commission and the Department of Justice have clearly laid out how employers’ use of AI and other automated systems can result in discrimination against job applicants and employees with disabilities

Disparity assessments identified harms to Black patients' healthcare access

Large employers have developed best practices to scrutinize the data and models used for hiring

Standards organizations have developed guidelines to incorporate accessibility criteria into technology design processes

NIST has released Special Publication 1270, Towards a Standard for Identifying and Managing Bias in Artificial Intelligence

A Annex

Real life examples (2/2)

3 Data privacy

The Privacy Act of 1974 requires privacy protections for personal information in federal records systems, including limits on data retention, and also provides individuals a general right to access and correct their data

NIST's Privacy Framework provides a comprehensive, detailed and actionable approach for organizations to manage privacy risks

A school board's attempt to surveil public school students—undertaken without adequate community input—sparked a state-wide biometrics moratorium

Federal law requires employers, and any consultants they may retain, to report the costs of surveilling employees in the context of a labor dispute, providing a transparency mechanism to help protect worker organizing

Privacy choices on smartphones show that when technologies are well designed, privacy and data agency can be meaningful and not overwhelming

4 Notice and explanation

People in Illinois are given written notice by the private sector if their biometric information is used

Major technology companies are piloting new ways to communicate with the public about their automated technologies

Lenders are required by federal law to notify consumers about certain decisions made about them

A California law requires that warehouse employees are provided with notice and explanation about quotas, potentially facilitated by automated systems, that apply to them

Across the federal government, agencies are conducting and supporting research on explainable AI systems

5 Human alternatives, consideration and fallback

The Privacy Act of 1974 requires privacy protections for personal information in federal records systems, including limits on data retention, and also provides individuals a general right to access and correct their data

NIST's Privacy Framework provides a comprehensive, detailed and actionable approach for organizations to manage privacy risks

A school board's attempt to surveil public school students—undertaken without adequate community input—sparked a state-wide biometrics moratorium

Federal law requires employers, and any consultants they may retain, to report the costs of surveilling employees in the context of a labor dispute, providing a transparency mechanism to help protect worker organizing

Privacy choices on smartphones show that when technologies are well designed, privacy and data agency can be meaningful and not overwhelming

MS

Management Solutions

Making things happen

International *One Firm*

Multiscope Team

Best practice *know-how*

Proven Experience

Maximum Commitment