EQUAL

Algorithmic Impact Assessment (AIA) Tool

(adapted from NIST Playbook for AI Risk Management)

Artificial Intelligence (AI) is an increasingly important and pervasive part of our daily lives and critical functions. As our reliance on this technology grows, the need increases exponentially to ensure it is trustworthy, meaning that the AI system is safe, inclusive, and tested thoroughly for unconscious biases and other potential harms.

While still a relatively nascent field, there is increasing alignment on best practices for AI development and deployment, which are important elements of responsible AI governance. A key resource for understanding responsible AI governance is the congressionally mandated AI Risk Management Framework (AI RMF), released on January 26, 2023 by the U.S. Department of Commerce's National Institute of Standards and Technology (NIST). The AI RMF provides actionable guidance on how to govern and manage systems' risks and limitations. NIST divides its guidance into four core stages in the AI lifecycle: (1) Govern; (2) Map; (3) Measure; and (4) Manage. NIST also released a companion guide, the draft AI RMF Playbook (Playbook) to help navigate and use the AI Risk Management Framework (AI RMF).

The **EqualAI Algorithmic Impact Assessment ("EqualAI AIA") Tool** was created to offer organizations a user-friendly tool to operationalize best practices offered in the NIST guidance materials. Relatedly, Algorithmic Impact Assessments (AIA) have gained increasing acceptance[1] as a useful approach for organizations to identify potential risks and proactively avoid harms and liability stemming from their AI systems. The EqualAI AIA was developed primarily to align with the guidance provided in the NIST AI RMF's Playbook. A small sample of relevant considerations that stem from NIST Special Publication 1270 ("Towards a Standard for Identifying and Managing Bias in Artificial Intelligence") and other relevant legal considerations have also been added to offer an initial inquiry into investigation for bias and discrimination[2]. The NIST AI RMF was selected as the primary template for the EqualAI AIA tool based on its

[1] Some examples of its use and application include: the Government of Canada developed an AIA risk assessment tool, a questionnaire that determines the impact level of an automated decision-system. The Chief Information Officers (CIO) Council has created an Algorithmic Impact Assessment tool to help federal government agencies begin to assess risks associated with using automated decision systems. Private companies are also developing their own version of Impact Assessment, including Microsoft's publicly available Impact Assessment template.

[2] **This tool does not provide legal advice nor offer a full scope of relevant bias, discrimination or legal considerations but rather, it offers an initial inquiry on laws that should be addressed with counsel before, while and following the use of this tool.** Such considerations include: data privacy regulations such as General Data Protection Regulation (GDPR) and California Privacy Rights Act; facial recognition and biometric information laws; civil rights laws such as the Americans with Disabilities Act (ADA), the Health Insurance Portability and Accountability Act (HIPAA), Title VII of the Civil Rights Act; unfair and deceptive practices enforced by Section 5 of the FTC Act; and consumer rights protections statues such as Fair Credit Reporting Act (FRCA) or Equal Credit Opportunity Act (ECOA)

law-agnostic, voluntary approach that includes a comprehensive focus on socio-technical considerations[3] and best practices for operationalization of AI principles. AI systems are "socio-technical" in nature, meaning they are influenced by societal dynamics and human behavior.

Timing. AI risk management is most effective when conducted throughout all stages of the AI lifecycle: it should be designed to identify potential risks from the early planning stages of the AI systems design or specification, through development and deployment. Of equal importance, the system must be retested routinely post-deployment in a regular cadence, which should be calculated depending on how fast new patterns will evolve and the sensitivity of the functions for which the system is used. Organizations are encouraged to start their impact assessment at the early stages, and complete the document iteratively as the AI project progresses or enters a new stage in the AI lifecycle. It is critically important to frequently assess whether there are new use cases for which the AI system will be used or how new users and subjects could be impacted, directly or indirectly, that were not envisioned in earlier stages of development and testing. Based on these routine assessments, the impact assessment documentation should be updated accordingly.

Multi-Stakeholder approach. AI governance benefits from including the broadest cross-section of stakeholders who can offer insight into use cases and individuals who will be impacted, directly or downstream. A diversity of perspectives can enhance an organization's ability to imagine use cases and potentially impacted individuals outside of their usual experience and expectations.

Legal Counsel. Legal counsel will also offer an important vantage point on potential risks and legal liabilities salient to your AI system. Legal counsel should help structure and monitor this process in general, and this tool highlights a few specific areas to be discussed and addressed with counsel, denoted with three asterisks (***).

Leadership matters. This process should start with organizations' senior leadership to ensure their buy-in and support for the AI governance strategy. The organization's AI use can either support or impede its core values[4] and planning should include related departments across the enterprise (e.g., general counsel, human resources director, chief innovation officer, chief data officer, head of product and sales) and should be supplemented with outside stakeholders to understand the ways AI is and *will be* used. For instance, there are AI systems readily used in human resources (HR), in evaluating candidate resumes or assessing employee productivity. AI tools are also commonly used in marketing efforts to identify target audience and test messaging. The AI RMF encourages organizations to "track and document existing AI systems held by the organization, and those maintained or supported by third-party entities." You will also

[3] In building the AI RMF, NIST took into account that "AI risks can emerge from the complex interplay of these technical and societal factors, affecting people's lives in situations ranging from their experiences with online chatbots to the results of job and loan applications."
[4] Also acknowledged in the Playbook (MAP 1.3: The organization's mission and relevant goals for the AI technology are understood and documented.)

want to involve your legal team early in the process to ensure they help shape the program and that documentation (including wording and retention protocols) follows their guidance.

Building trust. The results of implementing responsible AI governance are innumerable and range from economic gains and reduction in potential legal liability to significant impact in organizational culture and employee retention. If done thoughtfully and intentionally, including the support of senior management, the broader organization and affected stakeholders in a regular cadence to support and ensure the safety of your AI systems, or in other words: by implementing Responsible AI Governance, your organization will have taken an important step forward in building and sustaining trust among your employees, board, consumers and the broader public. It is not just the right thing to do to reduce potential harms and unconscious bias, it provides your organization with a competitive advantage and a path to build and sustain trust.

Algorithmic Impact Assessment Template Overview

As noted above, this risk assessment tool follows the NIST AI RMF Playbook with a particular focus on the Map function[5] with two additional sections: bias in AI (following NIST Special Publication 1270) and general considerations for legal compliance based on applicable laws and regulations[6].

This tool can be used either in sequence to provide an assessment of potential risks and harms in the AI system starting from the project inception throughout its lifecycle. Alternatively, users can jump to a particular section based on a specific concern or in response to a new development. **The assessment should be repeated in full as new key elements are added or changed in the AI system and repeated on a regular basis.**

The steps are as follows:

[5] The Map MAP is intended to enhance an organization's ability to identify risks and broader contributing factors and is divided into the following sections:
Map 1-context is established and understood
Map 2- Categorization of the AI system is performed
Map 3 AI capabilities, targeted usage, goals, and expected benefits and costs compared with appropriate benchmark is understood
Map 4 Risks and benefits are mapped for all components of the AI system including thor-party software and data
Map 5 Impacts to individuals, groups, communities, organizations, and society are characterized

[6] See FN 2.

Tool Assessment Sequence

System Description

Provide an initial description of the system, project overview, and points of contact who will be responsible for the audit or AI system.
Note: *You may want to include unique ID, tags or other project identifiers to facilitate identification and record keeping downstream.*

1. Working name of AI system and owner/organization:

2. Briefly describe the AI system under consideration:

3. List team members, their roles and responsibilities:

4. Projected (or past) date(s) of AI system release:

5. Who is the point of contact for the AI project? Who is the point of contact for the Impact Assessment?

Section 1 - Contextualizing the AI system

Section 1.1. Intended purpose(s) and limitations

It is important to provide context about the specific, envisioned use cases for which the AI system is designed to be deployed. In the absence of advanced knowledge about all potential settings in which a system will be deployed, examination of the bounds of acceptable deployment is instructive to identify unanticipated or untested downstream uses.

1. Describe the AI system's task, purpose, minimum functionality, and the benefits it offers:

2. Can any applicable non-AI systems solve the same problem without AI? Can any non-technical procedural changes or adaptations help solve the same problem without the need for deploying an AI system? Explain if and why the AI system is preferred:

3. List the prospective users of the AI system and/or data subjects[7] and their expectations for the AI system's use and limits:

4. Delineate the expected and acceptable AI system context of use, including: operational environment, user/operator characteristics, and social environment:

5. List the AI system's known limitations:

6. Describe the roles and responsibilities for human oversight of the deployed system:

Section 1.2. Interdisciplinary Collaborations

A team of AI actors with a diversity of experience, expertise, abilities and backgrounds, and with the resources and independence to engage in critical inquiry is critical to build a robust and comprehensive mapping system.

1. What interdisciplinary expert teams are you engaging to identify and manage risks in all stages of the AI life cycle? Have you included individuals with expertise in:
   - ☐ Systems design
   - ☐ Engineering
   - ☐ Ethics
   - ☐ Sociology

[7] The term 'data subject' refers to any living individual whose personal data is collected, held or processed by an organization. (See Article 4 of GDPR for more information). Data subjects are not necessarily the end-users of the AI system, but can still be impacted by the system.

- □ Psychology
- □ Privacy compliance
- □ Law and regulations
- □ Public policy
- □ Subject matter expertise for expected uses (domain experts)

2. Consider whether the teams responsible for developing and maintaining the AI system reflect diverse opinions, backgrounds, experiences, and perspectives. Identify gaps:

3. Review the demographics of those involved in the design and development of the AI system. This will help identify potential gaps in understanding and biases embedded during the development process.
    a. Which demographics are over-represented?
    b. Which demographics are under-represented?
    c. How were these metrics determined?

4. Which communities and potential end users were consulted in the development? At which stages in the design and development process were they engaged?

5. Have stakeholders expressed potential negative impacts of the AI system? If so, how have you incorporated these concerns to address negative impacts?

Section 1.3. The business value or context of business use

There are instances in which the AI solution is not the appropriate choice. The most significant example are situations where the AI solution would cause more harm than good. Another situation where the decision to terminate development could be appropriate is when AI systems do not present a business benefit beyond the status quo. Inherent risks and implicit or explicit costs should be weighed in the evaluation of whether an AI solution should be developed or deployed. Defining and documenting the specific business purpose of an AI system in a broader context of societal values helps teams evaluate risks and increases the clarity of “go/no-go” deployment decisions.

1. Have you reviewed the documented system’s purpose from a socio-technical perspective, i.e., considering characteristics such as explainability, interpretability, privacy, safety, and bias management, and in consideration of societal values?

□ yes
□ no

2. ***What potential latent business incentives may contribute to the AI system inflicting negative impacts?
   Note possible misalignments between societal values, stated organizational principles and code of ethics, as compared to business incentives could result in negative societal impacts.

Section 1.4. Organization's mission and goals

By establishing comprehensive and explicit enumeration of AI system purpose and expectations, organizations can identify and manage socio-technical risks and benefits that could be supported or jeopardized by the AI system.

1. What goals and objectives does the organization expect to achieve by designing, developing, and/or deploying the AI system?

2. Review organization's stated values, mission statements, social responsibility commitments, and AI principles. Are there misalignments between organization's goals and commitments and system's purpose/context of use?
   a. Have you established the organization's AI principles?

3. To what extent are the model outputs inconsistent with the values of fostering public trust? Broader equity?

***Section 1.5. Organizational risk tolerances

Risk tolerance reflects the level and type of risk the organization will accept while conducting its mission and functions. Deployment decisions should be the outcome of a clearly defined process that is reflective of an organization's values, including its risk tolerances. Go/no-go decisions should be incorporated throughout the AI system's lifecycle. For systems deemed "higher risk," such decisions should include approval from sufficiently senior technical or otherwise specialized or empowered executives (for more information on risk tolerance, see NIST AI RMF document *Section 3.2.2. Risk Tolerance*).

1. Provide justifications for the assumptions, boundaries, and limitations of the AI system including wh the system should (or should not) be deployed based on its limitations, boundaries, or assumptions.

2. What are the maximum allowable risk thresholds above which the system will not be deployed or will be prematurely decommissioned?

3. Establish risk tolerance levels for the AI system and allocate the appropriate oversight resources and authorities/ supervisors to each level:

Section 2 - Classifying the AI system

2.1. Learning Task

AI actors should define the technical learning or decision-making task that an AI system is designed to accomplish.

1. Which category of learning tasks does the AI system support?
    - □ Computer Vision
    - □ Natural Language Processing (NLP)
    - □ Recommender system
    - □ Classification system
    - □ Image or text generation
    - □ Other:

2. Have you defined technical specifications and requirements for the AI system? Provide a link to, or attach, the technical specifications document.

3. Have you documented the AI system's development, testing methodology, metrics, and performance outcomes? Provide a link to, or attach, the documentation.

4. How do the technical specifications and requirements align with the AI system's goals and objectives?

5. Did your organization implement accountability-based practices in data management and protection and, if so, which standard(s) (e.g. the PDPA and OECD Privacy Principles)?

6. What assessments has your organization conducted on data security and privacy impacts associated with the AI system? Consult NIST privacy and cybersecurity frameworks for a comprehensive assessment of these risk areas.

7. How are outputs marked to clearly show that they came from an AI system?

2.2. Operational Context

Once deployed and in use, there are times when AI systems can perform poorly, manifest unanticipated negative impacts, or violate legal or ethical norms. Human oversight and stakeholder engagement can provide important contextual awareness.

1. Risks can arise from deploying a system in an environment that differs from the original controlled or envisioned environment. Identify potential risks due to unanticipated or unintended deployment contexts or human-AI configurations:

2. What dependencies does the AI system have on upstream data and other AI systems? (See Section 2.3. and Section 5)

3. Does the AI system have connections to external networks (including the internet), financial markets, and critical infrastructure that have potential for negative externalities?
    - □ yes- If so, list the network(s)
    - □ no

***2.3. Data Collection and Selection

Many AI system risks and vulnerabilities can be traced to insufficient testing and evaluation processes as well as oversight in data collection and curation.

1. How was the data collected, cleaned, and curated? Who was involved in the data collection process?

2. What, if any, are known errors, sources of noise, or redundancies in the data?

3. Over what time-frame was the data collected?

4. [After extended use(s)] Is the training and testing data still representative of the current operational environment(s)?
    - □ yes
    - □ no

5. What is the variable selection[8] and evaluation process?

6. If the dataset relates to, or derives from individuals (e.g., their attributes) were they informed about the data collection?
    - □ yes

[8] Variable selection or feature selection refers to the process of identifying the important features from a set of features and removing the irrelevant or less important ones.

- □ no

7. Is the training and testing data representative of the demographic population with which the AI system will be used (e.g., age, gender, race, etc.)?

8. Do you know why the dataset was created? (e.g., what was the specific need or utility that it was created to fulfill?)

9. How do you ensure that the collected data are adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed ('data minimization')? See Article 5 of GDPR for more information.

10. Do the data collection processes adhere to organization policies related to bias, privacy and security for AI systems?
    - □ yes
    - □ no- If no, explain:

11. Do the data collection processes comply with relevant legal or regulatory requirements applicable to data or AI systems? (See Section 6 on Legal)
    - □ yes
    - □ no

Section 3 - Bias [9]

Trustworthy and Responsible AI has been evaluated to determine whether a given AI system is biased, fair and does what it claims to accomplish. Processes to ensure trustworthy and responsible AI often focus on computational factors such as representativeness of datasets and fairness of machine learning algorithms, which indeed are vital for mitigating bias. However, a robust process must include considerations of human, systemic, institutional and societal factors that can present significant sources of AI bias as well. An effective governance system requires

[9] Content of this section is developed based on NIST Special Publication 1270 "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence."

expanding the perspective beyond the machine learning pipeline to recognize and investigate how this technology is both created within and impacts our society.

Understanding AI as a socio-technical system acknowledges that the processes used to develop technology are more than their mathematical and computational constructs. A socio-technical approach to AI takes into account the values and behavior embedded in the datasets, the humans who interact with the AI systems, and numerous other factors that go into the design, development, and deployment of that system.

The importance of engaging in an evaluation process considering the transparency, datasets, and test, evaluation, validation, and verification (TEVV) cannot be overstated. Participatory design techniques and multi-stakeholder approaches, as well as human-in-the-loop processes are also important steps to help mitigate risks related to AI bias. However none of these practices provide a panacea against the introduction or scaling of bias in AI systems, and each can introduce potential pitfalls. It is critical to recognize the reality that it is not possible to achieve zero risk of bias in an AI system. Rather, we create and adopt AI governance to enable better identification, understanding, management, and reduction of bias, as well as other potential harms.

This section incorporates elements of SP1270 NIST taxonomy of bias in AI to help organizations initiate a probe for common types of biases that are likely to occur in the AI systems outcomes. Fairness evaluation and remediation is a fast-evolving, interdisciplinary field of research, which requires a variety of perspectives from different fields. Engaging a social scientist, an AI fairness specialist or similar an individual with expertise in this area could help facilitate responses to the questions in this section.

The NIST Special Publication on Bias categorizes AI biases into three categories:

- Statistical/computational bias
- Systemic bias
- Human bias

The Special Publication also identifies three broad areas that present challenges in addressing bias:

- Dataset factors
- Measurement and metrics related to TEVV (Test, Evaluation, Verification, and Validation)
- Human factors

1. Are the users of the AI system properly trained to interpret AI model output and decisions? Are the staff developing the AI system trained to detect and manage bias in data?

For the following biases **Statistical and Computational;Systemic; and Human Cognitive,** select the type of bias you tested the AI system for and describe each test you conducted. If you did not conduct a test, explain why.

2. **Statistical and Computational biases** stem from errors that result when a sample is not representative of the population. Consider whether your AI system should be tested for one or more of the following statistical/computational biases:

   - □ *Representation bias:* arises from non-random sampling of subgroups of population.
   - □ *Sampling bias:* Proper randomization is not used during data collection.
   - □ *Measurement bias*: arises when features and labels are proxies for desired quantities.
   - □ *Model selection bias:* introduced while using the data to select a single seemingly "best" model from a large set of potential models.
   - □ *Activity bias:* a type of selection bias that occurs when systems get their training data from their most active users.
   - □ *Temporal bias:* arises from differences in populations and behaviors over time.
   - □ *Automation bias:* tendency to favor results generated by automated systems over those generated by non-automated systems, regardless of accuracy.
   - □ *Other* (refer to enumerations in Appendix or see some examples here):

3. **Systemic biases** affect how organizations and teams are structured and who controls the decision making processes, which can result in certain social groups being advantaged or favored. Which types of systemic biases should your system be tested for?
   - □ historical
   - □ societal
   - □ institutional

4. **Human Cognitive biases** reflect systematic errors in human thought based on a limited number of heuristic principles and predicting values to simpler judgmental operations. These biases are often implicit and tend to relate to how an individual or group perceives information (such as automated AI output) to make a decision or fill in missing or unknown information.

   Consider which types of human bias should your system be tested for, including:
   - □ *Availability*: overly weighting a resource or response that comes easily or quickly to mind, e.g., which datasets are readily available.

- □ *Funding bias*: biased results are reported to satisfy the needs or interests of the funding agency, individual, or organization.
- □ *Groupthink*: individuals in a group can slant decisions toward a desire to conform to the group, or fear of dissenting with the group.
- □ *Other* (refer to enumerations in Appendix or see some examples here)

5. Could the AI system unfairly advantage or disadvantage a particular social group?
   a. Who?
   b. In what ways? (See, for example, Section 2.3 question 7)

6. Identify other sources of biases in your AI system that could lead to inequitable or discriminatory outcomes? (See NIST SP1270 Glossary, pg 58)

   Identify ways to reduce such biases:

7. List other potential impacts of bias from use of the AI system, including stereotyping or offensive outcomes for certain demographic groups:

8. What tools are available (and applicable to your system) for identifying and mitigating AI bias that you have identified above as a potential risk?

9. What procedures and mitigations have you put in place to reduce harmful outcomes of AI bias that you have identified above as a potential risk?

10. Post deployment check: Are additional or newly established procedures necessary to continue mitigating bias or inequity?

    a. Which types of bias will you test for and on what cadence going forward?

b. Have rechecks continued within stated time frames?

c. How are you documenting checks (e.g, is there an owner who is responsible in your c-suite, is this process overseen by your office of general counsel, where are testing results stored and who is in charge of verifying results for safe continued use of the AI system?)

Section 4 - Costs/Benefits Evaluations

Section 4.1. System Benefits

AI systems should be checked periodically to ensure benefits outweigh inherent risks as well as implicit and explicit related costs. To identify system benefits, organizations should define and document system purpose and utility, along with foreseeable costs, risks, and negative impacts.

1. What benefits does the AI system offer?

2. ***How are you quantifying and measuring AI costs and benefits:
   a. How are costs and benefits of your AI system defined?
   b. Who is the point of contact for monitoring the negative impacts from potential costs?

3. Have the benefits of the AI system been communicated to users? How?
   - □ yes
   - □ no

4. Have training material and disclaimers about how to appropriately use the AI system been provided to users?
   - □ yes
   - □ no

5. Has your organization implemented a risk management system to address risks involved in deployment and ongoing hazards that may arise from the AI solution? For example, have you implemented an incident management process?
   - □ yes

□ no

Describe the risk management plan:

6. Who is responsible for oversight of the risk management system?

7. Describe the process for collecting and incorporating stakeholder feedback on perceived system benefits:

8. ***Have any employees or users raised concerns about the systems' safety or inclusivity, or other potentially negative impacts arising from the AI system?
   - □ yes
   - □ no

   a. If so, what concerns have been raised and by whom (user/stakeholder/role within company)?

   b. What steps have been taken to remedy the negative impact?

***Section 4.2. Potential Costs

Negative impacts can be due to many factors, such as poor system performance, and may range from minor annoyance to serious injury, financial losses, or regulatory enforcement actions.

1. What constitutes an internal or external failure for this AI system? Are there various levels of failure?

2. Describe procedures for regularly evaluating the qualitative and quantitative costs of internal and external AI system failures:

3. Can users or parties affected by the outputs of the AI system test the AI system and provide feedback?
   - □ yes
   - □ no

Section 4.3. Application Scope

Systems that function in a narrow scope tend to enable better mapping, measurement, and management of risks in the learning or decision-making tasks and the system context. Areas that help narrow contexts for system deployment include:

- Length of time the system is deployed in between re-trainings
- Geographical regions in which the system operates
- Languages in which the system operates

1. Are there areas where the narrowing of application scope could help reduce risks exhibited by the AI system?
   - □ yes
   - □ no

2. Have you consulted with experts (e.g., legal and procurement experts) to identify whether the application scope needs to be further narrowed or refined?
   - □ yes
   - □ no

Section 5 - Third-party technologies

Section 5.1. Third-party risks

Technologies, such as pre-trained models, and personnel from third-parties are another source of risk to consider during AI risk management activities.

1. Did you acquire datasets from a third party for this AI system?
   - □ yes
   - □ no

2. Did you assess and manage the risks of using third-party datasets consistent with the process above?
   - □ yes
   - □ no

   To what extent were you unable to answer key questions about potential risks:

3. Did you acquire third-party material (open-source software, pre-trained models, open-source datasets, etc.) for developing the AI system?
    - □ yes
    - □ no

    If yes, inventory third-party material acquired or utilized for this project:

4. Evaluate the risks associated with third-party material inventoried above. For example, pretrained models could carry historical biases as they are trained on internet-scraped data, or certain demographic groups could be under-represented in an open-source dataset.

Section 5.2. Controls for third-party risks

AI actors often utilize open-source software, freely available datasets, or third-party technologies—some of which have been reported to have privacy, bias, and security risks.

1. Have you applied controls—such as procurement, security, and data privacy controls—to all acquired third-party technologies, for example, when procuring a third-party AI model?
    - □ yes
    - □ no

2. Have you reviewed any audit reports, testing results, product roadmaps, warranties, terms of service, end-user license agreements, contracts, model/system cards, or other documentation available for third-party resources used in your AI system? For example, if you are using models such as DALL-E 2 have you reviewed its system card?
    - □ yes
    - □ no

***Section 6 - Legal and Compliance

In addition to upcoming legal developments, such as the EU AI Act, there are numerous Federal and state laws in the US and across the globe currently on the books that are or could be applicable to AI systems. Legal professionals should review and assess AI systems for legal risks and identify mitigations at the earliest stages and throughout the lifecycle and assessment processes.

1. Have you considered whether General Data Protection Regulation (GDPR) or related U.S. state privacy laws (e.g., California Privacy Rights Act) apply to your system?
   - □ yes
   - □ no

2. Have you undertaken a privacy review?
   - □ yes
   - □ no

3. Does your system process personal data?
   - □ yes
   - □ no

4. Have you considered whether data used to train AI contains personal identifying information (PII) or other data that could create liability?
   - □ yes
   - □ no

5. Have you considered whether your system accesses private data after deployment?
   - □ yes
   - □ no

6. If applicable to your system, have you established appropriate mechanisms to notify the user (data subject) about their data being collected and to obtain required consent? (For an example see GDPR requirements of notice and consent)
   - □ yes
   - □ no

7. Does your system use facial recognition technology that could implicate state regulations (e.g., Washington Facial Recognition legislation SB2856)?
   - □ yes
   - □ no

8. Does your system use biometric information[10] that could implicate regulations (e.g., Illinois Biometric Information Privacy Act)?
   - □ yes
   - □ no

[10] Biometric identifiers are those that are biologically unique to an individual, such as a retina or iris scan, fingerprint, voiceprint, or scan of hand or face geometry. Since these identifiers are unique to an individual, when compromised, the person has no recourse.

9. Have you checked whether your AI system has biases that could result in discriminatory outcomes in violation of civil rights laws (e.g., Americans with Disabilities Act[11], HIPAA, Title VII of Civil Rights Act, etc.)?
    - □ yes
    - □ no

10. Have you checked whether your AI system exhibits unfair or deceptive practices in violation of Section 5 of the FTC Act or results in a discriminatory act that violates the Fair Credit Reporting Act or Equal Credit Opportunity Act?
    - □ yes
    - □ no

11. Have you consulted legal counsel on compliance with other applicable laws (e.g., privacy, security, intellectual property, discrimination, contract, tort liability) and whether the AI system could be used in a regulated domain, such as healthcare or finance?
    - □ yes
    - □ no

12. What other legal system (global, federal, state) could be implicated by the AI development or use:

***Section 7 - Harm Synthesis and Final Decision

Section 7.1. Identifying Impacts

Risk assessment enables organizations to create a baseline for system monitoring and to increase opportunities for detecting emergent risks.

1. What stakeholder engagement processes have you established to identify potential impacts from the AI system on individuals, groups, communities, organizations, and society?

[11] The Justice Department has recently launched a new and improved website (ADA.gov) to better assist people to comply with the Americans with Disabilities Act (ADA) as well as understand their rights and others' rights. You can find more details here.

2. Identify misalignments between organizational or societal values and system implementation and impact:

3. Are there systems or mechanisms in place to ensure continuous monitoring for impacts and emergent risks?
   - □ yes
   - □ no

4. If the AI system relates to people, could it expose people to harm or legal action (e.g., financial, social or otherwise)? What was done to mitigate or reduce the potential for harm?

5. If the AI system relates to subjects protected by international standards or bodies, have appropriate obligations been met (e.g., medical data might include information collected from animals)?
   - □ yes
   - □ no

Section 7.2. Likelihood and Magnitude of Impact

If an organization decides to proceed with deploying the system, the 'likelihood estimate' can be used to assign oversight resources appropriate for the risk level and triage the likelihood of the system's impacts. A 'likelihood estimate' includes:

1. Describe how you measure AI system impact. For example, you can establish qualitative assessment scales, such as red-amber-green, as well as simulations or econometric approaches. Apply scales uniformly across the organization's AI portfolio.

2. ***Identify the likelihood and magnitude of each identified impact based on expected use, past uses of AI systems in similar contexts, public incident reports, stakeholder feedback, or other data are identified and documented.

3. ***Identify the likelihood and magnitude of system benefits and negative impacts in relation to trustworthy AI characteristics.

Section 7.3. Final Decision

The final decision on go/no go should take into account the risks mapped from previous steps and the organizational capacity for their management.

1. Review and examine documentation, including system purpose and benefits, and mapped potential impacts with associated magnitude and likelihoods.

2. ***Document a summary of the system's estimated risk.

3. Document why a go/no-go determination was made based on magnitude and likelihood of impact and estimated risk. If a decision is made to proceed, assign the system to an appropriate risk tolerance and align oversight resources with the assessed risk.

Appendix A

NIST Special Publication 1270 provides: Systemic biases result from procedures and practices of particular institutions that operate in ways which result in certain social groups being advantaged or favored and others being disadvantaged or devalued. This need not be the result of any conscious prejudice or discrimination but rather of the majority following existing rules or norms. Institutional racism and sexism are the most common examples. Other systemic bias occurs when infrastructures for daily living are not developed using universal design principles, thus limiting or hindering accessibility for persons with disabilities. Systemic bias is also referred to as institutional or historical bias. These biases are present in the datasets used in AI, and the institutional norms, practices, and processes across the AI lifecycle and in broader culture and society. See VIGNETTE for more examples:

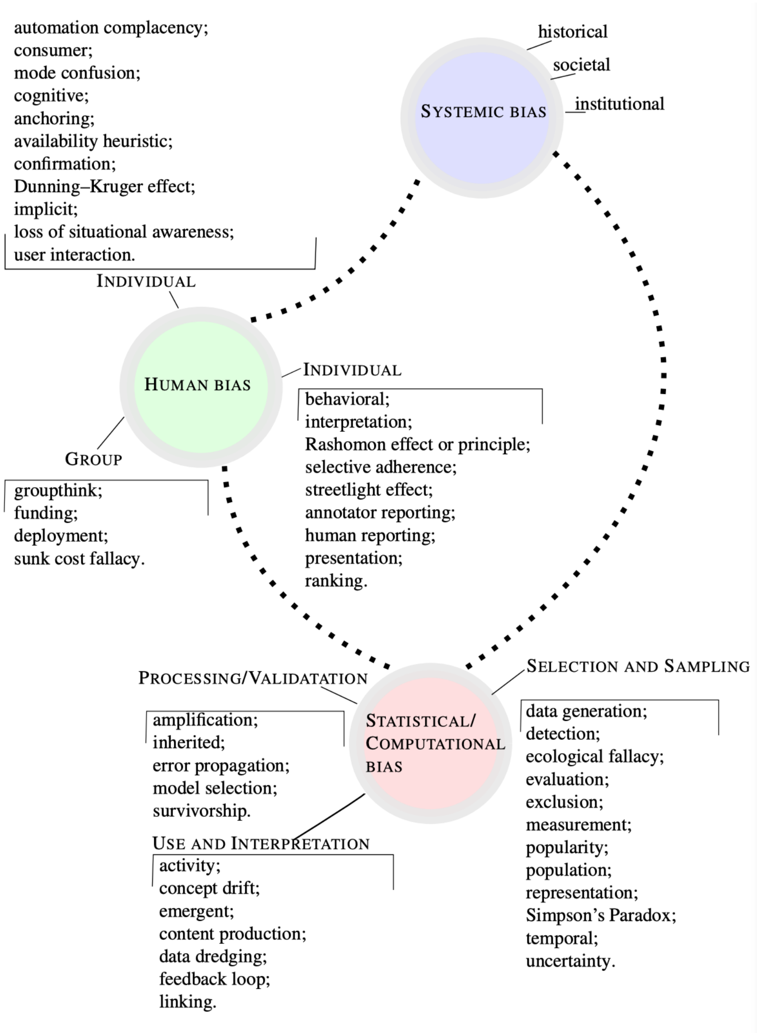

Fig. 1. Categories of AI Bias. source: NIST Special Publication on AI Bias