The following comments and proposed changes are in response to NIST's solicitation of feedback regarding NIST's "AI Risk Management Framework: Second Draft" published August 18, 2022. Each proposed change is accompanied by a corresponding comment to give NIST insight on why the proposed change was made. We appreciate to the opportunity to submit the following comments.

Location	Comment	Proposed Change
Section 1.1 (page 1)	It is important to highlight that with proper controls AI systems can have positive impacts.	Replace the second paragraph in Section 1.1 with: Managing AI risk is not unlike managing risk for other types of technology. Risks to any software or information-based system apply to AI, including concerns related to cybersecurity, privacy, safety, and infrastructure. Like those areas, effects from AI systems can be characterized as long- or short-term, high- or low-probability, systemic or localized, and high- or low-impact. However, AI systems bring a set of risks that require specific consideration and approaches. Without proper controls, AI systems can amplify, perpetuate, or exacerbate inequitable outcomes. With proper controls, AI systems can be used to fix inequitable outcomes. AI systems may exhibit emergent properties or lead to unintended consequences for individuals and communities. A useful mathematical representation of the data interactions that drive the AI system's behavior is not fully known, which makes current methods for measuring risks and navigating the risk-benefits tradeoff inadequate. AI risks may arise from the data used to train the AI system, the AI system itself, the use of the AI system, or interaction of people with the AI system.
Section 1.1 (page 1)	The list should be not exclusive. While the listed characteristics are important there may be other characteristics depending on the AI system. For example, ISO/IEC 22989 states that characteristics of trustworthiness "include for instance, reliability, **availability**, resilience, security, privacy, safety, accountability, transparency, **integrity**, **authenticity**, quality and **usability**." Several characteristics (in bold) are omitted from the NIST Framework. To achieve harmony with ISO/IEC 22989 and other frameworks there is a need to for the list to be not exclusive. It is also important to indicate that Trustworthy AI systems include balancing these characteristics. Balancing is called-out in other places of this document, but it is	Replace the third paragraph in Section 1.1 with: While views about what makes an AI technology trustworthy differ, there are certain key characteristics ~~of~~encompassing ~~trustworthy systems. Trustworthy AI is valid~~ accuracy and ~~reliabl~~reliability~~e~~, safety, fairness and bias is managementd~~d~~, security~~e~~ and resiliency~~t~~, accountability~~le~~ and transparency~~t~~, explainabliilty~~e~~ and ~~interpretabl~~interpretability~~e~~, and privacy-enhancement~~d~~. Creating a trustworthy AI requires balancing each of these characteristics based the use-case of the AI system.

	important that this call-out be made early for the readers clarity.	
Section 1.1 (pages 1-2)	Societal dynamics and human behavior/norms create the accepted parameters for bias, fairness, interpretability, and privacy.	Replace the fourth paragraph in Section 1.1 with: AI systems are socio-technical in nature, meaning they are a product of the complex human, organizational, and technical factors involved in their design, development, and use. Many of the trustworthy AI characteristics – such as bias, fairness, interpretability, and privacy – are ~~directly Connected~~influenced by societal dynamics and human behavior. AI risks – and benefits – can emerge from the interplay of technical aspects combined with socio-technical factors related to how a system is used, its interactions with other AI systems, who operates it, and the social context into which it is deployed.
New Section 1.4	The similarities and differences with ISO standards should be explained. This is important because someone reading only the NIST Framework may be unaware of the developments in the standards world, which NIST aims to leverage.	Create a new Section 1.4 entitled "Relationship to ISO Standards" with the following text: This Framework aims to harmonize efforts from other standards organizations. Both the NIST AI Risk Management framework and the the ISO-IEC JTC1 AI Risk Management standard (IS:23894 currently at FDIS) are derived from the ISO-IEC risk management framework 31000:2018. and both projects provide recommendations towards identifying and treating risks stemming from the use machine learning and artificial intelligence technology. They both target AI trustworthiness issues such as fairness, security, safety, privacy, robustness, explainability and data quality. They both establish a framework that incorporates organizational aspects such as leadership, governance, design, implementation, evaluation and continuous improvement that is active throughout the lifecycle of AI system development. Both projects also establish a set of processes for risk scoping, assessment, treatment, monitoring, review, and reporting. The NIST AI RMF structures its framework differently, with an objective towards being more risk based and pro-innovation and highlighting AI issues in a manner that makes the framework accessible to a broader audience. As it is structured into a reusable core with use-case specific profiles, it should be more appropriate and informative than a 1 size fits all framework. Finally, as the NIST AI RMF is a living document, it can adapt and evolve quickly to adjust to the rapidly evolving AI/ML technology context.

<table>
<tr>
<td></td>
<td></td>
<td><b>Framework and Process Recommendations Mapping between ISO/IEC 23894 and AI RMF</b>
<table>
<tr><td>Context</td><td>ISO</td><td>NIST</td></tr>
<tr><td>leadership and commitment</td><td>5.2</td><td>6.1 Core-Governance</td></tr>
<tr><td>Design: Understanding context of organization</td><td>5.4.1</td><td>6.2 Map</td></tr>
<tr><td>Assigning Org Roles</td><td>5.4.3</td><td>6.1 Core-Governance</td></tr>
<tr><td>Scoping Context</td><td>6.3.2-3</td><td>6.2 Core Map</td></tr>
<tr><td>Risk Criteria</td><td>6.3.4</td><td>6.3 Core Measure</td></tr>
<tr><td>Risk Identification</td><td>6.4.2</td><td>6.2 Core Map</td></tr>
<tr><td>Risk Analysis</td><td>6.4.3</td><td>6.2 Core Map</td></tr>
<tr><td>Risk Treatment</td><td>6.5</td><td>6.4 Core Manage</td></tr>
<tr><td>Risk Reporting</td><td>6.7</td><td>6 Core Map, Measure, and Manage</td></tr>
</table>
</td>
</tr>
<tr>
<td>Section 3.2.2 (pages 9-10)</td>
<td>Section 3.2.2 would benefit from an explicit call-out of low-risk and high-risk use cases. The risk tolerance heavily depends upon if the AI solution is a low-risk or high-risk solution. The balancing of the key characteristics of a trustworthy AI system will also differ based upon the use-case of the AI solution.</td>
<td>Add the following paragraph before the last paragraph in Section 3.2.2:

AI technologies can be applied to a diverse set of industries and contexts, and consequently, a one-size-fits-all approach is unlikely to be effective. AI systems could be considered low-risk depending on if the AI system's use-case poses especially no substantial harm to people or society. The costs of measuring reliability, robustness, resilience, explainability, and interpretability may not be warranted in such low-risk situations. An organization should utilize a framework to assess the general level of risk posed by using sector and use-case specific standards where available then determine the appropriate level of risk management that is warranted for the deployed AI solution.</td>
</tr>
<tr>
<td>Section 3.2.3 (page 10)</td>
<td>The defined AI actor terms "AI development" and "AI deployment" should be used into the AI RMF text in order to give clarity to the reader about the risks associated with those roles.</td>
<td>Add the following paragraph to the bottom of Section 3.2.3:

AI actors that have different roles within an AI system can have different risk perspectives. An AI developer who makes AI software available, such as pre-trained models, can have a different risk perspective than an AI deployer who implements the AI developer's pre-trained model in a specific use-case. The AI deployer has the responsibility to create a trustworthy AI system that is specific to the deployed use-case.</td>
</tr>
<tr>
<td>Section 4 (page 10)</td>
<td>The list should be not exclusive. While the listed characteristics are important there may</td>
<td>Replace the first paragraph in Section 4 with:</td>
</tr>
</table>

	be other characteristics depending on the AI system. There are also currently unforeseen characteristics that will emerge with the maturity of AI technology, so there is a need to provide flexibility. It is also important to indicate that Trustworthy AI systems include balancing these characteristics. Balancing is called-out in other places of this document, but it is important that this call-out be made early for the readers clarity.	Approaches which enhance AI trustworthiness can also contribute to a reduction of AI risks. This Framework articulates the following characteristics of trustworthy AI, and offers guidance for addressing them. Key characteristics of Trustworthy AI include ~~otrustworthy systems. Trustworthy AI is valid~~ accuracy and ~~reliabl~~reliability~~e~~, safety, fairness and bias is management~~d~~, security~~e~~ and resiliency~~t~~, accountability~~le~~ and transparency~~t~~, explainablilty~~e~~ and ~~interpretabl~~interpretability~~e~~, and privacy-enhancement~~d~~. Creating a trustworthy AI requires balancing each of these characteristics based the use-case of the AI system.~~Trustworthy AI is: valid and reliable, safe, fair and bias is managed, secure and resilient, accountable and transparent, explainable and interpretable, and privacy-enhanced.~~
Section 4 (page 11)	(1) There should be an explicit call-out that trustworthy characteristics involve tradeoffs. Although this call-out may appear in other places in the document, it is important to reiterate as it is a core concept of creating a trustworthy AI system. (2) The original second sentence is abstract and unclear. These tradeoffs should be stated in terms of their use-cases. For example, if there is a video game AI system that is highly secure but unfair that wouldn't mean that the video game AI system necessarily be untrustworthy according to its use-case. (3) Organizations that deploy AI systems have control and visibility into the risk. The AI system cannot control risk.	Replace the text in the blue box in Section 4 with: Trustworthiness characteristics explained in this document are interrelated and involve tradeoffs. ~~Highly secure but unfair systems, accurate but opaque and uninterpretable systems, and inaccurate but secure, privacy-enhanced, and transparent systems are all undesirable. Trustworthy AI~~Organizations ~~systems~~ should consider the balanced between ~~achieve a high degree of control over~~ risks associated with AI systems and other considerations such as ~~while retaining a high level of~~ performance quality, accuracy, privacy, and explainability~~.~~ Achieving this difficult goal requires a comprehensive approach to risk management, with tradeoffs among the trustworthiness characteristics. It is the joint responsibility of all AI actors to determine whether AI technology is an appropriate or necessary tool for a given context or purpose, and how to use it responsibly. The decision to commission or deploy an AI system should be based on a contextual assessment of trustworthiness characteristics and the relative risks, impacts, costs, and benefits, and informed by a broad set of stakeholders.
Section 4.2 (page 13)	The provided definition does not match the definition of safety in ISO/IEC TS 5723. Section 3.2.17 of 5723 states that "property of a system . . . such that it does not, under defined conditions, lead to a state in which human life, health, property, or the environment is endangered"	Replace the first paragraph of Section 4.2 with: AI systems "should not, under defined conditions, ~~cause physical or psychological harm or lead to a state in which human life, health, property, or the environment is endangered~~lead to a state in which human life, health, property, or the environment is endangered" (Source: ISO/IEC TS 5723:2022)
Section 4.3 (page 14)	The perception of fairness is highly dependent upon an organization's role within the lifecycle of an AI system. This is	Replace the first paragraph in Section 4.3 with:

	especially true when a designer or developer develops a general purpose AI product that is later deployed for a specific use-case. The designer or developer may not have the proper clarity into the deployer specific use-case to properly assess the fairness of the deployed product. In this instance what is fair for an AI customer may be more accurately assessed by the deployer.	Fairness in AI includes concerns for equality and equity by addressing issues such as bias and discrimination. Standards of fairness can be complex and difficult to define because perceptions of fairness differ among cultures and may shift depending on application. Systems in which biases are mitigated are not necessarily fair. For example, systems in which predictions are somewhat balanced across demographic groups may still be inaccessible to individuals with disabilities or affected by the digital divide. Fairness depends on an AI actor's position in the AI system. An AI designer AI developer for an AI product may have a different perception of fairness as opposed to an AI deployer who deploys the AI product.
Section 4.3 (page 14)	The proposed deletion section has extreme language that has an alarmist tone. We should be carefully balancing the potential negative aspects of improper AI with the benefits of properly implemented AI. Focusing too much on the potential negative aspects of AI is not advantageous to the progression of AI products and services. It is also debatable if AI can promote bias "at speed and scale far beyond the traditional discriminatory practices" given humans rich history of discriminatory practices.	Replace the third paragraph in Section 4.3 with: Bias exists in many forms, and can become ingrained in the automated systems that help make decisions about our lives. While bias is not always a negative phenomenon, certain biases exhibited in AI models and systems can perpetuate and amplify negative impacts on individuals, groups, communities, organizations, and society~~ – and at a speed and scale far beyond the traditional discriminatory practices that can result from implicit human or systemic biases~~. Bias is tightly associated with the concepts of transparency as well as fairness in society. (See NIST Special Publication 1270, "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence.")