Comment #	Submitted By (Name/Org):*	Type (General / Editorial / Technical)	Starting Page # *	Starting Line #	Comment (include rationale)*	Suggested Change*
1	Google	ed	1	N/A	Editorial - Grammatical adjustment to readability.	Modify from: "A useful mathematical representation of the data interactions that drive the AI system's behavior is not fully known, people with the AI system." Modify to: "A useful mathematical representation of the data interactions that drive the AI system's behavior ~~is~~ **may** not be fully known, **to the** people ~~with~~ **who developed** the AI system."
2		te	1	N/A	Remove "human-defined" as objectives can also be learned (e.g. reinforcement learning).	Modify from: The AI RMF refers to an AI system as an engineered or machine-based system that can, for a given set of human-defined objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. Modify to: The AI RMF refers to an AI system as an engineered or machine-based system that can, for a given set of ~~human-defined~~ objectives, generate outputs such as predictions,
3		te	1-2	N/A	Adding additional information and clarification of the concepts and terminology for increased technical accuracy. A brief description of 'inductive bias' as it is applicable for trustworthy and responsible AI and the idea that AI systems are biased.	Modify from: Responsible use and practice of AI systems is a counterpart to AI system trustworthiness. AI systems are not inherently bad or risky, and it is often the contextual environment that determines whether or not negative impact will occur. Modify To: Responsible use and practice of AI systems is a counterpart to AI system trustworthiness. **AI systems can be biased, which is called "inductive bias", however** ~~AI systems are not inherently bad or risky, and~~ it is often the contextual environment that determines whether or not a negative impact will occur. **Inductive bias is the set of a priori assumptions the machine learner must use to approximate the correct output (or label) for examples that have not been shown during training. Examples of these assumptions are that data are i.i.d.**
4		te	5	N/A	The TEVV (test, evaluation, verification, and validation) framing proposed does not fit the full range of activities that are important to risk management for AI (e.g. formulating your	The TEVV framing is fundamentally incorrect and needs to be revisited
5		te	6	N/A	Figure 2 omits Product Development which is a key step of the AI lifecycle	Add Product Development to the "deploy" phase of the AI life cycle in Figure 2
6		te	8	N/A	There are additional harms that should be considered beyond physical related harms. Expanding the example pool will help Actors realize that individuals, communities, and the environment can be harmed in other ways. For example, adding psychological harm i.e. losing out on opportunities such as loans, and other unfair allocation of resources.	The "Individual" bullet point within the "Harm To People" section should include "**Mental Injury or Psychological Safety**" and "**Opportunity or Economic Loss**" within Figure 3

Comment #	Submitted By (Name/Org):*	Type (General / Editorial / Technical)	Starting Page # *	Starting Line #	Comment (include rationale)*	Suggested Change*
7		te	10	N/A	"Accountable and Transparent" suggests that these equities are separate considerations when in reality they are overlapping, "Valid and Reliable" is a more accurate overarching AI trustworthy characteristic	Switch the position of "Accountable and Transparent" with "**Valid and Reliable**" within Figure 4
8		te	11	N/A	Additional clarity is needed around the scope, role, and responsibility of all actors in the value chain of AI development and deployment. As stated, the concept of "joint responsibility" does not sufficiently distinguish at what points each party may be responsible for what and may result in greater uncertainty as to each party's responsibilities. A good reference for this is the soon to be published ISO 5339 "Guidelines for AI Applications" Section 6 'Stakeholders' perspectives and AI application framework'.	Modify from: "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." Modify To: "It is the joint responsibility of all AI actors **at certain points** to determine whether AI technology is an appropriate or necessary tool for a given context or purpose, and how to use it responsibly."
9		te	12	N/A	Modify wording to remove the implication that the challenges detailed in this section are applicable to all AI systems.	Modify From: "These challenges are exacerbated by AI system opacity and the resulting lack of interpretability." Modify To: "These challenges, **which may be impactful but are not present in all AI systems**, ~~are~~ **may be further** exacerbated by AI system opacity and the resulting lack of interpretability."
10		ed	13	N/A	Update the section title to align with content. This section focuses on reliability and robustness while validity is only mentioned briefly.	Modify from "4.1 Valid and Reliable" Modify to: "4.1 ~~Valid and Reliable~~ Reliability"
11		te	13	N/A	While most "good" AI models interpolate fairly well, few models extrapolate well. Failure to extrapolate should not be considered to be "not robust", but failure to interpolate would be a "robustness failure".	Modify From: "Deployment of AI systems which are inaccurate, unreliable, or non-generalizable to data beyond their training data (i.e., not robust) creates and increases AI risks and reduces trustworthiness." Modify To: "Deployment of AI systems which are inaccurate, unreliable, or ~~non-generalizable~~ **poorly generalized** to data ~~beyond~~ **not in** their training data (i.e., not robust) creates and increases AI risks and reduces trustworthiness."
12		ed	14	N/A	Editorial - section title is grammatically incorrect	Modify from: "4.3. Fair – and Bias Is Managed" Modify To: "4.3. Fair**ness** ~~– and Bias Is Managed~~"

Comment #	Submitted By (Name/Org):*	Type (General / Editorial / Technical)	Starting Page # *	Starting Line #	Comment (include rationale)*	Suggested Change*
13		te	15	N/A	The usage of explainable and interpretable is not consistent with widespread use of these terms within industry. There are existing standards and frameworks already published, or being published soon that are internationally accepted, written by technical AI experts, addressing AI Explainability and AI terminology which would be beneficial to include in the NIST documents to promote cohesion and alignment between ISO and NIST frameworks. These are ISO/IEC 22989 AI Concepts and Terminology (Published May 2022) and ISO/IEC 6254 AI Explainability (Estimated Pub. February 2024)	Align the definitions of explainable and interpretable other with existing recognized standards including ISO/IEC 22989 AI Concepts and Terminology (Published May 2022) and ISO/IEC 6254 AI Explainability (Estimated Pub. February 2024)
14		te	16	N/A	Privacy related mitigation's have trade-offs with issues in addition to the ones listed, add verbiage to include an example of these trade-offs	Modify from "From a policy perspective, privacy-related risks may overlap with security, bias, and transparency." Modify to: "From a policy perspective, privacy-related risks may overlap with security, bias, and transparency **and come with trade-offs with these other equities e.g. restrictions on collection of demographic data can make it more difficult to do fairness testing.**"
15		ge	16	N/A	As NIST has multiple frameworks on applicable information in draft and published, along with various international standards, it is recommend to add a section in the annex which will include a non-exhaustive list of standards that may support AI risk management and promote cohesion and alignment between existing frameworks and standards.	Add a section in the annex to include a non-exhaustive list of standards that support AI risk management. Including the NIST Cybersecurity Framework and NIST Privacy Framework.
16		te	20	N/A	The ability to withstand attack and “fail gracefully” is crucial and should apply to more than third-party acquired software. More generally robustness can be interpreted as affirmatively and intentionally designing an AI system to cope with failure and adapt to new situations. For example: 1. Coding in hard constraints to prohibit unexpected system behaviours outside of the range deemed safe. Adding such constraints needs to be done judiciously so as to not undermine the system’s resiliency in adapting to new situations. 2. Formal pre- and post-launch vulnerability testing processes, as well as processes to support monitoring throughout the life of an AI system. No system will ever be perfect, and most failures that occur will be unexpected.	Modify from: GOVERN 6.2: Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk. Modify to: GOVERN 6.2: Contingency processes are in place to handle failures or incidents in ~~third-party~~ data or AI systems deemed to be high-risk.

Comment #	Submitted By (Name/Org):*	Type (General / Editorial / Technical)	Starting Page # *	Starting Line #	Comment (include rationale)*	Suggested Change*
17		te	22	N/A	Often the AI system expands its application or scope after deployment, the term "narrowed" is subjective. Narrowing uses of the application is not necessarily an effective risk mitigation measure.	Modify from: MAP 3.3: Targeted application scope is specified, narrowed, and documented based on established context and AI system classification. Modify to: MAP 3.3: Targeted application scope is specified, ~~narrowed,~~ and documented based on established context and AI system classification.
18		ge	22	N/A	Some guidance within the draft is currently limited to third party only. For instance, Govern, Measure, and Manage would all benefit from guidance on risk failure and internal risk controls and documentation. Controls assigned to third party risk management are also best practices for general AI risk management and should be included across the framework.	Remove the usage of "third-party" from the subcategories within Tables 2-5.
19		te	24	N/A	Accountability and transparency are key trustworthy characteristics and should be included here. Add wording to align with usage of these terms within section 4.	Add an additional subcategory to Measure 2 within table 4: **MEASURE 2.11 AI system is regularly evaluated for and maintains transparent and accountable measures**
20		te	24	N/A	Beyond documenting and conducting these measurements, there should be an assessment of effectiveness	Add an additional subcategory to Measure 2 within table 4: **MEASURE 2.12: Evaluate and determine the effectiveness of the above trustworthy measures**
21		te	24	N/A	An important part of measuring environmental impact of model training includes assessing energy use, source (i.e. how much carbon the source emits), and efficiency.	Modify from: MEASURE 2.10: Environmental impact and sustainability of model training and management activities are assessed and documented. Modify To: MEASURE 2.10: Environmental impact and sustainability of model training and management activities are assessed and documented including **energy consumption, source, and efficiency.**
22		te	24	N/A	This measure should be expanded to be inclusive of the product life cycle	Modify from: MEASURE 4.2: Measurement results regarding system trustworthiness in deployment context(s) are informed by domain expert and other stakeholder feedback to validate whether the system is performing consistently as intended. Results are documented. Modify to: MEASURE 4.2: Measurement results regarding system trustworthiness in deployment context(s), **as well as all other product lifecycle stages** are informed by domain expert and other stakeholder feedback to validate whether the system is performing consistently as intended. Results are documented.
23		te	27	N/A	Product design is not clearly incorporated within Appendix A, despite being essential to risk management	Product design should be defined within Appendix A

Comment #	Submitted By (Name/Org):*	Type (General / Editorial / Technical)	Starting Page # *	Starting Line #	Comment (include rationale)*	Suggested Change*
24		te	30	N/A	It is important to note it is the unwanted/harmful bias that is concerning, not bias in general. It is true that existing frameworks are unable to "adequately manage the problem of bias in AI systems" but it is important to point out further critical considerations: some of these are at odds with bias management. For instance, privacy and bias management may be in conflict; organizations will need to understand the trade-offs across all topics and make a decision in the best interest of managing risk and document it accordingly.	Modify from: Existing privacy, computer security, and data security frameworks and guidance are unable to: » adequately manage the problem of bias in AI systems; Modify to: Existing privacy, computer security, and data security frameworks and guidance are unable to: » adequately manage **unwanted/harmful** ~~the problem of~~ bias in AI systems;
25		te	31	N/A	Remove this sentence - not sure this statement is everyone's perception about AI systems, and may be over reaching.	Delete: ~~One major false perception is the presumption that AI systems work – and work well – in all settings.~~
26		te	31	N/A	System use across all settings should be addressed in a separate more robust section about managing such risks. For instance, in the case of General Purpose AI, developers should produce robust guidance and documentation about the settings in which the product was tested and validated for and in which settings it should not be used.	Introduce a new section on managing risks related to system use across all settings