

All comments will be made public as-is, with no edits or redactions. Please be careful to not include confidential business or personal information, otherwise sensitive or protected information, or any information you do not wish to be						
						Submit comments by October 15, 2020 to: explainable-AI@nist.gov
Comment Template for First Public Draft of Four Principles of Explainable AI						
Comment #	Commenter organization	Commenter name	Paper Line # (if applicable)	Paper Section (if applicable)	Comment (Include rationale for comment)	Suggested change
1	CSA AI Workgroup	Yanalitis, Mark	120	Introduction	sensational language "high-stakes"	Mundane Virtual Assistants, SEO and recommender systems use AI. Unsure if all are "high-stakes" decision making.
2	CSA AI Workgroup	Yanalitis, Mark	125	Introduction	Tortuous logic.	"AI-rendered decisions increasingly encounter calls for transparency regarding decision-making rationale and results explainability. Failing to articulate the decision rationale can affect the level of trust users grant to an AI system. Suspicions that the system contains biased or unfairness raises concern about harm to oneself and to society [40, 102]."
3	CSA AI Workgroup	Yanalitis, Mark	128-129	Introduction	misleading speculative statement – consumers often cannot opt out of AI decision making and use.	Strike out "This may slow societal acceptance and adoption of the technology, as [..]" Capitalize "Members. [new sentence]"
4	CSA AI Workgroup	Yanalitis, Mark	131	Introduction	Focus call to action, its not just about attribute awareness	Coming out the NIST Fairness in AI workshop, Data Scientists and system developers need to have more awareness and focus on project level attributes (Representative data, Fairness, good-faith statements), and have less (or on balance) parity with problem level attributes. DS and DEV needs to focus on metric alignment not exclusively model performance to reduce the impact of silent failures (Bias and fairness).
5	CSA AI Workgroup	Yanalitis, Mark	134	Introduction	Unsure if these attributes apply to the system, or the decision, or as a measure of trustworthiness. Is it that the AI decisions are consistently trustworthy under excessive system load (resiliency); Is it that the AI system makes reliable decisions over time, or is that a technical property of the system itself? Issues of bias and accountability are typically silent failures tied to governance. The descriptor munges architectural attributes with governance attributes.	Strike out "Other properties include resiliency, reliability, bias, and accountability."

6	CSA AI Workgroup	Yanalitis, Mark	165-171	2. Four Principles of Explainable AI	Are these principles or attributes of AI decision-making systems or AI governance programs? Line 153 sets the principles squarely in the HCI interaction. According to TOGAF 9.2 "Enterprise Principles provide a basis for decision-making throughout an enterprise, and inform how the organization sets about fulfilling its mission."	Strike In 165 to 171. Pick up with a full explanation in Section 2.1. Consider Restructure the principle in a manner such that it can be clearly distinguished between the principle name, principle statement, rationale, and implication(s).
7	CSA AI Workgroup	Yanalitis, Mark	245-312	Section 3 3. Types of Explanations	The definition and classification section seems come late, after the Explanation and Explanation Accuracy principle introduction occurs.	Consider moving this to a preamble section that defined terms and characteristics between the Introduction and Section 2.
8	CSA AI Workgroup	Yanalitis, Mark	209-210	2.3 Explanation Accuracy	Thesis or primary topic explanatory sentence occurs at the end of the introductory paragraph	Consider moving "The Explanation Accuracy principle imposes accuracy on a system's explanations." so as to have it be the introductory sentence following the subsection header.
9	CSA AI Workgroup	Yanalitis, Mark	209-211	3.3 Explanation Accuracy	Choppy introductory paragraph	Move line 209-210 to lead in sentence. Strike out "Together, the," Change "these" to "the," change word "only" by replacing with "specifically."
10	CSA AI Workgroup	Yanalitis, Mark	229-230	4.3 Explanation Accuracy	Unnecessary implicit logic.	Consider striking out "The previous principles implicitly assume that a system is operating within its knowledge limits."
11	CSA AI Workgroup	Yanalitis, Mark	277-312	Section 3 3. Types of Explanations	Should this be an end note? The section is a deep explanation of the interaction between explanation types and not necessarily in direct support of the principles.	Consider moving this to an Annex as in other NIST papers
12	CSA AI Workgroup	Yanalitis, Mark	313-544	Section 4	Typically, review of literature preceded the introduction of the derived or synthesized new content (the NIST AI principles).	Consider re-evaluating the location of the literature review in the paper.
13	CSA AI Workgroup	Yanalitis, Mark	313-545	Section 6	Consider recasting the subsection as a synthetic use case, perhaps something from the graph displayed earlier.	Section 6 reads like a veiled synthetic use case. Consider writing it as a theoretical, hypothetical synthetic use case to convey principle impact in context.