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 posted.

Comment Template for First Public Draft of Four Principles of Explainable Artificial Intelligence (Draft NISTIR 8312)

Submit comments by October 15, 2020 to:
explainable-AI@nist.gov

Comment	Commenter	Commenter name	Paper Line # (if	Paper	Comment (Include rationale for	Suggested change
	Monitaur.ai	Andrew Clark	134		Interpretability should be added as another property of trust in AI systems. Interpretability is the ability to ascertain the mechanics of how a model is constructed and decisions are made. Explainability is the ability to understand in an intelligible manner the result of a model outcome.	Include interpretability as a tenet of trust in AI systems.
	Monitaur.ai	Andrew Clark	134		Bias is not a tenet of trust in AI system. The *lack of bias* could be, but clarifty and care should be taken whenever bias is mentioned.	Bias should be removed or modified to say a lack of bias.

	Monitaur.ai	Andrew Clark	333		Counterfactuals can also be the ability to reperform a transaction to perform a "What if" analysis postfact. Interrogating a model to see where the inflection points are very helpful to provide end-users the explanation they need, especially for "black box" models where other explanation approaches may prove difficult. This approach is described as starting line 372. Provide a more concrete separation and explanation on the two different types of counterfactuals, the one described above and the one given in section 5.3. Define both around line 333	Add another sentence or two around the other definition of counterfactuals
	Monitaur.ai	Andrew Clark	358		Models that are self-explainable have the property of interpretability	

	Monitaur.ai	Andrew Clark			As a general comment, for a model to be truly trustworthy, the model needs to be audited and ideally monitored by an objective party. These Four Principles provide helpful guidance that helps to mitigate risks, but their success in mitigating those risks will only be realized if they are incorporated as part of a broad governance program overseen and validated by objective parties.	