Subject: Comment on AI RMF Initial Draft

Hi NIST AI RMF Team,

I think this is a good first pass, but there is an assumption stated on page 8, line 2 that I think you may want to reconsider:

"Since AI trustworthiness and risk are inversely related, approaches which enhance trustworthiness can contribute to a reduction or attenuation of related risks."

This relationship may not always hold. For example, a company could develop an AI system that passes trustworthy thresholds on accuracy, interpretability, etc., but the use of that AI can still be potentially harmful and unjust (i.e. law enforcement). Improving the accuracy or managing technical bias of the AI will be of no less risk to the people whom it is used on, and potentially could increase risk to those people by supporting an unjust or oppressive system.

While you state on page 1, line 12, that ensuring that a use case is not harmful "is fundamental to trustworthy AI", the framework doesn't seem to discuss how to determine the harmfulness of a use case. I wonder if you can expand on this.

Thanks,