Comments submitted by ISPE in response to a public consultation of National Institute of Standards and Technology (NIST) publication, A Proposal for Identifying and Managing Bias in Artificial Intelligence (NIST Special Publication 1270), September 2021.

Line number	Comments
1	In view of the document draft, there are a few points we believe are relevant to consider: 1) Achieves comprehensiveness of results The AI system-generated should be able to achieve comprehensiveness in the corresponding sector to prevent bias. 2) Deployment phase should also include reporting details of the design and development phase, ideally including the iterative process when developing the algorithm. 3) To develop best practices in pre-designing, designing, developing, and deploying AI systems. 4) To increase transparency and replicability 5) If possible involving multiple data sources throughout the process, especially in the validation process
4	From the title, the expectations for this proposal were to provide a solution to first, identify biases in AI, and second, to manage biases in AI. However, we are struggled to identify constructive and scientific recommendations in how to identify and manage bias, across different types of biases mentioned in the document. Instead, the document included extensive descriptions, yet in a little bit of spontaneous approach, of different types of bias with some examples. It would be very helpful to the readers if the documents could include some sensible approaches in both identifying and managing bias. Otherwise, it may be sensible to consider changing the title of the document.
101	The abstract sets the tone for this proposal, where the focus is directed towards establishing the credibility of AI without the mention of its actual validity for use. Biases not only affect AI trustworthiness but more importantly misinform stakeholders, with potentially serious consequences to its users and the broader public. Perhaps the authors wish to convey that biases will have diminished effects on misinforming the audience when properly identified and managed and output can be interpreted sensibly and productively. This link between AI biases and their effects on actual AI validity and consequently AI trustworthiness is currently missing. Overall, the NIST should be commended for providing recommendations to advance the science of bias identification and minimization significantly, particularly for the principles, practices and implementations. This document should add a statement that evaluation frameworks should be used to guide the design of the bias identification and minimization approaches and cite sources where different frameworks can be found.
207	Starting from the fourth paragraph, there are instances where the concepts of biases and trustworthiness were mixed together, e.g. line 209 “Managing bias is a critical but still insufficiently developed building block of trustworthiness.” The authors are advised to rethink the differences between the challenges brought by distrust in AI vs that by biases in AI as they are very different matters. Distrust in AI when existing biases are not addressed nor declared should not be discouraged as it reflects the critical

Line number	Comments
	thinking of its audience. This is exactly what is needed when using AI – human intelligence and leadership to ensure the output is informative and ethical, where AI system provides. (and also stated in Section 3 – Approach where the authors quoted Knowles and Richards). AI bias per se does not directly cause public distrust, invalid output and harmful consequences do.
213	“This deviation from the truth can be either positive or negative, it can contribute to harmful or discriminatory outcomes or it can even be beneficial.” – the first and second halves of the sentence refer to the same idea. However, is not immediately clear how biases can be beneficial. Examples are needed on how (1) biases in AI could be beneficial and (2) AI will be a valid and useful tool when biases are identified and managed. Alternatively, leave out sentences that refer to "beneficial bias"
219	"Not all types of bias are negative" may need some elaboration and seems this statement is only true in some sectors. For example when building an AI prediction model for health care, efforts should be made to minimize any types of bias.
231	Perhaps a succinct sentence to describe this sequential relationship of the inherent introduction of biases, mitigation of harms from AI bias, and productive use of AI is needed to conclude the introduction section: “By realizing the different type of biases and mitigating the harms from AI bias, this could better utilize the applications of AI in the modern society.”
233	Section 2 provided the problem statement that this paper is trying to address. The first paragraph would benefit greatly from digging deeper into the reasons behind the public’s concern. The current narrative provided may unintentionally polarise the public (the distrusting) and AI (the distrusted). Concerns from public are not solely directed towards AI technology per se, but more so towards those who misuse or even exploit AI for their personal gains. This has to do with regulatory monitoring, quality assessment, and an open feedback system.
237	"...that biases may be automated within these technologies, and can perpetuate harms more quickly, extensively, and systematically than human and societal biases on their own." The belief that “biases may be automated within these technologies” is not in itself untrue when considering how engineers and users have the ability to influence the systems. However, this is suggesting the importance of preventing biases. Although this premise may not be what the paper is trying to address, it should not be overlooked. Overall, there is a lack of mentions of the current understanding of AI technology in the public including the ways it operates, benefits, potential dangers, and precautions to be taken for its appropriate use in understandable layman terms. This perhaps is one of the key reasons that has led to public distrust and potentiates further biases of AI. When unaddressed, even upon scrutinization of the development process, will remain huge obstacles towards building reliable AI systems.

Line number	Comments
279	"The difficulty in characterizing and managing AI bias is exemplified by systems built to model concepts that are only partially observable or capturable by data." This difficulty potentially leads to an important concept: Transparency. Some AI systems or algorithms are being criticized as a "black-box" as they are lacking transparency and thus lead to challenges in building trust. It is recommended to increase the transparency of AI technology. Similar to the previous comment on public understanding of AI technology, the publicity work perhaps needed to be expanded to the scientific community as well.
356	It is unclear whether Section 3 aims to illustrate the approach that NIST is taking to mitigate AI bias or provide a direction to the broader audience of how they can manage these biases. It is advised that this section be broken down to highlight the roles of various parties in this “collaborative approach”, covering that of regulatory authority, decision-makers, developers, and users.
367	“...accompanying definitions are presented in an alphabetical glossary in Appendix A.” Appendix A provides a useful compilation of prominent biases and associated definitions. Where feasible, it may be beneficial to illustrate where/how these biases can be introduced/map to the 3 phases in the proposed framework stated in the following section.
397	Section 4 may benefit from rearranging the order of Figures 1 and 2, where the summary should be reframed to focus on specific action to be taken at each of the three-stage processes instead of the bias presentation which should be part of the problem statement. The comprehensive list of bias types included at the end will be extremely useful with examples of how they can be managed and when they arise during the AI lifecycle.
435	It is important to emphasize that investigators and developers need to be clear if the data is appropriate to address the study question/problem in the "pre-design" phase. In fact, it is essential to understand the data source in advance before carrying out any studies. Just as transparency is a critical concept for AI systems and algorithms as noted in our Comment in reference to Line 279, similar transparency in the characterization and assessment of the "fitness" and limitations/gaps of the source data/datasets being considered for use is essential during the "pre-design" phase.
441	This paragraph highlighted the need for practice guidelines or recommendations for best practices. We understand there could be unknown impacts, but investigators should have contingency approaches for the unexpected. This also reinforces the need for guidance/recommendations for best practices regarding broad stakeholder engagement and representation in the pre-design phase to elucidate potential unintended or unexpected uses and impacts in order to inform paths to mitigate potential biases and harms, or in certain cases, to determine that development of the proposed AI system or tool should not

Line number	Comments
	proceed.
462	In healthcare study that will involve human subjects, the study protocol needs to be reviewed and approved by a corresponding ethical or independent review board as a gatekeeper of the study. To a certain extent, this may also be applicable to other sectors as well in addressing problems that can occur in the pre-design phase.
512	"The stakeholders in this stage tend to include software designers, engineers, and data scientists who carry out risk management techniques..." As with the potential biases and unintended effects that can arise from restricting engagement during the predesign phase to a narrow set of stakeholder perspectives, it seems that similar considerations would be relevant for the Design and Development Phase. While the importance of a tighter connection between AI development teams and subject matter experts is noted later in the Design and Deployment Phase section of this document, it may be beneficial to highlight its importance in the introduction paragraph for this Phase.
521	"....always select the most accurate models" This may potentially lead to overfitting and thus the AI system cannot generalize or fit well on an unseen dataset, reducing the reliability in deployment.
540	"...It is also notable that, depending on the industry or use case, AI is typically marketed as an easy solution that does not necessarily require extensive support." It is quite clear that the advantage of deploying an AI system is that it does not require extensive human effort. However, it is the converse in building the system in designing and developing. Although most of the AI systems are data-driven, human effort is essential to make sure (1) the algorithm is set up properly and (2) to monitor its evolving uses and intended/unintended impacts over time.
690	Preventing bias in the pre-design phase, applying the best practice in the design and development phase and sensible deployment with proper recognition of the strengths and weaknesses of the system are all essential.
711	In all studies involving big data, transparency and replicability are also important. We recommend that to promote researchers to increase transparency in developing an AI system, and also to support "open science" to increase replicability.
730	Table 1 listed some common terminology of biases but some of them are quite similar to each other and it may be easier for the readers to understand if the list can be organized into a hierarchical table.