UCDAVIS
OFFICE OF RESEARCH

February 2, 2024

Subject: Request for Information (RFI) Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence

To Whom It May Concern:

The UC Davis Office of Research is pleased to share a response to this RFI from the National Institute of Standards and Technology from Scott MacDonald, MD, FACP, FAMIA. As an institution with an academic medical center at UC Davis Health, we have enclosed a healthcare-specific perspective relevant to NIST’s responsibilities in addressing artificial intelligence, referencing subtopic “Developing Guidelines, Standards, and Best Practices for AI Safety and Security”.

As a practicing physician and informatician, I have several years of experience evaluating and implementing ML models and now lead the deployment of several generative AI solutions. Furthermore, as a founding member of UC Davis Health’s Analytics Oversight Committee, I develop local processes for initial and ongoing evaluation of AI. Although informed about statistical evaluation, I am more interested in how we can identify and mitigate racial and other biases to ensure fair use of generative artificial intelligence to improve the efficacy and efficiency of clinical processes.

There is no standard, proven approach to identifying bias, especially for institutions with less sophisticated analytic expertise. Establishing best practices, with specific preference for those that can be implemented by users of many levels of expertise and resource availability, will be key to compensation for systemic biases in the underlying data used to train models. This will prevent amplification of existing disparities and may help alleviate existing disparities if done well. “Explainability” of large language models (LLMs) is a proxy for bias and other safety issues that are known or may arise. Given the complexity of LLMs, we need new metrics around aspects of models that can give users high-level understanding, as opposed to low-level understandings that may be less useful. Synthetic data use for disinformation purposes may be more concerning in general society than in healthcare, but ‘watermarking’ and ‘explain-ability’ are key for humans to understand the provenance and trustworthiness of model output.

The ‘nutrition label’ approach is conceptually useful, but the components of the label have yet to be well defined and agreed upon. Different levels of sophistication may be needed for different audiences (i.e. a front line clinician, an applied informaticist, a data science/AI expert).