Alliance for AI in Healthcare (AAIH)
1340 Smith Ave, Suite 400
Baltimore, MD 21209
Tele: (410) 779-1245

June 9th, 2019

Subject: AAIH Response to NIST Request for Information on AI Applications, Docket # 190312229-9229-01

Dear NIST,

The *Alliance for Artificial Intelligence in Healthcare (AAIH)* is an international multi-stakeholder advocacy organization based in Washington, D.C. that promotes scientific, legislative, and regulatory initiatives necessary to facilitate the development of, access to, and implementation of artificial intelligence (AI) powered healthcare solutions. AAIH is comprised of over 25 organizations that utilize AI in biomedical R&D and clinical applications including growth phase start-ups, biopharma, diagnostics and device manufacturers, and research institutions. AAIH takes the lead on the sector's most pressing and significant issues, fostering research, development, investment, and commercialization of transformational treatments and cures for patients worldwide.

It is out of that dedication to our mission that we are engaging with various entities regarding opportunities to bring together industry and government to assist in the development of multilateral standards in AI with focus on healthcare. In our preliminary discussions we have covered the need for open and collaborative government engagement with the technology development industry, to best facilitate continued growth and success of the sector. We look forward to continuing these interactions with other US government agencies and are eager to provide top level commentary on this RFI to help advance NIST's objectives in AI applications relating to the healthcare sector.

After review of the document in total, and conducting a survey of AAIH member organizations, we have compiled general comments on the RFI with a focus on high priority topics from a healthcare perspective. We see many positive features of this approach but believe that AI standardization must account for domain/sector specific dependencies and inter-dependencies particularly for healthcare, which we believe has sector-specific challenges. As such, our response covers topics pertinent to the healthcare arena. Notably, our comments are centered around the following themes:

1. Overview of Current Challenges of AI in the Healthcare Sector
2. Availability of Healthcare Data
3. Standardized Data Sets and Validation Methods of AI Approaches
4. Collaboration with the FDA for Emerging Regulatory Standards

Respectfully submitted *on behalf of AAIH,*

AAIH Technology & Standards Development Committee Members
Annastasiah Mudiwa Mhaka, PhD (Co-Founder, AAIH), Aaron Chang (Strategic and Technical Advisor, AAIH)
info@theaaih.org   www.theaaih.org

1. CURRENT CHALLENGES OF AI IN THE HEALTHCARE INDUSTRY

Despite recent advances and popularity of AI in healthcare, there are still many challenges facing the implementation, deployment, and adoption of AI in healthcare. A variety of existing use cases are already poised to disrupt healthcare across the continuum of biomedical discovery, R&D, and clinical applications. These include:

- **Biomedical Discovery, Research & Development:**
  - Use machine vision to analyze cellular assay images and video
  - Use machine vision to analyze animal behavior in in vivo testing for potential psychiatric drugs
  - Novel drug target discovery
  - Predict cost and time effective synthetic routes for new chemical entities
  - Drug discovery
  - Predict machine failures in pharmaceutical manufacturing processes
  - Patient selection for clinical trials
  - Clinical trial adherence
  - Categorize and analyze pharmacovigilance data
- **Clinical Care:** ***Workflow efficiencies, outcomes prediction and decision making***
  - Use machine vision in repetitive tasks like analyzing tests, X-Ray, CT scans, data entry, detecting skin cancer, analyzing eye scans, etc.,
  - Assist physicians in getting up-to-date medical information from journals, textbooks and clinical practices for proper patient care Predict susceptibility to hearing loss
  - Chatbots that can act as virtual nurses to provide basic health information and patient orientation
  - AI along with robotics to perform surgeries like cataract removal
  - Speech analysis to help predict depression and suicidal tendencies
  - Predict when a trauma victim might wake up from coma
  - Assist physicians in identifying treatment design
  - Predict and identify the outbreak of diseases such as Ebola
  - Predict blood glucose levels for diabetic patients
  - Applying AI to wearable device data to help physicians offer personalized and more immediate care
  - AI bots by Health insurance companies for faster and smarter claims management

We believe that the *collection, anonymization, organization, protection, compliance, accessibility, quantity, quality, and formatting* (structured and unstructured) of data are some of the key issues faced when implementing AI solutions in healthcare. There is, also, a lack of common data standards across healthcare stakeholders such as hospitals, clinics, biotech/biopharma companies, pharmaceutical companies, laboratories, academic institutions, federal agencies, insurance companies, managed care companies, healthcare professionals, doctors, patients, etc.

- **Legacy systems** common in the healthcare industry, especially those based on paper and other physical documentation, lead to the healthcare industry lagging behind many other sectors in digitization of records. The 'illegible doctor's handwriting' is more than a stereotype - this is one of the many sources of medical error, as well as, an impediment to digitization. Meaningful use guidelines have accelerated EMR adoption for this purpose, though the lack of interoperability between digital legacy systems merely shifts part of the 'illegibility' problem to the digital realm.
- **Data breaches** are one of the top issues which have increased dramatically over the last few years in healthcare due to improper data security measures and controls. For example, according to HIPAA Journal, "April was the worst ever month for healthcare data breaches. More data breaches reported than any other

AAIH

month since the Department of Health and Human Services' Office for Civil Rights started publishing healthcare data breach reports in October 2009."[1]

- **Regulatory requirements and restrictions** around healthcare specific data (e.g. GDPR, HIPAA, Part 11, etc.) may differ from region to region and may be incompatible with advancing use of AI/ML.
- **Explicit standardized permissions** or anonymization processes are lacking in the healthcare industry, which can lead to patients and healthcare providers (e.g., doctors, nurses, etc.) being uncomfortable with sharing their data for research and other purposes. Further, non-traditional healthcare companies such as Amazon, Google, Uber, Apple, and Microsoft, among others, are rapidly entering into the healthcare space, at a time when "universal standards" are yet to be developed.

Lastly, AI disruption in healthcare is in rapid growth mode; however, there are not enough qualified subject matter experts in both AI and industry-specific fields (e.g., neuroscientists, cardiologists, biologists, oncologists, etc.) to support the high demand.

2. AVAILABILITY OF HEALTHCARE DATA: *READY ACCESS TO DATA*

For AI technologies to succeed in healthcare, it is critical that data pertaining to patient health (e.g. electronic medical records and clinical trial data) be shared to enable the creation of large datasets that can be used to train and build models. At the same time, doing so must protect patient privacy and address any concerns of data being shared inappropriately and without appropriate consent. The standards for AI technologies in the healthcare realm have special needs due to both the sensitive, private information covering patient health as well as the significant impact of the outcomes which can impact an individual's health.

Availability of data remains a key challenge to innovation in the healthcare sector. Incentivizing and supporting appropriate measures to access health information and some drug related datasets is critical beyond simply establishing common standards to facilitate sharing of these data.

A government-led and industry informed strategic policy around data sharing, with appropriate privacy and security measures following recognized best practices in AI, is essential.

Best Practices for AI Technologies

Establishing an overarching strategic process and best practices for building and deploying AI technologies is one of the most critical components for a successful implementation of AI in healthcare. We have identified the following "top" best practices to enable healthcare companies and ultimately patients to realize the benefits of adopting AI technologies quickly and more efficiently. While not mandatory, demonstrating one's level of compliance to a set of best practices may enable comparison of different approaches:

- **Integrated and scalable data platform:** As part of a robust AI enterprise solution, healthcare organizations must have a unified, efficient and scalable data platform that allows them to accommodate their specific industry needs and continuous data challenges such as variety and volume, as well as providing data scientists the ability to deliver consistent AI at scale. Also, it's imperative that data is reliable, consistent, up to date, free of duplication, and fit for its purposes. *"Better data"* leads to *"better decisions"*, which ultimately leads to *"better AI solutions"* in healthcare.

- **Trustworthy AI machinery:** AI models should be designed to offer clear interpretability and serve as a strong foundation to support specific business and patient care needs (e.g. drug discovery work, treatment decision-making, etc). The following measures help in strengthening interpretability of results; *using exploratory data analysis* before model building; *identifying outliers* in the data and *generating a set of likely outcomes* from the training data set to verify it with model outputs; *storing data from each possible*

[1] HIPAA Journal (2019 Apr) "April 2019 Healthcare Data Breach Report," Retrieved from: https://www.hipaajournal.com/april-2019-healthcare-data-breach-report/

intermediate stage or layer of the machine learning process, and *utilizing model agnostic frameworks* to explain local results to identify decision-making mechanism by learning algorithms.

- **Data Privacy and Security:** Whether staying on-premise or utilizing a public cloud provider, healthcare organizations cannot afford to compromise existing security controls, data privacy standards and compliance policies associated with AI applications. A well-conceived policy helps with transparent data collection and usage to implement AI. For example, a data leakage protection system can be created using Machine Learning to protect Personally Identifiable Information (PII). Using Natural Language Processing, this system can detect PII such as name, home address, and email address.

- **Openness and Adaptability:** AI solutions must be able to be adapted to diverse environments and situations. For example, AI should work for and benefit people just as much in small, rural clinics as in larger, urban settings. Professional development and training should be planned for people tasked with implementing these solutions within various communities.

- **DevOps friendliness:** Agility and speed are critical to launching AI solutions in healthcare. An ideal AI solution should have automation for the *entire solution* (pre-processing, quality control, feature extraction, model selection, model optimization, model interpretability, etc.). Automation also empowers data scientists to collaborate, avoid repetition, and focus more time on developing effective fit for purpose AI models.

3. Standardized Datasets and Validation of AI Methods

When people talk about standardized datasets in deep learning, one often refers to 'ImageNet' as being the dataset that started the revolution. The success of 'ImageNet' led to the creation of additional datasets for image segmentation, speech recognition, video analysis, translation, etc. This touches upon the problem that a single standard dataset is not enough to solve all of the (modeling) problems within the healthcare industry. With data involving a range of situations and representing a variety of informational sources from molecular properties, x-ray scans, patient information cards, gene hierarchies, etc., *there is no one size fits all dataset or data format*. Here we advocate for a way to define and standardize datasets at a higher level that defines the way datasets can be used and shared. This should be done without enforcing rules that are too strict on the particular data format such that it would hinder the willingness of organizations to share the data because it requires investment of too much time and money into reformatting (legacy) data. For a dataset to be useful *at least* the following information should be available:

- **Metadata** such as statistics on the data (size, number of (validated) records, number of missing items/cells, number of categories/labels, experimental background, etc. What is applicable will differ per dataset.

- **File format description and data dictionary**, preferably including example software (e.g in Python, Java, C++, etc.) on how to read the data and ways to decode any binary data that might be included (e.g. custom image formats). This is particularly important when sharing internal datasets that use a custom data format.

- **Underlying distribution of dataset properties** are required to understand the applicability, limitations, and possible biases in the data (see below).

- **Enabling data access APIs** (where applicable) that can abstract complex layers of the data. This will allow data scientists to focus on data science activities instead of sifting through mountains of datasets that are perhaps not applicable to their needs. This will also encourage collaboration while allowing data scientists to repurpose data access solutions without recoding.

- **Uniform way to cite the data**. This will encourage researchers (and companies) to share the data and still get credit for sharing. This also works as a reverse citation so one can get an overview of which (published) works are based on the data. This initiative is already taking place in other fields outside healthcare and AI as well and NIST might want to encourage a global/nationwide standard on how to accommodate this. For example, it would be useful to have well-defined data standards to access healthcare data such as electronic medical records, imaging data, vocal, endoscope, laparoscope, surgery robot, ePROs, Clinical, eDROS, via FHIR, etc. A further example for accessing imaging data is a specification called "The Brain Imaging Data Structure (BIDS) Specification"; however, BIDS is not the standard in the healthcare industry. The federal government can promote and support the use of some of these data standards.

- **An open and readily available platform** where appropriate. This involves making datasets available on an open, non-profit platform, such as Figshare.[2]

Uniform Metrics

When comparing the results of various methods, it is always important to set an equal playing field. For example, when everyone uses the same standard dataset (see above), they should also use the same metric to compare across methods. To make sure that appropriate comparisons are made (apples to apples), we at AAIH are looking to develop a fixed set of evaluation metrics that could be used by all partners and collaborators. These metrics have a mathematical basis and are clearly defined in the literature. For the NIST AI initiative we encourage NIST to consider something similar. Without uniform metrics there is a danger that press releases, promotional talks, year-end reports, etc. will present results in the most positive light rather than in the most significantly comparable and transparent way. This approach carries consequences for public opinions, willingness of investors to invest in and support new businesses, and the quality of research in general.

The exact metrics used will depend on the problem and application at hand. Referring back to the ImageNet network, the metrics are the top 1 and top 5 percentage of accurately labelled images, nevertheless some companies instead focus on the number of images processed. This can be beneficial to a company in the business of making and selling hardware that wants to show how well their hardware performs. If we pull this into the healthcare industry you could look at, for example, the number of properties of a molecule that are properly predicted or the number of molecules processed per second. Whereas the latter is not as useful if you do not maintain the level of accuracy reported before, this type of metric is something that is often omitted in ImageNet performance numbers. Another healthcare example would be the number of novel molecules that are generated by a General Adversarial Networks (GANs). Here the metrics could be the number of requested properties that the novel molecule has or even if the molecule would be a valid and synthesizable molecule. At the same time, it again can be the number of molecules generated per second, a metric which would not be useful if the generated molecules are all invalid.

These examples show that without having clearly defined metrics in place, reporting of results risks being biased to the reporter.

Controlling for Bias and Trustworthiness

There must be standards and processes in place for reviewing and approving AI models in the healthcare industry, which could directly affect people's wellbeing and health. For example, the data used in creating an ML model must comply with data privacy policies in place and the metadata accompanying the model must state the property distribution of the underlying dataset, such as, race, sex, gender, socioeconomic background, education, age, country of origin, physical appearance, etc. While all models should contain data that is diverse enough to represent all segments of the population, this is often not possible. Therefore, the metadata should contain the information

---

[2] Figshare. (2019 Jun). Retrieved from: https://figshare.com/

needed for the proper application of the model avoiding biases. This approach will help to ensure accuracy and trustworthiness.

The target population for AI solutions in healthcare includes but is not limited to race, sex, gender, socioeconomic background, education, age, country of origin, physical appearance, etc., and these variables must be taken into account when developing AI solutions.[3] Historically, many ML models in healthcare have been biased due to the lack of data collected from minorities (women, members of the LGBT community, children, racial minorities, etc.).[4] For example, "Users discovered that Google's photo app, which applies automatic labels to pictures in digital photo albums, was classifying images of black people as gorillas. Google apologized; *it was* unintentional."[5] AI solutions should make reasonable efforts to consider the diversity of the population to which the model will be applied while at the same time respecting clearly delineated parameters determining the dataset. Furthermore, provisions for explainability (i.e. Shapley values, etc.), especially related to black box solutions play a crucial role in both informing solution architects throughout development and clinical decision makers at point of care in order to provide transparency in model results while engendering trust at the human-AI interface.

A collaborative government-industry strategic policy around data standards, validation metrics, and control for bias in AI poses as an important opportunity for public private partnership going forward.

4. Collaboration with FDA on Emerging Standards

Where standards are defined, their application and use in healthcare applications and devices would likely fall under the oversight of the FDA. Ensuring that the FDA is suitably staffed and informed to support the application of AI in healthcare devices, protocols and drugs is necessary. The existing focus of the FDA on AI and healthcare software has been in the realm of medical devices and diagnostics.[6,7] However, we believe we are approaching a time where algorithm outputs may form part of an IND filing augmenting or even replacing some aspects of the IND package. It is possible that within five years some aspects of *in-vitro* testing could be replaced by *in-silico* testing. Eventually, a combination of AI/ML models with emerging experimental methods such as single cell sequencing and emerging pre-clinical models such as organoids may reduce or even remove some elements of *in-vivo testing* in animal models. Further, AI algorithms will have a greater role to play in clinical trial design and real-world evidence. Such approaches could lead to more efficient drug pipelines and reduced costs for drugs.

In order for these AI approaches to be adopted in a timely fashion, ongoing dialogue around precedents with industry to promote understanding on how to navigate within the context of regulatory approvals is necessary, and we look forward to contributing to this process through AAIH's Technology & Standards Development Committee in concert with our Federal Engagement and Regulatory Approval Committee.

Concluding Remarks

In summary, AAIH sees many challenges and opportunities for applying AI to healthcare. The main areas for government agencies, including NIST, and industry partners to collaborate lie in *facilitating access* to healthcare data, *standardizing healthcare datasets* including validation and bias control, *increasing dialogue around challenges* unique to healthcare, and *promoting ongoing industry understanding and dialogue with regulatory agencies* such as the FDA. We at AAIH look forward to further collaboration with NIST and other governmental agencies at this pivotal time in history to provide tools that enable healthier populations both here in the U.S. and beyond.

---

[3] Gershgorn, Dave (2018, Sep 6). If AI is going to be the world's doctor, it needs better textbooks. Retrieved from https://qz.com/1367177/if-ai-is-going-to-be-the-worlds-doctor-it-needs-better-textbooks/

[4] American Heart Association News (2019, Feb 20). Why are black women at such high risk of dying from pregnancy complications?. Retrieved from https://www.heart.org/en/news/2019/02/20/why-are-black-women-at-such-high-risk-of-dying-from-pregnancy-complications

[5] The New York Times (2019, Jun 26). "Artificial Intelligence's White Guy Problem," Retrieved from https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

[6] FDA (2019 Jan). Software Precertification Program. Retrieved from https://tinyurl.com/y4dk7yq2

[7] FDA (2019 Jun). Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD). Retrieved from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device