Machine Intelligence in Healthcare: Perspectives on Trustworthiness, Explainability, Usability and Transparency

July 12, 2019, Neuroscience Center, National Institutes of Health, Bethesda, Maryland

Background:

Machine Intelligence (MI) is rapidly becoming key to biomedical discovery, clinical research, medical diagnostics and devices, and precision medicine. In the context of this meeting, MI is defined as the ability of a trained computer system to provide rational, unbiased guidance to humans in such a way that achieves optimal outcomes in a range of environments and circumstances. MI tools can uncover new possibilities for both physicians and patients, allowing them to make more informed decisions and achieve better medical outcomes. When deployed, these outputs can enhance efficiency at every level of the healthcare system.

The challenges are: how do we trust that what the computer tells us is correct when we don't understand how it arrived at the output/answer? How do we ensure that these outputs are safe and beneficial for human health? And, if we change the data or environment, how does this affect the output? These questions are especially relevant to clinical care decision making – are the risks of using such tools understood and how can the technology be deployed for maximal benefit?

Goal:

To invite the community to share their perspectives with us on current issues associated with incorporation of MI tools into healthcare. Meeting outputs from this workshop will be used to develop a whitepaper on translating MI for clinical applications and the associated process improvement needed when implementing MI tools in healthcare environments.

Expected Meeting Outputs:

1. What lessons have already been learned?
2. Ways we can better stimulate data sharing and open access to training data and MI system development.
3. A framework for identifying and preventing bias in MI healthcare tools.
4. Input on approaches to address quality control and use of standards.
5. Ideas about what tools/protocols are needed to ensure usability of MI systems in multiple environments.
6. Tools and methods for introducing data updates without altering trustworthiness of outputs.
7. Ideas about tools needed for evaluating output reliability and safety.

In the context of this meeting, we will use the following definitions and frame each of the sessions as listed below:

Session 1: Trustworthiness – the ability to accept the validity and reliability of a result, given a change in input or algorithmic parameters, without necessarily knowing why. Healthcare professionals need to be able to determine when a result is wrong (or the probability that a result is wrong) and ensure that the result is interpreted correctly without needing to know what happened "under the hood".

a. What tools can help us to evaluate output validity and reliability?
b. How can we define appropriate benchmarks for the quality of MI outputs?

c. What lessons have we learned from MI systems in other areas in which MI systems are more advanced?

Session 2: Explainability – the ability to understand and evaluate the internal mechanics of a machine or deep learning system in human terms. As these MI systems are being built, additional steps will need to be added in the development process to account for: data quality, metrics for the system's functioning and impact, standards for applications in the healthcare environment, and future updates to the data/system.

a. Is the logic on which the algorithm is based transparent and reproducible?
b. What is the effect of new or different data on the algorithm? How do we introduce "updates" to MI systems without altering outputs?
c. What are qualities of datasets that make them useful for MI development? What do we need to do to establish standards for data quality?

Session 3: Usability – the extent to which an MI system can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in multiple healthcare environments. How useful are these systems going to be to both doctors and patients in multiple settings, particularly when compared to systems and standards of care that are already in place?

a. How do we ensure utility, adaptability and generalizability of MI systems?
b. How do we evaluate the comparative effectiveness of these MI systems against usual care?
c. How can the average person make an informed decision based on MI systems?
d. What is an appropriate output/meaningful metric for clinical MI systems?

Session 4: Transparency and Fairness – the right of a user to know and understand the aspects of a dataset/input that could influence outputs (clinical decision support) made by algorithms. Such factors should be available to the people who use, regulate, and are affected by the systems that employ those algorithms. For MI systems in healthcare, we will need to identify ways of evaluating and preventing bias, encouraging data transparency, and ensuring open access to MI system development, each of which have unique challenges associated with them in the context of healthcare.

a. What data informed the algorithm, did it introduce bias, and how do we control for or prevent bias perpetuation? What are the unique challenges associated with this in healthcare settings?
b. How can we stimulate data sharing and open access during MI system development?