Concept Clearance for Artificial Intelligence for BiomedicaL Excellence (AIBLE)

Program co-Chairs: Patti Brennan (NLM), Eric Green (NHGRI), Bruce Tromberg (NIBIB)

In December of 2018, a Working Group on Artificial Intelligence (AI) was established within the Advisory Council to the NIH Director (ACD) and charged with making recommendations on harnessing the emerging power of AI for the NIH mission. This group delivered their recommendations at the December 2019 ACD meeting. That presentation may be viewed here, beginning at one hour and two minutes into the session.

The Working Group's recommendations were unanimously accepted by the Advisory Council and the NIH Director. The Office of Strategic Coordination (OSC) and Office of Data Science Strategy (ODSS) were then charged with implementing them. OSC has coordinated NHGRI, NIBIB, NLM, and other interested ICOs in the development of the program concept "Artificial Intelligence for BiomedicaL Excellence (AIBLE)". That concept is presented here and will be discussed on May 15th. Sustained workforce development/training initiatives fit within the ODSS mandate and may be addressed by ODSS programs.

The ACD recommended that NIH invest in the generation of new biomedically relevant data sets amenable to analysis using machine learning methods. Their key insight was that many data sets that are typical of biomedical research, including almost all large data sets currently in existence, are *not* amenable to machine learning analysis at scale for reasons including but not limited to size, provenance, content, or accessibility. The Working Group argued that by defining a question and collecting the data expected to answer it (arguably the essence of scientific research), traditional researchers make assumptions and choices that ultimately limit the ability of machine learning tools to generate surprising insights. For this reason, the AIBLE program starts from the principle that we need to intentionally design a different approach. The program concept dedicates time and resources to articulating ethically and analytically sound design principles and building data-handling tools to enable AI, before proceeding to data generation. The plan includes five interdependent initiatives with staggered start times:

- Initiative 1 | Data design centers | *Convert ML-friendliness attributes into rubrics and standards that allow planning and evaluation.*
- Initiative 2 | Tools to accelerate generation of AI-ready data | *Create software and hardware to speed annotation and structuring.*
- Initiative 3 | Instant interactive testbeds for Initiatives 1 and 2 | *Immediately initiate collaboration with existing projects.*
- Initiative 4 | Gold data generation | *Generate large multimodal, metadata-complete, available data that exemplify ML-friendliness.*
- Initiative 5 | Evaluate and update | *Use the rubrics to assess and improve select public data sets of biomedical importance.*

The NIH working group will programmatically balance data types to ensure that many subdisciplines advance together. In this way the AIBLE program will generate standards, tools, and data that will be of lasting broad value to biomedical research.