U.S. Department of Health and Human Services

National Institutes of Health (NIH)

Office of the Director

Division of Program Coordination, Planning, and Strategic Initiatives

Bridge to Artificial Intelligence (Bridge2AI) Grand Challenge Team Building Expo

June 23, 2021

Summary
Discussion Points, Highlights, and Action Items

I. Welcome

Jocelyn Tejeda, facilitator, Knowinnovation (KI), welcomed the participants and provided an overview of the Bridge2AI Grand Challenge Team Building Expo agenda.

Grace C. Y. Peng, Ph.D., Program Director, Division of Discovery Science and Technology (Bioengineering), National Institute of Biomedical Imaging and Bioengineering, welcomed participants and presented a review of the full Bridge2AI event prior to introducing the final meeting for the event, the Grand Challenge Team Building Expo. She explained that the purpose of the Team Building Expo is to create networking activities around the Grand Challenges. Grand Challenges respond to big biomedical and behavioral research problems that cross research domains; that require new, complex data sets and cannot be addressed with existing data sets; are hypothesis-agnostic; and employ diversity to explore ethical issues. Some Grand Challenge themes illustrated by NIH Institute director co-chairs of the program included digital twins, artificial intelligence (AI) and machine learning (ML) in clinical care, functional genomics, movement phenotyping, precision public health, and salutogenesis, but applicants are welcome to propose other grand challenge ideas.

- Many Grand Challenge ideas were discussed during Microlabs 1, 2, and 3. Dr. Peng emphasized that Grand Challenge ideas submitted for presentation and discussion during this Expo are expected to only start conversations. Applications will not be limited to these ideas. Grand Challenges that are not presented can still be developed. NIH will not assess responsiveness during Bridge2AI team building activities.
- Dr. Peng encouraged biomedical and behavioral researchers to lead a Grand Challenge or collaborate on multiple Grand Challenges. She encouraged all participants to offer their expertise in developing tools, addressing ethical issues, establishing standards, engaging data acquisition processes, team building, and developing workforce skills and to participate in multiple modules across multiple Grand Challenges as appropriate.
- Grand Challenge teams should demonstrate equal partnerships among diverse data and biomedical/behavioral scientists.
- Grand Challenge ideas and their modules should be generalizable across other Grand Challenge projects. Once funding decisions are made, team members might be placed in a Grand Challenge that they did not propose because of the generalizability of the individual modules. The Bridge2AI Integration, Dissemination, and Evaluation (BRIDGE) Center will coordinate across the multiple

Grand Challenge projects and integrate, disseminate, and evaluate Grand Challenge projects and the resulting ethically-sourced data.

- Dr. Peng presented NIH Bridge2AI coordinator responses to questions most commonly asked by participants.
    - Foreign applications are allowed from Data Generation Projects, and the BRIDGE Center can have foreign components.
    - Bridge2AI funding guidelines supersede all other guidelines and instructions.
    - Applicants have the next 2 months to continue to form teams, finalize ideas, and develop proposed project budgets (in an editable spreadsheet, organized by task).
    - Bridge2AI is an NIH Common Fund initiative, rather than an individual NIH Institute initiative (i.e., there is no need to find an institute home for a project).
    - After a review of OT2 proposals, NIH has the ability to combine modules; adjust milestones, timelines, and tasks; and conduct interactive pairing processes to form final teams.
    - Two funding opportunities are available as part of the Bridge2AI initiative: (1) the Data Generation Research Opportunities Announcement (ROA), and (2) the BRIDGE Center U54.
- Participants also asked about the use of new versus existing data in Grand Challenge applications. Shurjo K. Sen, Ph.D., Program Director, Division of Genome Sciences, National Human Genome Research Institute, explained that many participants had ideas for Grand Challenge projects that would involve working on existing data. Dr. Sen recognized that existing and new data cannot be completely separated when creating AI databases and existing data could be used to improve the quality or completeness of new data sets or to create simulations. NIH prioritizes the creation of new AI-ready or ML-amenable data sets for the Bridge2AI Data Generation Grand Challenges, so applications focused on existing data sets likely would not do well in review. When considering their Grand Challenge ideas, potential applicants should consider whether the bulk of the project is focused on creating new data.
- Dr. Sen encouraged participants to continue asking questions via Slack or email.
- Dr. Peng delivered a "Pop-up" questions quiz that allowed participants to test their knowledge about the Bridge2AI Grand Challenge. She presented the following responses to the pop-up questions:
    - The spirit of the Bridge2AI program is to instill a culture of ethical inquiry from diverse perspectives and generate new data that apply FAIR principles (Findability, Accessibility, Interoperability, and Reuse).
    - Grand Challenge ideas should be much larger in scale than a typical R01.
    - Bridge2AI will not create new AI models to solve biomedical/behavioral Grand Challenges. The purpose of the Grand Challenge projects will be to create new data. After the Bridge2AI Consortium is formed, NIH will invite AI modelers to assess and evaluate these data using AI models to extract knowledge for scientific discovery (e.g., during annual jamborees with the Consortium).
    - Bridge2AI products are considered generalizable, so applicants might end up being awarded to work on a Grand Challenge proposed by a different team. During a post-review negotiation phase, applicants might be put on a different team to study a different Grand Challenge.
    - Bridge2AI Grand Challenge ideas should have all of the following characteristics: (1) require AI/ML-friendly data that do not currently exist to answer the question; (2) respond to a question that cannot be answered using existing non-AI analytical tools; (3) use data sets that

incorporate some existing data streams (e.g., surveillance or electronic medical record data); and (4) incorporate ethical challenges.

- Dr. Peng explained that an expanded version of this quiz will be available on KIStorm to allow participants to further test their knowledge about Bridge2AI.
- Dr. Peng thanked participants and reminded them to use breakout discussions to develop and build on any Grand Challenge ideas they have and connect with potential new team members. Participants who are interested in a Grand Challenge idea presented during the Expo should attend the breakout session for that idea to learn more about it, ask questions, and share their expertise with the person(s) presenting the idea. Participants who are unclear about their role in Bridge2AI should attend various breakout sessions, ask questions, and consider joining a team, forming a team around a new idea, or developing a BRIDGE Center application instead.

II. Grand Challenge Presentations

Ms. Tejeda introduced the Grand Challenge presentations. One member of the current team developing the Grand Challenge idea delivered a brief presentation of that idea. A total of 24 Grand Challenge ideas were presented during this session.

- Pinaki Sarder, Ph.D., from the University at Buffalo presented an idea titled "A Multi-scale, Multi-modal Approach to Data Annotation in Renal Pathology Informatics." He proposed developing databases of slide images of kidney tissues linked to other types of data (e.g., genetic, serological, molecular pathology, transcriptomic, and clinical pathologic, treatment, outcome) from populations worldwide. The project would develop a distributed cloud-based annotation technology through which human annotators could interact with AI and other human annotators in an active learning framework to generate seamless, large-scale annotated renal tissue image data, which will be AI-ready and can be used to answer a wide range of research questions.
- Les Sztandera, Ph.D., from Thomas Jefferson University presented an idea titled "Exploration of Additive, Antagonistic, and Synergistic Effects of Heavy Metals in Pediatric Health." He proposed developing new data sets for infants born at 14 hospitals. The data sets would include information on maternal age, race/ethnicity, insurance, ZIP code for geolocation, body mass index, medical morbidities, pregnancy complications, maternal medical morbidities, duration of hospital stay, and other neonatal outcomes. The project also would develop a diverse database of blood samples from mothers and infant umbilical cords to identify toxic elements and their potential transfer and examine mother and newborn nutritional status.
- Jiaqi Gong, Ph.D., from The University of Alabama presented a Grand Challenge on a "Movement Phenotyping Study for Generating AI/ML-Friendly Datasets." A unified understanding is needed of the basic mechanisms of movement signatures across a broad range of factors, such as neural and musculoskeletal drivers, clinical conditions, the complexities of human development and aging, and signifiers of physical activity for health and wellness. The goal of Dr. Gong's proposed project would be to systematically develop rigorous, mechanism-focused methods and generate hypothesis-agnostic, ethically sourced, trustworthy, shareable, and AI/ML-friendly data sets via a more unified science of movement phenotyping.
- Homayoun Valafar, Ph.D., from the University of South Carolina presented the idea of "Human Activity Monitoring Using Wearable and Smart Devices." He proposed generating data on what occurs during the healthy state and the transition from a healthy to an unhealthy state, which is likely to benefit human health care more than data on the end-stage of disease. Research on daily human health behaviors (e.g., physical activity, sleep) will require a comprehensive collection of sensor data

(e.g., accelerometer, gyroscope) that the project could acquire from commercially available, off-the-shelf wearable health devices (e.g., watch, phone, ring).

- Vicki Hertzberg, Ph.D., from the Nell Hodgson Woodruff School of Nursing at Emory University presented "Health Care Records from Womb to Grave." In many areas, pediatric care is distinctive from adult care. Dr. Hertzberg proposed a project that would generate new data sets that integrate maternal and paternal electronic health records (EHRs) with pediatric records and that integrate individuals' pediatric records with their adult records when they age out of the pediatric health care system. AI would be required to link pediatric records to parental and adult records.
- Joel Berendzen, Ph.D., from The University of New Mexico presented an idea titled "Biofield in Global Public Health." The proposed project would provide data to answer the question, how does the distribution of viruses and virus-like particles influence public health? The project would generate new data sets, collecting 10 to 80 terabytes of DNA sequence data from 16,000 filtered stool samples from individuals, families, schools, and clinics at 2,000 global locations (ideally, employing citizen science that would allow data collection to continue after the project ends).
- Ebenezer Tolman of Tufts Medical Center presented the "Augmented Physiological Signal Network" Grand Challenge idea. The proposed project would generate invasively sourced perioperative high-frequency, multimodal physiological waveform data obtained from multiple cardiac catheterization laboratories, creating source data for AI to build a variability metric (VM) as well as a feature-derived index registering autonomic function, cardiac response, and vascular compliance. The VM would allow the public to examine their health status relative to an age- and gender-matched cohort, thus generating data from healthy individuals.
- Nathan Hotaling, Ph.D., of the National Center for Advancing Translational Sciences presented the idea titled, "Human Organoid Image Atlas Microscopy." Currently, no reference database in microscopy (light, fluorescent, or electron) exists for training or benchmarking AI algorithms. The proposed project would develop an exabyte-scale 2-D and 3-D image data set of diverse human cells (healthy and pathological) taken from organoids (to observe cell types from a variety of tissues in the human body and perform native confirmations) treated with approved drugs and derived from induced pluripotent stem cells from a diverse patient population (i.e., all people, all organs, all drugs) to examine morphology.
- James Glazier, Ph.D., of Indiana University presented "Bridge2AI to Digital Twins of the Immune System." Infectious disease, cancer, autoimmune diseases, transplant rejection, and sepsis all result from pathological behaviors of the immune system. Dr. Glazier emphasized the need for data sets to enable digital twins to predict individual-to-individual rapid dynamic changes in immune state. Current clinical and animal model data sets are inadequate. Dr. Glazier proposed a longitudinal data set to develop a digital twin for predicting changes in the immune state. He indicated that he was interested in brainstorming ideas for ways to collect the data needed to develop the digital twins for the immune system.
- Jessica Binder, Ph.D., from The University of New Mexico presented "Molecular Basis of Salutogenesis." Dr. Binder raised the possibility of identifying diseases associated with specific proteins through a project that would examine the molecular pathways that are engaged in healing. Dr. Binder proposed tracking 500 young athletes over 4 years of performance, injury, and healing to study this question by collecting proteomics, motion-capture, EHRs, microbiome, and performance data as well as patient-reported outcome diaries.
- Faramarz Valafar, Ph.D., from San Diego State University presented a functional genomics project to generate data needed to support AI applications for a wide range of diseases titled "From Specimen to Comprehensive -Omics Data for AI Applications in Broad Disease Domains." He proposed developing -omics data sets using the same tissue sample for each domain collected at the same stage

without amplification or culturing, which inject bias. These data would be used to develop standardized protocols for sequencing and collecting -omics data directly from clinical specimens.

- Zeeshan Ahmed, Ph.D., from Rutgers Institute for Health, Health Care Policy and Aging Research presented "A FAIR Approach for AI/ML-Ready Genomics Data Acquisition." The proposed project would develop a personalized approach to improve traditional symptom-driven medical practice with disease-causal genetic variant discovery. The project would collect sequenced data of diagnosed cases and controls from publicly available data sets, such as the Database of Genotypes and Phenotypes, The Cancer Genome Atlas, Sequence Read Archive, and European Nucleotide Archive; it also would generate its own data set.
- Bruce Y. Lee, M.D., of the City University of New York presented "Obesity Prevention and Treatment Worldwide." The project would tackle the challenge of understanding the worldwide obesity epidemic by developing data sets on genetics, dietary intake, physical activity, change in obesity/overweight prevalence over time, and health outcomes over time, in addition to a range of factors that have been linked to obesity from the molecular level up to the sociopsychological level.
- Oded Regev, Ph.D., from New York University presented "Regulation of Gene Expression at the RNA Level." The proposed project would examine how variants affect gene expression at the RNA level in different cell types. The project would employ synthetic test sequences in primary cells to prove RNA processing and generate larger, higher-quality AI-ready data sets than those generated in the past.
- Jake Chen, Ph.D., from The University of Alabama at Birmingham presented "'ImageNet' for Deep Learning Applications in Biology and Medicine." He proposed using an existing standardized ontology to annotate the more than 5 million samples in the ImageNet database. The project would transform public functional genomic data into disease ontology or network-organized standard representations of GeneTerrain objects in a large GeneTerrain KnowledgeBase to make it AI-ready for numerous applications.
- Raghu Machiraju, Ph.D., from The Ohio State University presented "DIY Frameworks for Ubiquitous, ML-Based Model Commons for Outcome-Driven Studies Grand Challenge" on behalf of a team from Northwestern and The Ohio State universities. The proposed project would foster communities of practice to focus on both rare and less-characterized cancers and provide a means to assemble *in silico* outcome-driven studies to address specific questions. The project also would allow resource-poor laboratories to access a network of ML-based model commons for data analysis.
- Cody Bumgardener, Ph.D., from the University of Kentucky presented "Foundations Supporting Self-Learning and Clinically Validatable AI." He proposed a project that would bridge transdisciplinary data sets in support of non-hypothesis-driven discovery (AI-to-AI) systems and direct clinical use of AI through targeted models developed as laboratory tests, which would make human-actionable data computationally actionable. The project would build metadata in targeted areas, including pathology, laboratory medicine, and radiology data sets obtained from clinical workflows.
- Julia Perederiy, Ph.D., from Om Biome, Inc. presented "Healing from Cancer Treatments (Salutogenesis): A Multidimensional, Repeated-Measures Analysis of Oral/Gut/Neuro Complications." She proposed a project that would create biomarker profiles for tissue regeneration by combining existing data from metabolic and oncology blood panels; quality of life and behavioral assessment data; and new biomarker data on the immune response, stem cells, and microbiota (representing the microenvironment that underlies most tissue regeneration processes). Biospecimens would be collected from patients before, during, and after treatment, as well as from healthy controls, to examine the chronic conditions that might drive adverse effects after cancer treatment.

- Joseph Brady, M.S., from the University of Denver presented "The Effect of Complementary and Integrative Medicine Approaches to Salutogenesis in Community Settings." To make health promotion programs developed in research settings effective in real-world settings, Dr. Brady proposed obtaining data from health promotion programs and community clinics, as well as patient-reported outcomes on global health risks and changes in health behavior. The project would focus on both underserved and general populations, with a focus on healthy populations to better understand what inspires people to engage in healthy behavior.
- Xiao Hu, Ph.D., of the Duke University School of Nursing presented "Patient CHoRUS for Equitable AI." The proposed project would create a new data set that responds to the limitations of current EHR data sets by adapting methods available across acute care settings to capture more data across a wider geographic area while ensuring that formatting and metadata are standardized. The project would acquire both EHR and high-resolution physiological data from bedside devices at multiple centers, then harmonize and annotate these data to generate a large flagship data set with unprecedented diversity for clinical AI/ML solutions focused on improving recovery from acute illness.
- William Hersh, M.D., of Oregon Health & Science University presented "Translational AI for Better Health and Better Care: From Bench to Bedside to Community." He proposed a project that would move AI out of basic research and into translational science. The new data set required for such a project would be scalable and based on clinical problems with specific randomized controlled trials aiming to use the full spectrum of data from EHRs, other clinical data, and patient-generated and public health data.
- Jun Deng, Ph.D., of Yale University presented "Developing Digital Twins for Predictive Personal and Public Health." He proposed developing high-performance and AI-empowered digital twins computational frameworks that learn from multiscale, multimodality, real-world data collected in real-time. This project would require linked clinical, mobile health, molecular profile, social media, environmental, lifestyle, dietary, family history, and medication data at both spatial and temporal scales.
- Tim Haithcoat at the University of Missouri presented "Geospatially Enabled Convergence Health." The proposed idea would involve the creation of a geospatially enabled big data table and the quantification of context to understand health challenges. The proposed project would build a robust, unbiased, locationally integrated social, environmental, health care, and infrastructural data warehouse and knowledge base.
- Jeffery Lotz, Ph.D., at the University of California, San Francisco presented "Chronic Low Back Pain (cLBP)." He explained that this Grand Challenge project would focus on chronic low back pain and take advantage of the metrics, protocols, and other data collection infrastructure developed through the NIH Back Pain Research Consortium Research Program (BACPAC). The proposed project would use and develop new and extended deep phenotyping measures concurrent with and following BACPAC Phases 1 and 2 and expand the data collection efforts to the international level.

After the 24 Grand Challenge idea presentations, Ms. Tejeda asked participants to go to the Expo tab on KIStorm page to select and sign up for no more than three breakout sessions on the Grand Challenge ideas that interested them. The sign-up allowed the presenters to know approximately how many people would attend their sessions. The attendees were welcome to move around between breakout rooms.

III. Grand Challenges Breakout Sessions

- Three rounds of breakout sessions were held to allow participants to engage in full discussions of a maximum of three different Grand Challenge ideas. The first breakout session was the only in which all 24 Grand Challenge idea presenters/teams were available to discuss their ideas. Breakout

discussions were held for only 22 of the 24 Grand Challenge ideas presented during the second and third breakout sessions. KI staff set up a separate Zoom call for each Grand Challenge idea during each of the three breakout sessions. Each breakout session lasted approximately 45 minutes.

- NIH Bridge2AI coordinators were available for anyone who had questions for NIH. Presenters leading a Grand Challenge discussion were allowed to continue that discussion into the second and third breakout sessions, but KI staff prompted discussion leaders to announce when the next session was starting, and new people might be joining.

IV. Theme Breakout Sessions

Lanay M. Mudd, Ph.D., Program Director, Division of Extramural Research, National Center for Complementary and Integrative Health, explained the purpose of this portion of the Expo, which was dedicated to breakout sessions to discuss thematic areas of the Grand Challenge ideas. The thematic areas for the Grand Challenge ideas included digital twins, AI/ML in clinical care, functional genomics, movement phenotyping, precision public health, and salutogenesis, and "other ideas." Participants with new Grand Challenge ideas were encouraged to join a breakout room for the thematic area that most closely aligned with their idea. Other participants were encouraged join the breakout room that most interested them although they also could visit different breakout rooms. The breakout sessions were not moderated, and the participants were asked to conduct the sessions in the manner they thought would be productive. Dr. Mudd suggested asking all breakout participants to share their ideas and then examine the similarities and dissimilarities of the ideas within the theme.

Andy Burnett, Managing Director, KI, explained how participants could submit new ideas using digital sticky notes. Ms. Tejeda added that participants should only submit one idea or thought per sticky note.

V. Action Planning

In closing, Dr. Mudd remarked that many connections had been made during the meeting and several Grand Challenge ideas had been expanded. She reinforced that the team building and idea generation processes do not end with this meeting, which was designed to springboard participants into further discussions with colleagues. Participants will be able to continue accessing the KIStorm platform, Slack channels, and Wonder platform to connect with colleagues, refine ideas, and build teams. Dr. Mudd encouraged participants to make use of Bridge2AI resources, including the Bridge2AI listserv where members will receive updates, and noted that interested individuals can sign up for the listserv through the NIH Common Fund website. The Common Fund website also includes Bridge2AI FAQs, instructions for applying for Bridge2AI funding opportunities, and resources for the Plan for Enhancing Diverse Perspectives (PEDP). The Plan should be a key component of both applications for Data Generation Grand Challenge projects and the BRIDGE Center. Dr. Mudd also referred participants to a Bridge2AI wiki page, where evolving Bridge2AI resources are posted, such as a glossary of terms, references, and publications.

- New resources will be available on KIStorm, including a survey for feedback on June Team Building Activities, an expanded quiz to test knowledge about Bridge2AI, and an office hour sign-up for teams and individuals to meet with NIH program staff to discuss Data Generation Projects or BRIDGE Center ideas in more detail. NIH staff cannot support or reject ideas during these discussions but will provide a consultation during which they might ask probing questions or suggest expanding or further developing the idea. Individuals who sign up for consultations with NIH staff should submit a description of topics they would like to discuss prior to the consultation.

- Dr. Mudd presented a timeline up through the Bridge2AI application due date on August 20, 2021. Letters of intent are due July 20, 2021. These letters are required for the Data Generation Projects ROA and recommended for BRIDGE Center U54 applications. NIH will not provide feedback on letters of intent. Instructions for the letters of intent are included in the funding announcements. Letters of intent for ROA and U54 applications differ somewhat but both require a proposal title and contact information for the principal investigator and key personnel. Letters of intent for Grand Challenge ROA applications also must include a description of the Grand Challenge and the data sets that will be generated. These letters should be sent to the Bridge2AI email address.

Dr. Mudd reminded participants of the overall goals and structure of Bridge2AI. NIH wants to avoid silos by bridging different fields and backgrounds to create new flagship data sets motivated by the Grand Challenge concept. The project will generate AI/ML-friendly data by developing software to standardize data attributes from multiple sources across data types, creating automated tools to describe those data sets, producing cross-training materials for workforce development, and disseminating products and best practices. Ethical considerations will be a core part of the project.

Projects must include all components covered in the Microlabs, including teaming, ethics, standards, tools, skills and workforce development, and data acquisition. Participants interested in building infrastructure that could support all Data Generation Projects should consider submitting a U54 Core BRIDGE Center application. BRIDGE Center applications must include an administrative core plus a core for at least one of the following: team building, ethics, standards, tools, and/or skills and workforce development. BRIDGE Center applications might be funded to work together as a single Bridge Center that includes all of the cores. The purpose of the BRIDGE Center is to integrate, disseminate, and evaluate activities, knowledge, and practices across all Data Generation Projects using input from external stakeholders with the goal of producing best practices for using AI/ML in biomedical and behavioral research to address Grand Challenges.

Products of Bridge2AI will include novel, quality data sets that follow FAIR principles and tools that accelerate the creation of these data sets and foster a culture change for the research community to embrace data preparation for AI/ML analysis across disciplines.

Dr. Mudd asked other NIH colleagues to add closing thoughts. Dr. Sen remarked that diversity in viewpoints, applicants, and types of institutions is an important goal of Bridge2AI. Dr. Peng encouraged participants to view the video entitled “Other Components”, which covers the Plan to Enhance Diverse Perspectives. She emphasized the importance of diversity, ethics, people, and data in the Grand Challenge Projects. James Gao, Ph.D., Division of Extramural Science Programs, National Eye Institute, emphasized the need for data and tool (hardware, software, firmware) sharing, particularly for the Data Acquisition module. Applicants must provide a timeline and milestones for data and tool sharing and sharing must begin in 2022. Dr. Peng added that the tools must be reusable.

VI. Adjournment

Dr. Mudd thanked the participants and encouraged them to continue building teams and using the resources described and stated that NIH looks forward to receiving applications. All NIH staff and Ms. Tejada thanked participants and adjourned the meeting.

VII. Action Items

Dr. Peng explained that NIH will—

- Obtain an updated list of people who have signed up for the Bridge2AI listserv.

Mr. Burnett and Dr. Mudd added that KI will—

- Examine the reason for the misnumbering of Grand Challenge ideas on the KIStorm Bridge2AI website and correct the problem.
- Reactivate the original Grand Challenge page after this meeting so that people can continue to add ideas. Dr. Mudd suggested that participants review the questions/answers that will continue to be posted on the KIStorm page.
- Make archived videos available on the KIStorm platform.