

Data Generation Projects for the NIH Bridge to Artificial Intelligence (Bridge2AI) Program (OT2)

Other Transaction Opportunity Announcement

Overview Information

Participating Organization(s)	National Institutes of Health
Components of Participating Organizations	This Research Opportunity Announcement (ROA) is developed as a Common Fund initiative (https://commonfund.nih.gov/) through the NIH Office of the Director, Office of Strategic Coordination (OD-OSC). All NIH Institutes and Centers participate in Common Fund initiatives. The ROA will be administered by OD-OSC on behalf of the NIH.
Research Opportunity Announcement Title	Data Generation Projects for the NIH Bridge to Artificial Intelligence (Bridge2AI) Program (OT2)
Activity Code	<p>OT Other Transactions (OT2)</p> <p>This Funding Opportunity will use the Other Transactions Authority governed by 42 U.S. Code § 282 (n)(1)(b). Other Transactions (OT) are not grants, cooperative agreements, or contracts. They are used by the NIH to provide considerable flexibility in establishing policies for the awards; policies and terms for individual OT awards may vary between awards, each negotiated with a specific agreement, which may be expanded, modified, partnered, not supported, or later discontinued based on program needs, changing research landscape and or availability of funds.</p> <p>The Bridge2AI program is using the OT2 activity code to combine Modules from different Data Generation Project proposals to be selected and funded to work collaboratively, to achieve the goals of the Bridge2AI program (see Funds Available below).</p>
Research Opportunity Announcement (ROA) Number	OTA-21-008
Related Notice	Notice of Intent to Publish a Funding Opportunity Announcement for Research Opportunity Announcement for the Data Generation Projects of the NIH Bridge to Artificial Intelligence (Bridge2AI) Program (OT2), NOT-RM-21-022

<p>Research Opportunity Purpose</p>	<p>This Research Opportunity Announcement (ROA) invites proposals to the NIH Bridge to Artificial Intelligence (Bridge2AI) Program, soliciting Data Generation Projects to create flagship datasets based on ethical principles, associated standards and tools, and skills and workforce development to address biomedical and behavioral research grand challenges that require artificial intelligence and machine-learning (AI/ML) analysis.</p>
<p>Objective Review</p>	<p>NIH will engage appropriate expertise to evaluate Bridge2AI Data Generation Projects. See the Objective Review section of this opportunity for further details.</p>
<p>Eligibility</p>	<p>See the Eligibility section of this opportunity.</p>
<p>Funds Available and Anticipated Number of Awards</p>	<p>The planned budget for this effort is approximately \$96 million over a 4-year period. NIH anticipates making 5-8 data generation awards. However, NIH Common Fund procedures and OT mechanisms allow for significant flexibilities and adjustments as may be necessary to pursue catalytic and transformative initiatives. Award levels may increase or decrease over time based on programmatic needs, funding availability, and recipient performance. Each proposed Module may be combined with Modules from other data generation projects in the final awards to create the Bridge2AI Consortium.</p> <p>In addition to the due date given below, additional due dates may be made available in the future and will be announced by posting on the program website.</p>
<p>Award Project Duration</p>	<p>Initial project duration is anticipated to be four years, but individual projects may be extended or curtailed based on programmatic objectives, performance, and available funds.</p>
<p>Research Module Submission Instructions</p>	<p>All Bridge2AI Data Generation Project proposals must be submitted via the NIH eRA ASSIST System by 5:00 p.m. local time on the due date (see Key Dates below).</p> <p>To submit a proposal via ASSIST, the proposer organization must be registered in eRA Commons (see Submission Instructions). Organizations already registered in eRA Commons do not need to register. Once the organization is registered, the individual(s) with the roles of Authorized Organizational Representative (AOR) and Principal Investigator must be affiliated with the organization and have eRA Commons credentials to complete the submission process.</p> <p>Complete proposals must be submitted via ASSIST by the Authorized Organizational Representative. Use OTA-21-008 in</p>

	the field requesting Funding Opportunity Announcement. Here are instructions for submitting via the NIH eRA ASSIST system . Technical help is available at the eRA Service Desk .
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Key Dates

Release Date of this Funding Opportunity Announcement	May 13, 2021
Grand Challenge Team Building Activities	June 1, 2021 to July 20, 2021 To facilitate team building across communities and ensure responsiveness of proposals, NIH strongly encourages potential proposers to participate in the Grand Challenge Team Building Activities, according to the due dates listed in the Proposal Timeline . Detailed Instructions about the Bridge2AI Grand Challenge Team Building Activities, including informational webinars, will be posted on the Bridge2AI website, http://www.commonfund.nih.gov/bridge2ai .
Data Generation Project Letter of Intent (LOI) Due Date	July 20, 2021 The Data Generation Project Letter of Intent (LOI) is required. The LOI must be emailed to bridge2ai@od.nih.gov by 11:59 PM ET on or before the due date by the institution's Authorized Organization Representative/Signing Official. The contact Principal Investigator and other relevant institutional officials must be cc'd. See Submission Instructions below.
Data Generation Project Full Proposal Due Date	August 20, 2021 Full proposals must be submitted via ASSIST (see How to Apply section).
Earliest Start Date	February 1, 2022
Kickoff Meeting	A Bridge2AI Program kickoff meeting will be held in February 2022 in Bethesda, MD.

Agency Contacts

NIH encourages inquiries concerning this announcement and welcomes the opportunity to answer questions from potential proposers.

Scientific/Research Contact(s):	Bridge2AI program staff will respond to inquiries sent to this email: bridge2AI@od.nih.gov
Financial/Agreements Officer Contact(s):	Linda Shariati - Division of Program Coordination, Planning, and Strategic Initiatives, NIH Office of the Director (OD); Common Fund Telephone: 301.402.4113 Email: linda.shariati@nih.gov

Sections in this ROA

OVERVIEW INFORMATION	1
KEY DATES	3
AGENCY CONTACTS	3
SECTIONS IN THIS ROA	4
THE BRIDGE2AI PROGRAM	5
THE DATA GENERATION PROJECTS (THIS OPPORTUNITY)	7
<i>Teaming Module</i>	<i>10</i>
<i>Ethical and Trustworthy AI/ML Module</i>	<i>11</i>
<i>Standards Module</i>	<i>12</i>
<i>Tool Development and Optimization Module</i>	<i>13</i>
<i>Data Acquisition Module</i>	<i>14</i>
<i>Skills and Workforce Development Module</i>	<i>15</i>
HOW TO APPLY	15
PROPOSAL TIMELINE	15
ELIGIBILITY.....	16
GRAND CHALLENGE TEAM BUILDING	17
LETTER OF INTENT (LOI)	18
PROPOSAL SUBMISSION INSTRUCTIONS	18
FULL PROPOSAL FORMAT	19
RESEARCH PLAN	21
OTHER ATTACHMENTS:.....	25
OBJECTIVE REVIEW	27
REVIEW CRITERIA.....	29
SPECIAL AWARD TERMS AND INFORMATION	31

The Bridge2AI Program

Purpose

The Bridge2AI Program seeks to bridge the biomedical and behavioral research communities with the rapidly growing community of experts developing AI/ML models by producing flagship datasets that adhere to the [FAIR principles](#) (Findable, Accessible, Interoperable, Reproducible) and critically integrate ethical considerations in preparing data for computation. The program will use biomedical and behavioral research grand challenges (see illustrative list below) to drive the development of ethics, standards, tools, data sets, and skills and workforce development strategies for linking scientific workflows, protocols, and other information about the data collection process into computable knowledge. Data sets may be linked to existing clinical, environmental and surveillance data as required by the chosen grand challenges. The overall goal of the Bridge2AI Program is to generate flagship datasets and best practices for the collection and preparation of AI/ML-ready data to address biomedical and behavioral research grand challenges. The Bridge2AI program will require multiple scientific domains to come together with data science, data management and analytic experts to unlock the potential of AI/ML for the scientific community.

Primary Objectives

Planned objectives of the Bridge2AI program include: 1) New biomedical and behavioral datasets, which are ethically sourced, trustworthy, well-defined and accessible; 2) Software to standardize data attributes across multiple data sources and across data types (establishing new standards as needed); 3) Automated tools to assist the creation of FAIR and ethically sourced datasets (e.g.: through intelligent workflows, sensorized instruments, etc.); 4) Resources to disseminate data, ethical guidance, tools and best practices; and, 5) Cross-training materials and activities for workforce development that bridges the AI/ML and biomedical/behavioral research communities. Completion of these objectives will establish a culture for rigorously generating FAIR, credible, ethical and generalizable data which will enable AI/ML methods to address key biomedical and behavioral grand challenge problems.

Background

Rapid advancements in AI/ML analysis in recent years show promise in generating new biomedical and behavioral research insights (that are often surprising). In 2018, [members of the AI/ML and computational biomedicine research communities](#), and in 2019, the [AI Working Group of the Advisory Committee to the NIH Director](#), presented visions to propel progress in biomedical and behavioral research through AI/ML analysis. These groups encouraged the NIH to stimulate widespread adoption of AI/ML models that can tackle cross-cutting biomedical and behavioral research grand challenges, with the following recommendations:

1. Support flagship data generation efforts to propel progress by the scientific community.
2. Develop and publish criteria for AI/ML-friendly data sets.
3. Design and apply “datasheets” and “model cards” for biomedical AI/ML.
4. Develop and publish consent and data access standards for biomedical AI/ML.
5. Publish ethical principles for the use of AI/ML in biomedicine.
6. Develop curricula to attract and train AI/ML-BioMed experts.
7. Expand the pilot for AI/ML-focused trainees and fellows.
8. Convene cross-disciplinary collaborators.

The AI Working Group of the Advisory Committee strongly emphasized the need for the NIH to generate

flagship (AI/ML-friendly) data sets that can be used by experts in the AI/ML community. To truly harness the emerging power of AI/ML models for human health, NIH must begin with an “analysis-first” approach which values the machine-understandability of data and incorporates ethical principles surrounding data collection and use.

Currently, the classical instantiation of the scientific discovery process dominates the biomedical and behavioral research enterprise. It begins with models of varying sorts generating hypotheses that are tested in experimental studies. Research protocols are designed, data are collected from a variety of instruments and subsequently analyzed to address defined hypotheses, and new knowledge is reported based on human interpretation of the analyses. While useful, this process does not generate data that is maximally accessible to new methods in AI/ML modeling. By defining a question and collecting the data expected to answer it, researchers typically make assumptions and choices that ultimately limit the ability of analytic tools to generate surprising insights.

Machines (computers) need to “understand” or capture the conditions in which the data are collected and the questions the data are intended to answer. Modern AI/ML models place priority on the analysis and the data requirements. All possible information about the data generation process must be captured, including human assumptions and inferences. Typical data sets generated from biomedical and behavioral research, including almost all large data sets currently in existence, do not include this level of detail and are *not* amenable to AI/ML analysis at scale. In addition, variability exists between scientific study designs, data sources, expert labels, and data processing methods. Moreover, uncertainties, biases and errors occur due to ethical issues, hardware, software, and operator attributes. The systematic translation of this highly heterogeneous information into a format that is understandable by machines will maximize the potential of new AI/ML methods to make scientific discoveries in health.

Close interdisciplinary collaboration will be key to this endeavor. The Bridge2AI program is dedicated to bringing together biomedical and behavioral science communities with AI/ML data science experts to incorporate FAIR and ethical principles into data-handling tools for generating data sets that will transform science through AI/ML models.

Structure

The Bridge2AI program will support several interdisciplinary Data Generation Projects and one complementary cross-cutting Integration, Dissemination and Evaluation (BRIDGE) Center to generate flagship data sets and best practices for the collection and preparation of AI/ML-ready data to address biomedical and behavioral research grand challenges. In the long-term, we expect the community to build on these efforts to further assess and evaluate the utility of the flagship data sets and to create new data sets using the Bridge2AI framework. The program will strive to provide sustainable long-term resources for the community to train and replicate generalizable Bridge2AI tools for ethical and reproducible scientific discoveries using AI/ML models.

The Bridge2AI Program will support interdisciplinary collaborations to merge biomedical and behavioral research and AI/ML data science communities. It is expected that the Data Generation Projects and the BRIDGE Center will work together to accomplish the goals of the Bridge2AI program. Teams will include relevant expertise in biomedical and behavioral research domains, ethics, data science and data management expertise, as well as teaming and logistical expertise to facilitate interdisciplinary collaboration.

It is critical that the Bridge2AI program proactively recruits and develops individuals from diverse social,

cultural, economic, academic, industrial and technical/non-technical backgrounds and communities. NIH is particularly interested in Bridge2AI data generation projects that have the potential to characterize the diversity of society and health problems and for Bridge2AI teams to include members from groups (e.g. researchers, research cohorts and institutions) that are underrepresented in biomedical and behavioral research (see [NIH's Interest in Diversity](#), information about [IDeA states](#), and the [NIH UNITE](#) initiative). Individuals with training in different disciplines, and from varied backgrounds, tend to focus on different project details and constraints and contribute unique experiential knowledge. This diversity increases the likelihood that all critical ethical considerations will be addressed, and provides a holistic approach to identifying solutions, while combating against biases, such that the AI/ML system will be more effective with the best efficiencies for all population groups, including those who are underrepresented in biomedical and behavioral research. This requirement of team science across the Bridge2AI program will lead to revolutionizing the production, dissemination, acquisition, and impact of scientific knowledge through AI/ML.

This ROA requires a Plan for Enhancing Diverse Perspectives (PEDP) as part of the application (see further below). Applicants are strongly encouraged to read the ROA instructions carefully and view the available [PEDP guidance material on the Bridge2AI website](#), www.commonfund.nih.gov/bridge2ai/programresources.

The BRIDGE Center which will be supported through a U54 Funding Opportunity Announcement (see [NOT-RM-21-021](#)), separate from and complementary to this one, will be responsible for integrating activities and knowledge across Data Generation Projects, disseminating products, best-practices, and training materials/activities, and evaluating all aspects of the Bridge2AI program with input from external stakeholder communities. The BRIDGE Center will combine cores of expertise in administration, team science, ethics, standards, tools optimization, and skills and workforce development to harmonize the data generated across all the Bridge2AI Grand Challenges. The BRIDGE Center must be positioned to be well-informed of the changing landscape of ethical principles, standards, AI/ML model requirements, automated software tools, and skills and workforce development methods. Each BRIDGE Center core will work across the corresponding Modules of the Data Generation Projects to facilitate interdisciplinary team science and to create cross-cutting products, such as ethical AI/ML best practices, standards harmonization, workflow and tool testbed analyses, and skills and workforce development opportunities for Bridge2AI Consortium as well as broader communities.

The Data Generation Projects (this opportunity)

Each Data Generation Project will be centered around a biomedical and/or behavioral research “grand challenge” chosen by the project team to produce new data sets amenable to broad AI/ML analyses. Data sets may be applicable to multiple grand challenges. The data collected may include multiscale, multi-modal, and multi-stream data that span the entire spectrum of biomedical and behavioral methods, from atoms to populations, and should be hypothesis-agnostic, ethically sourced, trustworthy, sharable and machine readable. Data sets may incorporate existing data streams, such as clinical, environmental and surveillance data, as needed. Data collected along with annotations and metadata will be curated and shared with the community as it is collected. Projects are envisioned as multi-component undertakings utilizing loci of expertise, or Modules, in team science, ethics, standards, software tools, relevant data acquisition methods, and skills and workforce development.

Applicants will identify and form project teams around a biomedical and/or behavioral grand challenge. Applicants are encouraged to consider the following illustrative examples of biomedical and/or behavioral grand challenges and may also develop their own grand challenge:

- **Digital Twins:** A digital twin, [first introduced in 2002](#), is a virtual representation of a physical system that is continuously updated with the system's performance, maintenance, and health status throughout its lifetime. In general, a digital twin virtually incorporates sensor and/or imaging data from its physical system into the context of its operations, which allows for real-time monitoring and analysis of historical data to address, predict and prevent problems with the physical system. Digital twin simulations are being used in the device and manufacturing industries and have much promise for understanding life processes in complex biomedical and behavioral systems. Digital twins in the biomedical research environment can use AI/ML modeling to standardize experimental processes and integrate multiple disparate and heterogeneous sources of data to infer mechanisms to discover emergent properties. A digital twin of a person could help to monitor health history, predict physical and psychiatric phenotypes, design personalized treatment interventions for multiple health conditions, facilitate virtual clinical trials, improve lifestyle and prevent potential health issues in the future. AI/ML modeling may also serve to overcome many of the ethical challenges (e.g. privacy, ownership, moral distinctions in the data) associated in the development of personalized human digital twins.
- **Expanding AI/ML in Clinical Care:** With the rapid adoption of AI/ML analysis in the biomedical arena, there has been a rapid rise of AI/ML medical devices and algorithms, especially as intended for diagnostic devices. For example, using AI/ML analysis of retinal images, individual diseases such as diabetic retinopathy can be detected with high sensitivity and specificity. Research has also found that the same types of retinal images can be used to predict cardiovascular risk factors and other diseases. To move beyond diagnosis of individual diseases, generation of imaging and clinical data that is annotated for multiple disease indications will help to advance AI/ML applications to a broader system for detecting multiple diseases. Furthermore, integration of AI/ML systems to analyze multiple data types, such as genomic, phenotypic, imaging, electronic health record, and social factors data will propel AI/ML analysis to learn faster from the extended data types. This will in turn bring more advanced, in-depth predictive models to assist in identifying new biomarkers for more accurate disease diagnosis and risk predictions.
- **Functional Genomics:** One of the central problems in biology is understanding the mechanisms by which genomic variation affects human disease and health. Since the early days of the Human Genome Project, NIH has supported the generation of massive amounts of genome sequencing data which has enabled catalogs of human variation and improved knowledge about the genomics of human disease and health. Moving into a new phase, the goal is to develop a framework for systematically understanding disease mechanisms, including an improved understanding of gene pathways and networks. Progress in this area will benefit from integrating genomic data (DNA, transcriptomes, epigenetic data) and other multi-omics data with environmental and clinical data. Development of AI/ML-based predictive models linking functional genomic and environmental perturbations to changes in phenotypes have potential to advance this area, since existing analytical methods are not feasible to scale to the complexity inherent in decoding genomic, cellular, phenotypic and clinical data. Further, using machine learning to search through high-dimensional spaces in large-scale multi-omics and human health

data sets, intelligent feedback loops may be created from results of existing measurements to prioritize the gathering of new measurements and data types.

- **Movement Phenotyping:** Each overt human behavior (movement) has a unique signature, with variations reflecting key features of neural and musculoskeletal drivers, clinical conditions, the complexities of human development and aging, and signifiers of physical activity for health and wellness. Increasingly, objective, quantitative measurement devices are being used to measure movement especially during free-living activity and in real-world environments. However, we are lacking reliable and physiologically verifiable measures of these movements. In addition, no standard outputs currently exist to enable comparability of movement signatures derived from different devices. Creating a suite of validated, reproducible, accessible signatures for movement would increase the use of reproducible and objective metrics associated with a host of conditions that include movement as their central indicator. Validated, real-time measures of these movement phenotypes would dramatically change our ability to predict, prevent, and treat serious conditions. AI/ML modeling can serve a critical role in developing automated techniques to curate and validate the data to predict diagnostic markers of human movement.
- **Precision Public Health:** There is a need to understand how behavioral, social, economic, and environmental factors and structures integrate with individual biology and in the context of health care to better understand the impact and influence on an individual's health and that of populations. A precision public health approach integrates cumulative knowledge to enhance understanding of how multiple societal dimensions and contexts interplay with biological and behavioral factors to influence health in specific populations. The analytic power of AI/ML models have clearly demonstrated utility in addressing complexities, such as precision public health. Initial flaws in AI/ML models caused by predictive and restrictive biases can be mitigated by improving the quality of the data, the development of inclusive algorithms, and diverse training of the systems. AI/ML modeling has a strong potential to provide the framework needed to integrate social determinants, which influence health and wellbeing across the life course, reducing structural injustices, social harms and health inequities. A rapid advancement of AI/ML models in health care can both foster more comprehensive technical attention to data and social infrastructures and offer the promise of identifying and applying more effective treatments to populations, which will improve precision public health.
- **Salutogenesis:** Vast resources and efforts are invested towards understanding the pathogenesis of disease. In contrast, salutogenesis, the process by which individuals move from a less healthy to a healthier state, is understudied and poorly understood. Salutogenesis may occur via "pathogenesis in reverse", and/or may engage other yet-to-be-defined pathways in the process of restoring health. In either case, the process of salutogenesis is likely to involve multiple physiological systems (e.g., cardiovascular, digestive, metabolic, immune), domains (bio-psycho-social) and spatiotemporal scales. Unlike pharmacological treatments that target specific molecular pathways, health restoration may incorporate behavioral and homeostatic physiological mechanisms, as well as complex endogenous cellular processes such as repair, resolution and regeneration. The unprecedented "grand challenge" of understanding salutogenesis will require innovative analytical tools using AI/ML modeling to develop multiscale computational models that explain how physiological and biopsychosocial networks dynamically change over time during the process of human health restoration.

Prospective applicants are **strongly encouraged to participate in Grand Challenge Team Building activities** (see Proposal Timeline) to ensure their proposal is responsive to the Bridge2AI Program. **Data Generation Project proposals must include all Data Generation Project Modules described below.** The order of Modules that follows reflects the sequence of activities expected to be involved in building Bridge2AI datasets. Each proposed Module will be reviewed individually and may be separately awarded to work with other Modules proposed by other Data Generation Project proposals to form the NIH Bridge2AI Consortium. Applicants should consider how to promote diversity, equity and inclusion across all Modules of the Data Generation Project (see [NIH's Interest in Diversity](#), information about [IDeA states](#), and the [NIH UNITE](#) initiative), and include Module-specific details in the required **Plan for Enhancing Diverse Perspectives (PEDP)** – see [Other Attachments](#) for PEDP instructions.

Teaming Module

Team science is critical to the success of the Bridge2AI program, which will bring together teams with broad arrays of expertise, as well as technical and non-technical backgrounds in a multi-team system. It is anticipated that the Bridge2AI program may begin as a *multidisciplinary* endeavor, with each team member bringing individual expertise; however, the Teaming Module of each Data Generation Project will work with the Teaming Core of the BRIDGE Center to transition the Bridge2AI program to *interdisciplinary* collaboration that integrates these different disciplines, sectors and viewpoints into a synergistic team to accomplish project goals. The primary purpose of the Teaming Module is to enable interdisciplinary and inclusive team science within the Data Generation Project. This will include:

- Task integration, creating a governance structure and shared common workflows within the Data Generation Project.
- Relationship Building and Social integration, from individual relationships to establishing a culture in which a diverse array of participants can thrive, to promote optimal team formation and functioning, efficiency in communication and decision making, and leadership development of participants throughout career stages.
- Team Formation, identifying needed areas of expertise and brokering new partnerships as needed, including partnerships with individuals from diverse social, cultural, economic, academic, industrial, and technical/non-technical backgrounds.
- Executing Key Elements of the PEDP, such as recruitment strategies, career development and mentoring opportunities, outreach and partnering with advocacy groups, and other strategies to advance inclusivity in alignment with the program goals.

The Teaming Module must include the expertise and diverse viewpoints needed to support team science across All Modules for the proposed Data Generation Project.

The Teaming Module of each Data Generation Project will be expected to interact with the Teaming Core from the BRIDGE Center to accomplish cross-cutting teaming activities across the Bridge2AI Program.

Ethical and Trustworthy AI/ML Module

Bridge2AI defines the field of ethical and trustworthy artificial intelligence (ETAI) as the study of the influence and behavior of humans and/or AI/ML systems during the design, development, and implementation of AI/ML-ready data, tools, and technologies as well as the effects and implications of those behaviors on users and society. ETAI research encompasses the unique ethical, legal, and social implications (ELSI) of biomedical and behavioral AI/ML research in addition to principles of trustworthy AI/ML. Issues include but are not limited to: accuracy, reliability, and FAIRness of AI/ML-ready data and tools; accounting for bias in sampling, data measurement, analysis and implementation of AI/ML models; transparency, accountability, and auditability of machines and humans; consent, safety, privacy and responsibility for data; and stakeholder involvement and trust.

The goals of this Module are to instill a culture of ethical inquiry (rather than compliance) to: (1) Identify, assess, and help address ETAI issues raised as the data generation project creates and releases AI/ML-ready datasets; and (2) Incorporate ETAI scholars as we create a new biomedical and behavioral AI/ML research community working together across technical, social and health disciplines.

Each Data Generation proposal must contain an embedded ETAI research project focused on identifying and addressing issues raised by the grand challenge goals. The proposal should demonstrate how work from the ETAI Module will inform the proposed data generation project. ETAI projects should also look ahead to anticipate questions likely to be faced as their data and tools are shared, used, and adapted in research.

The ETAI research aims should be described. Relevance and significance of the Data Generation Project to the broader biomedical AI/ML field should be clearly explained. Since the Bridge2AI program is prioritizing a culture of ethical and trustworthy research throughout the data generation process, proposals should describe how the proposed ETAI aims will augment and complement the efforts of the other Modules in the Data Generation Project.

Methods appropriate to address the ETAI research goals and inform the larger study should be clearly explained and choices justified. Proposals may propose multi-disciplinary studies using single or mixed research methods. Proposed methods may include but are not limited to: qualitative and quantitative approaches, legal and normative analyses, and other types of analytical and conceptual research methodologies, including those involving the direct engagement of stakeholders.

Proposals should explain how the ETAI project will directly inform planning, implementation, and evaluation of the Data Generation Project. To this end, planned interactions and overlap between other Data Generation Modules and ETAI project staff should be emphasized and described.

Examples of research topics include but are not limited to:

- Addressing potential unique challenges of open-ended informed consent for AI research, within and across diverse communities.
- Assessing and informing data protection technologies and policy frameworks that address a broad range of identifiability, safety, efficacy, and privacy issues from newly merged data types (e.g., social and contextual variables, the genome, exposome, biometric identifiers such as retinal scans).

- Addressing and ensuring the project will not exacerbate disparities and inequities in downstream health decisions and outcomes, and contributing to guidelines and standards for the field.
- Separately identifying and addressing societal biases that significantly influence the data generation research (e.g., integrating social determinants of health and psychosocial factors influencing biology, health and disease for AI development; also ensuring diverse lived experiences inform data plans, sampling and measurement).
- Identifying and addressing the Data Generation Project team needs with respect to ETAI tools, training and culture.
- In collaboration with the Teaming Module, assessing and incorporating community and stakeholder views on the use of AI/ML to address the biomedical or behavioral focus of the data generation project.
- In collaboration with the Standards, Tools and Data Acquisition Modules, incorporating ETAI factors throughout the data design and collection processes.

Expertise in the ETAI Module should form multidisciplinary teams that bridge AI/ML/data science fields and ethics/trust-related fields including but not limited to: bioethics, law, health disparities research, health services and implementation research, public health, computer science, AI/ML, bioinformatics and health information sciences, behavioral and social sciences (e.g., psychology, sociology, anthropology, political science, economics, communication science, and humanities).

From the outset and throughout the life of the Data Generation projects, ETAI staff from each of the Data Generation Projects will come together to create synergy by identifying needs and common problems and solutions applicable to the broader biomedical AI field. Therefore, it is expected that personnel funded through this Module will work closely with the other Modules of the proposed Data Generation Project and with the BRIDGE Center ETAI Core to identify ongoing and emerging questions, needs and solutions. The ETAI component is also expected to interact and share resources and practices with other Data Generation Projects through the BRIDGE Center and participate in advisory board meetings of the BRIDGE center.

Standards Module

The purpose of the Standards Module for each Data Generation Project is to bring together multi-disciplinary experts to unify data attributes across multiple data sources and data types required to address the chosen Grand Challenge. These experts will draw from their domain background to create a flexible strategy for addressing the chosen Grand Challenge, identify potential AI/MI analysis methods that may be used in the future (after the data are curated), determine the requirements for employing the associated analysis, and produce the data attributes needed to fulfill the analysis requirements. The Standards Module experts should assess currently available standards (refer to Bridge2AI website, www.commonfund.nih.gov/bridge2ai/programresources) and determine whether new standards should be established. For example, applicants can consult the [UMLS Metathesaurus](#) for its adequacy as a standard terminology basis; if it is determined to be inadequate, proposals should demonstrate the inadequacy and propose an alternative structure that can be aligned with the UMLS Metathesaurus.

The Standards Module should possess expertise in data interoperability, and familiarity with the standards required to implement AI/ML approaches for distributed learning systems to address biomedical and behavioral grand challenges. Experts of the Standards Module will organize themselves according to their roles and functions as the Bridge2AI program proceeds, adapting to the changing needs of each Data Generation Project as follows:

- Use principles established in the ETAI Module to formulate and establish the required standards for semantic interoperability, inclusion and data sufficiency
- Work with the Tool Development and Optimization Module to incorporate standards into the automated algorithms being developed for data processing and management
- Review the implementation of standards in the data being collected from the Data Acquisition Module, through quality assessment of the metadata
- Interact with the Data Acquisition Module and Skills and Workforce Development Module to create and disseminate *datasheets* and *model cards* (refer to Bridge2AI website, www.commonfund.nih.gov/bridge2ai/programresources) and other outputs from the Standards Module

It is expected that the Standards Module from each Data Generation Project will engage with the Standards Modules from the other Data Generation Projects and with the BRIDGE Center to ensure the Bridge2AI standards are generalizable to other biomedical and behavioral applications. By adopting generalizable standards, data generated for different grand challenges may become exchangeable for AI/ML analyses. The standards are required to produce reproducible, credible outputs that can be reused by the broader NIH community. Each Standards Module must propose a sustainable process for continually updating standards as the biomedical AI/ML community matures.

Tool Development and Optimization Module

The purpose of the Tool Development and Optimization Module is to build and provide to the biomedical and behavioral research community software, hardware and firmware tools that will enable: (a) automated annotation and structuring of [FAIR](#) and ethically sourced data to facilitate its use in machine learning analysis; (b) future AI/ML-based analysis frameworks to extract knowledge from such data; and (c) lowering of barriers for researchers currently not working in AI/ML to be able to use AI/ML-based tools in their future research. Examples of such tools include research instruments that are sensorized and controlled to identify batch effects and error, automated platforms for imputing missing data and performing *active learning* (refer to Bridge2AI website, www.commonfund.nih.gov/bridge2ai/programresources) across workflows, software that speeds annotation and metadata completion at the point of capture, as well as tools for linking/mapping between any new standards created through Bridge2AI Standards Modules and established standards (e.g. SNOMED, LOINC). The scope of tools built in this Module encompasses not just data annotation, structuring and analysis tools, but also those that would be relevant for development of ethical AI/ML, such as tools to navigate inherent bias in AI/ML algorithms or data sets being used for training AI/ML models. Further, tools currently used to build and deploy AI/ML applications and analyze results are often developed by distinct scientific communities in isolation. Breaking these silos down and facilitating the creation of combined workflows that have relevance across multiple Data Generation Projects in

Bridge2AI will be a high priority for this program. All tools developed through Bridge2AI will be expected to follow open access principles.

The experts forming the Tool Development and Optimization Module should possess substantial prior experience with building computational tools to incorporate AI/ML for multiple stages of the biomedical research lifecycle (data collection and annotation, analysis, modeling and visualization).

It is expected that the Tool Development and Optimization Module from each Data Generation Project will engage with the Tools Modules from the other Data Generation Projects and with the Tool Development and Optimization Core in the BRIDGE Center, so that the overall resource of tools built in Bridge2AI evolves in the direction of a *composable*

www.commonfund.nih.gov/bridge2ai/programresources) AI/ML infrastructure, whereby the appropriate tools for addressing biomedical grand challenges can be brought together under software control to build workflows that match the analysis requirements of each challenge.

In addition, each Tool Development and Optimization Module will be expected to have a synergistic interaction with the Skills and Workforce Development Module within their Data Generation Project (as well as with the corresponding Skills and Workforce Development Core in the BRIDGE Center), for generalizing use of tools developed and providing technical support and user training in these tools for non-AI/ML experts and underserved populations.

Data Acquisition Module

The primary purpose of the Data Acquisition Module is to generate new, high-value, high-quality, AI/ML-ready datasets that will address one or more biomedical and behavioral research grand challenges and advance the field of AI/ML analysis. Biomedical and behavioral data are multimodal and dynamic, and they are often collected from interconnected networks of systems. Applicants are encouraged to collect multimodal data from a cohort of human and/or animal subjects as appropriate, which can be used by more than one AI/ML model to understand the dynamic nature of biological and behavioral systems. The Data Acquisition Module must include a clearly defined data acquisition plan that describes the domain expertise, data sources and data types that will be collected to address a chosen grand challenge. Transparent data collection, processing and curation workflows are required to identify the tools and methods that will be used by the Data Generation Project. The Data Acquisition Module will abide by the ethical framework as specified in the ETAI Module, and work with the Standards Module to incorporate data and model standards applicable for the planned data types to be collected. The Data Acquisition Module will work with the Tool Development and Optimization Module to automate the data collection workflows as much as possible and will be expected to implement experimental designs and protocols to minimize environment or human operator variabilities and reduce subject population selection biases. Data quality control and analysis processes should be clearly defined to ensure the completeness of data and metadata with minimal data biases, batch effects, and other artifacts.

Data sets are to be the primary product of the Bridge2AI program. All data sets generated should be made available in a timely manner and disseminated as broadly as possible to the research community, in collaboration with the Skills and Workforce Development Module. The Bridge2AI program has overall goals and timelines regarding data availability. Projects that do not meet these goals may be restructured or terminated. A data sharing plan describing how these goals will be met must be part of

the resource sharing plan (see How To Apply, below). As NIH Common Fund (CF) awardees, Bridge2AI awardees will be expected to commit to [Common Fund Data Ecosystem \(CFDE\)](#)-compliant data/metadata models, formats, and ontologies appropriate to their data type(s)/problem domain. Where applicable, projects are encouraged to deploy scalable and adaptive cloud-enabled AI/ML data architectures to better accommodate collaboration, data sharing, and data protection.

Skills and Workforce Development Module

The Bridge2AI Data Generation Projects will require interdisciplinary collaboration across a broad array of expertise, including both technical and non-technical backgrounds. The primary purpose of the Skills and Workforce Development Module is to bridge expertise across biomedical and behavioral research domains, ethics, AI/ML data science and data management, and team science in order to enhance skill development and attract and develop an AI/ML-biomedical and behavioral research workforce. The Skills and Workforce Development Module must include the following two primary components:

- 1) **Skills Development:** Develop integrated and coordinated curricula and educational materials/activities to support use of the AI/ML-ready datasets created by the Data Generation Project, as well as to support use of products and cross-cutting best practices developed by all the Data Generation Project Modules, with ethics and diversity as a central theme incorporated throughout.
- 2) **Workforce Development:** Support career development and mentoring within the Data Generation Project, with emphasis on the recruitment, nurturing, and retention of prospective participants, at all career stages, from underrepresented backgrounds.

A key aspect of skills and workforce development involves outreach and involvement of individuals and communities who are underrepresented in biomedical and behavioral research (see [NIH's Interest in Diversity](#) and information about [IDeA states](#)). Diversity of AI/ML engineers is especially critical for the development of tools that represent the backgrounds and needs of all groups. This Module should be tasked with executing key elements of the PEDP, for example recruitment, nurturing, and retention of prospective participants, at all career stages, from underrepresented backgrounds, outreach and partnering with advocacy groups, and other strategies to advance inclusivity in alignment with the program goals.

How To Apply

Proposal Timeline

Key Events	Date(s)	Action needed by Proposers
Funding Opportunity posted	5/13/2021	
Grand Challenge Team Building (not required)	6/1/2021-7/20/2021	Participate in virtual team building activities.

		Follow instructions on Bridge2AI website, www.commonfund.nih.gov/bridge2ai
Data Generation Project Letters of Intent (required)	7/20/2021	Email: Bridge2AI@od.nih.gov
Full Proposal Due	8/20/2021	Submit to ASSIST
Review	10/2021	
Award Negotiation	11/2021 to 1/2022	
Kickoff meeting/annual meetings	2/2022, yearly	Allocate budget and attend

Eligibility

Successful applicants may or may not have received NIH funding in the past. All entities public and private, small or large, for-profit or not-for-profit, are eligible to apply. Unaffiliated individuals who are U.S. citizens are also eligible to apply.

Organizations

Non-domestic (non-U.S.) Entities (Foreign Applicants) **are** eligible to apply. Non-domestic (non-U.S.) components of U.S. Organizations **are** eligible to apply. Foreign components **are** allowed.

The following entities are eligible to apply under this ROA:

Higher Education Institutions

- Public/State Controlled Institutions of Higher Education
- Private Institutions of Higher Education

The following types of Higher Education Institutions are strongly encouraged to apply for NIH support along with Public or Private Institutions of Higher Education:

- Hispanic-serving Institutions
- Historically Black Colleges and Universities (HBCUs)
- Tribally Controlled Colleges and Universities (TCCUs)
- Alaska Native and Native Hawaiian Serving Institutions
- Asian American Native American Pacific Islander Serving Institutions (AANAPISIs)

Nonprofits Other Than Institutions of Higher Education

- Nonprofits with 501(c)(3) IRS Status (Other than Institutions of Higher Education)
- Nonprofits without 501(c)(3) IRS Status (Other than Institutions of Higher Education)

For-Profit Organizations

- Small Businesses
- For-Profit Organizations (Other than Small Businesses)

Governments

- State Governments
- County Governments
- City or Township Governments
- Special District Governments

- Indian/Native American Tribal Governments (Federally Recognized)
- Indian/Native American Tribal Governments (Other than Federally Recognized)
- Eligible Agencies of the Federal Government
- U.S. Territory or Possession Other
- Independent School Districts
- Native American Tribal Organizations (other than Federally recognized tribal governments)
- Faith-based or Community-based Organizations
- Regional Organizations

Financial and Risk Assessment

Proposers may be subject to financial analysis and risk assessment conducted by NIH staff.

Cost Sharing

Cost Sharing is not required; however, applicants proposing to develop commercial applications or who are using other state or government resources may consider identifying a cost share percentage. Applicants may voluntarily choose to propose a financial plan that includes non-federal resources. The budget submission must clearly identify and justify the use of these resources. Any voluntary cost share must be supported in the proposal by including a letter of support from the providing organization(s)/individual(s).

Grand Challenge Team Building

Applicants are strongly encouraged (though not required) to participate in the Grand Challenge Team Building activities to bring together the diverse expertise necessary to build their Bridge2AI Data Generation Project Teams. During the Grand Challenge Team Building period, listed in the [Proposal Timeline](#), potential applicants will participate in activities to 1) understand the spirit of the Bridge2AI program, 2) understand the components of the funding opportunities, and 3) find potential collaborators to complete the diversity of expertise required to prepare and collect data around biomedical and/or behavioral grand challenge(s). The Bridge2AI program will conduct the Grand Challenge Team Building activities in the spirit of the Bridge2AI program to provide a venue that reduces human biases and maximizes diversity during the team formation process. It is expected that participants will evolve and develop the Grand Challenge ideas during the Team Building. Detailed Instructions for the Grand Challenge Team Building activities can be found on the Bridge2AI website, www.commonfund.nih.gov/bridge2ai/programresources.

Participation in Grand Challenge Team Building activities is not mandatory. It is acceptable for individuals to join teams after these team building activities have been completed and/or for teams that did not participate in the team building activities to apply.

Letter of Intent (LOI)

All applicants who wish to submit a Data Generation Project to the NIH Bridge2AI Program are **required** to submit an LOI, on or before the due date listed in the [Proposal Timeline](#).

The LOI must include the following information [**no more than 1 page**]:

- a. Data Generation Project Title
- b. Contact Principal Investigator first and last name, title, institution, mailing address, email address, and phone number
- c. List of team members, affiliations, and associated expertise
- d. Brief description of biomedical/behavior grand challenge and the associated ethically sourced datasets that will be generated

The NIH will not review and will return full proposals without a submitted LOI.

Proposal Submission Instructions

Institutions

Participating organizations must complete and maintain the following registrations to be eligible to receive an award. There should NOT be any cost associated with ANY of these registrations. All registrations must be completed prior to award issuance. Registration can take 6 weeks or more, so proposers should begin the registration process as soon as possible.

- Dun and Bradstreet Universal Numbering System (DUNS) – All registrations require that proposers be issued a DUNS number. After obtaining a DUNS number, proposers can begin both SAM and eRA Commons registrations. The same DUNS number must be used for all registrations, as well as on the grant proposal.
- System for Award Management (SAM) (formerly CCR) – Proposers must complete and maintain an active registration, which requires renewal at least annually. The renewal process may require as much time as the initial registration. SAM registration includes the assignment of a Commercial and Government Entity (CAGE) Code for domestic organizations which have not already been assigned a CAGE Code.
- eRA Commons - Proposers must have an active DUNS number and SAM registration in order to complete the eRA Commons registration. Organizations can register with the eRA Commons as they are working through their SAM registration. eRA Commons requires organizations to identify at least one Authorized Organizational Representative and at least one Program Director/Principal Investigator account in order to receive an award. Unaffiliated individuals will be registered as “independent scholars” and will also act as the Authorized Organizational Representative, with the same authority in eRA Commons that the Authorized Organizational Representative(s) has in Grants.gov.

Principal Investigators

The Bridge2AI Program requires each Data Generation Project to identify a Principal Investigator/Project Director (PI/PD) team with at least one biomedical/behavioral domain expert and one data science/data management expert, who are responsible for the overall leadership of the Data Generation Project. This

team may be drawn from the leadership of any of the six modules. The application may be submitted by any member of this team. That member's institution will then be the proposing institution.

The submitting Principal Investigator/Program Director should work with their organizational officials to ensure that their eRA Commons account is affiliated with the proposing organization in eRA Commons. If the Principal Investigator is also the organizational Authorized Organizational Representative, they must have two distinct eRA Commons accounts, one for each role. Obtaining an eRA Commons account can take up to 2 weeks.

Full Proposal Format

The proposal should clearly and fully articulate the proposer's capabilities, knowledge, and experience and the budget proposed.

Proposals will be accepted from organizations and individuals listed in the [Eligibility](#) section of this Announcement and must be submitted by the due date listed in the [Proposal Timeline](#). The format should be text-recognizable PDF (Adobe) with single-spaced Arial 10-point font and 1" margins, and the file size must be no greater than 20 MB.

Proposals must include a Cover Page and a separate Research Plan and Budget for each Data Generation Project Module (6 required). The Research Plans are limited to a maximum of **10 pages for each Module**. After review, Modules from different proposals may be selected and funded to work collaboratively with the BRIDGE Center in the Bridge2AI Consortium.

The **Cover Page** for each Module should include **[1-page limit]**:

- A. The proposal title [refer to Grand Challenge chosen]
 - i) Brief description of biomedical/behavioral grand challenge and the associated ethically sourced datasets that will be generated
- B. The **Module** title [use: Teaming Module, Ethical and Trustworthy AI Module, Standards Module, Tool Development and Optimization Module, Data Acquisition Module, Skills and Workforce Development Module]
- C. The Awardee's:
 - i) Legal entity name
 - ii) Address and contact information
- D. The name and contact information for the Awardee's Principal Investigator (with NIH Commons Account information)
- E. The name and contact information for the Awardee's Business Official, the person authorized to negotiate and bind the Awardee as a signatory to the Other Transaction agreement.
- F. The total cost proposed for this Module
- G. Other key personnel names and organizations (multiple Principal Investigators, co-Investigators, collaborators, etc.), expertise/role in Module
- H. Confirmation that the work involves human/animal subjects or data from human/animal subjects
- I. Agreement that select proposal materials may be shared, at NIH's discretion, among the final group of proposal candidates, on a confidential basis, as part of the last stage of agreement construction

The **Research Plan** for each Module must address the following four elements [**10-page limit, not including biosketches and budget pages**]:

A. Technical Approach

Describe how the work of the proposed Module will be accomplished. The proposer must demonstrate understanding of the Bridge2AI Program and the Module being proposed by clearly showing a grasp of the range and the complexity of the work. This section should include a detailed project plan that includes milestones and deliverables, including the interactions between Modules and with the BRIDGE Center. Proposers should demonstrate a conceptual understanding of the challenges specific to the tasks required in the ROA and plans for overcoming these challenges. Applicants must address specific processes and procedures for how they will achieve the required integration with the other components of the Bridge2AI program and for resolving any areas of disagreement.

B. Key Personnel Experience

Proposers must demonstrate experience of key personnel supporting the planning and implementation of activities described in the ROA. Please provide biosketches describing key staff who will be assigned to manage performance and supervise the work for each task and subtask (as appropriate). These biosketches will be reviewed to evaluate whether the individuals possess the required experience to successfully complete the work of the Module. **Biosketches should be no more than two (2) pages in length and shall not count toward the page limits.** At a minimum, the information in the biosketch should include the name and position title, education/training (including institution, degree, date (or expected date), and field; list of positions and employment in chronological order (including dates); Other Research Support; and a personal statement that briefly describes the individual's role in the project and why they are well-suited for this role. The biosketch format (<https://grants.nih.gov/grants/forms/biosketch-blankformat.docx>) used for an NIH grant application is acceptable, but not required, as Other Transactions are not grants.

C. Management/Staffing Plan

Proposals should detail how the proposer will provide the necessary project administration, organization, and staff to ensure quality control, compliance with ROA expectations, and necessary staffing adjustments. In addition, proposers must demonstrate the ability to simultaneously manage multiple tasks within set time periods.

D. Past Experience

Proposers should provide an example(s) of prior experience serving in a similar capacity as the proposed Module. Each example should include a brief description of the project and how the project was analogous to the needs identified in this ROA with respect to the Module being proposed. Applicants will need to demonstrate prior work AND competency associated with the Module being proposed.

The **Budget** must address the following:

The Budget section must provide a realistic, fully justified annual budget and cost proposal for performing the work specified in the ROA over a period of 4 years. Applicants must complete a [SF424](#) budget. **Budget information and any related administrative documentation shall not count toward the total proposal page limit.**

The Budget should provide the overall expected cost for each of the following categories:

- Personnel
- Equipment (including compute resources)
- Workforce development resources
- Travel
- Subawards/subcontracts/consultants
- Funds reserved to engage new partners that may be required to accomplish the goals
- Other direct costs
- Total costs (with indirect costs included)
- Proposed Cost Share contribution, if any

Proposers need to budget for attending an in-person kickoff meeting of the Bridge2AI program and annual Bridge2AI Consortium meetings as noted in the [Proposal Timeline](#). Proposers are required to interact with all the Bridge2AI Data Generation Projects and BRIDGE Center through their regular virtual meetings and to attend special program meetings as needed during the program period.

Research Plan

The Research plan of each Data Generation Project must include **all 6 Data Generation Project Modules**. All Modules must include a timeline and a clear set of measurable milestones. Milestones may be revised as needed for the Bridge2AI program, per OT Authority.

Each Data Generation Project must identify a biomedical and/or behavioral grand challenge that will motivate the data generation (see list of illustrative [example grand challenges](#) above).

Instructions for each Data Generation Project Module:

Teaming Module

The **Teaming Module** should build upon known teaming strategies and resources, such as <https://www.inscits.org/> and <https://www.teamsciencetoolkit.cancer.gov/default.aspx>. Applicants should address the following four main components as part of the Research Strategy to enable interdisciplinary team science within the Data Generation Project.

The Teaming Module must include:

- **Task Integration:** the process to establish a governance structure for the Data Generation Project, creating common workflows across all Modules to interact with each other. Applicants should include plans for performance assessment of all Modules.
- **Relationship Building and Social Integration:** processes to develop trust and a shared vision for the grand challenge, as well as how shared responsibilities, interpersonal interactions, conflict resolution, and professional credit will be managed within the Data Generation Project team. Applicants should consider the need for coordinating varied expertise and cultivating diversity and cultural awareness in individual relationships and throughout the Data Generation Project. Applicants should proactively identify potential barriers to optimal team formation and consider strategies to mitigate against them.
- **Team Composition:** processes to involve diverse disciplines, demographics, and perspectives, which includes biomedical and behavioral research, ethics, AI/ML data science and data management, teaming, and technical and non-technical experts. Describe a continual process to evaluate and update needed expertise across the Data Generation Project team.
- **Executing Key Elements of the PEDP:** processes to include elements, such as recruitment strategies, career development and mentoring opportunities, outreach and partnering with advocacy groups, and other strategies to advance inclusivity in alignment with the program goals.

Ethical and Trustworthy AI Module

The **Ethical and Trustworthy AI (ETAI) Module** should describe an integrated plan for research on the ethics and trust implications of the biomedical and/or behavioral grand challenge being proposed.

The ETAI Module of the proposal must include:

- A clear statement of the specific empirical ETAI research questions related to the biomedical and/or behavioral grand challenge.
- A justification of the multidisciplinary expertise involved in the ETAI Module.
- Methods and approaches that will be used to study those questions.
- A concise description of how this component of the research will be integrated into the development and implementation of the overall data generation design.
- Descriptions of how the ETAI Module and its work will augment and complement work in the Teaming, Standards, Tool Development and Optimization, Data Acquisition, Skills and Workforce Development Modules.
- Planned outcomes of the work.
- Plans to share and disseminate findings and ETAI tools (see **Resource Sharing Plan** section below).

While the research strategy should focus on ETAI issues raised by the specific data generation project, description of the generalizability/applicability of ETAI research questions and work products to the broader field of biomedical and behavioral AI are strongly encouraged. This is an ideal interface point to the ethics locus of the BRIDGE Center.

Standards Module

The research plan for the **Standards Module** should include a detailed strategy for assessing current standards, as well as a strategy for developing new standards as needed.

The strategy must include methods to collaborate with the other Bridge2AI Modules in creating a continuous process for adopting and updating standards as more information is produced in the Bridge2AI program, including:

- Methods for infusing relevant standards being developed in other domains (outside of Bridge2AI) into the efforts of the Bridge2AI Data Generation Project.
- Justification of the expertise in the Standards Module to address data interoperability issues for distributed learning systems.
- Processes to work with the Ethics (ETAI), Tool Development and Optimization, Data Acquisition, and Skills and Workforce Development Modules to incorporate Bridge2AI standards coherently and consistently into the Data Generation Project.
- Evaluation and assessment of the standards used in the metadata and data produced from the Data Acquisition Module.
- A plan for creating and disseminating with the Skills and Workforce Development Module the data sheets and model cards based on the data and model requirements produced from the Data Acquisition Module.
- Consortium efforts with the BRIDGE Center to harmonize the standards process to be generalizable, reproducible, reusable, credible and sustainable to continually adapt to the changing needs of the biomedical AI/ML community.

Applicants are encouraged to include letters of support from other communities developing standards that are relevant to the Bridge2AI program goals.

Tools, including models, algorithms and software, to be developed in the Standards Module must be represented in the overall Resource Sharing Plan for the proposal (see Resource Sharing, below).

Tool Development and Optimization Module

The **Tool Development and Optimization Module** should include the following components:

- Brief overview of existing biomedical AI/ML tools relevant to the specific biomedical grand challenge being addressed by the Data Generation Project, along with justification for why new tool development is needed in that scientific area as part of Bridge2AI.
- Justification of the expertise in the Tool Development and Optimization Module to create computational tools to manage data throughout the project lifecycle.
- A description of how this Module will incorporate the ethics, standards, and data acquisition needs of the Data Generation Project into the development of the proposed tools and workflows.
- A plan for providing detailed documentation for each tool being developed, along with user-guides that are accessible to non-AI/ML biomedical scientists.

- A statement of how tools being developed in the Module adhere to principles of ethical and trustworthy AI.

Tools, including models, algorithms, and software, to be developed in the Standards Module must be represented in the overall Resource Sharing Plan for the proposal (see Resource Sharing, below).

Data Acquisition Module

The **Data Acquisition Module** should include a detailed data acquisition plan to address the biomedical and/or behavioral research challenge(s) targeted by the overall Data Generation Project. The data acquisition plan should incorporate the tools and standards that will be used for data collection, and the ethical and trustworthy principles as defined in the other Modules of the Data Generation Project.

The Data Acquisition Module should include the following components:

- Justification of the domain expertise, data sources and data types needed to collect the grand challenge data.
- Plan to generate ethically sourced data in collaboration with the ETAI Module.
- Justification of the data collection sample size to allow adequate training and evaluation of future AI/ML models to generate reliable outcomes.
- Strategy to work with the Standards Module to allow data integration and data harmonization with minimal batch effects and artifacts, especially for longitudinal data or data that grow over time.
- Strategy for producing data and model requirements for the Standards Module for integrated AI/ML analyses using multiple models when data collection involves multimodal data types from complex systems.
- Strategy to work with the Tool Development and Optimization Module to incorporate data collection workflows and processes, including machines or human operators as needed.
- A plan for data quality control and analysis processes to ensure the completeness of data and metadata, and to minimize data variabilities caused by subject population selection biases and batch effects.
- A detailed data management and sharing plan to ensure data is available in a timely manner and as broadly as possible, deploying scalable and adaptive cloud-enabled AI/ML data architectures as needed - See **Resource Sharing Plan** instructions below.

Data acquisition involving human subjects research, clinical research, and/or vertebrate animal research must follow all associated instructions in the [SF424 \(R&R\) Application Guide](#).

Skills and Workforce Development Module

The **Skills and Workforce Development Module** should address the following two main components as part of the research strategy. For each component, investigators should address important considerations such as inclusion of broad career stages (academically and within industry), challenges of incorporating and developing broad technical expertise, and appropriate outreach, recruitment, and retention methods to reach a wide variety of interested communities, including underrepresented populations. Applicants should address how each component supports the implementation of the PEDP.

Applicants are expected to incorporate existing educational and mentoring resources from the NIH Office of Data Science Strategy (<https://datascience.nih.gov/>) or from other relevant sources.

Component 1: Skills Development

The Module should describe processes to develop, implement, and assess skill development activities. Such activities should be developed for a wide array of career stages, including students, post-doctoral trainees, junior scientists, early-stage and established investigators, and technical and non-technical backgrounds. Applicants are expected to provide plans for outreach to diverse communities to include underrepresented populations in the development and use of skill development activities.

Applicants should describe plans to develop integrated curricula and/or educational materials/activities to support use of the flagship data sets that are being collected by the Data Generation Project. Proposed skill development materials/activities should utilize a broad variety of educational methods with respect to type of instruction (e.g., didactic vs. experiential), manner of content delivery (in-person vs. online), and terms of engagement (self-directed vs. instructor-led, short- vs. long-term). Applicants should provide details on methods to integrate ethical use cases into all skill development activities to minimize bias and health disparities.

Component 2: Workforce Development

Applicants should describe plans to support career development and mentoring within the Data Generation Project. This Module is expected to help develop transdisciplinary expertise in AI-Biomedical and Behavioral Research, with a special emphasis on recruiting scientists from underrepresented backgrounds and a focus on ethical considerations. Applicants should include plans to provide Data Generation Project team members with a breadth of exposure to and hands-on experiences with a myriad of Bridge2AI activities. Activities might include mentoring, onboarding, research experiences and related activities that will enhance the development of relevant expertise across career stages. Applicants should provide interdisciplinary mentoring plans that address frequency, level (e.g., peer-to-peer, senior-to-junior, junior-to-senior) and type (e.g., formal vs. informal, one-on-one vs. group) of mentoring tailored to career stage and background, and include critical elements of team science and development of “soft skills”. Proposals should describe plans for formal recognition of accomplishments from students, trainees and early career investigators involved in the Data Generation Project to support career longevity of these transdisciplinary scholars.

Other Attachments:

Resource Sharing Plan

All proposals must include plans for sharing the resources produced from the Bridge2AI program. Plans from different Modules, as described below, should be grouped together into one overall Resource Sharing Plan in the proposal. This plan should, for each relevant Module, include policies for access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. As NIH Common Fund (CF) awardees, Bridge2AI awardees will be expected to commit to Common Fund Data Ecosystem (CFDE)-compliant data/metadata models, formats, and ontologies appropriate to their data type(s)/problem domain.

For Data: In the body of the text, the section should begin with the heading "**Data Sharing Plan**". Early and frequent data release, independent of traditional scientific publishing, and above and beyond what may be required by other NIH programs and policies, is a core goal of the Bridge2AI program. All proposals must describe a data release schedule and plan for how the data to be acquired will be continuously made available for use and re-use by a broad variety of researchers, beginning no later than the second year of funding. Considerations in this plan might include but are not limited to: choice of repository or cloud platform, budgeting for data hosting and transfer, de-identification methods as appropriate, and plans for controlling access to protected data. A non-exhaustive list of data repositories of interest may be found here: https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html.

For Tools (conceptual, theoretical, algorithmic, modeling, etc.): In the body of the text, the section should begin with the heading "**Tool Sharing Plan**". Tool Sharing Plans should provide details including minimum requirements for tool documentation, validation, and availability. Applicants are also expected to include plans to link proposed tools with other relevant tools. Software tools should be addressed in the Software Sharing Plan (below).

For Software: In the body of the text, the section should begin with a heading indicating "**Software Sharing Plan**". There is no prescribed single license for software produced through grants responding to this announcement. Software tools should be developed using standard open source practices and their availability should in no way be limited by timelines or processes associated with traditional scientific publication. Proposals should include plans for software dissemination consistent with these principles:

- The software should be freely available to, at minimum, biomedical, biological, behavioral, environmental, and clinical researchers and educators in the non-profit sector, such as educational and research institutions, and government laboratories.
- The terms of software availability should permit the commercialization of enhanced or customized versions of the software, or incorporation of the software or pieces of it into other software packages.
- To preserve utility to the community, the software should be transferable such that another individual or team can reproduce the analytical tool and continue development in the event that the original investigators are unwilling or unable to do so.
- The terms of software availability should include the ability of researchers to modify the source code and to share modifications with other colleagues. An applicant should take responsibility for creating the original and subsequent official versions of a piece of software, and should provide a plan to manage the dissemination or adoption of improvements or customizations of that software by others. This plan should include a method to distribute other user's contributions such as extensions, compatible Modules, or plug-ins.

NOTE: Applicants are encouraged to look at example data management and data sharing plans on the Bridge2AI website, www.commonfund.nih.gov/bridge2ai/programresources.

Early and frequent data release, above and beyond what may be required by other NIH programs and policies, is a core goal of the Bridge2AI program. All proposals must describe how this will be done.

[Plan for Enhancing Diverse Perspectives \(PEDP\)](#)

In an "Other Attachment" entitled "Plan for Enhancing Diverse Perspectives," all proposals must include a summary of strategies to be used to advance the scientific and technical merit of the proposed project through expanded inclusivity. The PEDP should provide a holistic and integrated view of how enhancing diverse perspectives is viewed and supported throughout the application and within each Module, particularly the Teaming and Skills and Workforce Development Modules. The PEDP will vary depending on the scientific aims, expertise required, the environment and performance site(s), as well as how the project goals are structured. The PEDP may be no more than 2 pages in length and should include a timeline and milestones for relevant components that will be considered as part of the review. Examples of items that advance inclusivity in research and may be part of the PEDP can include, but are not limited to:

- Discussion of engagement with different types of institutions and organizations (e.g., research-intensive, undergraduate-focused, minority-serving, community-based).
- Description of any planned partnerships that may enhance geographic and regional diversity.
- Plan to enhance recruiting of women and individuals from groups traditionally under-represented in the biomedical, behavioral, and clinical research workforce.
- Proposed monitoring activities to identify and measure PEDP progress benchmarks.
- Plan to utilize the project infrastructure (i.e., research and structure) to support career-enhancing research opportunities for diverse junior, early- and mid-career researchers.
- Description of any training and/or mentoring opportunities available to encourage participation of students, postdoctoral researchers and co-investigators from diverse backgrounds.
- Plan to develop transdisciplinary collaboration(s) that require unique expertise and/or solicit diverse perspectives to address research question(s).
- Publication plan that enumerates planned manuscripts and proposed lead authorship.
- Outreach and planned engagement activities to enhance recruitment of individuals from diverse groups as research participants including those from under-represented backgrounds.

For further information on the Plan for Enhancing Diverse Perspectives (PEDP), please see resources on the Bridge2AI website: www.commonfund.nih.gov/bridge2ai/programresources.

Additional information to include in the submission:

- A letter of support from the proposer's organization indicating institutional commitment to the project, which may include but is not limited to: human and capital resources, support for training activities or consortium meetings, licenses, and preparations to negotiate Other Transactions agreements.
- Letters of support from proposed collaborators who are not included as Key Personnel.

Objective Review

Proposals to Other Transactions Research Opportunity announcements such as this one are not reviewed using the standard NIH peer review process, but using custom processes referred to as Objective Review. Objective Review will involve the submission of written critiques by subject matter experts against the Review Criteria described below, and interactive individual discussions between

those experts and NIH program staff. Those discussions may also include proposing teams in later stages of the review.

The Objective Review of the Research Strategy will consider:

- The potential of the Data Generation Project to address the biomedical and/or behavioral research grand challenges
- The ability of the proposed research teams to implement each of the Modules, and their potential impact if successfully implemented
- Past performance and experience of the team members and complementarity with other recipients
- The adequacy and appropriateness of the Multiple PI/PD plan, budget, resources, data and resource sharing, and collaboration plans
- The extent to which the efforts described in the Plan for Enhancing Diverse Perspectives further the significance of the project, strengthen and enhance the expertise required for the project, meaningfully contribute to innovation, have a well-developed and feasible timeline and milestones, and will contribute to the success of the project through an enhanced environment (e.g. collaborative arrangements, geographic diversity, institutional support)

Note that past performance and experience includes not only traditional measures of scientific productivity such as publication counts, invited presentations, or past funding success, but also proposers' demonstrated track record of particular behaviors (community participation, collaborative efforts, openness to sharing data and resources, etc.).

Other Transactions agreements will be negotiated with eligible entities whose proposals are determined to be the most advantageous and provide the best value to the NIH. Funding decisions will be based on the outcome of the Objective Review and on the desired portfolio balance within the Bridge2AI program. Up to eight Data Generation projects will be funded. The level of funding for awards made under this solicitation has not been predetermined but will depend on (1) the objectives proposed by the applicants and their fit to the goals of the Bridge2AI Program, (2) the quality of the proposals received, and (3) availability of funds.

Following the review of proposals, NIH may recombine subteams from different proposals in the construction of Other Transactions awards to be made, or individual Modules from distinct proposals might be selectively funded to achieve the program goals. Additionally, if, over the duration of the project, some of the Modules either gain or lose relevance to programmatic goals, the funding for such Modules may be increased, decreased, or discontinued.

NIH reserves the right to:

- Invite all, some, one, or none of the Principal Investigators submitting proposals in response to this solicitation to present their proposal in a web-based videoconference;
- Share proposals between and among any proposer(s) as necessary for configuring teams, economizing work, and prioritizing activities;
- Select for negotiation all, some, one, or none of the proposals received in response to this solicitation;
- Accept proposals in their entirety or select only portions of plans for award.

Appeals of the objective review will not be accepted for plans submitted in response to this ROA.

Review Criteria

Review Criteria for Teaming Module

- Are the members of the Teaming Module highly qualified to successfully lead and manage the proposed teaming and team science activities? Is there a diversity of backgrounds? Are the planned efforts appropriate and sufficient for the work proposed?
- Is the team's history of facilitating collaboration adequate as determined by references and/or letters of support?
- Are the proposed leadership approaches, governance, and organizational structure appropriate for the proposed Data Generation Project?
- Is there a clear process for brokering new partnerships? Are adequate plans provided for how gaps in expertise will be identified, filled, and integrated into the Data Generation Project?
- Is there a description of methods and activities to be used to help define a collective identity, shared purpose, and patterns of constructive communication across disciplines and career stages? Are there methods and activities to encourage input from junior personnel and/or individuals from underrepresented and disadvantaged groups, who may be hesitant to contribute?
- Is there a clear process for implementing the PEDP?

Review Criteria for ETAI Module

- Is successful execution of the ETAI plan likely to increase the acceptability, trustworthiness and equitable ethical use of the Data Generation Project's data sets, tools, and resources? Will the Module contribute ETAI knowledge, tools, and approaches that are generalizable to other biomedical and behavioral AI/ML research areas?
- Does the ETAI Module detail unique or unsolved problems that the Data Generation project is likely to encounter? Does the Module detail the research which will integrate existing scholarship, tools, and guidance with development of new resources to address ETAI issues? Does the approach make clear how the ETAI Module's work will address unique or unsolved problems?
- Do the team members have complementary expertise? Is the leadership approach, governance and organizational structure appropriate for the project?
- Does the proposal detail how the ETAI component will both contribute to and benefit from formal and informal collaborations with other Data Generation projects and BRIDGE Center Cores? Is the ETAI Module fully integrated with the other Modules of the project plan? Is it clear how the ETAI component will inform and be informed by the other Modules of the Data Generation Project? Does the management plan ensure the integration of researchers from the different Modules into the development and implementation of the overall research plan?

Review Criteria for Standards Module

- To what extent will the proposed standards activities address the current imbalance across biomedical/behavioral domains of data for AI/ML analysis?
- Is the research team addressing the most relevant standards for the grand challenge chosen?
- Is any strategy for developing new standards appropriate and justified?

- To what extent is the project team qualified to use, modify, and develop standards related to the proposed grand challenge datasets?
- Have the applicants proposed adequate personnel (time and effort) to harmonize standards across the Bridge2AI Consortium?
- To what extent does the research environment provide adequate support for Bridge2AI standards to be sustainable long-term to the broader community?

Review Criteria for Tool Development and Optimization Module

- Are the tools from this Module compatible with a composable AI infrastructure, whereby they can be effectively brought together under software control with tools from other Data Generation Projects to build workflows that can be used for multiple biomedical grand challenges?
- To what extent will the proposed tools enable annotation and structuring of data to facilitate its use in AI/ML applications and, after annotation and structuring, enable AI/ML-based analysis frameworks to be applied to extract knowledge from such data?
- To what extent does the tool development proposed include ethical AI/ML considerations, such as tools to navigate inherent bias in AI/ML algorithms or data sets being used for training AI/ML models?
- Will the tools facilitate speedy annotation and metadata completion at the point of data capture, as well as tools for linking/mapping between any new standards created through Bridge2AI and established standards?
- Will the tools be openly accessible and reusable by the broader community?
- Will the proposed tools lower the barriers for researchers currently not working in AI/ML to be able to use AI/ML-based tools in their future research?

Review Criteria for Data Acquisition Module

- Does the proposed Data Generation Project provide a sufficient data acquisition plan for the proposed data sources and data types to be collected?
- Is the project team qualified to lead the data acquisition efforts?
- Is the strategy sufficient for working with the other Modules of the Data Generation Project – to collect ethically sourced data with harmonized standards, and utilizing automated experimental workflows, processes, and validation procedures?
- Are data quality control and analysis processes implemented into the data collection adequate to ensure completeness of data and metadata, and to reduce data biases?
- Is the data acquisition Module prepared to make the data set readily available to meet the FAIR principles for AI/ML analysis?
- To what extent will the proposed Module work with the Standards and Skills and Workforce Development Module to deliver Data Sheets and Model Cards for the community?

Review Criteria for Skills and Workforce Development Module

- Do the investigators have a track record in: Curriculum and/or educational materials/activities development, assessment, and dissemination? Culturally aware career development and mentoring? Incorporation of ethical considerations in skills and workforce development activities?
- Are skill development plans provided to enhance use of data sets from Data Generation Projects? Do the plans incorporate activities for different career stages and different technical expertise? Have they considered type of instruction (e.g., didactic vs. experiential), manner of

content delivery (in-person vs. online), and terms of engagement (self-directed vs. instructor-led, short- vs. long-term)? Do skill development plans incorporate use cases illustrating ethics concepts?

- Are workforce development plans provided for different career stages, backgrounds, and levels of technical expertise? Do they incorporate hands-on interdisciplinary activities and opportunities to develop “soft skills”?
- Does the Module specify details about the strategies that will be employed to recruit and retain diverse participants in accordance with the PEDP?

Special Award Terms and Information

The awardees from the Bridge2AI program must participate actively (allocate time and effort) to work with the Bridge2AI Consortium (e.g., in monthly calls with other Bridge2AI awardees) and respond in a timely manner to requests from the BRIDGE Center. In addition, the Bridge2AI Program is expected to interact with the [Common Fund Data Ecosystem \(CFDE\)](#) to ensure the interoperability and compatibility of common software standards and tools with the CFDE and other Common Fund datasets and resources.

NIH Discretion

The OT award mechanism allows significant ongoing involvement from NIH Program and Project Managers and provides the NIH the flexibility to alter the course of awarded activity in real-time to meet the overarching goals. This may mean that an awarded activity could be expanded, modified, partnered, not supported, or discontinued based on program needs, emerging methods or approaches, performance, or availability of funds. Performance during the award period will be reviewed on an ongoing basis and course corrections will be made as necessary. As a result, the NIH reserves the right to:

- Fund projects in increments and/or with options for continued work at the end of one or more phases;
- Fund projects of two or more entities (potentially across different proposals) as part of a reorganized collaboration, teaming arrangement, or other means acceptable to the government;
- Request additional documentation (certifications, etc.); and
- Remove participants from award consideration should the parties fail to reach a finalized, fully executed agreement prior to a date determined by the NIH, or the proposer fails to provide requested additional information in a timely manner.

Proposals selected for award negotiation may or may not result in the issuance of an OT award, dependent on the outcome of negotiations, the nature of the work proposed, changing external conditions, and other factors. The NIH reserves the right and sole discretion to engage in negotiation with the selectees applying under this solicitation during all phases of the proposal lifecycle.

Award Governance

The NIH will actively engage with recipients to establish a vision and capabilities for the Bridge2AI program and to oversee the effort of individual recipients to achieve the vision.

NIH Roles and Responsibilities:

1. **Agreements Officer:** NIH individual responsible for legally committing the government to an OT award and to the agreement through which terms and conditions are established, and for the administrative and financial aspects of the award. The Agreements Officer (AO) is the focal point for receiving and acting on requests for NIH prior approval and is the only NIH official authorized to change the funding, duration, or other terms and conditions of award.
2. **Agreement Specialist:** A designee of the AO for administrative and financial aspects of the award.
3. **Program Officer:** Individual within NIH who provides day-to-day programmatic oversight of individual awards, working closely with the AO and with the Office that manages the Common Fund.

OT Agreement Governance

OT awards are not grants, cooperative agreements, or contracts. They are used by NIH for particular purposes as authorized by Congress, including in the execution of certain programs supported by the Common Fund. They provide considerable flexibility in establishing policies for the awards. Each award is therefore issued with a specific Agreement, which is negotiated with the recipient and details specific terms and conditions for that award. Policies and terms for individual OT awards may vary between awards, which may be expanded, modified, partnered, not supported, or later discontinued based on program needs, changing research landscape and or availability of funds. Program and administrative policies and the terms and conditions of individual awards are intended to supplement, rather than substitute for, governing statutory and regulatory requirements. Awards or a specified subset of awards also may be subject to additional requirements, such as those included in executive orders and appropriations acts, including the Other Transactions-authorizing legislation cited in the Notice of Award (NoA), as well as all terms and conditions cited in the NoA and its attachments, and conditions on activities and expenditure of funds in other statutory or regulatory requirements, including any revisions in effect as of the beginning date of the next funding segment. The terms and conditions of the resulting OT awards are intended to be compliant with governing statutes.

For the awards funded under this Research Opportunity Announcement, the NIH will engage in negotiations (before, during, and at the end of award) and all agreed upon terms and conditions will be incorporated into the Agreement.