

# ARTIFICIAL INTELLIGENCE SAFETY INSTITUTE CONSORTIUM COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

#### **Article 1. Introduction**

This Cooperative Research and Development Agreement ("Agreement" or "CRADA") is entered into by and between ("Collaborator") and the National Institute of

Standards and Technology ("NIST") within the Department of Commerce (Collaborator and NIST may be referred to herein individually as a "Party" or collectively as the "Parties"). This Agreement is effective as of the date of the last authorized signature hereto ("Effective Date") for the duration stated in the attached Appendix A, which by this reference is made a part hereof and incorporated herein.

### **Article 2. Purpose and Authority**

- 2.1 The purpose of the NIST Artificial Intelligence Safety Institute Consortium ("Consortium") is to establish a new measurement science that will enable the identification of proven, scalable, and interoperable measurements and methodologies to promote safe and trustworthy development and use of Artificial Intelligence (AI), particularly for the most advanced AI ("Purpose"). NIST does not endorse any of Collaborator's products or services that are used in the course of the Consortium. Notice of the Consortium was published in the Federal Register on November 2, 2023.
- 2.2 NIST enters into this Agreement pursuant to the authority granted to NIST under 15 U.S.C. 3710a. Collaborator enters into this Agreement with the understanding that each participant in this Consortium (each individually, "Consortium Member," and two or more, "Consortium Members") that is legally permitted to enter into a CRADA will be bound by the terms and conditions enumerated herein. If Collaborator is an agency or department of the Federal Government (such Consortium Members referred to individually and collectively herein as the "Federal Government"), to the extent that the terms of the attached Appendix B conflict with the terms of this Agreement, the terms of Appendix B shall control and by this reference are made a part hereof and incorporated herein. Entities that are legally prohibited or not legally authorized to enter into a CRADA may, at NIST's discretion, be permitted to participate in the Consortium under an agreement other than a CRADA with terms that will differ, as necessary, from the terms of this Agreement. Foreign governmental entities may, at NIST's discretion, be permitted to participate in the Consortium under an appropriate international agreement.

#### Article 3. Membership and Collaborative Research

- 3.1 **Membership.** Collaborator shall be a Consortium Member as of the Effective Date. Collaborator acknowledges that NIST may publicly display Collaborator's name as a Consortium Member, including on NIST's website.
- 3.2 **Research Plan.** The intended research and development activities of the Consortium are detailed in the research plan of Appendix A ("**Research Plan**"). The Research Plan shall be performed on a reasonable efforts basis.
- 3.3 Contributions and Personnel. NIST shall provide administrative and scientific supervision for the Consortium and the Research Plan, and the NIST individual responsible for managing the Consortium and the Research Plan ("NIST Consortium Manager") is identified at Appendix A. NIST's and Collaborator's respective contributions of material and equipment to the Consortium and the Research Plan (respectively, "NIST Contributions" and "Collaborator Contributions") are listed in Appendix A, as well as the respective personnel of each Party who will contribute to the Research Plan (for each Party, its "Project Team"). NIST cannot contribute funds to Collaborator under this Agreement; however, Collaborator may contribute funds under this Agreement.

- 3.4 **Contractors**. Collaborator acknowledges and agrees that non-federal researchers ("**Contractors**") who are not employees of NIST may work with NIST in the performance of the Research Plan.
  - i. The Parties acknowledge that a portion of the research activities described in the Research Plan may be performed by the employees, subcontractors, or consultants of non-federal organizations funded by NIST to perform such activities ("Contractors"). Contractors are not employees of NIST. Collaborator hereby permits NIST to share with Contractors any information, not including Proprietary Information, as that term is defined herein, that Collaborator provides pursuant to this Agreement. Collaborator and Contractors may enter into separate non-disclosure agreements to protect such Proprietary Information. Collaborator agrees that NIST shall not be responsible to Collaborator for any loss, claim, damage, or liability resulting from any unauthorized disclosure of such Proprietary Information by any Contractors.
  - ii. Contractors' individual personnel may vary throughout the performance of the Research Plan. The NIST Consortium Manager may periodically request an updated list of Contractors' personnel throughout the performance of the Research Plan, and such information shall be available to Collaborator at Collaborator's request. NIST shall be responsible for the employees of NIST only to the extent permitted by the Federal Torts Claims Act. NIST is not responsible for the conduct of any Contractors or their employees.
- 3.5 **Membership Fees.** Membership fees for each Consortium Member that enters into a CRADA are one thousand U.S. Dollars (\$1,000) per year. Collaborator will be invoiced according to the schedule in Appendix A. In addition to payment of the membership fees, Collaborator's intellectual contribution to the performance of the Research Plan is required.
- 3.6 **Working Groups**. The Research Plan will be performed through the concerted efforts of various working groups, as identified in Appendix A ("**Working Groups**"). The specific foci of the Working Groups may be shifted, expanded, or otherwise revised by amendment to this Agreement or at the discretion of the NIST Consortium Manager in response to, *inter alia*, feedback from Consortium Members, progressing priorities and interests in the AI field, and the results obtained by the Consortium in the performance of the Research Plan. After joining the Consortium, Collaborator will have an opportunity to designate the Working Groups in which it wishes to participate, but the NIST Consortium Manager shall have final authority in all assignments. It is the expectation of the Parties that each Consortium Member will participate in the activities of at least one of the Working Groups.
- 3.7 **Steering Committee.** The planning and coordination of the Working Groups' performance of the Research Plan, as well as the monitoring and reviewing of the Working Groups' technical progress, shall be overseen by NIST and the NIST Consortium Manager. For ease of administration, each of the Working Groups will be guided by a committee composed of members of the NIST Project Team and such other Consortium Members as the NIST Project Team deems necessary and beneficial ("**Steering Committee**"). Each Steering Committee will meet and confer with the NIST Consortium Manager in order to aid in the planning and coordination of the Working Groups' performance of the Research Plan and the monitoring and reviewing of the Working Group's technical progress. The NIST Consortium Manager shall have final authority in all decisions.
- 3.8 **Conduct.** Collaborator agrees that each member of its Project Team will abide by all applicable regulations, policies, and procedures relating to safety, security, and conduct and adhere to all applicable building and restricted area access controls while on NIST premises.

#### **Article 4. Proprietary Information and Publication of Results**

4.1 **Proprietary Information.** Collaborator agrees to limit its disclosure of Proprietary Information to NIST and/or the Federal Government under this Agreement to that which is necessary for the performance of the Research Plan. Collaborator agrees to place a Proprietary Information notice on all Proprietary Information that it delivers to NIST and/or the Federal Government under this CRADA. NIST may share Proprietary Information of Collaborator with the Secretary of Commerce or the Secretary's delegees only as necessary for purposes of the Consortium and in the performance of the Research Plan. Except where NIST and/or the Federal Government are legally obligated to release information pursuant to the Freedom of Information Act (5 U.S.C. 552) or other requirement of law, Collaborator's Proprietary Information shall not be disclosed or otherwise made available in any form to any person, firm, corporation, partnership, association, or other entity that is not NIST or the Federal Government

without the written consent of Collaborator. NIST agrees to use best efforts to maintain the confidentiality of Collaborator's Proprietary Information. NIST and/or the Federal Government will promptly notify Collaborator of requests for Collaborator's Proprietary Information. Collaborator agrees that NIST and the Federal Government are not liable for the disclosure of information designated as proprietary which, after notice to and consultation with Collaborator, NIST and/or the Federal Government determine may not lawfully be withheld or which a court of competent jurisdiction requires to be disclosed. For the purpose of this Agreement, "**Proprietary Information**" means the confidential and proprietary scientific, business, or financial information of Collaborator, which may embody trade secrets, when such information is developed exclusively at private expense, except if such information:

- was in the possession of NIST and/or the Federal Government before receipt from Collaborator with no attached obligation of confidentiality;
- ii. is or becomes a matter of public knowledge through no fault of NIST and/or the Federal Government;
- iii. is received by NIST and/or the Federal Government from a third party without a duty of confidentiality;
- iv. is disclosed by Collaborator to a third party without a duty of confidentiality on the third party;
- v. is independently disclosed by NIST and/or the Federal Government with Collaborator's prior written approval; or
- vi. is independently developed by NIST and/or the Federal Government without reference to information disclosed hereunder.
- 4.2 **Sharing of Proprietary Information Among Consortium Members.** Except as otherwise provided herein, NIST will not share Collaborator's Proprietary Information with any Consortium Members that are not Federal Government Consortium Members. Should Collaborator find it useful or necessary in the performance of the Research Plan to share its Proprietary Information with another Consortium Member that is not a Federal Government Consortium Member, Collaborator may coordinate with such Consortium Member to enable such sharing, including entering into an agreement with such Consortium Member to protect the confidentiality of Collaborator's Proprietary Information.
- 4.3 Whistleblower Protection Act. These provisions are consistent with and do not supersede, conflict with, or otherwise alter the employee obligations, rights, or liabilities created by existing statute or Executive order relating to (1) classified information, (2) communications to Congress, (3) the reporting to an Inspector General or the Office of Special Counsel of a violation of any law, rule, or regulation, or mismanagement, a gross waste of funds, an abuse of authority, or a substantial and specific danger to public health or safety, or (4) any other whistleblower protection. The definitions, requirements, obligations, rights, sanctions, and liabilities created by controlling Executive orders and statutory provisions are incorporated into this Agreement and are controlling.
- 4.4 **Research Results**. The Parties agree that all recorded data first produced by Collaborator and/or NIST in the performance of the Research Plan and during the term of this Agreement ("**Research Results**") shall be exchanged between Collaborator and NIST. NIST and Collaborator shall each have the right to use and disclose the Research Results in accordance with the terms of this Agreement and agree only to those delays in the public disclosure of the Research Results that are provided for herein.
- 4.5 **Publication of Research Results.** The Parties intend to publish the collective Research Results of all Consortium Members ("Collective Results"). Until the date of the first public disclosure of the Collective Results, Collaborator agrees not to disclose to any third party that is not a Consortium Member the Research Results, the Collective Results, or any other data or report provided to Collaborator under the terms of this Agreement. NIST will give each Consortium Member thirty (30) days to review the proposed publication of the Collective Results to ensure that no Proprietary Information is contained therein. Collaborator may publish Collaborator's Research Results after NIST publishes the Collective Results, provided that Collaborator's publication references the Consortium.
- 4.6 **National Security.** The Parties do not intend to undertake or make available the results of testing of specific AI models that may raise national security concerns, which fall outside the Purpose of the Consortium. If the Parties wish to undertake or develop results involving specific AI models for capabilities relating to areas of national security concerns, the Parties may enter into a separate agreement outside of the Consortium for such purpose.

#### **Article 5. Intellectual Property**

5.1 CRADA Inventions. The Parties do not intend to conceive of any invention in the performance of the Research Plan. The Parties agree that any invention conceived by the Parties and/or the Federal Government in the performance of the Research Plan ("CRADA Invention") shall be dedicated to the public domain to be freely used by all. The Parties agree to neither seek intellectual property protection for nor attempt to any enforce intellectual property rights, including copyright, in either any CRADA Invention or any work authored in the performance of the Research Plan, including any publication of the Research Results or Collective Results, or in any other intellectual property resulting from the performance of the Research Plan. Collaborator hereby acknowledges that NIST and/or the Federal Government, pursuant to any applicable requirements of 15 U.S.C. 3710a(b), has offered Collaborator the option to obtain a license to NIST's and/or the Federal Government's ownership in any CRADA Invention conceived by employees of NIST and/or the Federal Government, and that Collaborator affirmatively declines the option to license or acquire any interest in any such CRADA Invention. Although neither Party is permitted to seek patent protection for any CRADA Invention, Collaborator grants to NIST and the Federal Government a nonexclusive, nontransferable, irrevocable, paid-up license to practice or have practiced any CRADA Invention that is conceived solely by Collaborator's Project Team, throughout the world by or on behalf of the U.S. Government.

Contractors have the right under 35 U.S.C. 200 et seq. to elect to retain title to their interest in any invention created by Contractors' employees in accordance with 35 U.S.C. 202(c). Ownership rights stemming from any Contractors' elections are not subject to the provisions of this Agreement.

- 5.2 Ownership and Use of Collaborator Contributions. Collaborator grants to NIST and the Federal Government, and to other Consortium Members as NIST determines necessary, the right to use Collaborator Contributions for the purpose of the Consortium and in the performance of the Research Plan. Collaborator agrees to collaborate with other Consortium Members as necessary for the purpose of the Consortium and in the performance of the Research Plan. Except for the rights explicitly granted to NIST and the Federal Government and the Consortium Members herein, no other rights or licenses are granted by Collaborator. Collaborator Contributions are and will at all times remain the property of Collaborator. At the earlier of the conclusion of the Consortium, the expiration or termination of this Agreement, or the relevant amendment or updating of Appendix A, NIST and/or the Federal Government will return to Collaborator or request Collaborator's retrieval of all Collaborator Contributions not expended in the performance of the Research Plan at Collaborator's retrieval expense. The U.S. Government shall not be responsible for damage to Collaborator Contributions or other property acquired by NIST for the purpose of the Consortium and in the performance of the Research Plan. Equipment or material purchased by NIST with funds paid by Collaborator under this Agreement shall be the property of NIST.
- 5.3 **Contributions of Consortium Members.** Collaborator and NIST agree not to sell, distribute, sublicense, modify, translate, disassemble, reverse engineer, or otherwise alter the equipment and/or material contributed to the Consortium by NIST, Collaborator, and/or other Consortium Members.

#### **Article 6. Modifications and Termination**

- 6.1 Amendments. Should a need arise to modify Collaborator's Project Team, Collaborator Contributions, Collaborator's contact information, or any other Collaboration information contained in Appendix A, Collaborator shall notify the NIST Consortium Manager of the need for such modifications so that NIST may prepare an amendment to this Agreement. Should a need arise for NIST to modify the terms and conditions of this Agreement, the details of the Research Plan, or any other NIST information contained in Appendix A, NIST may propose to each Consortium Member that has signed a CRADA identical modifications in the form of a CRADA amendment. NIST will require all Consortium Members that have signed a CRADA to accept the same modified terms and conditions. NIST may terminate this Agreement immediately if such an amendment is not signed by Collaborator within the time prescribed by NIST. No amendment to this Agreement will be effective until fully signed by both Parties.
- 6.2 **Termination Notices**. Collaborator and NIST shall each have the right to terminate this Agreement, with or without cause. Termination by one Party will be effective upon thirty (30) days' prior written notice to the other Party.

6.3 **NIST's Right to Terminate.** NIST may terminate this Agreement immediately in the event of either of the following: (i) direct or indirect control of Collaborator is transferred to a foreign company or government; or (ii) if Collaborator is already controlled by a foreign company or government, such control is transferred to another foreign company or government.

#### Article 7. Miscellaneous

7.1 **Entire Agreement.** This Agreement constitutes the entire agreement of the Parties with respect to the matters set forth herein and supersedes and replaces in the entirety any prior understanding, written or oral, between the Parties concerning such matters. The Parties understand and agree that this Agreement supersedes and replaces in the entirety any subsequent software license or other type of license or evaluation agreement, including but not limited to any end-user license agreement (EULA), "shrink wrap," "click through," or any other type of license or other agreement provided with or required by Collaborator in connection with the use of Collaborator Contributions.

### 7.2 **Indemnification.**

- i. Acts by the Parties and/or the Federal Government. Collaborator shall indemnify and hold harmless the U.S. Government for any loss, claim, damage, or liability of any kind caused to or by Collaborator's Project Team arising in connection with this Agreement, except to the extent that such loss, claim, damage, or liability arises from the gross negligence or wrongful acts of NIST and/or the Federal Government and/or their employees. NIST's and the Federal Government's responsibility for payment of tort claims in connection with the performance of the Research Plan is governed by the Federal Tort Claims Act.
- ii. **Use by Collaborator of Consortium Research Results.** Collaborator shall indemnify and hold harmless the U.S. Government for any loss, claim, damage, or liability of any kind arising out of the use by Collaborator, or others acting on its behalf or under its authorization, of the research results of NIST and/or the Consortium Members, the Collective Results, or any other research and/or technical development or product arising from the Research Plan and/or received by Collaborator under this Agreement or out of any use, sale, or other disposition of the same by Collaborator or others acting on its behalf or with its authorization.
- 7.3 NO WARRANTY. ANY PROPERTY PROVIDED BY EITHER PARTY UNDER THIS AGREEMENT IS PROVIDED "AS IS" WITHOUT EXPRESS OR IMPLIED WARRANTY AS TO ANY MATTER WHATSOEVER, INCLUDING THE CONDITION OF ANY RESEARCH OR PRODUCT, WHETHER TANGIBLE OR INTANGIBLE, MADE OR DEVELOPED UNDER THIS AGREEMENT, OR AS TO THE OWNERSHIP, MERCHANTABILITY, NON-INFRINGEMENT, OR FITNESS FOR A PARTICULAR PURPOSE THEREOF.
- 7.4 **Use of Name or Endorsements.** Each Party agrees not to use the name of the other Party on any commercial advertisement or promotional material for any product or service that is directly or indirectly related to this Agreement or to the Consortium without prior written approval by the other Party, except that NIST may use, disclose, and publicly display Collaborator's name to identify the Consortium Members to third parties. For the avoidance of doubt, Collaborator shall be free to display or reference any Research Results made publicly available by NIST as long as such display or reference by Collaborator does not imply endorsement by NIST, the Department of Commerce, the U.S. Government or any subunit of the foregoing entities.
- 7.5 **Public Statements by Collaborator.** Collaborator may use the following text on its websites, publications, and promotional material without further approval by NIST.

"[Collaborator] is collaborating with the National Institute of Standards and Technology (NIST) in the Artificial Intelligence Safety Institute Consortium to establish a new measurement science that will enable the identification of proven, scalable, and interoperable measurements and methodologies to promote development of trustworthy Artificial Intelligence (AI) and its responsible use. NIST does not evaluate commercial products under this Consortium and does not endorse any product or service used. Additional information on this Consortium can be found at: [link to FRN]."

CRADA Identification Number:

Collaborator:

- 7.6 **Export Control.** NIST and the Federal Government comply with, and Collaborator shall comply with, all applicable export laws and regulations, including, but not limited to, the International Traffic in Arms Regulations (22 C.F.R. Part 121 *et seq.*) and the Export Administration Regulations (15 C.F.R. Part 730 *et seq.*), for all equipment, materials, and information shared under this Agreement. Collaborator agrees that it will not disclose any technical data created in the performance of the Research Plan that is controlled by U.S. export laws and regulations to any foreign national, firm, or country, including foreign nationals employed by Collaborator, without Collaborator first obtaining any necessary licenses or approvals.
- 7.7 **Assignment**. Neither this Agreement nor any rights or obligations of either Party hereunder shall be assigned or otherwise transferred by either Party without the prior written consent of the other Party.
- 7.8 Liability, IN NO EVENT WILL THE PARTIES BE HELD LIABLE FOR ANY LOST REVENUES, LOST PROFITS, OR ANY INCIDENTAL, INDIRECT, CONSEQUENTIAL, SPECIAL, OR PUNITIVE DAMAGES OF ANY KIND ARISING OUT OF THIS AGREEMENT, PARTICIPATION IN THE CONSORTIUM, OR THE PERFORMANCE OF THE RESEARCH PLAN. Collaborator enters into this Agreement with the understanding that each Consortium Member has agreed or will agree to the preceding statement.
- 7.9 **Governing Law**. The construction, validity, performance, and effect of this Agreement for all purposes shall be governed by the laws of the United States.
- 7.10 **Counterparts.** This Agreement may be signed in one or more counterparts, each of which shall be deemed to be an original and all of which when taken together shall constitute the same Agreement. Any signed copy of this Agreement made by photocopy, facsimile, or PDF Adobe format shall be considered an original.

[Signatures follow on next page]

| Collaborator:  |  |
|--|--|
| IN WITNESS WHEREOF, the Parties have caused this Agreement representatives as follows:   | at to be executed by their duly authorized |
| SIGNATORY FOR COLLABORATOR:  |  |
|  |  |
| Dat  | te   |
| Contact Information for Notices:   |  |
| SIGNATORY FOR NIST   |  |
| Jeffrey DiVietro, PhD Deputy Director, Technology Partnerships Office  | te   |
| 2 - p.u.y 2 incolor, 1 commonegy 1 incoloring common   |  |
| Contact Information for Notices:   |  |
| NIST Technology Partnerships Office<br>Attn: Consortia Officer<br>100 Bureau Drive, Mail Stop 2200, Gaithersburg, MD 20<br>agreements@nist.gov | 1899-2200                                  |
|  |  |

CRADA Identification Number:

# Appendix A Research Plan and Related Information

#### Part I. Project Information

- 1. Consortium Title: Artificial Intelligence Safety Institute Consortium (AISIC)
- 2. **NIST's Principal Investigator(s) and Consortium Manager(s)** (NIST PIs may change at NIST management's sole discretion.):

Kathleen Roberts

NIST Information Technology Laboratory

Office of Emerging Technologies

9700 Great Seneca Hwy, Rockville, MD 20850

Telephone: (202) 515-6213

E-Mail: kathleen.roberts@nist.gov

Jacob Taylor

NIST Physical Measurement Laboratory

**Quantum Measurement Division** 

100 Bureau Drive, Gaithersburg, MD 20899

Telephone: (301) 975-8586 E-Mail: jacob.taylor@nist.gov

- 3. **Duration of the Agreement:** From the Effective Date to January 15, 2029
- 4. NIST Project Team, Services, Facilities, Intellectual Property, and/or Equipment Contributions:

NIST Project Team (NIST management reserves the right to replace these staff members at its sole discretion.): Kathleen Roberts (PI), Jacob Taylor (PI), Elham Tabassi, Alycia Jayson, Timothy McBride, Reva Schwartz, Mark Przybocki, Yooyoung Lee, George Awad, Martin Stanley, Bilva Chandra, Mark Latonero, Jesse Dunietz, Kyra Yee

#### NIST Contributions:

- Personnel and expertise
- Hardware: devices, servers, switches, disks, racks, power units, UPS, and connectivity to the internet and server capacity to accomplish goals
- Software: virtualization, operating systems, and appropriate software applications

#### 5. Contractors:

To be determined and added via amendment

following foreign country:

#### Part II. Collaborator Information

### COLLABORATOR MUST CHECK ALL THE APPROPRIATE BOX(ES) BELOW.

| 1. | <b>Collaborator Eligibility.</b> Collaborator certifies the following to NIST, and the Collaborator agrees to notify NIST within thirty (30) days of any change in the following: |
|----|---|
|    | Collaborator certifies that it is incorporated or organized under the laws of one of the states or territories of the United States.  |
|    | Collaborator certifies that it is <b>not</b> subject to the control of any foreign government or foreign company.   |
|    | Collaborator certifies that it is partially or wholly owned by the following foreign government:  |
|    | Collaborator certifies that it is not owned by any government, but organized under the laws the   |

| CRADA Identification Number:  |
|---|
| Collaborator:   |
| Collaborator certifies that it is subject to the control of the following foreign parent company                        |
| (please also specify country):  |
|   |
| Collaborator certifies that it has a manufacturing presence in the United States.                                       |
| 2. Participation in other Federally Funded Projects. Collaborator certifies that:                                       |
| Collaborator's participation in this Consortium is not related to any research supported by othe Federal or NIST Funds. |
| Collaborator's participation in this Consortium is related to research supported by Federal                             |
| funding, which is identified as follows:  |
| (attach additional pages if necessary).   |
|   |
| 3. Collaborator's DUNs/UEI/TIN:   |
| 4. Collaborator's Principal Investigator(s):  |
| 5. Collaborator's Project Team and Materials and/or Equipment Contribution:   |
| 6. Collaborator Financial Contribution:   |
| Collaborator will contribute to NIST:   |
| \$1,000 due on Effective Date,  |
| \$1,000 due on January 15 <sup>th</sup> of the following calendar year, and   |
| \$1,000 due on January 15 <sup>th</sup> of each subsequent year thereafter, for the term of the CRADA.                  |
| Financial Contribution Information:   |
| Payment Instructions shall be sent to:  |
| Billing Contact Name/Payables Department:   |
| Phone Number:   |
| Mailing Address (if applicable):  |
| Email Address (if applicable):  |
| Special Instructions or description to be included on correspondence:   |

#### Part III. The Research Plan

#### **Approach**

NIST and Consortium Members will collaboratively work to establish a new measurement science that will enable the identification of proven, scalable, and interoperable measurements and methodologies to promote safe and trustworthy development and use of AI. The work will particularly focus on evaluations, red-teaming, and safety mitigations for advanced AI models, particularly for capabilities involving risks related to security, misuse, and control, with initial activities focused in support of the Executive Order of October 30, 2023 (The Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence). The AISIC will stimulate activities for knowledge transfer, information sharing, and an array of basic and applied research and evaluation activities to spur innovation and advance trustworthy and responsible AI. The AISIC will publish results and guidelines as appropriate throughout the duration of the Consortium.

## Consortium Member Contributions:

- Technical expertise in one or more of the following areas
  - o Data and data documentation
  - AI Metrology
  - o AI Governance
  - o AI Safety
  - Trustworthy AI
  - o Responsible AI
  - o AI system design and development
  - o AI system deployment
  - o AI red-teaming
  - o Human-AI Teaming and Interaction
  - o Test, Evaluation, Validation, and Verification methodologies
  - Socio-technical methodologies
  - AI Fairness
  - o AI Explainability and Interpretability
  - Workforce skills
  - Psychometrics
  - Economic analysis
- Models, data, and/or products to support and demonstrate pathways to enable safe and trustworthy AI systems
- Infrastructure support for Consortium projects in the performance of the Research Plan
- Facility space and hosting of Consortium Members' participants, workshops, and conferences

#### Activities of the AISIC:

#### **Project Direction**

- Establish a knowledge and data sharing space for AI stakeholders
- Engage in collaborative and interdisciplinary research and development through the performance of the Research Plan
- Prioritize research and evaluation requirements and approaches that may allow for a more complete and effective understanding of AI's impacts on society and the U.S. economy
- Identify and recommend approaches to facilitate the cooperative development and transfer of technology and data between and among Consortium Members
- Identify mechanisms to streamline input from federal agencies on topics within their direct purviews
- Enable assessment and evaluation of test systems and prototypes to inform future AI measurement efforts

Technical work will initially be organized around the following Working Groups, with revisions directed as appropriate by the NIST Consortium Manager:

#### CRADA Identification Number:

#### Collaborator:

- Working Group #1: Risk Management for Generative AI
  - Develop a companion resource to the AI Risk Management Framework (AI RMF) for generative AI
  - o Develop minimum risk management guidance geared toward federal agencies
  - Operationalize the AI RMF
- Working Group #2: Synthetic Content
  - O Identify the existing standards, tools, methods, and practices, as well as the potential development of further science-backed standards and techniques, for authenticating content and tracking its provenance; labeling synthetic content, such as using watermarking; detecting synthetic content; and preventing generative AI from producing child sexual abuse material or producing non-consensual intimate imagery of real individuals; testing software used for the above purposes; and auditing and maintaining synthetic content
- Working Group #3: Capability Evaluations
  - Create guidance and benchmarks for evaluating and auditing AI capabilities, with a focus on capabilities through which AI could cause harm, such as in the areas of chemical, biological, radiological, and nuclear (CBRN), cybersecurity, autonomous replication, control of physical systems, and other areas
  - Develop and aid in ensuring the availability of testing environments, such as testbeds, to support the development of safe, secure, and trustworthy AI technologies
- Working Group #4: Red-Teaming
  - Establish appropriate guidelines, including appropriate procedures and processes, to enable developers
    of AI, especially of dual-use foundation models, to conduct AI red-teaming tests to enable deployment
    of safe, secure, and trustworthy systems
- Working Group #5: Safety & Security
  - Coordinate and develop guidelines related to managing the safety and security of dual-use foundation models.

# Appendix B FEDERAL AGENCY & DEPARTMENT COLLABORATOR ADDENDUM

If Collaborator is an agency or department of the Federal Government, Collaborator and NIST enter into this Cooperative Research and Development Agreement subject to the following additional terms:

- 1. The NIST Consortium Manager is responsible for the scientific and technical conduct of the Research Plan on behalf of NIST. Collaborator's designated Principal Investigator(s) is responsible for the scientific and technical conduct of the Research Plan on behalf of Collaborator.
- 2. Neither NIST nor Collaborator may contribute funds to Consortium Members that are not part of the U.S. Government under this Agreement. Collaborator may either provide funds or in-kind contributions to NIST. Collaborator will be invoiced according to the schedule in Appendix A. At NIST's discretion, an in-kind contribution of equivalent value may be substituted for all or part of the membership fee. An in-kind contribution is a non-monetary contribution that may take the form of personal property (e.g., equipment and supplies), capital equipment, real property, work to be performed at either Party's facilities, and/or services that are directly beneficial, specifically identifiable, and necessary for the successful performance of the Research Plan.
- 3. Collaborator agrees to notify NIST and the relevant Consortium Member if Collaborator receives a request for such Consortium Member's Proprietary Information from an entity that is not NIST or the Federal Government.
- 4. Collaborator may terminate this Agreement immediately upon written notice to NIST.
- 5. The obligations to indemnify and hold harmless NIST identified in Section 7.2 do not apply to Collaborator.
- 6. Collaborator need not provide responses at sections 1, 2, and 3 of Part II of Appendix A.