ANNEX
BETWEEN
THE NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
GEORGE C. MARSHALL SPACE FLIGHT CENTER
AND
INTERNATIONAL BUSINESS MACHINES (IBM) CORPORATION THROUGH ITS
THOMAS J. (T.J.) WATSON RESEARCH CENTER
UNDER SPACE ACT UMBRELLA AGREEMENT
NO. SAA8-2233097, DATED 11/2/2022
ANNEX NUMBER SAA8-2233097.2

ARTICLE 1. PURPOSE

This Annex shall be for the purpose of research and collaboration to address data discovery, access and use challenges specifically related to large volumes of NASA science data. Under this annex, both parties seek to address these challenges in four focus areas:
a. development of a Geospatial Foundation Model (GFM);
b. expansion of Large Language Models (LLMs) referenced in the Agreement No. SAA8-2233097 using open source architectures;
c. deployment of those developed models (GFM and LLMs) on IBM Foundation Model Platform (watsonx.ai); and
d. evaluation of those developed models (GFM and LLMs) on IBM AI System One technologies.

Research goals for each of the subject areas are as follows:

Geospatial Foundation Model (GFM)
1. Build a GFM using science data provided by NASA (e.g., Harmonized Landsat Sentinel-2 (HLS);
2. Evaluate, benchmark and fine-tune the model for specific downstream tasks; and
3. Publicize the collaboration and its results to the scientific community as an open source version of the model

Large Language Models (LLMs) Based on Open Source Architectures
1. Train a LLM using science data provided by NASA from multiple science domains to perform multiple natural language processing tasks;
2. Evaluate open source architectures for use and then benchmark and fine-tune the model for specific downstream tasks including but not limited to Name Entity Recognition (NER), document classification, sentiment analysis, entity relations, and question answering; and
3. Publicize the collaboration and its results to the scientific community in the question-answering pipeline to showcase the models interactive search capabilities

GFM & LLMs on IBM Foundation Model Platform (watsonx.ai)
1. Deploy an open-source version of GFM and LLMs on the IBM Foundation Model Platform (watsonx.ai) for evaluation on NASA's Science Managed Cloud Environment (SMCE); and
2. Train NASA on how to use the IBM Foundation Model Platform (watsonx.ai) for both GFM and LLM models training and validation

GFM & LLMs on IBM AI System One
1. Compile the developed models (e.g., one pre-trained or fine-tuned GFM and LLM) for IBM Artificial Intelligence Unit (AIU);
2. Evaluate and validate the performance and accuracy of model inferencing (e.g., key performance indicators including but not limited to latency, throughput per Watt at small and large batch sizes); and
3. Demonstrate the complied and optimized models on the IBM Foundation Model Platform (watsonx.ai)

For each subject area listed, Parties agree to meet regularly, and corporately prepare status reports, and a final report summarizing the work performed pursuant to the work completed under this Annex. The meeting times, frequency, format, and contents of the reports will be as agreed upon by both Parties..

The legal authority for this Annex, consistent with the Umbrella Agreement, is in accordance with the Space Act, Other Transactions Authority (OTA), 51 U.S.C. § 20113(e).

ARTICLE 2. <u>RESPONSIBILITIES</u>

A. NASA MSFC will use reasonable efforts to:

1. Provide access to science datasets (e.g., Harmonized Landsat Sentinel-2 (HLS) and on other domain specific science datasets as appropriate) for IBM to develop a GFM, and train a LLM provided as part of Annex SAA8-2233097.1;
2. Work with IBM to establish evaluation criteria and benchmarks to assess the performance of the GFM, LLM, and the optimization of the models once deployed on the the IBM Foundation Model Platform (watsonx.ai);
3. Work with IBM to evaluate the performance and accuracy of the developed models on the IBM AI System One; and
4. Work with IBM to conduct outreach and promotion activities regarding the work accomplished under this Annex.

B. Partner will use reasonable efforts to:

1. Develop a GFM and train a LLM provided as part of Annex SAA8-2233097.1 based on the science datasets provided by NASA;
2. Work with NASA to establish evaluation criteria and benchmarks to assess the performance of the GFM and LLMs;

3. Conduct performance evaluations and fine tuning of the GFM and LLMs on downstream tasks;

4. Deliver the initial models (e.g., GFM as an open source version and LLM in the PrimeQA pipeline) and provide training and insights;

5. Deploy the GFM and LLMs on the IBM Foundation Model Platform (watsonx.ai) on the science managed cloud environment;

6. Work with NASA to conduct outreach and promotion activities regarding the work accomplished under this Annex;

7. Evaluate the developed models for performance assessment on IBM AI System One; and

8. Deliver final report(s) to NASA for the work accomplished under this Annex.

ARTICLE 3. SCHEDULE AND MILESTONES

The planned major milestones for the activities for this Annex defined in the "Responsibilities" Article are as follows:

1. NASA will provide access to the science datasets	Within three (3) months of the Effective Date and continuing until Agreement ends
2. NASA and IBM will establish evaluation criteria and benchmarks	Within three (3) months of the Effective Date
3. IBM will complete preliminary development of GFM and LLM	Within 6 months of the Effective Date
4. Parties will jointly participate in outreach and dissemination activities (e.g., joint peer-reviewed conference journal articles) to share the impact and results	Upon completion of Milestone 3 and continuing until Annex ends
5. IBM will deploy and evaluate the GFM and LLM on the IBM Foundation Model Platform and on the IBM System One	Upon completion of Milestone 3 and continuing until Annex ends
6. IBM will deliver the final report for the completed GFM and LLMs	Within 24 months of the Effective Date

ARTICLE 4. FINANCIAL OBLIGATIONS

There will be no transfer of funds between the Parties under this Agreement and each Party will fund its own participation. All activities under or pursuant to this Agreement are subject to the availability of funds, and no provision of this Agreement shall be interpreted to require obligation or payment of funds in violation of the Anti-Deficiency Act, (31 U.S.C. § 1341).

ARTICLE 5. <u>INTELLECTUAL PROPERTY RIGHTS - DATA RIGHTS</u>

A. Data produced under this Annex which is subject to paragraph C. of the Intellectual Property Rights - Data Rights Article of the Umbrella Agreement will be protected for the period of two years.

B. Under paragraph H. of the Intellectual Property Rights - Data Rights Article of the Umbrella Agreement, Disclosing Party provides the following Data to Receiving Party. The lists below may not be comprehensive, are subject to change, and do not supersede any restrictive notice on the Data provided.
1. Background Data:
The Disclosing Party's Background Data, if any, will be identified in a separate technical document.
None
2. Third Party Proprietary Data:
The Disclosing Party's Third Party Proprietary Data, if any, will be identified in a separate technical document.
None
3. Controlled Government Data:
The Disclosing Party's Controlled Government Data, if any, will be identified in a separate technical document.
None
4. The following software and related Data will be provided to Partner under a separate Software Usage Agreement:
None

ARTICLE 6. <u>TERM OF ANNEX</u>

This Annex becomes effective upon the date of the last signature below ("Effective Date") and shall remain in effect until the completion of all obligations of both Parties hereto, or two years from the Effective Date, whichever comes first, unless such term exceeds the duration of the Umbrella Agreement. The term of this Annex shall not exceed the term of the Umbrella Agreement. The Annex automatically expires upon the expiration of the Umbrella Agreement.

ARTICLE 7. <u>RIGHT TO TERMINATE</u>

Either Party may unilaterally terminate this Annex by providing thirty (30) calendar days written notice to the other Party.

ARTICLE 8. <u>POINTS OF CONTACT</u>

The following personnel are designated as the Points of Contact between the Parties in the performance of this Annex.

Technical Points of Contact

NASA George C. Marshall Space Flight Center	International Business Machines (IBM) Corporation through its Thomas J. T.J. Watson Research Center
Dr. Rahul Ramachandran Senior Research Scientist Mail Suite: ST11 Marshall Space Flight Center, AL 35812 Phone: 256-961-7620 rahul.ramachandran@nasa.gov	Raghu Ganti Principal Research Staff Member 1101 Kitchawan Road Yorktown Heights, NY 10598-0001 Phone: 914-945-2164 rganti@us.ibm.com

ARTICLE 9. MODIFICATIONS

Any modification to this Annex shall be executed, in writing, and signed by an authorized representative of NASA and the Partner. Modification of an Annex does not modify the terms of the Umbrella Agreement.

ARTICLE 10. SIGNATORY AUTHORITY

The signatories to this Annex covenant and warrant that they have authority to execute this Annex. By signing below, the undersigned agrees to the above terms and conditions.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION GEORGE C. MARSHALL SPACE FLIGHT CENTER

INTERNATIONAL BUSINESS MACHINES (IBM) CORPORATION THROUGH ITS THOMAS J. T.J. WATSON RESEARCH CENTER

BY: *David Burns*

David M. Burns, Ph.D.
Manager, Science and Technology Office

BY: *Jan Garrett-Hoffman*

Jan Garrett-Hoffman, PhD
STSM, Manager Research Government Programs

DATE:_____7/12/23_____   DATE:_____7/10/2023_____