NIST GenAI FAQ

(Updated: June 21, 2024)

Please note that this FAQ is by no means comprehensive and will be updated as needed and based on the questions we receive from the community.

NIST GenAI

What is NIST GenAI?

NIST GenAI is an umbrella program that supports various evaluations for research in Generative AI in different modalities. It is designed to measure and understand the capabilities and limitations of generative AI state-of-the-art (SOTA) technologies. The initial effort of the program aims to learn the capabilities of SOTA generative AI systems for creating content that appears plausible and indistinguishable from human-produced content. Simultaneously, the program also aims to learn the capabilities of SOTA detection systems for discriminating between AI-generated and human-produced content, as well as between plausible and implausible content. The information gathered from this program will help stakeholders (e.g., government, private sector, and academia) develop approaches for ensuring a trustworthy information ecosystem. Our goals are that the evaluation results and findings will promote information integrity and guide the responsible and safe use of digital content.

The program will be carried out in multiple stages. The first stage is the GenAI Pilot (testing text modality). During the pilot stage, our primary focus is to implement and test the necessary pipeline or infrastructure for the evaluation series. A secondary objective is to make a preliminary assessment of the ability of SOTA generative AI systems to create content that is indistinguishable from human-produced content, as well as the ability of SOTA detection systems to differentiate between AI-generated or human-produced content. The next stage will address both the "indistinguishability" and "plausibility" aspects of AI-generated content. Details for subsequent stages of evaluation series will be decided based on lessons learned from the previous stages and feedback from participants and stakeholders.

Why are there two types of participants?

GenAI evaluation series is a generative adversarial (cat-and-mouse) test framework between content generators and discriminators. “Generator (G)” teams will be tested on the ability of their systems to generate synthetic content that appears credible and is indistinguishable from human-produced content. “Discriminator (D)” teams will be tested on the ability of their systems to detect synthetic content created by generative AI models such as large language models (LLMs) or deepfake tools. Each type of participant will be able to improve the performance of their model(s) by this evaluation framework and performance measure (e.g., AUC). For example, Generator participants will attempt to improve their models in order to lower the AUC value of the discriminator models while discriminator participants will attempt to improve their models to increase the AUC value of their models when applied to G-participants' data. This is an ongoing iterative process.

What are the primary goals for the NIST GenAI Pilot?

The goals of NIST GenAI Pilot include, but are not limited to:

(a) Develop necessary pipelines or systems for both NIST GenAI team and G&D participants
(b) Develop performance metrics for measuring the capabilities and limitations of SOTA *generative* AI models.
(c) Develop performance metrics for measuring the capabilities and limitations of SOTA *discriminative* AI models.
(d) Understand attributes of AI-generated content that are helpful in discriminating it from human-produced content.
(e) Understand attributes of AI-generated content that can help discriminate between content that appears plausible from content that appears implausible.
(f) Evolve benchmark datasets in the generative adversarial testing domain.
(g) Promote the development of technologies for identifying the source of fake or misleading information.
(h) Share the information gathered and lessons learned with participants and stakeholders.

How will participants know how their system performed relative to another system?

For G-participants, the G-performance metrics will be displayed on a leaderboard to inform each participant (anonymized) on their performance relative to others.
For D-participants, the D-performance metrics will be displayed on a leaderboard to inform each participant (anonymized) on their performance relative to others.

What are the modalities being tested under GenAI?

In the first pilot stage, the program only addresses the “text” modality. We plan to add more modalities in the coming evaluation cycles (e.g., image, audio, video, code, multi-modality)

How can the community know where the program is and what is happening over time?

Please check the schedule page on our website: https://ai-challenges.nist.gov/t2t#tab_schedule

Will GenAI coordinators meet with teams to clarify issues and answer any questions?

Yes, the GenAI team will hold a webinar after registration closes to answer any questions from the community and teams registered. Also there will be a workshop after each evaluation cycle ends to present overall conclusions and give opportunity to teams to communicate their learned lessons.

Can the GenAI team and/or NIST endorse top Discriminator (D) teams?

No. NIST and the GenAI team can not endorse any team per NIST policy

Can top teams use the published leaderboard results to advertise their technology and imply NIST and the GenAI endorsement and recommendation?

No. NIST and the GenAI team can not endorse any team per NIST policy. Teams are allowed to include their results as part of academic publications and reports. None of these publications should imply NIST or the GenAI program endorsement or recommendations for the team's system. All communicated results in these publications have to report the exact experiments' conditions the team's system operated at (e.g. dataset, metrics) as well as other participating team's results.

What is Out of Scope for NIST GenAI?

Although the GenAI results and findings are expected to inform or contribute to standards and best practices development or risk and impact assessment in AI systems, the program, at this stage, *does not* directly address bias, fairness, or uncertainty measurement in AI systems or content provenance (e.g., watermarks). These topics and objectives may be added in subsequent evaluation cycles depending on the interest of the research community and stakeholders.

Registration

What are the prerequisites for participation in a GenAI challenge?

- Individuals may only participate on behalf of an organization.
- Organizations need to be legally registered/incorporated entities.
- Both domestic and foreign organizations may request to participate.
- Organizations interested in participating should submit their qualification information upon request and international organizations may require IAAO approval.

How do I access the GenAI evaluation web server?

Follow these steps to access the GenAI evaluation web server:

1. Go to https://ai-challenges.nist.gov/users/sign_in.
2. If you do not already have one, create a login.gov account.
3. Sign in with your login.gov account.

What are the steps to register for GenAI challenges?

Follow these steps to register for a GenAI challenge:

1. **User profile**: Complete user profile. All fields are required. Remember you can not participate as an individual, only on behalf of an organization.
2. **Site**: Create a site, or request to join a site that a collaborator has already created. The site name may appear in results; choose a name you are comfortable with.
3. **Tasks**: Register for all tasks you wish to participate in. This is a necessary step that is separate from the data agreements for the tasks.
4. **Data agreements**: See separate section on data agreements for specific instructions.

NIST will review the request to participate and let you know of the outcome or any issues that need to be addressed.

Data Agreements

How do I complete the necessary data agreements for the Generator Challenge?

Proceed as follows to complete the **two** data agreements necessary for the Generator Challenge:

1. GenAI Generator DUA (data usage agreement):
   - 1.1. Download DUA template via your account on web server.
   - 1.2. Complete form.
   - 1.3. Upload completed DUA via your account on web server.

2. GenAI Generator DTA (data transfer agreement):
   - 2.1. Download DTA template via your account on web server.
   - 2.2. Complete form.
     - 2.2.1. Leave identifier in the header empty (will be filled by NIST).
     - 2.2.2. Make sure to provide your organization’s name in the field in the first paragraph.
     - 2.2.3. ***See note below on who is authorized to sign the DTA!***
   - 2.3. Return via email to NIST Technology Partnerships Office (TPO) at agreements@nist.gov, cc’ed to genai-poc@nist.gov.

3. Wait for confirmation and further instructions from NIST’s GenAI team.

Can I modify the Generator DTA?

No. NIST cannot accept any changes to the terms of the Data Transfer Agreement. All organizations are participating under the same terms.

Who is authorized to sign the Generator DTA?

General guidelines:

- **U.S. universities**: The signatory is usually the director of one of the following or a senior contracts officer:
    - Office of Programs
    - Sponsored Research Office
    - Research Administration Office
    - Contracts and Grants Office
- **Non-US universities**:
    - Similar to above (US Universities), *or*
    - If a Professor is authorized to sign a DTA, they should include a statement in their request noting that they are authorized to sign
- **Commercial or non-profit entities (please include title)**:
    - Director
    - President
    - Vice-president
    - …
- **Government**:
    - Varies

***** If in doubt, please contact NIST TPO at agreements@nist.gov with your organization's information to receive guidance on who is authorized to sign. *****

How do I complete the necessary data agreement for the Discriminator Challenge?

Proceed as follows to complete the data agreement necessary for the Discriminator Challenge:

1. GenAI Discriminator DUA (data usage agreement):
    1.1. Download DUA via your account on web server.
    1.2. Complete form.
        1.2.1. Make sure to indicate task selection (texts, images, or both).
    1.3. Upload completed DUA via your account on web server.
2. Wait for confirmation and further instructions from NIST's GenAI team.