

Statistical Analysis and Tool Development for the NIST GenAI Evaluation Program

(U.S. Citizenship Preferred)

Overview

This position involves statistical analysis and analysis tool development supporting the NIST GenAI evaluation series (<https://ai-challenges.nist.gov/genai>) within NIST's Information Technology Laboratory. The primary project is a Testing and Evaluation (T&E) framework for generative AI watermarking, where the associate will lead the statistical analysis of evaluation results and develop the statistical engine behind NIST-CARAT (**Calibrated Risk Assessment Tool** for authentication technologies), an interactive tool enabling policymakers to explore the empirical performance consequences of compliance-threshold choices. This work centers on characterizing the tradeoff between content quality and watermark resilience under routine image handling, using rigorous detection-performance analysis, calibration assessment, uncertainty quantification, and Bayes-risk estimation to produce internationally defensible threshold claims.

Beyond this project, the associate will contribute statistical and analytical support across the wider NIST GenAI evaluation portfolio, which spans the evaluation of generative AI technologies across multiple modalities (text, voice, image, video, and code). This includes experimental design, metric development, analysis of evaluation outputs, and the development of reproducible analysis pipelines and interactive reporting tools. The associate will actively participate in NIST measurement science and contribute to cutting-edge research and evaluation in generative AI.

Duties include but are not limited to:

- Analyzing generative AI evaluation results using detection and classification performance methods (ROC, partial AUC, equal error rate, Brier score) and Bayes-risk characterization
- Performing uncertainty quantification and calibration assessment to support defensible threshold and compliance claims
- Developing the statistical backend of NIST-CARAT, including the models mapping compliance-threshold choices to expected operational consequences
- Designing and implementing reproducible analysis pipelines from raw evaluation outputs through to estimates with calibrated uncertainty
- Building interactive analysis and reporting tools (e.g., dashboards, interactive plots) to communicate evaluation results to technical and policy audiences
- Processing large evaluation datasets, including work in GPU-accelerated, high-performance computing environments
- Exploring data through descriptive statistics and graphical displays
- Contributing to internal reports, workshop white papers, and submissions to international standards bodies (e.g., ISO/IEC JTC 1/SC 42 and SC 29)
- Explaining statistical and metrological concepts to non-statisticians, including policy and standards stakeholders

Required Skills, Expertise and Qualifications

- A Ph.D. in statistics, biostatistics or a closely related quantitative field
- Mastery of statistical analysis methods, including experimental design, detection/classification evaluation (ROC, partial AUC, EER, Brier score), calibration, uncertainty quantification, and Bayes-risk characterization
- Experience with Bayesian modeling and Monte Carlo / Markov Chain Monte Carlo methods
- Proficiency in a statistical computing and scripting language (e.g., R, Python) and in shell scripting, with version-controlled, reproducible analysis workflows
- Experience building interactive analysis tools and dashboards (e.g., R Shiny, interactive plotting libraries, Jupyter notebooks)
- Familiarity with generative AI tools, including large language models, and with AI test and evaluation
- Experience with high-performance or GPU-accelerated computing environments, and with containerization and workflow tooling (e.g., Docker, Argo Workflows) in a data-analysis context
- Strong communication skills in speaking, writing, and graphical display, including the ability to explain statistical concepts to non-statisticians

Employment Terms: This opportunity is to be an associate researcher in the NIST Statistical Engineering Division for a term of 1 year, with options to renew and/or pursue longer-term federal employment. Associate researchers are NOT Federal Employees, but they work aside NIST researchers. Relocation expenses will not be provided.

Salary: \$100,000-\$100,000

How to express interest: Interested candidates, U.S. Citizens preferred, who meet all of the required qualifications are invited to express their interest in the position by sending an updated CV to Julia Sharp at julia.sharp@nist.gov or apply at <https://engineering.gwu.edu/post-doctoral-fellow-statistical-analysis-and-tool-development-nist-genai-evaluation-program>.