

Uncertainty characterization of Artificial Intelligence and Machine Learning models

(U.S. Citizenship Preferred)

Overview

Building a Python-based numerical sandbox that abstractly simulates a step-by-step sequential process representing modern manufacturing. The simulations will involve complex and irreversible systems, where early choices constrain later ones and outcomes are delayed. Distribution-free statistical theory and metrological discipline will be directly integrated into the evaluation of black-box deep learning architectures. The main goal is to design independent statistical oversight layers to monitor AI behavior and catch systemic errors under dataset shift. Specifically, the research will target two major AI failure modes: 1) **Silent Overconfidence**: This occurs when an AI operates on shifted out-of-distribution data but continues to output incorrect predictions with high mathematical certainty. You will design and evaluate distribution-free calibration and uncertainty quantification (UQ) frameworks to force deep architectures to output honest, mathematically guaranteed coverage intervals. 2) **Rapid Failure Velocity**: Because automated systems execute instantly, a mis-calibrated AI agent can propagate a continuous string of systematic errors at runtime speed before human operators can intervene. You will develop independent tracking layers using advanced time series and multivariate monitoring methods to isolate small, sustained process drifts before they hit catastrophic boundaries.

Duties include but are not limited to:

- **Conduct a comprehensive survey** of state-of-the-art uncertainty quantification methods for AI models and tools, and software implementation of these methods. Develop test examples for evaluating their performance, conduct relevant simulation experiments, and publish the results.
- **Construct a parameterized, mathematical Python testbed** simulating a multi-stage sequential decision framework (inspired by manufacturing workflows) characterized by time-delays, irreversibility, and endogenous data generation loops (i.e., Abstract Process Simulation).
- **Develop functional, lightweight AI agent architectures** to interact with the simulated environment, establishing a controlled subject for statistical stress testing (i.e., Agent-Environment Implementation). Design independent statistical tracking infrastructure to model and monitor the in-control dynamics of streaming data independently of the agent's internal decision-making assumptions (i.e., Advanced Temporal and Multivariate Monitoring).
- **Implement and validate distribution-free uncertainty quantification frameworks** and evaluate scoring rules to bound neural network overconfidence. Formulate statistical methods to account for correlation structures within generated data streams. Develop automated techniques to inspect internal latent states and intermediate network layer activations during execution to flag out-of-distribution inputs.

- **Apply formal statistical decision theory** to balance the trade-offs between different types of AI errors, setting rules for when the agent can act on its own versus when it must escalate to a human operator.
- **Present results at meetings**, mostly internal and occasionally with external stakeholders; Ensuring that results, protocols, software, and documentation have been archived or otherwise transmitted to the larger organization.

Required Skills, Expertise and Qualifications

- Ph.D. in Statistics or related field
- Expertise in statistical uncertainty quantification, including simulation-based uncertainty propagation methods and conformal prediction.
- Expertise in time series analysis, spatial statistics, multivariate statistics, and hierarchical mixed-effects modeling.
- Expertise in statistical decision theory (Bayesian loss analysis, cost modeling)
- Expertise in Python
- Hands-on experience with deep learning libraries (such as PyTorch or TensorFlow), including the technical capability to extract and evaluate hidden-layer tensor activations. Ability to create and experiment with Agentic AI.

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-uncertainty-characterization-artificial-intelligence-and-machine-learning>.