Flamelit

We build data science systems

10308 Megan Ct
Frisco, TX 75035
https://www.flamelit.tech

27 September 2022

AI Risk Management Framework c/o Mark Przybocki
U.S. National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899
AIframework@nist.gov

Flamelit Consulting, LLC, offers the following public comment in response to NIST's forthcoming Artificial intelligence (AI) Risk Management Framework (RMF) companion Playbook. Flamelit is a consultancy that enables businesses, government agencies, and institutions to strategically adopt AI, unlock value from novel data sources, and engineer data-driven systems and processes.

AI is at Flamelit's core, and we applaud NIST's initiative in establishing the AI RMF and companion Playbook. In our comment, we identify a nuanced, but currently missing, concept for both the RMF and the Playbook. This comment addresses the missing concept, as well as offers a set of recommendations both to address its lack and to enhance Playbook adoption.

MEASURE is missing a crucial concept

AI can enhance decision capacities, which in turn reduces costs, decreases negative outcomes, and improves lives. The AI RMF drafts to date readily approach this concept through recommended principles and identifies plays to enhance governance of AI systems. What is not apparent in the current framework is guiding principles for correlated, interconnected systems; this comment is to ensure good practices for integrating measurement of systemic risk are included in the AI RMF Playbook planning.

Often, the information produced by artificial intelligence flows from the system's owner to other organizations and entities, creating in turn emergent and interdependent information systems. Such an emergent system's mistakes, errors, or other incidental gaps flow through the same channel as the system's outputs. These mistakes can compound. In the context of developing AI models, this idea is known as ***error propagation***. Error propagation has a real impact on everyday lives through medical mistakes, misjudgments in credit underwriting, performance gaps experienced in self- and assisted-driving technologies, and so on.

As AI systems become more complex and integrated into everyday life, good governance will require consideration beyond a single organization's governance boundary. System design accounting for error propagation ensures that errors attenuate rather than compounding. As an example of how to monitor for attenuation, a blunt approach measures where interoperating models are jointly wrong. Systems which are wrong the same way will compound errors.

It is crucial that test, evaluation, verification, and validation (TEVV) functionality and findings be accessible to systems that utilize AI system outputs. Methods of identifying, handling, testing, and measuring error propagation should be included in the Playbook.

Playbook Recommendations

This section reviews Flamelit's recommendations for ensuring error propagation is considered within the Playbook, as well as a recommendation for easier adoption of the Playbook.

Recommendation #1: Measure error propagation. Principles and plays that assist the Measure principle and enhance mitigation of error propagation include:

1. Make it easy to test and validate your predictions
    a. Provide elements of the training and validation data to downstream systems for performing propagation and integration testing. Options include
        i. Full datasets used in training process
        ii. Samples of the training dataset
        iii. Synthetically generated samples
        iv. Statistics describing the joint probability distributions
    b. Key questions:
        i. How can your downstream users guarantee your work?
        ii. What liability will be reduced if data is not available to users of the AI system?
2. Regularly perform and document end-to-end system testing with information consumers
    a. Elements include
        i. Making system testing available to system users and people impacted by those decisions
        ii. Model explainability, both at aggregate and individual output levels
        iii. Bias assessment
        iv. Reporting concept drift and data drift
        v. External system equilibrium impact: feedback should be taken into account where large-scale systems can impact more than an individual's outcome.

   b. Key questions include
      i. What systems provide input?
      ii. How have any recent changes in an input system impacted AI system performance?
      iii. How has the AI system context changed?
      iv. What are the key findings?
      v. How are the findings documented?
      vi. How can the findings be accessed?
3. Provide system uncertainty information as a standard feature
   a. Elements include
      i. Using conformal prediction and other best practices to standardize prediction and classification uncertainty
      ii. Sharing measures and fit metrics used in system training and development
   b. Key questions include
      i. How can users distinguish usefulness and trustworthiness of system outputs?
      ii. How will uncertainty be communicated?
4. Implement and appropriately staff feedback mechanisms
   a. Knowing where and when errors are occurring helps to limit negative human interaction with AI systems.
   b. Elements of the play include
      i. Using automated systems to collect downstream feedback as input for product enhancements, incident information, and bug reports
      ii. Ensuring a broad channel is available for non-standard feedback.
   c. Key questions include
      i. How are you collecting user feedback for bugs and issues?
      ii. If using an API, who uses it? How is it documented?
5. Automate measurement and TEVV through platform use
   a. Elements include
      i. Identifying core performance metrics before beginning training
      ii. Choosing metrics to be deployed within or alongside production frameworks
   b. Key questions include
      i. Are your development metrics collection integrated into quality assurance or automated testing frameworks?
      ii. Can your system's testing framework be a shared service among similar systems or downstream systems?

These plays are starting points to account for error propagation, part of ensuring that AI systems avoid unnecessary complexity and negative interactions for people and for society.

Recommendation #2: Name the Plays in the Playbook. This is a general recommendation for the Playbook. Naming the plays ensures that AI operators, developers, and users have a common phrasing for each play, which will decrease instances of miscommunication and enhance the Playbook memorability. A good example of naming plays can be found in CIO.gov's digital services Playbook (https://Playbook.cio.gov/).

Conclusion

We at Flamelit are excited to see NIST developing and organizing the AI Risk Management Framework. AI has the capacity to improve life and society in fundamental ways. We wholly support the development of the RMF and Playbook and look forward to seeing its completion and adoption.

Regards,

Thomas Roderick
Principal Consultant