Request for Information (RFI)

National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework (AI RMF)

Response to Request for Information (RFI)

Submission Deadline:
September 15, 2021

Submitted to:
Mr. Mark Przybocki
National Institute of Standards and Technology (NIST)
MS 20899
100 Bureau Drive
Gaithersburg, MD 20899
Phone: 301.975.3347
Email: AIframework@nist.gov

Submitted by:
KeyLogic Technologies Corp.
7927 Jones Branch Drive, Suite 5100
McLean, VA 22102
POC: Michele Smith, VP of Contracts
Email: michele.smith@keylogic.com
Phone: 443.539.9062
Fax: 304.296.9300

KEYLOGIC

TABLE OF CONTENTS

LIST OF TABLES

COMPANY OVERVIEW (RFI NOTICE)

1.1 Brief Description of the Company

KeyLogic Technologies Corp. (KeyLogic), is a rapidly growing midsize company focused on enabling data-driven decisions. We have developed a keen understanding of the role of data within and among organizations and how to unlock the potential of data as a strategic resource. We employ over 700 information and data systems designers, data scientists, software engineers, cybersecurity experts, science and technology (S&T) professionals, and mission experts serving government agencies with data-intensive missions from our main offices in McLean, VA, Alexandria, VA, Oak Ridge, TN; Morgantown, WV, and Dayton, OH. On staff, we have over 60 PhD scientists and engineers. IIA Tech will leverage our expertise in information science, big data analytics, artificial intelligence (AI), machine learning (ML), large-scale data management, agile software development, and emerging technologies to support the NIST AI Risk Management Framework (RMF).

1.2 Company Information

Table 1 provides KeyLogic's requested company information.

Table 1. KeyLogic's Company Information

Requested Information	KeyLogic's Response
Company Name	KeyLogic Technologies Corp.
Socio-Economic Status	Large Business
DUNS Number	117120644
Government and Commercial Clients	KeyLogic is a federal contractor providing professional services and IT services and solutions to a broad range of federal civilian and defense agencies. Our contracts range from a few hundred thousand dollars to over $330M. The latter contract provides hybrid multi-cloud support to the US Patent and Trademark Office (USPTO), supporting over 250 enterprise systems, 10s of petabytes of data, and 15,485 users plus the general public. Our customers for AI work include the Departments of Energy, Commerce, Defense, and Homeland Security.
Website Address	www.keylogic.com

RESPONSES TO RFI QUESTIONS

Based on our experience with implementing risk management systems (e.g., NIST 800-53) and artificial intelligence (AI) systems at enterprise scale, we believe that NIST's proposed foundation for an AI risk management framework (RMF) is solid. NIST's proposed, broadly two-dimensional foundation of attributes (accuracy, explainability, etc.) and management (e.g., identifying, assessing, etc.) resembles NIST's highly successful 800-53 two-dimensional framework of risks and controls. With the proposed AI framework, NIST has essentially created sequential control phases (*identifying, assessing, responding to, and communicating*). In this response, we build upon your solid foundation and offer some supplementary suggestions and caveats.

SUGGESTIONS

1. Integrate the AI foundation for use with conventional risk management calculations

Conventional risk management calculations have served enterprises well in problem domains as varied as civil engineering and human space flight. The conventional Risk calculation considers several variables:

- Threat (an event that can result in harm) (categorical)

- Vulnerability (a state or condition of the system that could allow the threat to cause harm) (categorical)
- Likelihood (the probability that the threat will occur) (numerical)
- Impact (the cost, usually ultimately expressed in dollars, if possible, of the harm) (numerical).

For every intersection of a Threat T_i ,Vulnerability V_j, and Asset A_k, there exists Risk $R_{i,j,k}$ such that:

$$R_{i,j,k} = L_i * I_k$$

Controls are designed to remove Vulnerabilities, reduce Likelihoods, and/or reduce Impacts. Although coarse-grained, this formula delivers vital governance and management benefits. By quantifying risk, it allows risk (and risk mitigation) to be measured, monitored, scheduled, budgeted, and compared. Integrating NIST's attribute/management foundation with conventional Risk Calculation requires only four strategies:

1.a. Convert attributes into threats
Although it may be instinctive to think of attributes as positive measures and instinctive to combine these measures into a positive performance envelope, RMFs work best with negative attributes. Thus, our suggestion would be to convert NIST's positive attributes (e.g., accuracy, interpretability, etc.) into negative attributes (inaccuracy, uninterpretibility, etc.).

1.b. Establish a project-centric orientation and define vulnerability categories for both the data universe and the processing steps
Processing steps include both feature engineering decisions and algorithm selection. For example, available data may already be biased, and certain algorithms, such as decision trees, are inherently more transparent and explainable than other algorithms, such as singular value decomposition. In this scenario, projects become the Assets in the Risk Calculation and data and processing characteristics become the Vulnerabilities.

1.c. Generate likelihoods via extended exploratory data analysis (EDA), preferably via peer review
Current EDA methodologies rarely consider such issues as to whether it is necessary to combine PII elements (such as name and birth date) in the same place (standard DevSecOps practices are to separate them in order to mitigate the impact of a confidentiality compromise). Another example issue is whether to stratify training samples by demographics instead of by class label.

1.d. Maintain project orientation when calculating Impacts
Although it may be tempting to adopt a more global orientation (e.g., customer, public, or legal), our experience persuades us to recommend maintaining a project orientation for Impacts; and making other perspectives (legal etc.) categories of Impacts. This maintains a consistent grain and permits simple summing up of Impacts, as well as permitting Impacts to slide in time just as actual projects do.

2. Make automated demonstrable compliance an official goal of the framework
Frameworks that generate text-based, stand-alone documents (such as FEA and DoDAF)—although well-conceived—do not actually hasten delivery, improve performance, or reduce the level of effort of demonstrating compliance with policy. Furthermore, they are typically out of date before they are even published. They require significant engineering hours that are outside of the value chain, which is why most organizations resist them. Automating demonstrable compliance

requires the use of ontologies, ontology matching, and declarative grammars in all aspects of DevSecOps (described below).

3. Expect that there will never be "one framework to rule them all"

Even a widely adopted and hugely successful standard such as 800-53 is not a single standard to rule them all. Indeed, 800-53 includes three Baselines, multiple Revisions, and optional integration touchpoints with FedRAMP. Furthermore, most organizations "tailor" it for their own reasons. Ontology matching can overcome the issues arising from being subject to multiple frameworks.

4. Make a machine-to-machine (M2M) declarative grammar (such as RDF) an official part of the Framework

In our experience, committing from the very beginning to make a declarative M2M grammar (such as RDF or OWL) an official part of the framework, greatly simplifies the process and increases the probability of success in achieving several important goals. First, it solves overlapping framework issues. Several semantic technologies come into play here. Ontology engineering allows us to translate relevant portions of each framework into a machine-readable declarative representation of concepts and the relationships between them (an ontology). Ontology matching allows us to define relationships between equal-rank ontologies, rather than trying to create an uber-ontology that subsumes all applicable ontologies (usually an impossibly complex task). Additionally, many challenges to the "fairness" of AI algorithms will be viewed under the auspices of consumer protection laws, which will cause different obligations in different States (and countries), as well as substantially identical obligations expressed in very different terminology.

The generation of competency questions assures that the networked ontologies are sufficiently expressive to demonstrate compliance. Ontologies integrate with declarative grammars and artifacts from the DevSecOps world (such as Kubernetes manifests), so that a semantic query language (such as SPARQL) can drill all the way down to deployment implementations to demonstrate compliance in real time with fast-executing queries, rather than a lengthy and laborious set of stakeholder meetings, conversations, and white boarding sessions. Furthermore, machines keep the system constantly up to date, instead of humans struggling to provide a snapshot in time periodically (usually every two years).

Finally, ontologies conquer cultural divides. Ontologies support multiple perspectives simultaneously (e.g., mission, business, data, services, applications, infrastructure, and security). In our experience, overall, when demonstrating compliance, "drilling down" involves crossing a cultural divide from "policymaker" to "engineer"; "aggregating up" involves a reverse process of crossing the cultural divide of "engineer" to "policymaker". When humans demonstrate compliance, the cultural divide results in delay, human translation, and out-of-band asynchronous communications that often escape peer review. Semantic queries, however, cross the cultural divide instantly.

5. Make reproducibility of results a first-class citizen of the AI RMF

In our experience, reproducibility does not receive the attention it deserves. Reproducibility is even more challenging in an AI context than in a typical Big Data context, because AI algorithms are often re-tuned periodically against streaming data to check for model drift. An allegation of demographic bias several years ago is difficult to reproduce and therefore difficult to refute. Most legal causes of action have statutes of limitation of four years or more; most cases take two years to go to trial; most e-Discovery rules permit pretrial discovery of all information that is relevant or likely to lead to relevant information. Many AI projects contain thousands of engineered

features. They consist of ensemble model stacks of hundreds of models. They undergo hourly, daily, or weekly re-tuning, using enormous data populations streaming at ultra-high rates.

CAVEATS

Our experience with enterprise-scale AI and enterprise-scale policy frameworks convinces us that any AI RMF faces a number of inescapably asymptotic hard limits, analogous to the speed of light or the Heisenberg Uncertainty Principle. These limits largely concentrate on four of NIST's proposed attributes: Explainability, Fairness, Ease of Adoption, and Agility.

1. Explainability

Scientists constantly struggle with issues of transparency, explainability and interpretability. Scientists employ an array of mature techniques (the "scientific method") to advance our knowledge: null hypotheses, confidence intervals, experimental design, etc. However, some of the drivers for Explainable AI originate outside the scientific community. Politically active groups with various agendas are advocating legal and quasi-legal requirements and prohibitions in AI. Our experience suggests that there are several caveats that apply to Explainable AI:

1.a Juries don't know math

Imagine being the unfortunate litigator who needs to explain to a jury that “support” vector machines classify observations by expanding them to a potentially infinite number of dimensions, and dividing them with hyperplanes, without actually computing that infinite-dimensional space, by using the “kernel trick.” Likewise, juries are not going to understand linear algebra, word embeddings, convolutional neural networks, stochastic gradient descent, etc. Some experienced litigators may counter that litigators have always done this, but our estimation is that the complexity of explaining AI to a jury is several orders of magnitude more complex than past or existing complex litigation.

1.b All the data, models, and results must be made available and reproducible during discovery

Imagine a class action lawsuit where the plaintiffs allege a racial bias, and:

- The Defendant cannot provide the data (and engineered features) as they existed at the time of the alleged discrimination.
- Neither the Plaintiff's nor the Defendant's Expert Witness can generate the same (or even similar) model ensembles from the data.
- Neither the Plaintiff's nor the Defendant's Expert Witness can generate the same inference result for the Class Representative Plaintiff.

1.c. Routine scientific skepticism becomes a liability and results in a chilling effect

As stated above, scientists usually disagree on the explanatory power of models and their interpretation. In a legal context, such routine skepticism (e.g., “science”) becomes a liability and tends to result in a chilling effect unless protected by law. For example, evidence of subsequent improvements to a product cannot be introduced as evidence of a defect in a product liability case, because to allow such evidence would discourage the manufacturer from making safety improvements to the product.

2. Fairness

2.a. It's not about just removing demographic fields from the data and then you're done

“Fairness” does not mean simply removing all demographic observations and then your model is “fair.” There are many “proxy variables” for demographic variables. Furthermore, a vast number

of features are engineered, not observed. Massive online retailers such as Amazon have entities (such as "Customer") that contain thousands of features, mostly engineered. This means that even after ruthlessly removing "demographic" features, in the process of engineering synthetic features, you are probably re-introducing demographic proxies in a very hard-to-trace way. Even natural language processing techniques that employ sequence-based techniques such as skip-grams detect ethnic-based nuance in language and preserve them in the word embeddings they generate.

2.b. Census-aware models are inherently demographic and inherently unreproducible at micro-scale

Census data are inherently demographic. Because ethnicities are not uniformly distributed across census tracts, any modeling which uses census data will be vulnerable to allegations of ethnic bias. For example, site analysis (predicting the optimal place to establish a new hospital, police station, or Big Box Store) will always have a difficult time convincing critics that the site analysis was not ethnically biased. The more detailed American Community Survey (and many health-related surveys that use census polygons) compound this problem through anonymization. Ironically, anonymization makes allegations of bias harder to refute, not easier! This is because when the number of observations is small in a polygon, out of fear that people could guess the identities of those people, the statistic is either truncated to zero, or fuzzed using the Laplace Method. This adds a stochastic element and inter-statistic aggregation inconsistencies.

2.c. Fairness constraints still must work mathematically

Whatever constraints or burdens are applied to ensure "fairness" must still work mathematically. AI is math! Linear algebra, tensor calculus, approximation, propagation and feedback, regularization, activation functions, etc. still must work. We cannot inject if-then-else logic into latent Dirichlet allocation. We cannot take the derivative of a normative rule.

2.d. New protected classes are emerging rapidly

Just a decade ago, there were generally viewed to be two sexes and gender was a language concept. Now some argue there are over a hundred genders, and they are fluid. Not only is the hyper-subdivision breathtakingly rapid, but it will be unevenly acknowledged or recognized in different jurisdictions and by different advocacy groups.

2.e. Scientists and advocates are working at cross-purposes: the Great Equality v. Equity Debate has reached AI

In our experience, most AI practitioners are extremely conscientious. They are highly dedicated to removing ethnic bias, maximizing anonymization, and minimizing the amount of PII in their work. They are hard at work developing Best Practices for this. They may not be aware, however, of a growing divide between political groups seeking to safeguard equality (equal opportunity; selections based on some concept of "merit," not identity) and those seeking to promote equity (equality of outcome based on group identity). AI practitioners do not currently have a method to achieve these competing objectives.

3. Risk Reduction and Ease of Adoption are always inversely proportional

Most engineers are acutely aware of this. All RMFs exhibit this inverse relationship, and any AI RMF will be no exception. Right now, AI software frameworks and architectures are becoming spectacularly user-friendly (e.g., keras or the tidymodels R ecosystem), and spectacularly scalable (e.g., "[Fill-In-The-Blank] on Spark"), making *ad hoc* AI a real possibility. Ironically, introducing any AI RMF at this moment is going to re-introduce administrative burden into the work.

The obvious and inevitable offspring of any AI RMF will be audits and certifications. Data Quality Management guidelines will be complemented by Data Fairness guidelines (or some similar appellation). In the open-source software world, software licenses will be joined by AI RMF Reports, and developers will have to whitelist their open-source dependencies for both their license dependencies and their AI RMF dependencies.

4. Agility

For most of its history, IT has evolved far more rapidly than law or policy. Only very recently has the legal refrain ceased to be "I'm sorry, but there is just no caselaw applicable to your IT issue yet." In our view, AI is likely to turn this tortoise and hare relationship on its head. In the AI realm, policy demands, especially with regard to explainability and fairness, are developing at a much faster pace than mathematical and software solutions for them.

5. Conclusion

KeyLogic believes the AI RMF is a solid foundation for moving forward and we have indicated several areas that might benefit from additional consideration and refinement. KeyLogic is able and willing to participate in this continuing improvement process and can help design and demonstrate some of the applicable semantic technologies that would support greater automation and document integration, as well as algorithmic approaches to tackling challenging and sensitive topics.