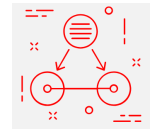


# PROTOFECT



99 Hudson Street, 5th Floor  
New York, NY 10013  
(573) 228-8497  
roy@protofect.com

September 12, 2021

Protofect, based in New York City, is a small-business focused on providing data science solutions that can turn AI products into sustainable businesses in real-world situations and prioritize ethical use - whether it impacts the life of one person deeply or thousands superficially and agnostic of whether consumers are aware of an AI's presence when interacting with software.

We are pleased to provide comments to the National Institute of Standards and Technology (NIST) regarding Artificial Intelligence Risk Management Framework (RMF). The focus of this letter is three-fold: (1) discussing the RFI topics in the structure of comment template, (2) building on our [previous suggestions](#) to NIST, and (3) introducing how we developed a method to rate software products that are AI-powered, and where it aligns in the context of this RMF. Our mission is to understand the DNA of software products that use artificial intelligence, to inform consumers, businesses and governments regarding liabilities involved in using the product.

According to reports, AI is expected to be a \$15.7 trillion industry by 2030. As the AI revolution booms due to this business opportunity, potential risks in comprehension, safety, malfunction and misalignment will evolve just as fast as the technology and deployment. Unintentional failures will spark the need for urgent oversight – in creation, deployment and proper functioning of AI systems. In the future, we envision that before AI products launch or are adopted in mainstream markets - they will need to be safety rated just as our food products and securities are rated today.

Here are our comments on the main topics as listed in the template:

*1. The greatest challenges in improving how AI actors manage AI-related risks – where “manage” means identify, assess, prioritize, respond to, or communicate those risks.*

One of the fundamental issues with the current situation is the presence of enormous verbiage and not enough computable units that can interface with software systems in managing the risk. Breaking it down to the necessary best-practices in “manage” criteria:

**Identify:** The biggest challenge to identification is efficient and ethical monitoring tools that can quickly interface with software systems and determine a relative threat that may exist in its normal operating procedure.

**Assess:** Some of the NIST work here has been good as we must determine what is mission critical and whether the effects are physical or psychological, instant or long-term, quickly patchable or necessary to take the software offline. A further question is determining cascaded prediction, i.e. AI systems that generate prediction results which serve as data and variables for downstream systems.

**Prioritize:** The truth here is that engineers and product managers are under constant pressure to deploy new features and deal with fires in systems. This means that AI security and accountability is often quite low in priority lists, unless there are specific compliance requirements.

**Response:** We need something akin to a “fire-brigade” squad to come in and respond quickly to identified and prioritized risk situations or incidents brought about by deployed AI systems. However, both bureaucratic processes and competing business goals might significantly slow the process down.

*2. How organizations currently define and manage characteristics of AI trustworthiness and whether there are important characteristics which should*

*be considered in the Framework besides: accuracy, explainability and interpretability, reliability, privacy, robustness, safety, security (resilience), and mitigation of harmful bias, or harmful outcomes from misuse of the AI*

The current characteristics are sound and a decent baseline in forming trustworthy AI. However, they might have relative thresholds and benchmarks in different industries. Further, their utility could be of variable importance given the kind of software. And possibly, there is a strong chance that universal agreed-upon “metrics” for these characteristics will never organically form. Therefore, a standardized threshold will have to be developed and iteratively improved with time.

Something to note is techniques such as Network Science are widely used in complex scenarios to determine “risk” or “load” on a system. We can borrow some ideas from this field to develop metrics that can produce “measurable” aspects of the characteristics and attributes. Perhaps, these characteristics should be layered/structured in order to better comprehend how they affect each other.

An overlooked characteristic here is “framing” - which revolves around the question: what final utility do we want this AI to maximize? An example is the balance between “engagement” vs. “digital-break” in feed algorithms. This needs to be explicitly called out by creators, whenever possible.

*5. Standards, frameworks, models, methodologies, tools, guidelines and best practices, and principles to identify, assess, prioritize, mitigate, or communicate AI risk and whether any currently meet the minimum attributes described above*

In addition to the suggestions in the other letters, we would like to recommend data loggers and activity trails that can help in audit to understand the “telemetry” of AI systems during normal or stress modes of operation. As some others have mentioned, Canada and the EU have suggested their own rubrics for AI safety. The best tool is something like an “On-board Diagnostics” module that modern cars possess as an in-built component, which

can plug into software and come up with diagnostic values for these characteristics.

*6. How current regulatory or regulatory reporting requirements (e.g., local, state, national, international) relate to the use of AI standards, frameworks, models, methodologies, tools, guidelines and best practices, and principles*

The current state of regulatory reporting in this specific context is less than satisfactory at the moment. We need something akin to the SEC, FTC or Consumer Reports to determine an effective way to do compliance and reporting, that builds trust both with the consumers and builders/manufacturers of AI products. Further, while there are some existing “standards” or “guidelines”, there is nothing to enforce it - thus it is left to manufacturers to self-police or consumers to fend for themselves.

*7. AI risk management standards, frameworks, models, methodologies, tools, guidelines and best practices, principles, and practices which NIST should consider to ensure that the AI RMF aligns with and supports other efforts*

Protofect is interested to share with NIST the “Airate” system, a rating system that is our opinion of the relative risk that the AI product will fail to align with the original intention of the creators, operators and beneficiaries in that market. It addresses the possibility that the AI software’s obligation to the business, users, markets and society will not be honored. Our ratings reflect both the likelihood of failure and the corresponding loss suffered in defaulting.

We have developed an algorithmic rating system for products that employ AI or predictive/prescriptive data modules in any capacity for its functioning. It enables companies to become more transparent to consumers, investors and oversight committees when using AI, by providing clarity about components of AI software, evolution from traditional technology and its direct or indirect impact on digital ecosystems.

Airate has a dynamic evaluation tool to help data professionals build or procure effective and scalable AI systems. It addresses 8 distinct AI alignment segments, which cover information across 70 sub-dimensions in

aggregate, including Data Acquisition, Databases, Warehousing, Learning Models Deployment, Performance Monitoring, Key Performance Indicators and Human Interactions.

*8. How organizations take into account benefits and issues related to inclusiveness in AI design, development, use and evaluation - and how AI design and development may be carried out in a way that reduces or manages the risk of potential negative impact on individuals, groups, and society.*

Myer's interesting quote should be of note here: "diversity is being invited to the party, inclusion is being asked to dance". Most machine learning developers agree that a biased dataset leads to biased models. However, we have to do better - because bias will always be a plague in the field of AI. What we pay attention to matters.

Much of the recent research has shown not just a lack of gender and racial diversity, but a strong lack of geographical diversity among AI researchers. In a recent paper by Chi et. Al titled "Reconfiguring Diversity and Inclusion for AI Ethics", an interesting conclusion is reached - that the responsibility of defining who diversity and inclusion are meant to protect and where it is relevant is pushed downstream to the AI product's customers. The situation is far from ideal, and more often than not, diversity in the workforce starts with diversity in leadership of the product and manufacturing team.

*9. The appropriateness of the attributes NIST has developed for the AI Risk Management Framework. (See above, "AI RMF Development and Attributes")*

While NIST's RMF attributes are commendable, a real effort must be undertaken to hunt down the bottlenecks and obstacles when implementing at the ground-level. An excessively broad attribute set could mean every institution will be open to interpreting it to the best of their ability and synchronize with their business goals and technological prowess, which might dilute the efficacy of the RMF.

Further, one of the biggest gaps right now is a lack of communication between researchers, developers and policy makers. Ethics sounds "uncomputable" to programmers, which makes it difficult to encode into software. There needs to

be a bridge between the judgement of ethics and the implementation of ethical and safe rules into AI systems.

Finally, many of the technical artifacts that are being judged for sufficient and necessary risk assessment are strongly coupled together computationally and in life-cycle. This means data, process, glue-code, algorithms, thresholds and stochastic boundaries must be individually yet holistically analyzed. It might not be easy to make this understandable to a broad audience, because it needs a specific and in-depth understanding of the architecture and AI process in certain scenarios. However, it should of course be aligned to stakeholder's vision of the product.

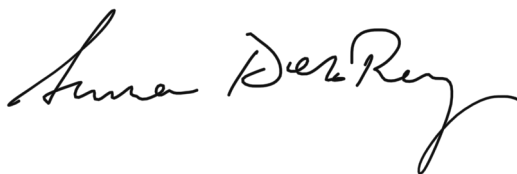
*11. How the Framework could be developed to advance the recruitment, hiring, development, and retention of a knowledgeable and skilled workforce necessary to perform AI-related functions within organizations.*

We would strongly recommend dividing the attributes and scope into specific task forces. Following that, if certain institutions gel into the task force and take responsibility for hiring, developing and providing the engineering cycles necessary to incorporate it into their organizations, that would be a seamless way to adopt the standard.

---

We are excited to see the outcome and impact of NIST's leadership for this RMF. We humbly request that NIST contact the undersigned at any time with any questions, especially ways to be involved or contribute to upcoming strategies regarding RMF or otherwise in building a sustainable AI ecosystem.

Sincerely,



SUMAN DEB ROY