VIANAI

VIAN RESPONSE TO NIST

AI RMF Risks – Vian Response

Prompt: Whether the AI RMF appropriately covers and addresses AI risks, including with the right level of specificity for various use cases.

The AI RMF framework overall covers many of the most important risks in AI today. And there are many risks that regulators have authority over already, such as understanding the financial risk inside of a model.

The emphasis on consumer protections makes sense from a policy framework standpoint, as policymakers are most likely to act in reaction to threats against their constituents, the consumers. There is however, a risk in overlooking "hidden" parts of the economy, such as business-to-business companies that may seek to use AI in non-obvious ways. The follow-on effect of such use cases could both impact the broader economy and the same consumers, without clear understanding as to why.

For example, an algorithm in the already regulated financial industry caused the Knight Capital Group to lose $440 million in less than an hour[1]. This took place in 2012, and while it did not directly lead to the many existing OCC regulations today to audit financial risk to the company and the broader economy, it exemplifies just what can happen with unregulated AI. Such regulations in manufacturing, life sciences, healthcare, aerospace, defense, would help mitigate unintended consequences.

There is also a sense in the report that AI today should be regulated to protect the broader public. Missing, though, is a sense that AI is lacking in many capabilities precisely because there is no clear framework for assessing and understanding risk. The ostensibly least risky economic areas for AI — marketing, digital social media data, etc. — became weaponized tools for foreign countries to manipulate our democracy. This then begs the question of what repercussions may arise from the use of AI for more critical areas of the economy such as infrastructure, health care or transportation.

Without comprehensive frameworks, and then the necessary tools to ensure compliance, AI will not be able to properly work inside the greater economy, due to significant concerns around the associated risk. We are already seeing this play out in business today.

[1] https://www.henricodolfing.com/2019/06/project-failure-case-study-knight-capital.html

AI RMF Flexibility – Vian Response

Prompt: Whether the AI RMF is flexible enough to serve as a continuing resource considering evolving technology and standards landscape.

Many of the guidelines in the AI RMF document appear sufficiently flexible. The key, though, to making AI both safe and beneficial, comes in human judgment. This is currently referenced in the "socio-technical characteristics" mentioned on page 10.

> Unlike technical characteristics, socio-technical characteristics require significant human input and cannot yet be measured through an automated process. Human judgment must be employed when deciding on the specific metrics and precise threshold values for these metrics.

We call this type of AI, "human-centered AI," in the vein of Stanford's Institute for Human-Centered AI. This type of AI is crucial for not only the metrics and thresholds, but for having a "human-in-the-loop" to make the final decision.

The algorithms can then learn from the decision, but the ultimate decision should remain with a human, as AI will likely not have the capabilities of human judgment for many decades.

AI RMF and Managing AI Risks – Vian Response

Prompt: Whether the AI RMF enables decisions about how an organization can increase understanding of, communication about, and efforts to manage AI risks.

In general, yes. The AI RMF puts a strong framework around the need for an organization to make a comprehensive effort. There needs to be a concerted effort to include all relevant stakeholders within an organization, including, and perhaps especially, those not considered technical.

Legal or Human Resources groups, for example, have an acute sense for risk, but often would struggle in understanding traditionally technical methodologies, such as SHAP values, which data scientists use to understand the risk in their models. Appropriate tools would enable non-technical understanding of the risk, or at least bring the methodologies of the tool to their attention with human friendly explanations.

AI RMF Functions + Categories – Vian Response

Prompt: Whether the functions, categories, and subcategories are complete, appropriate, and clearly stated.

In general, yes. See the previous note about the risk of focusing too heavily on consumers/voters and therefore not including a wide berth on the type of companies, i.e., not focusing enough on the companies that are not directly consumer-focused, it could result in skewed regulation.

AI RMF Alignment with Frameworks + Standards – Vian Response

Prompt: Whether the AI RMF is in alignment with or leverages other frameworks and standards such as those developed or being developed by IEEE or ISO/IEC SC42.

No comment

AI RMF Alignment with Existing Practices– Vian Response

Whether the AI RMF is in alignment with existing practices, and broader risk management practices.

The key for mitigating any risk, including the risk of AI, is to have a contingency set up in the chance of a failure. AI today is exceptionally good at quantity and speed, but it is completely surface level. The meaning, intention and nuances underneath the surface absolutely need human judgment and will for the foreseeable future.

AI RMF Gaps– Vian Response

Prompt: What might be missing from the AI RMF.

The AI RMF has provided a strong framework for many components of the risk associated with AI, particularly in how it may impact various communities and consumers. It does not adequately focus in too great a detail on the technological risk of AI. AI today essentially functions on data, and the data often suffers from bias, lack of context, oversampling of some groups, under-sampling of others, and other issues. These issues then lead to inherent faults in the outputs from the AI. There are many examples of this in the real world. Recently, Dall-E the text-to-image generator, would routinely return white men for words such as “lawyer” or “doctor,” while creating

images of Asian women in response to “flight attendant.[2]” The AI has no understanding of the meaning of these words, rather it’s pulling from an inherently biased dataset.

AI Risk + Companion Document– Vian Response

Prompt: Whether the soon to be published draft companion document citing AI risk management practices is useful as a complementary resource and what practices or standards should be added.

With AI, there must be a method to publish understandable information that a non-technical person, such as a regulator or consumer, can then understand the risks associated with the models. This doesn’t necessarily have to be exhaustive (i.e., creating a “slippery slope effect”), but it should give a sense for the size of an impact, the intention for the model and the safeguards taken to ensure that the model will perform as expected.

The act of understanding AI risk today often falls on a Data Scientist, but with appropriate tools and methodologies, which work in concert with the humans using the technologies, all members of the AI RMF stakeholders guide can begin to understand AI. This demystification is crucially important to growing the technology as a beneficial service to humankind.

About Vian

Vianai Systems, Inc. is a Human-Centered AI platform and products company launched in 2019 to address the unfulfilled promise of enterprise AI. Vian's customers include many of the largest and most respected businesses in the world, to which it delivers AI, ML and data science platforms and products. Vian helps its customers amplify the transformational potential within their organizations using its H+AI Platform and a variety of advanced AI and ML tools with a distinct approach in how it thoughtfully brings together humans with technology. This human-centered approach differentiates Vian from other platform and product companies and enables its customers to fulfill AI's true promise for the benefit of humanity.

[2] https://www.vox.com/future-perfect/23023538/ai-dalle-2-openai-bias-gpt-3-incentives