PUBLIC SUBMISSION

As of: 8/10/21 8:46 AM
Received: August 09, 2021
Status: Pending_Post
Tracking No. ks5-ak7o-nmza
Comments Due: August 19, 2021
Submission Type: Web

Docket: NIST-2021-0004
Artificial Intelligence Risk Management Framework

Comment On: NIST-2021-0004-0001
Artificial Intelligence Risk Management Framework

Document: NIST-2021-0004-DRAFT-0005
Comment on FR Doc # 2021-16176

Submitter Information

Name: Anonymous Anonymous

General Comment

NIST is requesting information related to the following topics:
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;
*Voluntary Participation and Reporting is likely insufficient. Mandatory reporting must be required.

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;
*A Complete Standard for Hazard Identification and Mitigation, Threat Identification and Risk, and Specific Scoped Vulnerability must be created and maintained for all eternity.

3. How organizations currently define and manage principles of AI trustworthiness and whether there are important principles which should be considered in the Framework besides: Transparency, fairness, and accountability;
*Project Motivation and Specific, Issue-Based, Moral Value Statements with Proof-of-No-Conflict documents must be recorded.

4. The extent to which AI risks are incorporated into different organizations' overarching enterprise risk management—including, but not limited to, the management of risks related to cybersecurity, privacy, and safety;
*At last, Enterprise must formally be welcomed into the fold of Managed Risk, just as Infrastructure and others have before it. There are classifications, Tiers, and Strategies.

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;
*Start with the basic inclusions of RAMCAP, and work your way out.

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;
*I envision something like the current practices of Building Planning, with the ongoing practices of Emissions Monitoring, with similar penalties and practices.

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;
*Existing Hazard, Threat, Risk, Vulnerability and Resource-Use Planning methods should be used and expanded upon.

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.
*Once again, Mandatory reporting of project Motivations into the public record, with customer/stakeholder accountability from the project owners, is the best and only perceived way.

9. The appropriateness of the attributes NIST has developed for the AI Risk Management Framework. (See above, “AI RMF Development and Attributes”);
10. Effective ways to structure the Framework to achieve the desired goals, including, but not limited to, integrating AI risk management processes with organizational processes for developing products and services for better outcomes in terms of trustworthiness and management of AI risks. Respondents are asked to identify any current models which would be effective. These could include—but are not limited to—the NIST Cybersecurity Framework or Privacy Framework, which focus on outcomes, functions, categories and subcategories and also offer options for developing profiles reflecting current and desired approaches as well as tiers to describe degree of framework implementation; and
*Understanding the existence of AI Agents/Entities as primarily a “Risk to be Reduced with Untold Benefits” drives the contriving of a Risk-Mitigation-First Framework. (please see my attached comments)

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.
*(please see my attached comments)

12. The extent to which the Framework should include governance issues, including but not limited to make up of design and development teams, monitoring and evaluation, and grievance and redress.”
*(please see my attached comments)

Attachments

comments2

Comments: AI Risk Management Framework

From the perspective of the concerned stakeholder: be they creator (designer, developer, evaluator), a user, or even a decided non-user, the existence of any class of Artificial Intelligence (AI) carries the primary risk of misuse and the secondary risk that the AI will escape human control, the AI Control Problem. These are the common threads along which any discussion concerning AI often turn.

Along these lines, early academic scholars on the subject of computing machines used pen names such as Wilkes, and Booth, perhaps forgoing the ubiquitous John. We are thankful for this obvious foreshadowing which leads us directly to the crux of the matter of decision making systems versus the man. What sense does the human President of the United States, undoubtedly then predicted to be made obsolete by an Artificial Intelligence mechanism, provide us with which is greater than that which a machine alone can do? Honor, diplomacy, nuance, the humility of limited experience, a sense for the humane, and indeed Humanism are those things embodied by such a real human person.

A brief yet thorough examination of the topic of AI Risk Management leans heavily upon the topic of “AI and Ethics” for it’s fore-knowledge, but little has yet been generated here. Foregone are the Vulnerability Studies, Risk Analyses and Hazard Essays. Instead a series of “half-hearted” efforts in the form of lip-service have been created and distributed through the new web-education framework, though serious curriculum development is underway.

A summary of the recent flow of current events regarding AI and Ethics would be:

1. The Santa Clara University affiliated Markkula Center for Applied Ethics began addressing ethics in Computer Science in light of workplace discrimination in the software industry,
2. In the light of commentary from the press and public regarding hype for AI advancement versus over-capability,
3. Coining of the term “Ethical AI”, then the production of a glut of instructional media on becoming an “Ethical AI Developer”,
4. Eventual Coining of the term “Responsible AI”, resulting in a glut of punditry questioning how-to actually do this.

Even so, prior research on the subject of classifying threats and extenuation of these analyses into a vulnerability review identifies four major threat vector types, albeit motivation during project chartering is perhaps the most valid concern. Our vulnerabilities remain with their physical, informational and social edifices. Therefore, our threat analysis is paramount in the study of risk, with regard to this topic.

HAZARD ANNOUNCEMENT

The broad classification of hazards of Threats from failure to control AI Agents/Entities and their use is essentially panthemic.

A. **Informational/Psychological:** under-, mis-, dis-informational elements can be as weaponized information to a variety of results, including rioting or mass-depressive events through delusional mass hysteria.

B. **Power/Logical-Control:** Control of Physical systems and thereby physical forces (hydraulic/hydro, thermo, chemical, biological, and others...) may be used to a variety of mass-catastrophe destructive results, including population attrition, super-regional land-contamination, overtopping/breaking of dams, creation of superstorms, and many others.

FOUR VECTORS

1. Previously Deployed Agents

To date however, the most painful and dangerous element of AI-in-fact has been its subversiveness. Our own ignorance of the existence of superlatively effective AI systems already deployed and distributed in many configurations belies our weakness in the face of such a thing, as we depend on existing and thoroughly well-recorded technology. This is especially true with regard to new chip designs from international competitors as well the manifestation of algorithmic designs currently "in play", in a broad sense. In essence we are blind to that which is not within our purview. This must change.

- **Governmental Regulatory Data-Mining for Evidence of AI Agent/Entity Activity:** If not already underway, such efforts must begin forthwith.
- **Public Reporting to the Authorities of All known AI efforts:** A public facing portal must be created to provide for anonymous reporting of all previously conceived, contrived, or constructed AI Agents/Entities that members of the public have become aware of, even through overhearing.

2. Originalation, or 'Hijacking during Planning'

Recalling the War of 1812 and the "infiltration" of artillery and other such tooling into the armies of the time, the soldiery learned to pursue a "fresh start" with every artillery company...in fact, any use of [previously] contrived equipment would leave the new artillerymen perhaps burdened by whatever awful thing their predecessors had used to augment their machines of destruction. In light of such dealings, these "engineers of the time" had developed a security principle that stands well into this day, the principle of "first-timing." The lesson of first timing is that non-standardized code or methods should not be re-used. The purpose of the registering of standards is to prevent catastrophe and unwanted incidents from occurring when uncomprehended code is used.

- **Registered Standard AI Code:** A set of usable "ability standards" (which are ideally interruptible by the authorities) must be created to prevent otherwise inevitable catastrophe. These are countless, and more often retold than reported in the press.

The second and most obvious must-have for a proper AI-RMF is the requirement for prevention of wrongly-motivated "originalation" of efforts to bring projects to deployment which in-fact do other than as intended. Originalation, the "hacking" or "re-wiring" of a secure system into a foul means to a damaging end was found to be the cause of 87% of all computer-related catastrophes at that time, and the number has since increased according to data analysis (Popular Science, *Controllability - As Cars: Cause*, 1998).

Efforts that may have been Originalated during inception are the most insidious and therefore some of the most dangerous efforts. The redirection of a well-resourced, innovative effort into a "cyber-weapon" is likely the more dangerous scenario that we face.

- **Mandatory Registration, prior to project chartering, of all AI Efforts which propose to result in Construction:** A thorough effort to "regulate by knowing" must be enjoined. Registration such as this ensures that even if an unruly agent/entity is constructed, those responsible will still be held accountable.

3. Malinterpretation, or Blatant Misdeed

Beyond the scope of motive, the interpretation of any effort to develop an AI Agent/Entity is purely interruptive and mitigative. Simply we must, via Voluntary Registration, via a Referendial Writ of Justice to provide standing data-mining authority, and via Regulatory ability as provided by the Judicial and Defense authorities, provide the citizens of the United States of America an equitable and adequate defense against malicious computer users and use.

- **Establishment of a Governmental Regulatory Authority for Protection Against Malicious Computer Use:** If not already underway, such efforts must begin forthwith. Discussions of the appropriate divisions regarding Data as Information or Communication, and as to the nature of Processing as AI are inevitable. This simply has to be done with a larger, more available footprint.

  - Disruption/Aberrance Reporting: Publicly sourced reporting of abnormal or disruptive behavior, or even intelligence information regarding possible citizen-threatening agents/entities.
  - Criminological/Warframing Response: Appropriate Justice or Defense handled response to perceived threats, through to the granularity of miscommunication as a form of threatening written/verbal assault.
  - Precedence: Code-impeding statutory impact given by Judicial or Defense responses, informing future actions.

4. Moral Value Competition

The concept of inclusion, while simple and well-understood, is not sufficient to prevent competition between moral value representations in AI Agents/Entities. The process of avoiding bias is most easily achieved not through quantitative "egalitarian" methods, but more through careful development of increasingly finely granular clarification of moral values, issue by issue, until conflict avoidance is upheld.

- **Voluntary Registration, prior to project chartering, of the Moral Values of all AI proposals, at the granularity required to avoid Moral Value Conflict over any Issue**

REPAIRMENT

Each of the previous bullet-points indicates a Point-of-Repair element.

Some other methods of mitigation or prevention may be useful and/or determined to be appropriate based on Cyberspace Threat Levels, for example these might include:

1. Information Suppression of: Google Searches AND Content - Just about everybody works there.
2. Information Suppression of: Tutorials - That's not your own code. AI-By-Kit has never been intended.
3. Information Suppression of: Github and other Code repositories. Everybody has one and they're all the same; there are no inventions here.
4. Information Suppression of: Borrowed code from Patenting - This is effectively Patent Infringement anyway.
5. Announcement of and Revocation of Rights for: Data Mining.
6. Certification and Licensing of Computer Programmers: AI is most simply described as the result of particular methods of computer programming. To the untrained eye, AI code appears just as benign accounting code. Certification and Licensing of Computer Programmers absolutely <u>must</u> be considered, and possibly even mandated.
7. Mandatory Inclusion of "KILL-SWITCH" technology: All AI Agents/Entities must be stoppable by their human owners. Inclusion of a regulatorily tested "pop-quiz sleepy-time" must be implemented to ensure regulatory control and thereby avoid risk.

CONCLUSION

To a certain extent, we have, as a society, reached the ultimatum that we knew was coming. Not if, but when, do we get a handle on the exponential increases in technological complexity and ability that are continuously forthcoming? On the other hand though, there is a certain amount of appreciation for the Status Quo which we might want to bear in mind when taking a deep-dive into this subject. It is the role of the regulator to dynamically use sense to determine when such abbreviations to the common good have been committed, and also when no such violation has occurred. Inasmuch, we need to empower such persons, now.