RTI International
3040 E. Cornwallis Road, P. O. Box 12194
Research Triangle Park, NC 27709-2194

Response to NIST RFI 2019-08818: Developing a Federal AI Standards Engagement Plan

In this response we do not lay out a specific plan for AI standards but rather point to some considerations related to standards for both the process and ethics/oversight of AI.

Process Over Products: A Place for Open Source Software in NIST AI Standards

As acknowledged in the NIST S 6106.01 standard on Open Source Code, custom-developed software solutions for the Government often rely heavily on open source code. The current AI software ecosystem is no different, with open source libraries for machine learning, data analysis, and broader scientific computing playing a crucial role in the growth and development of the field and in innovation itself. While commercial software can certainly play a role in AI-related government projects, continued support of open source libraries and frameworks for use in custom-developed software and analyses will help the federal government innovate while exhibiting leadership in openness, transparency, and accessibility that the public and scientific community expects. For example, the major deep learning frameworks (Tensorflow, PyTorch, MxNet, etc) are open source, with thousands of organizations and millions of users relying on them daily. Importantly, due to the burden of developing and maintaining a modern deep learning framework, most commercial software that trains and deploys deep learning models relies on these open source frameworks and uses them for model development.

Benefits of Including Open Source in AI Standards

- Underlying code is shipped with software, providing *transparency* to users. Users can check, modify, and/or extend the codebase for specific project needs.
- Open source machine learning libraries and frameworks implement new methods faster than proprietary software, which is important in rapidly developing fields like AI.
- There is a vibrant ecosystem and community around large open source projects to learn, develop, contribute to, and improve the projects.
- Open source projects integrate more readily with other open source projects, allowing for innovative solutions that may often not be possible within a single proprietary software framework.
- Tests in open source projects are transparent and are shipped with the codebase. In addition, if code is hosted in a popular code repository, such as GitHub, there are direct and transparent ways of posting issues discovered about the project. By contrast, commercial software may fail ways that users cannot troubleshoot, since they lack access to source code.
- Open source software entails neither licensing costs not vendor lock-in, as well as makes transfer of deliverables developed by contractors much easier, because there is little risk of violating licensing terms.

Challenges for Open Source Software Projects

- Most AI open source projects are libraries and frameworks designed for software developers and scientific programmers, rather than applications designed for end-users without

programming experience. Technical sophistication, programming, package management, and version control may be required to engage with software.
- Customer support may be limited.

AI Uncertainty

AI systems operate in a world of uncharacterized and unquantified uncertainty. Data on which they depend are imperfect, incomplete, erroneous, not timely and of uncertain provenance, not to mention subject to deliberate corruption. Yet, results and decisions are often conveyed as if they carry no uncertainty. With its unparalleled history of understanding and promulgating the consequences of uncertainty in physical systems, NIST is uniquely positioned to force attention to uncertainty in the context of AI.

AI Bias

AI systems face problems of bias through a variety of channels such as the underlying societal systems that affect the population a dataset represents, design decisions, and implementation in the real world. Biased AI technologies can reinforce existing societal biases and can harm individuals in identifiable communities. Both the technical and the popular literatures are rife with assertions that AI systems are "biased against" subgroups defined by age, gender or race. (Not all of these claims are credible, however.) As with most of the other issues raised here, NIST's AI standards for bias will need to: *Define* what bias is; Specify, develop or lay out a framework to develop, both broad and domain-specific, *quantified measures* of bias; and Form the *basis for enforcement*.

AI Security

The development of standards for AI Security requires a clearly delineated definition of what constitutes security in the context of AI. AI security differs from privacy and safety, even though the same components of AI system may be implicated in all three domains. AI security falls into two categories:

- *Security risks experienced by an AI system*: robustness in the face of external threats that have the potential to cause intentional damage to the context (organizational, public, governmental) they operate in; and
- *Security risks posed by an AI system*: Intentional use of the system to cause damage to people or systems (e.g., city government, financial institutions, or physical infrastructure).

AI Security standards need to be developed as part of large collaborative efforts between many different actors. AI systems are being developed by many different countries and deployed worldwide. Given potentially antagonistic relationships between countries, international enforceable standards may only be possible between allied countries. Where such enforceable standards are not possible, more general security concerns should drive standards of non-deployability. Within the US, the private sector and federal agencies need to collaborate on enforceable standards that prevent the exploitation of AI for nefarious purposes. In this context high level standards need to be developed that guide the balancing of national security versus security concerns for the general public or private organizations and institutions. Given the pervasiveness of AI systems, different governmental agencies need to collaborate to develop cross-domain as well as domain-specific standards. Standards that apply to self-driving cars or flight control may primarily concern the NHTSA (National Highway Traffic Safety Administration) or the FAA Federal Aviation Administration, but law enforcement agencies might also be implicated and should participate in defining standards. Seemingly more specific domains such as smart TVs or hardware-based secure routers may require more narrow standards. However, the underlying AI algorithms may be subject to the same vulnerabilities as AI systems where a security breach would be much more consequential (such as self-driving cars).

As noted previously, many widely employed AI packages (TensorFlow, PyTorch, Keras, language models such as BERT or UlmFIT) are open source software. Any AI standards need to apply to commercial software as much as they do to open source software. If an open source package is developed and published by a large cooperation (such as Google, Facebook, Amazon), standards can be injected into the development process so that deviations from standards can be minimized. This situation becomes much more complicated with community-driven open source software. Assuming the same standards apply, who is responsible for adapting workflows and software development processes to account for the adherence to standards? Even if security standards are only tested by external authorities it would be too costly to respond to un-met standards once a project has reached a stable beta state. Any agency testing for (or even enforcing) Security standards should thus actively work with the open source community to develop feasible workflows and best practices.

One key instrument of making sure that security standards are met is standardized tests of AI systems. Unlike stable end products such as tables, software and in particular AI software, is frequently updated and modified. Do tests have to be applied with every update of that software? Does every update warrant a new round of tests or should tests only apply to significant updates? Who defines what a significant update is and how the concept of "significant" translates across different domains? One way to tackle this problem would be to make these standardized tests integral parts of the AI system with the ability to continuously monitor the system very much like the human immune system. However, tests themselves may become victims of adversarial attacks, and so tests should also adhere to security standards. To avoid conflicts of interest, should different authorities be responsible for different aspects of the process? Similar to adversarial neural networks, should software and testing standards be pitted against each other to guarantee maximum adherence through continuous improvement?

AI can no longer be considered solely a software product. More and more efforts focus on embedding AI in hardware architecture to maximize performance. Examples here are Google's Tensor Processing Units (TPU) and Intel's Loihi neuromorphic chip. AI Security standards need to apply to these AI units as well. The challenge with hardware embedded-AI is to clearly differentiate between software-and hardware-specific components of the system. Do more complex standards need to be developed for such embedded systems? What about systems that run AI algorithms on top of these embedded systems? Should separate standards for embedded hardware systems and software be applied or should the entire system as whole be subject to security standards?

AI Ethics

This section seeks to provide information to help NIST understand the current state, plans, challenges, and opportunities in ethics and AI.

Opportunities

In 2016 the European Union (EU) created the General Data Protection Regulation (GDPR) that would expand protections around EU citizens' personal data beginning in 2018. Meanwhile, China has extensively integrated AI technologies into their government and social structure via the China Social Credit System. Under this system, the Chinese government employs widespread data collection and analytics to monitor the lives of its citizens. Citizens are scored based on everyday decisions like what they purchase and what they do in their free time, and their scores are affected by the scores of their family and friends. An individual's score amounts to a social ranking, and it affects a variety of aspects of life, like the ability to travel throughout the country, acquire loans, and receive an education[i]. The

Chinese Social Credit System is at odds with U.S. perspectives on the values of liberty and freedom as well as other countries' values on personal data privacy and protection as seen with the GDPR. As China continues to heavily integrate AI into daily life, the country is setting an international precedent for what it considers ethical AI.

The U.S., in collaboration with other nations, can position itself as a leader in defining ethical and responsible practices around AI and AI-related technologies. The growing ubiquity of data collection continues to enhance the power of predictive analytics and AI to an extent that most U.S. laws and regulations have not accounted for. The federal government has nonetheless made steps towards being in an informed position for regulating AI. For example:

- 2016 formation of the National Science and Technology Council Subcommittee on Machine Learning and Artificial Intelligence, to help coordinate Federal activity in AI and create the National Artificial Intelligence Research and Development Strategic Plan.[ii]
- Introduction of the *FUTURE of Artificial Intelligence Act of 2017*.[iii]
- 2019 launch of ai.gov, which features AI initiatives from the administration and federal agencies.

Local governments are beginning to create regulations on their own, as seen with San Francisco's May 2019 ban on use of face-detection software by city agencies and law enforcement[iv]. Nevertheless, most regulation around AI currently depends on those who are creating the AI technologies, and these creators may often have incentives that conflict with the privacy or long-term interest of their users as well as non-users who are nonetheless affected by externalities of AI technologies. In these situations, the U.S. government may have an opportunity to expand AI regulation that protects citizens' interests.

Challenges and Ethical Issues

The ethical issues facing AI have a wide scope. These issues include data collection, storage, and distribution as well as machine or statistical use of data and how a subsequent AI product interacts with the world. AI systems face problems of bias through a variety of channels such as the underlying societal systems that affect a sample's population, design decisions made by an AI engineer, or an AI system's actual implementation in the real world. Biased AI technologies can reinforce existing biases in society and can harm individuals in marginalized communities. For example, a 2016 analysis by ProPublica found that the criminal risk assessment system, COMPAS, was biased against black Americans such that "black defendants were twice as likely to be incorrectly labeled as higher risk than white defendants" and "white defendants labeled low risk were far more likely to end up being charged with new offenses than blacks with comparably low COMPAS risk scores"[v]. There are many stakeholders in ethical AI including individuals, governments, countries, ethnic groups, cultures, and corporations.

Given the human-specific nature of most AI technologies, the ethical frameworks that have emerged out of medical and biological research are good reference points for ethical evaluation of AI. In particular, there are key concepts that could extend to ethics in AI:

- *Autonomy* – Being "free from both controlling interference by others and from limitations, such as inadequate understanding, that prevent meaningful choice."[vi]
- *Informed Consent* – Providing permission for or agreeing to an activity in which one is involved or directly affected by (e.g., a medical procedure). A threshold for informed consent is not always clear, but it typically requires that the consenter have sufficient information about the activity. Additionally, it requires capacity and maturity for understanding the decision being make. Finally, the individual must also be free from outside influence, like coercion, in their decision making.[vii] [viii]

- *Coercion* – Compelling one to act against one's will by threatening violence, reprisal, or other intimidated behavior that puts one in fear of the consequences of not complying.[ix]
- *Undue inducement* – Influencing one's decision or action by offering one an incentive that is so large or enticing that it undermines one's ability to rationally consider the costs and benefits of such a decision or action.[x]

Most individuals interact with AI technologies created by corporations, and corporations define the rules for using their technologies via Terms of Service agreements. Many issues can begin to emerge when evaluating these agreements under the ethical concepts enumerated above. For example, companies often retain the right to change their Terms of Service agreements, so a user may agree to certain data retention conditions but then find that the originally agreed to terms are no longer the active terms of use. Depending on the situation, the company may not provide clear communication about the change and the user could have false expectations. In situations where companies are using an individual's data in research or to predict something about them, this could be problematic under the lens of expectations in informed consent. As shown in Figure 1, an individual may knowingly disclose certain personal pieces of information, but they may do so without knowing that the disclosed information will be used to predict something undisclosed.

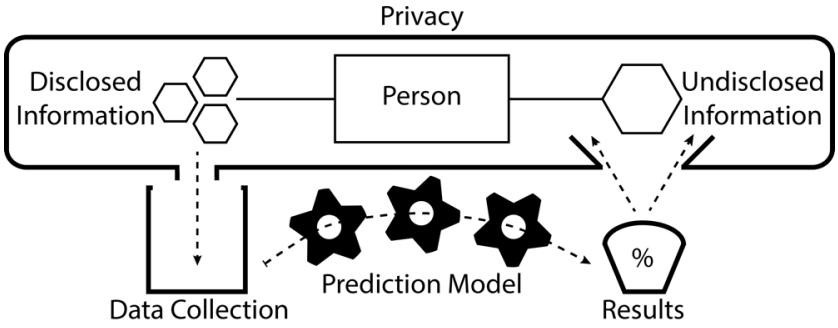

Figure 1: Illustration of disclosed information being used to predict undisclosed information.[xi]

Medical information is typically protected under regulation of Personally Identifiable Information (PII) data.[xii] Yet, the growing sophistication and adoption of AI technologies is making it easier to extrapolate undisclosed medical or psychological information using out-of-domain data like social media behavior.[xiii] Understanding how AI technologies actually work requires a deep level of education in and experience with statistics and AI implementation. Given the requirements of informed consent and that understanding AI is not immediately straightforward, it is questionable whether an average citizen could consent in an informed way to data collection that will involve downstream AI research, especially if that research is not clearly defined.

Some leaders at the technical end of AI have more formally defined guiding ethical practices. For example, Google has outlined principles for developing AI responsibly[xiv]:

1. Be socially beneficial
2. Avoid creating or reinforcing unfair bias
3. Be built and tested for safety
4. Be accountable to people
5. Incorporate privacy design principles
6. Uphold high standards of scientific excellence
7. Be made available for uses that accord with these principles.

Companies like Google face inherent limitations in their ability to self-monitor ethical issues, as was seen when Google canceled its AI Ethics Board less than two weeks after its launch due to internal and external criticisms of board member conflicts of interest.[xv]

With the growing availability of AI-powered autonomous and semi-autonomous vehicles there are underlying concerns in how machines will react in situations that require decision-making that could affect pedestrian or passenger lives. Additionally, when accidents occur that involve autonomous vehicles, questions around culpability arise, which could involve pedestrians, passengers, vehicle companies and their leaders, AI software companies and their leaders, and AI engineers.

U.S. Laws & Regulations

When making any ethical evaluation of AI one must also acknowledge relevant laws, regulations, community values, and corporate policies. Moreover, law embodies ethical values, and in order for the U.S. to establish an ethical position in AI it must have its position reflected in laws and regulations. With AI being driven primarily by academia and corporations, there are few U.S. laws and regulations specifically aimed at AI and data in its modern context. That said, there are well-established laws and regulations that provide guidance for dealing with privacy and citizen rights that extend to the recent data-driven software technologies.

At the U.S. federal level, a keystone for guidance and philosophical perspective is the *Fourth Amendment*:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.[xvi]

In *Katz v. United States* (1976), the U.S. Supreme court established an extension of the Fourth Amendment called the *Reasonable Expectation of Privacy Test*. The test consists of two requirements: an individual has exhibited an actual (subjective) expectation of privacy and the expectation is one that society is prepared to recognize as reasonable. Under this test, if these two requirements have been met and the government has taken an action that violates the expectation in question, then the government's action has violated the individual's rights under the Fourth Amendment.[xvii]

Later, in *United States v. Miller* (1976)[xviii] and *Smith v. Maryland* (1979)[xix], the courts established what is known as the *Third-Party Doctrine*, which holds that information one voluntarily gives to a third party does not carry a reasonable expectation of privacy (e.g., bank records, dialed phone numbers). However, in *Carpenter v. U.S.* (2018)[xx], the Supreme Court found that the Third-Party Doctrine did not apply to cell phone location data.

Specific laws that define the extent of an individual's right to privacy from other individuals can vary by state, county, and city. For example, taking photos while on a public space (e.g., a sidewalk) is typically unrestricted even if the photos are of a private space (e.g., a person's yard) – with exceptions of private spaces like bathrooms and bedrooms – but a property owner can bar one from taking photos while on their property.[xxi] Additionally, recording audio of individuals can face certain restrictions, such as requiring consent from various parties who are being recorded.[xxii] [xxiii]

Conclusion

The U.S. has an opportunity to lead the way in defining guidelines, regulations, and laws that address ethical issues in AI. With differing perspectives on what constitutes ethical AI across nations, it is important that the U.S. clearly defines its position in such a way that is consistent with the values defined in the U.S. constitution and body of law. The ethical frameworks developed in medical and biological research are good reference points for ethical evaluation in AI, given the human-focused nature of many AI systems. In particular, notions of informed consent may need to be more thoroughly evaluated when considering data collected on individuals and the variety of predictive information that can be garnered from that data. While the U.S.'s legal position on AI and associated practices and technologies is nascent, there are well-established legal precedents that can serve as a backbone for more specific AI regulations and law.

References

[i] https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say

[ii] https://www.nitrd.gov/news/national_ai_rd_strategic_plan.aspx

[iii] https://www.whitehouse.gov/ai/

[iv] https://www.washingtonpost.com/technology/2019/05/14/san-francisco-becomes-first-city-us-ban-facial-recognition-software/

[v] https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say

[vi] Beauchamp, Tom L., and James F. Childress. *Principles of biomedical ethics*. Oxford University Press, USA, 2001.

[vii] Beauchamp, Tom L., and James F. Childress. *Principles of biomedical ethics*. Oxford University Press, USA, 2001.

[viii] Faden, R. R.; Beauchamp, T. L. (1986). *A History and Theory of Informed Consent*. New York: Oxford University Press. ISBN 978-0-19-503686-2.

[ix] https://www.merriam-webster.com/dictionary/coercion

[x] https://bioethics.yale.edu/research/irb-case-studies/irb-case-payments-subjects-who-are-substance-abusers/fair-compensation-or

[xi] https://thommiano.com/#dialogue

[xii] https://www.gsa.gov/reference/gsa-privacy-program/rules-and-policies-protecting-pii-privacy-act

[xiii] https://www.pnas.org/content/111/24/8788

[xiv] https://ai.google/principles/

[xv] https://www.forbes.com/sites/jilliandonfro/2019/04/04/google-cancels-its-ai-ethics-board-less-than-two-weeks-after-launch-in-the-wake-of-employee-protest/#2ff34a1f6e28

[xvi] https://www.law.cornell.edu/constitution/fourth_amendment

[xvii] https://www.law.cornell.edu/wex/expectation_of_privacy

[xviii] https://www.oyez.org/cases/1975/74-1179

[xix] https://www.oyez.org/cases/1978/78-5374

[xx] https://www.oyez.org/cases/2017/16-402

[xxi] http://www.krages.com/ThePhotographersRight.pdf

[xxii] http://www.dmlp.org/legal-guide/recording-phone-calls-and-conversations

[xxiii] https://www.upcounsel.com/audio-surveillance-laws-by-state