U.S. National Institute of Standards and Technology (NIST) AI Risk Management Framework

Visa Inc. Response to NIST Request for Input

September 29, 2022

Visa welcomes the opportunity to provide input on the second draft of the NIST AI Risk Management Framework ("RMF"),[1] published on August 18, 2022.

Please find our comments below:

Part 1: Trustworthiness Characteristics

Structured in two parts, the NIST AI RMF examines (1) core AI trustworthiness characteristics that should guide the design and application of AI systems, and (2) governance controls in the Map, Measure, and Manage approach that can provide practical guardrails to manage risk across a model's lifecycle.

NIST AI RMF defines trustworthy AI as: *"valid and reliable, safe, fair and bias is managed, secure and resilient, accountable and transparent, explainable and interpretable, and privacy-enhanced."* These characteristics have been widely accepted as necessary constituents of AI ethics, for example by the OECD,[2] U.S. Department of Defense,[3] and various European data protection authorities – most notably in the algorithmic accountability reports published by the French CNIL[4] and the UK Information Commissioner's Office.[5]

- **Trustworthy taxonomy lacks quantitative guidance for organizations to implement**

Page 11 of NIST AI RMF discusses the inextricable ties of these trustworthy characteristics with social and behavioral contexts which are inherently qualitative, rather than quantitative. Indeed, organizations should account for context-specific risks, such as the potential differences in linguistic or cultural interpretations of fairness. However, in order to operationalize these values, organizations require guidance beyond understanding these principles as social concepts, to be equipped to break them down as objective goals, parameters, and thresholds that can be instituted into existing governance frameworks. For that reason, we believe certain taxonomies in this guidance, such as 'fairness,

[1] https://www.nist.gov/itl/ai-risk-management-framework
[2] https://oecd.ai/ai-principles
[3] https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF
[4] https://www.cnil.fr/en/algorithms-and-artificial-intelligence-cnils-report-ethical-issues
[5] https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/

accountable, and safe' require more quantitative recommendations that could lead to action in implementing, rather than interpreting.

For example, **Section 4.3 'Fair – Bias is Managed'** should expound on the concepts of false positives and negatives, differential treatment, as well as emerging metrics for AI fairness in anti-classification, statistical parity, and calibration.[6]

- **AI should be fit-for-purpose**

Section 4.2 'Safe' emphasizes measures to mitigate threats to safety posed by AI systems, but does not address whether certain models or use cases should be approved for deployment in the first place. A more proactive approach would be in striving to prevent harm rather than reacting to harm; by ensuring that stakeholders assess whether a model's proposed output and potential impact are fit-for-purpose. If the model poses high risks, stakeholders should consider whether it should proceed to deployment, and what guardrails are necessary and proportionate to address the potential impact.

- **RMF does not adequately address governance of B2B models**

Sections 4.5 'Transparent and Accountable' and **4.6 'Explainable and Interpretable'** both address consumer-facing AI systems where the end users are directly interacting with the model. These terms may be interpreted differently in B2B models, where it would be important to drive transparency and explainability for internal stakeholders who have the capacity to check for other trustworthy characteristics (i.e. secure, privacy-enhanced).

- **AI should be privacy-enhanced**

Section 4.7 'Privacy-enhanced' should be a critical underpinning of all AI systems, as AI relies on good data to produce more accurate and representative outcomes. Unfortunately this principle is not adequately explored in this guidance, although there are various concrete examples of privacy-enhanced AI that also benefit other trustworthy characteristics such as accountability and security.

NIST AI RMF should mention the use of privacy-enhancing technologies ("PETs") for machine-learning and AI, as well as data minimizing methods such as de-identification and aggregation for certain model outputs. There are a myriad ways in which privacy governance frameworks and emerging PETs can produce a net-positive design for a privacy-enhanced AI system, rather than focusing merely on avoiding harms.

Part 2: Core Governance – Map, Measure, Manage

- **Third Party Risks and AI Vendor Management**

Vendors and suppliers that provide AI products require more careful review. Third-party systems that build models or integrate white-labelled AI products into a business should be governed with standardized AI vendor risk management, from which a chain of accountability is clear on operationalizing this guidance; monitoring vendor model degradation; and ensuring that third-party

[6] *See for example,* Davis, J. L., Williams, A., & Yang, M. W. (2021). Algorithmic reparation. Big Data & Society, 8(2). https://doi.org/10.1177/20539517211044808

models are built with sound methodology that is not insulated from risk assessments due to trade secret protections.

- **Privacy governance and legal should be listed as stakeholders in the AI lifecycle**

The chart below (NIST AI RMF, Page 6) outlines a comprehensive view of the cross-functional stakeholders whose input should be provided during the AI lifecycle.

Lifecycle	Activities	Representative Actors
Plan & design	Articulate and document the system's concept and objectives, underlying assumptions, context and requirements.	System operators, end-users, domain experts, AI designers, impact assessors, TEVV experts, product managers, compliance experts, auditors, governance experts, organizational management, end-users, affected individuals/communities, evaluators.
Collect & process data	Data collection & Processing: gather, validate, and clean data and document the metadata and characteristics of the dataset.	Data scientists, domain experts, socio-cultural analysts, human factors experts, data engineers, data providers, TEVV experts.
Build & use model	Create or select, train models or algorithms.	Modelers, model engineers, data scientists, developers, and domain experts. With consultation of socio-cultural analysts familiar with the application context, TEVV experts.
Verify & validate	Verify & validate, calibrate, and interpret model output.	
Deploy	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	System integrators, developers, systems/software engineers, domain experts, procurement experts, third-party suppliers with consultation of human factors experts, socio-cultural analysts, and governance experts, TEVV experts, end-users.
Operate & monitor	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives and ethical considerations.	System operators, end-users, domain experts, AI designers, impact assessors, TEVV experts, product managers, compliance experts, auditors, governance experts, organizational management, end-users, affected individuals/communities, evaluators.
Use or impacted by	Use system/technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.	End-users, affected individuals/communities, general public; policy makers, standards organizations, trade associations, advocacy groups, environmental groups, civil society organizations, researchers.

Early stages of the AI lifecycle, such as Plan & Design, should solicit input from non-technical stakeholders such as legal, regulatory, and compliance – in order to create a shared understanding of the legal or regulatory frameworks that the AI model would be operating in, and whether its purpose aligns with organizational policies and existing regulations.

Currently, regulatory input is assigned at the end of the proposed lifecycle in the chart, in "Use or Impacted By" – where industry groups and civil society organizations may advocate for the end users. Putting these actors in the early stages of the AI design and planning, rather than at post-deployment, would help govern the model to be fit-for-purpose and compliant with existing laws (not all of which may be AI-specific, but may still govern certain regulated industries' use of AI, like the Fair Credit Reporting Act).

Visa appreciates the opportunity to provide comments on the NIST AI Risk Framework. We look forward to continuing to work with NIST as the framework is finalized, and we are happy to make ourselves available to discuss our comments in greater detail.