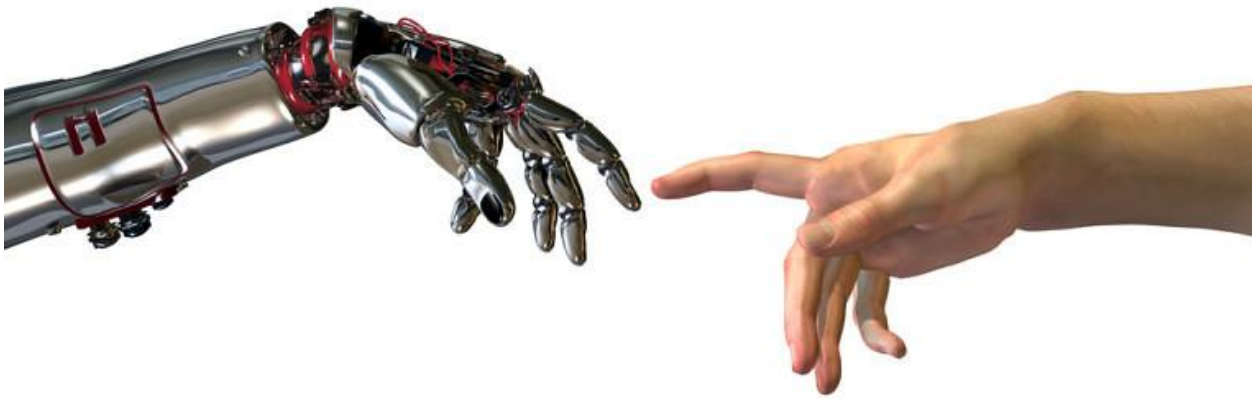


Comments of

ForHumanity[1]
Ryan Carrier, *Executive Director*
Mark Potkewitz, *General Counsel*
Sarah Clarke, ForHumanity Fellow
Markus Krebsz, ForHumanity Fellow
Sundar Narayanan, ForHumanity Fellow

In the Matter of

Artificial Intelligence Risk Management Framework

National Institute of Standards and Technology, Department of Commerce
Docket Number: [210726-0151]

September 15, 2021

[1] ForHumanity (https://forhumanity.center/) is a 501(c)(3) nonprofit organization dedicated to addressing the Ethics, Bias, Privacy, Trust, and Cybersecurity in artificial intelligence and autonomous systems. ForHumanity uses an open and transparent process that draws from a pool of over 400+ international contributors to construct audit criteria, certification schemes, and educational programs for legal and compliance professionals, educators, auditors, developers, and legislators to mitigate bias, enhance ethics, protect privacy, build trust, improve cybersecurity, and drive accountability and transparency in AI and autonomous systems. ForHumanity works to make AI safe for all people and makes itself available to support government agencies and instrumentalities to manage risk associated with AI and autonomous systems.

Letter from the Executive Director of ForHumanity -

It is a distinct honor, as the Executive Director of ForHumanity, to introduce NIST to ForHumanity, an unfunded, and rapidly growing non-profit collection of 500+ AI Ethics experts and practitioners from around the globe banded together for a specific mission - to mitigate downside risk from AI and Autonomous systems. Our collective offers our singular knowledge, expertise and crowdsourced, human-centric perspective to NIST, as we have done for the EU, ICO, Federal Reserve, EEOC and others, always on behalf of humanity. We submit our comprehensive and operational risk management framework to NIST as it may be deemed valuable, additive and complementary to NIST's goals in AI and risk management. It is my sincerest hope that we can help. - Ryan Carrier

Risk and Artificial Intelligence

ForHumanity is a mission-driven non-profit organization. That mission is *To examine and analyze the downside risks associated with the ubiquitous advance of AI & Automation, to engage in risk mitigation, and ensure the optimal outcome… ForHumanity*. Our human-centric approach is one of risk control, mitigation, and management, comprehensively. Proper management of downside risks generates better results for everyone. To that end, we have identified five key areas of risk to humans/citizens from applications of artificial intelligence:

1) Ethics
2) Bias
3) Privacy
4) Trust
5) Cybersecurity

We have developed Independent Audit of AI Systems, a transparent, crowdsourced service model for governments, regulators and authorities. We craft audit rules and criteria, submitting them to authorities (such as NIST) for consideration and/or approval (e.g. the UK Information Commissioner's office approval of certification schemes). ForHumanity currently trains and certifies AI auditors and intends to plug into a network of teaching centers (Universities and Executive education entities) to mobilize and scale. We license qualified entities to engage in audits or pre-audit compliance partnering with National Accreditation bodies when they exist (e.g., United Kingdom Accreditation Service).

Potential Risks Associated with AI

Each of the risks listed below is addressed in detail through crowd-sourced audit criteria and will be submitted for review to governments and regulators.

1) Types of bias
    a) Data Bias[2]
        i) Labelling Bias

[2] For more information about bias in data, see Shea Brown, Ryan Carrier, Merve Hickok and Adam Leon Smith: Bias Mitigation in Datasets found here: https://static1.squarespace.com/static/5ff3865d3fe4fe33db92ffdc/t/60d22a1c1e57e33b08de0cd8/1624386076920/biasInDatasets+%281%29.pdf

ii) Sample Bias
iii) Representativeness
iv) Cognitive Bias
v) Non-accessibility Bias
vi) Confirmation Bias and Sunk Cost Bias

2) Legal/regulatory risk — violation of fairness rules, anti-discrimination laws, unfair and deceptive practices, data privacy rules
3) Ethical risk[3] — failure to abide by a proper Code of Ethics, misuse of data, lack of transparency, lack of disclosure, lack of explainability
4) Model risk — bad validity, bad reliability, lack of accuracy, misalignment with scope, nature, context and purpose (both within and outside the model), model concept drift
5) Cybersecurity risk
6) Control and Safety Risk
7) Privacy breaches and failures

Given the global, jurisdictionally-sensitive, comprehensive and operational nature of Independent Audit of AI Systems, our commitment to NIST is to ensure the AI risk framework that NIST settles upon will be mapped comprehensively in a timely manner to our framework, training, and audit criteria in development for various governments and regulatory agencies.

Our mission is people-centric, and therefore, we bring a unique perspective to the NIST AI Risk framework discussion and can provide the insights of a global, diverse team of 450+ AI ethics researchers and practitioners coming from 46 different countries building independent, third-party audit criteria and infrastructure to assure governance, oversight and accountability.

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;

The greatest challenge facing AI actors in regards to managing risk is a cultural one. The US technology sector often approaches new challenges with a "move fast and break things" strategy that focuses on quick, short-term gains and profit rather than contemplating long-term risk and legal/regulatory compliance. More importantly, the "things being broken" are increasingly - people. From ForHumanity's perspective, a continuation or worse, acceleration of this culture is unacceptable. Therefore, we have embarked on a comprehensive, human-centric risk management approach - centered around a comprehensive risk management framework (Independent Audit of AI Systems) designed to

[3] For more information about Ethical Risks - ethics committees and ethical choice, see Ryan Carrier, Rise of the Ethics Committee April 2021, found here: https://static1.squarespace.com/static/5ff3865d3fe4fe33db92ffdc/t/60767acef0d59e782d2af79b/1618377424910/The+Rise+of+the+Ethics+Committee.pdf

bring governance, accountability and oversight to AI and Autonomous systems in the name of building an infrastructure of trust for people.

ForHumanity is collaborating with nation-states that sympathize with this perspective such as the United Kingdom (Children's Code and UK GDPR), the EU (GDPR and Artificial Intelligence Act) and many other countries implementing or drafting legislation designed to counteract this culture in defense of their citizens. Despite this regulatory push, many of the purveyors of AI and autonomous systems continue to aggressively avoid regulation, oversight and accountability. Risks are managed by AI designers and developers with a corporate, profit-oriented perspective, at least foundationally.

In general, they believe they can manage their own risks. Historically, independent corporate self-regulation has a poor track record - not dissimilar to the state of accounting and financial reporting pre-1973 and the formation of Generally Accepted Accounting Principles and the SEC's subsequent mandate for public reporting and accountability through independent, third party audits performed on a normalized set of independent third party rules (Financial Accounting and Standards Board).

Artificial intelligence systems have an equally important if not more far-reaching impact on the lives of Americans, than financial accounting and reporting, and thereby, it follows that independent third party audit is a natural evolution in governance, accountability and oversight for these systems.

NIST's role, as a standard-setting body, helps harmonize approaches to protect those engaging in science and commerce throughout the United States. ForHumanity believes the need for consistency requires consideration of specific risks to people in the United States and their data as users of AI systems and inputs to it. We would encourage all participants to expand their notion of risk beyond function, operation and protection for entities and consider impacts and risks directly to human beings. As the federal government considers rules and regulations for these markets, it will turn to NIST for technical guidance. Guidance on risk management must include consideration and mitigation of risks to people from these systems.

We submit that ForHumanity's unique expertise and perspective will aid NIST as it formulates an AI Risk Management framework that balances American corporate interests with the needs and protections owed to the American people and humanity in general.

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;

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;

ForHumanity's approach to risk management centered on Ethics, Bias Privacy, Trust and Cybersecurity and while that focus has not waivered, we have wrapped those pillars with a risk management, moral framework that is: ethical, human-centric, accountable, governable, overseeable, transparent, documentable, proveable, evidence-based, and independently auditable. Below we summarize our perspective on a comprehensive risk management framework considered from initial system design to decommissioning execution.

Human-centric, Ethical and FAIR	**Actionable, Operational and ACCOUNTABLE**	**Auditable, Certain and TRANSPARENT**
Privacy-by-design	Cybersecure-by-design	Auditable-by-design
Accessible-by-design	Objectivity/Governance/ Oversight	Pre-production evidence
Data Quality/ Representativeness	Data Control	Disclosure
Active Bias mitigation	Controllability of autonomous systems	Third-party independent audit on entire process
Ethical oversight	Post-market monitoring	
Explainability		

- Privacy-by-design
- Auditable-by-design
- Accessible-by-design
- Cybersecure-by-design
- Objectivity/Governance/Oversight
    - Committee structures for objective oversight
    - Committee structure for accountability
    - Management of single-point-of-failure risk

    - Board culpability
    - Corporate cultural risk-awareness adoption
- Data Quality/Representativeness
    - Labelling Bias
    - Sample Bias
    - Protected category representativeness
    - Cognitive Bias
    - Non-accessibility Bias
- Active Bias mitigation
    - At the data layer
    - At the Protected Category Variable layer
    - At the architectural input layer
    - Over outputs for disparate impact and discrimination
- Data Control
    - Mitigation of Data Entry Point Attacks
    - Mitigation of Human-in-the-loop Attacks
    - Pseudonymization and Anonymization
    - Special Category and Biometric enhanced protections
- Pre-production evidence
    - Contextually sufficient validity
    - Contextually sufficient accuracy
    - Contextually sufficient reliability
- Ethical oversight
    - Segmentation of Ethical choices to trained professionals
    - Algorithm Ethics
    - Reasonable and fiduciary considerations
    - Diverse inputs and multi stakeholder feedback on design, development and risk assessment
    - Human oversight or integration considerations ([in or on]-the-loop)
    - Elevated Children's considerations
- Controllability of autonomous systems
    - Resilience
    - Robustness
    - Model Obedience
- Disclosure
    - Ethical choices
    - Soft law interpretations
    - Data Disclosure Documentation
    - Privacy and data usage
- Post-market monitoring
    - Validity stability
    - Accuracy stability

    - Reliability maintained
    - Assess for Concept Drift within predetermined KPI and KRIs
    - Interpretability
- Explainability
    - Reasonableness standards
    - Education for negative outcomes for improvement
- Third-party independent audit on entire process
    - Binary (compliant/non-compliant certainty)
    - Certified practitioners
    - Regular reviews and real-time monitoring
    - Third party rules
    - Upholding Independence

ForHumanity advocates for a risk management framework that is omni-directional - considering corporate risks (which damage employees and shareholders), risk to humans (to users/clients/prospects and unwitting participants), societal risks (to our systems, groups, communities, markets and collectives) and environmental risks (to nature and sustainability considerations). All of these vectors result in a residual risk after maximized risk mitigations. These residual risks, well disclosed and considered, will empower an increased ability to deal with emerging risks, concentrated research on novel mitigations and the informed acceptance of consequences when residual risk manifests itself.

The characteristics listed in the question are reasonably comprehensive, but the key challenge is consolidating a view of residual risk. How much inherent risk does improved explainability, or accuracy, or security, or privacy have potential to mitigate? This thought process naturally leads to more thorough considerations such as, documentation, governance, oversight and accountability. Subsequently, these characteristics accumulate to an endgame - Independent Audit.

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;

Current efforts in adapting Enterprise Risk Management (ERM) to AI Risk Frameworks have been sporadic, limited and temporary. Issues of trustworthiness, explainability and transparency have been largely glossed over or as an extension to cyber security and many of them are control monitoring approaches. Control and monitoring frameworks such as

the Committee of Sponsoring Organizations of the Treadway Commission (COSO) and System and Organization Controls (SOC 2) have been viewed as sufficient. Internal control perspectives have become unsustainable and now feedback loops including stakeholder, civil society are necessary for proper enterprise risk management and are creating demand for adaptations to existing models including KPI design and post market monitoring activities.

Most organizations currently take a siloed approach to management of AI related risk. This frequently breaks down into more traditional functional areas: quality, safety, cybersecurity, data protection, software development, and data analytics. This was evident in McKinsey's Global Board Survey for 2017[4], and 2021 FICO report "The State of Responsible AI"[5]. Collective ForHumanity fellow and contributor experience backs up reports of a lack of cohesion in attempts to manage these risks.

Organizations will occasionally engage external specialists to provide insight into Ethics and Bias in data sets, but outputs are rarely combined with other risk insights from other specialist areas. Board members, who frequently lack training and understanding certainly cannot form a coherent view of AI related risks, and effectively manage them on behalf of either the organisation, or their customers. A key blockage to improving risk assessment and risk management coherence is immaturity in associated governance and oversight functions, often underpinned by lack of clarity about accountability across functional areas and up through reporting lines.

Any risk management framework needs to address those governance and risk management challenges. There needs to be a path to maturity for AI governance and oversight and a satisfaction of the omni-directionality of risk associated with AI.

Independent Audit[6] provides a comprehensive process that classifies all AI risk into ethics, bias, privacy, trust, and cybersecurity, breaking down each classification into thousands of binary (compliant or non-compliant), implementable and measurable normative statements that are crowdsourced by hundreds (and soon to be thousands) of AI experts globally. These criteria are iterated over and over again until they are submitted to governments and

[4] Value and Resilience Through Better Risk Management, 2018, Daniel Guis et al for McKinsey, reporting on McKinsey Board Survey for 2017
https://www.mckinsey.com/business-functions/risk-and-resilience/our-insights/value-and-resilience-through-better-risk-management

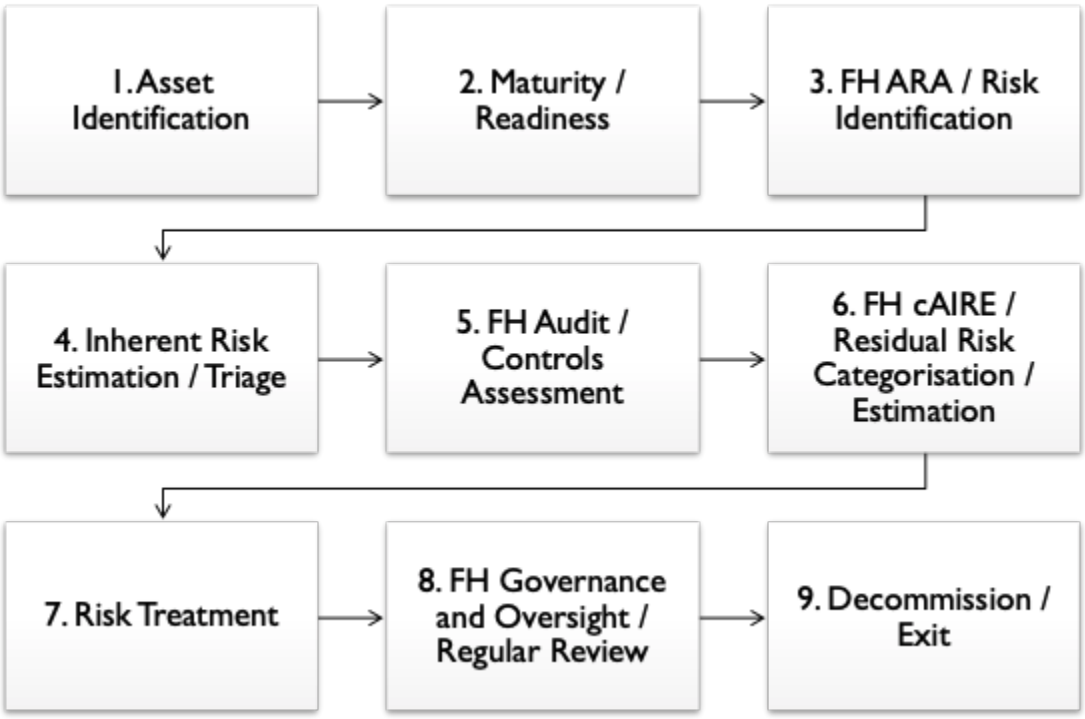
[5] It's 2021. Do You Know What Your AI Is Doing?, 2021, Scott Zoldi for FICO reporting on FICO and Corinthian report: The State of Responsible AI
https://www.fico.com/blogs/its-2021-do-you-know-what-your-ai-doing

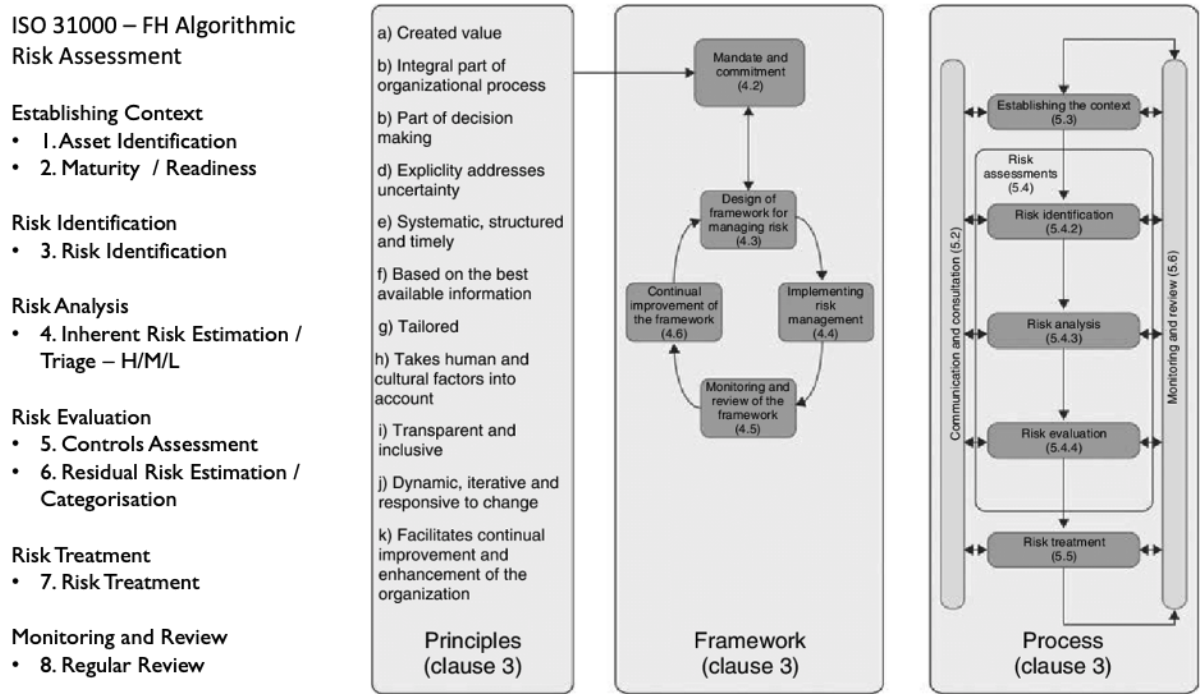
[6]Governing AI Safety through Independent Audit please review from Greg Falco et al in Nature Machine Intelligence .
https://www.nature.com/articles/s42256-021-00370-7?error=cookies_not_supported&code=636ac32e-1dfa-4267-8b2d-8d5b9f8bd6bf

regulators for approval. Each set of criteria are jurisdictionally sensitive, system-specific and consider risk omni-directionally.

A specific example is the responsible feedback loop requiring organizations to add to the ERM approach, considerations for the human-centric layer: ethics-by-design, privacy-by-design, data ethics, and bias mitigation at design, development and deployment phase.

How we mapped risk management activity to ISO31000





Independent Audit of AI Systems is a comprehensive risk management system, designed to manage and mitigate omni-directional risk in Ethics, Bias, Privacy, Trust and Cybersecurity. For this question, it is important to further breakdown how we treat Trust. #InfrastructureOfTrust[7] explains eight different ways that trust is built, includes transparency, accountability, disclosure, explainability but also control and safety.

Safety, in the case of AI and autonomous systems, includes sufficient auditable testing to determine validity and reliability for a system combined with KPI designed to ensure that the system stays consistent with its predetermined scope, nature, context and purpose. We advocate for post market environmental and output based testing, similar to those present from the National Transportation Safety Board (NTSB) but those testing and evaluation measures (beyond audit compliance of reliability, validity and KPI maintenance) of production-ready systems are beyond the scope of the Independent Audit of AI Systems criteria.

With regard to Privacy, Independent Audit of AI Systems has a comprehensive GDPR certification scheme submitted to the UK's Information Commissioner's Office. Regarding cybersecurity, NIST has been a global leader with its gold standard framework and ForHumanity is launching a program to comprehensively map that framework to auditable

[7] See explanations of how humans trust from #InfrastructureOfTrust by Ryan Carrier https://cacm.acm.org/blogs/blog-cacm/250260-auditing-ai-and-autonomous-systems-building-an-infrastructure-of-trust/fulltext

criteria. Cybersecurity is one of the five pillars of Independent Audit of AI Systems and audit compliance includes compliance with the NIST cybersecurity framework.

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;

Compliance with Independent Audit of AI Systems requires a robust risk management strategy that complies with exacting criteria. As designed, the criteria require risk management systems to assess, monitor and mitigate the sum total of risks that our global, crowdsourced approach has been able to identify. Independent Audit of AI Systems adapts best practices and the law into jurisdictionally sensitive auditable criteria around the world. When best practices are developed ForHumanity maps them to operationalize local terminology, law, regulations and best practices.

Independent audits performed by third parties using statutory criteria, regulations, or industry custom, such as the Generally Accepted Accounting Principles (GAAP) or International Financial Reporting Standards (IFRS), create a system that encourages proactive legal and regulatory compliance. ForHumanity believes this system represents a more trustworthy environment for the processing of personal data using emerging technologies.

Independent Audit of AI systems is an entire ecosystem already being operationalized. Our Taxonomy[8] differentiates characteristics of audit/assurance/assessments and how these valuable services are delivered to the marketplace. Infrastructure of Trust - A Guide to Roles and Responsibilities[9] lays out the roles for governments/regulators, accreditation bodies, ForHumanity, certified practitioners, auditors, pre-audit service providers, auditees and the public. Everyone has a role to play in managing the risk from AI and autonomous systems. These criteria, crafted to be auditable (binary, measurable, implementable) adapt the law and best practices. This process is done in a crowdsourced manner where we concentrate AI experts, AI Ethics experts, attorneys, practitioners, philosophers, designers and developers to ensure that the criteria protect humans (by maximizing risk mitigation). These criteria are submitted to government authorities for approval.

[8] To read more - see Taxonomy by Ryan Carrier and Shea Brown https://static1.squarespace.com/static/5ff3865d3fe4fe33db92ffdc/t/60329e0a4cfbaa172691f7e6/1613929999802/Taxonomy+of+AI+Audit+%282%29.pdf

[9] To read more about Roles and Responsibilities - see Infrastructure of Trust a Guide to Roles and Responsibilities by Ryan Carrier https://static1.squarespace.com/static/5ff3865d3fe4fe33db92ffdc/t/60afa2c08b921273acba3a24/1622123209761/Infrastructure+of+Trust+for+AI+-+Guide+to+Entity+Roles+and+Responsibilities.pdf

Pre-audit service providers will prepare auditees for audits. Auditors, however, put themselves at risk for false assurance of compliance, bearing the risk on behalf of society to ensure that AI and Autonomous systems are compliant with the aforementioned criteria. This service (risk-taking) is indispensable towards building governance, accountability, oversight and trust. This system is a comprehensive assessment of risk ranging from internal controls and risk management, to external assessments of risk impact and on to systemic riskiness.

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;

Regulators around the world have been tackling AI risk only so far as they relate to another primary target area, such as data privacy, employment bias, consumer protection, or health. Examples include privacy (CCPA, GDPR), bias (NYC Bias Audit), fairness, accountability and risk-based governance (EU Artificial Intelligence Act) and data storage and transfer (Court of Justice of the European Union's *Schrems II* ruling). ForHumanity is working with regulators in most of these jurisdictions and would provide that mapped and harmonized effort to the NIST AI Risk framework process.

Independent Audit of AI Systems adopts and improves upon the model of public accounting. Independent Audit of AI Systems mimics this system with minor enhancements, such as a globally, crowd-sourced system to ensure best-practices are readily shared across national borders and those regulators may choose to avail themselves of these best practices provided with the singular mission of protecting humans from risk from AI and Autonomous systems.

We expect the documentary evidence produced by Independent Audit of AI Systems audit compliance will greatly enhance regulatory oversight and reporting requirements. By normalizing standards across levels of government and even across nation-states, ForHumanity can help maximize the global interoperability of the audit criteria in a superior way than any specific government agency that does not abide by the same raison d'etre as ForHumanity's global mission to benefit humans.

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;

It is common and efficient to adopt AI principles defined by respected independent bodies, rather than defining principles internally. For example;

- The EU High Level Expert Group Ethics Guidelines for Trustworthy AI 2019;[10]
- The OECD Recommendation of the Council on Artificial Intelligence 2019;[11] and,
- The Alan Turing Institute Understanding artificial intelligence ethics and safety 2019.[12]

Each of these collections of AI principles incorporates transparency, fairness, and accountability either explicitly or in description of coverage. However, they focus more on the potential impact on individuals, society and the environment in complement to Draft NISTIR 8312 Four Principles of Explainable Artificial Intelligence and other NIST papers. Our primary recommendation is a shift in perspective to the omni-directionality of risk, rather than an inward-focused, corporate-centric model.

NIST is an important beacon for best practice synthesis in artificial intelligence — attracting many of the best and brightest minds to provide input and guidance for the United States. ForHumanity will ensure that NIST conclusions (best-practices and guidance) are mapped onto its already operationalized framework.

Where new best practices are developed or law enacted on the back of that guidance, ForHumanity will translate it into executable, audit criteria quickly once it becomes law.

ForHumanity does not aim to replace traditional Standards Development organizations (SDOs). However, since SDOs tend to rely largely on industry funding, we aim to provide some balance to further standardisation in the name of humanity. ForHumanity will drive grass-roots standards into areas that SDOs do not prioritise or do not reach consensus at sufficient pace. The art of ForHumanity's work and our primary mission however is distinct from all standards-making bodies specifically because our end goal is always binary (compliant/non-compliant) audit criteria.

At the present time, ForHumanity does not consider that any of the normative standards published by International Standards Organisation (ISO) /International Electrotechnical Commission (IEC), European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC), and British Standards Institute (BSI) on Artificial Intelligence are relevant to our certification schemes. However, ForHumanity Fellows participate in each of these organizations at a working level and aim to align in three ways.

---

[10] Ethics guidelines for trustworthy AI 2019
https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1
[11] Recommendation of the Council on Artificial Intelligence 2019
https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
[12] Understanding artificial intelligence ethics and safety 2019
https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

1. Where fundamental terminology is defined that can be used to drive consistency, we will adopt it (e.g. ISO/IEC DIS 22989 — Information technology — Artificial intelligence — Artificial intelligence concepts and terminology).

2. Where informative guidance is produced, it will be reviewed and considered (e.g. ISO/IEC DTR 24027 - Information technology — Artificial Intelligence (AI) — Bias in AI systems and AI aided decision making).

3. Where normative standards are produced that can be used for conformity assessment (e.g. by CEN/CENELEC JTC 21), they will be considered with a view to adoption. However, it is noted that international standards require international consensus. ForHumanity will not automatically weaken or lower standards due to a lack of international consensus.

As alignment naturally grows between ForHumanity and SDOs over time, ForHumanity will increasingly focus on the certification schemes and professional skills required to deliver independent audits.

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.

In 2019, after repeated criticisms of a "profit-only" *raison d'etre*, the Business Roundtable addressed the role of corporations in society with this amendment to the statement of purpose of a corporation:

> *the release of a new Statement on the Purpose of a Corporation signed by 181 CEOs who commit to lead their companies for the benefit of all stakeholders – customers, employees, suppliers, communities, and shareholders.*[13]

This change in business attitudes combined with the ability of society to coalesce around key issues has resulted in the desire and need for inclusivity. Risk management frameworks must incorporate this change in the interface between corporations and society.

[13] For further explanation see the press release from the 2019 Business Roundtable meeting https://opportunity.businessroundtable.org/ourcommitment/

The NIST Risk Framework focuses on the first two lines of defense — operational and legal/compliance. Independent Audit of AI Systems provides a third line of defense from the assurances provided by independent, third party auditors acting as verification proxies for the public. This third line of defense has become increasingly necessary as the impact from AI and autonomous systems has grown. AI and autonomous systems have substantial impact outwardly on humans, communities, society and the environment that need to be assessed and mitigated at the design and development stage to minimize downside risk to humans.

An example of the human-centric layer of IAAIS, requires all AI and autonomous systems that impact humans to go through an Algorithm Risk Analysis (ARA), detailed in our Body of Knowledge Repository,[14] which documents over 50 different areas of compliance documentation with sufficient and maturity satisfaction. The ARA by definition is a diverse input and multi stakeholder risk assessment approach. Governed by the Algorithmic Risk Committee, an accountability structure reporting to the CEO in collaboration with the Ethics Committee who, objectively, assures that sufficient diversity, in accordance with the diversity policy of the entity, exists amongst the multi stakeholders providing feedback. This risk assessment system assures human agency at the design, development and risk assessment phase specifically identifying adverse impact to groups by examining severity and likelihood or risk. This assessment must be thoroughly documented.

By engaging this process at the design and development phase, the ARA helps to manage both of those phases and avoid confirmation bias and sunk cost bias overriding input from a wide array of stakeholders. Furthermore, it provides accountability, governance and oversight to the design and development process ensuring that those teams have objective input that is devoid of conflict of interest innate in the teams working for the entity building the system.

Another risk ForHumanity has identified is the market impact and market dynamics associated with AI systems can change quickly. We are developing a framework to assess these market dynamics which translate into new or different risk profiles for AIs. This system, called the Systemic Societal Impact (SSI) Assessment, incorporates a combination of diverse inputs and multi stakeholder feedback with quantitative, market-based, customer/user/client feedback designed to measure and track these features of the system. The quantitative aspect of the SSI is designed to create automatic triggers for reassessment when thresholds are reached. These triggers are necessary to combat the speed of change in scope in some of these systems or features of the system. Viral growth rates of adoption and acceptance of systems are a risk unto themselves. The elements of an SSI consider the

[14]For further information on ForHumanity's Body of Knowledge - Knowledge Stores https://forhumanity.center/body-of-knowledge-bok/

following features or aspects attributable systems such as large language models and facial recognition which might be adopted ubiquitously:

1) Market penetration
2) Perceived system authority
3) Control of participation
4) Nature of the system
5) Environmental/Sustainability Impact

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

NIST development and Attributes

1. Flexible
2. Risk-based
3. Outcome-based
4. Cost-effective

We would encourage the following additions:

1) Human-centric
2) Ethical
3) Actionable
4) Operational
5) Auditable
6) Certain

*and all of the subcomponents of those overarching principles.

We support flexible systems because of the number and variety of AI designs and applications. The audit criteria contemplate two vectors: 1) Top-down accountability, governance and oversight 2) laterally, AI system by AI system. The top-down approach creates accountability systems for ethics, bias, privacy, trust, and cybersecurity for the Board of Directors, Chief Executive Officer and Chief Data Officer. Committee structures are required such as Algorithmic Risk, Children's Data Oversight, and Ethics to manage the audit/compliance responsibilities. All of these top-down criteria apply to every AI and every autonomous system in the organization. The system-specific audit criteria is designed to ensure legal and best practice compliance tailored to the specific impact of each system

on humans. This comprehensive approach ensures consistency across the organization combined with complete risk management coverage of each unique system.

Independent Audit of AI Systems is risk-based, considering outcomes and cost in the creation of auditable criteria, but puts the needs to mitigate risk to humans before the compliance costs of entities. The last point however is not simply Boolean, but rather a collection of tensions and tradeoffs balancing between the need to mitigate downside risk to humans against the needs of humans to benefit from the use of systems and benefit as employees/ shareholders of the companies producing these systems.

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

Independent Audit of AI Systems is a comprehensive risk management framework that builds in governance, oversight and accountability to create an infrastructure of trust in AI and autonomous systems. The system is a traditional audit system that allows for significant pre-audit compliance work to build compliance capabilities across the areas of ethics, bias privacy trust and cybersecurity. Third-party pre-audit service providers will work with companies to build internal risk controls and governance systems to enable compliance with the audit criteria. Independence declares that pre-audit service providers may not be the auditor for the company they advised.

Audits have three primary characteristics that make them distinct from assessments. First, they are done by certified practitioners. Second, they use independent, third party rules to ensure objectivity with respect to the criteria. Third, audits are not done for auditees, but rather over auditees. Audits are conducted for the public's benefit, by auditors acting as a proxy for society. Auditors therefore have to wear the risk of false assurance of compliance and their only remuneration may be fair market value of their audit services. These characteristics ensure that when compliance is asserted, then it can be trusted. No system is perfect and fraud and malfeasance will defeat most transparent systems of risk management and accountability, but through transparency, disclosure and governance of auditors, eventually even fraud and malfeasance will be caught out as Enron and Worldcom proved in the early 2000s in financial audits.

Critically, the system will require certified practitioners and criteria that are auditable. Standards, laws and anything short of auditability, short of binary criteria that is compliant or non-compliant will result in auditor uncertainty and a subsequent "non-compliance" decision. Auditable criteria is the critical foundation by which an auditor can issue compliance decisions and assure the public that maximum risk mitigations have been conducted on the audited system.

Regarding tiers, the states of regulatory frameworks, guidance state and federal law have already created the tiers. Some compliance is already required, others are considered best practices and still others are ahead of their time. Barring a comprehensive mandate that might only come from the Securities and Exchange Commission for publicly traded companies (or some other equivalent catch all for large companies), the ubiquitous nature of AI touches many different agencies already and will likely result in fragmented implementation, piece-meal law making, specific regulatory guidance and sporadic mandates for the near future.

ForHumanity will create a complete audit criteria mapping for the NIST Cybersecurity framework, allowing for the creation of independent audit of cybersecurity on a standalone basis and the adoption, as appropriate of NIST framework criteria into all appropriate ForHumanity audit criteria around the world.

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.

Capacity building needs to be structured as part of the operational framework. Currently, there are not many specific risk management courses or certifications with such particular focus on AI-related risks. However, ForHumanity is training individuals on all of our government approved audit criteria. These certifications ensure auditors knowledge of criteria and understanding of satisfactory audit compliance. All persons demonstrating comprehensive expertise on the audit process and criteria can pass an examination and become ForHumanity Certified Auditors (FHCA). We expect over time that this certification process will expand to resemble other fields where expertise proved through examination is complemented by some practical experience in the field under the guidance of practitioners. As this is a new field, mentorship will have to grow as a certification criteria.

ForHumanity provides forums for discussion, a Body of Knowledge (BOK) repository maintained for over 50 different criteria of compliance. The BOK provides two resources to the auditor. First it is a source of visual and document sufficiency and maturity around items of compliance, allowing the FHCA to check the compliance of the company against the repository. Second, as this is a dynamic marketplace, the BOK provides a forum for

discussion where new developments, new best practices, new methods of audit satisfaction can be considered, discussed and/or implemented as they are brought forward.

This training and education starts with ForHumanity, where the criteria are approved by governments and regulators around the world. ForHumanity is developing a global infrastructure of trainers who can help FHCA get certified on new criteria. FHCA maintain their knowledge through continuing education and they abide by the ForHumanity Code of Ethics and Professional conduct ensuring that their work and behavior is held to the highest standards. Embedded in the code is a responsibility of continuing education - to remain current when criteria, laws, guidelines and best practices changes - a critical responsibility in this dynamic marketplace.

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.

A critical aspect of ForHumanity's certification scheme is robust governance, oversight and accountability at the Corporate Officer and Board of Directors level to ensure trust. It is not sufficient to require "the organization" to comply with aspects of this certification scheme. Simply requiring an organization to comply with certification criteria without defining specific accountability in governance structures creates ambiguity regarding delegation of authority and/or segregation of duties. Further, organizational compliance alone does not eliminate conflicts of interest, and its general nature could offer 'cover' to persons designing, developing and maintaining the data controlling/joint-controlling/processing system without the requisite experience to satisfy all criteria. The same ambiguity is often the excuse offered when non-compliance occurs and blame is shifted. Clarity of responsibility and accountability are hallmarks of ForHumanity certification schemes.

The certification scheme requires the establishment of an Algorithmic Risk Committee, Ethics Committee and Children's Data Oversight committees to ensure conflict free, objective decision-making. This approach mitigates against any individual gaps in expertise or training. Examples of expertise include: 1) awareness of instances of Ethical Choice 2) discipline to execute Ethical Choice 3) awareness of the special needs of children 4) bias mitigation techniques including for cognitive biases 5) establishment of Key Performance Indicators to manage Concept Drift. This design deals with the expanse of skills needed to effectively manage risk in data controller/joint-controller/data processing works as well as acting as a check on the analysis, outputs and/or decisions rendered including algorithmic systems, artificial intelligence or autonomous systems. This governance structure is especially valuable in areas of emerging risk.

The Board endorses the formation of these committees and makes them culpable for systemic failure in the organization, which enhances the likelihood of robust systems of compliance. Liability at the CEO level, where each of these committees report, ensures responsible structural compliance across the entire organization. The result is a culture of compliance — from design to decommission — with clearly delineated responsibilities, reporting lines and final accountability that increases transparency and advances governance.

The most tangible expression of a changed culture can be measured in resource allocation. Resource allocation (time, investment, expertise) is a strong signal to both internal and external stakeholders that a culture of governance, oversight and accountability is valued by the Board and Officers. The committee structure, a proven model of risk management, further ensures that checks and balances exist when these critical decisions and reviews are executed. Committees are responsible for having trained members and appropriate documentation to meet the demands of the audit criteria.

Yet risk remains omni-directional, thus a comprehensive system of governance, accountability and oversight combined with a market-based expectation of audit compliance and associated transparency and disclosure also lends itself to a robust ecosystem for supply chain risk management as audit criteria and the expectation of review drive compliance as well. This normalization should result in increased transparency and the expectation of robust risk management that can minimize vendor integration risks and acquisition risks.

As NIST seeks to gather inputs for an AI Risk management framework, we strongly encourage an omni-directional approach to risk assessment, but one that starts and finishes with humans. NIST's original mandate was to serve Congress's mandate to govern standard weights and measures, but the impetus for that was "to protect people in their commerce". A human-centric, ethical, fair, actionable, operational, accountable, auditable, certain and transparent risk management approach will ensure robust mitigation and thus achieve the maximum benefit from these systems. ForHumanity offers our work to the service of that mission and stands ready to assist NIST in its endeavors.