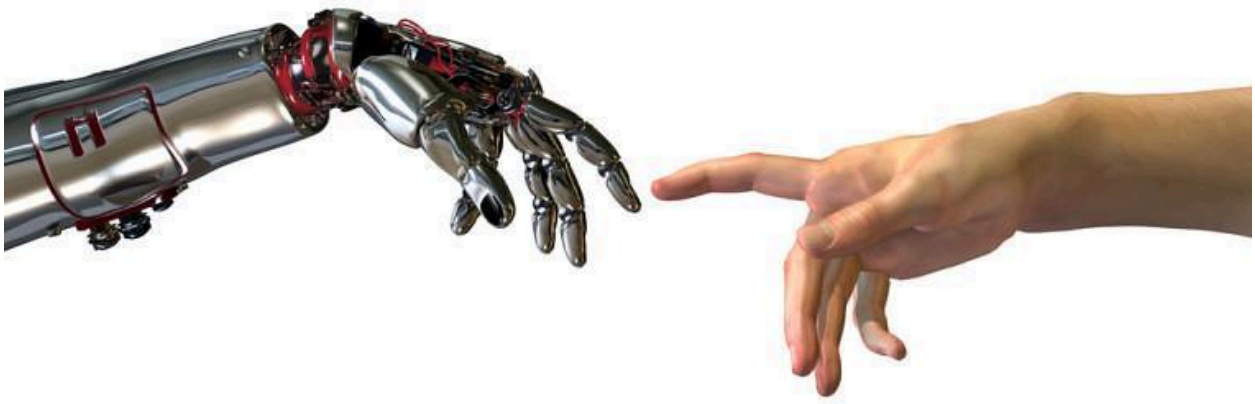


ForHumanity[1]

Ryan Carrier, *Executive Director*
Sundar Narayanan, Fellow, ForHumanity

with valuable foundational contributions from Sarah Clarke, Fellow, ForHumanity

Risk Management Process

Feb 24, 2022

[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 850+ 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.

Introduction and understanding of GRC

Governance, Risk Management and Compliance (GRC) can be a confusing term in the AI, algorithmic and autonomous systems (AAA Systems) space. Sometimes it is thought of as a function, other times as a process and still other times including assurance and performance management. ForHumanity's risk management framework and processes cover the Governance, Risk management and compliance aspects of GRC with Ethics, Bias, Privacy, Trust and Cybersecurity as key pillars (reflecting instead negative impacts to humans as the focal point), regardless of the organization's silos.

ERM & FH Risk Management

Enterprise Risk Management is the process and method adopted to manage risks that may impact the organization's objectives. Enterprise risk management is an umbrella term for overall risk efforts of the organization. Typically, the Chief Risk Officer along with the Operational Risk Management function hold responsibility in driving Enterprise Risk Management efforts. FH AI Risk management is intended to be a feed to ERM and operational risk management. Furthermore, FH risk management requires functional risk efforts to be complemented by representatives from operational risk management enabling better integration of organization wide efforts.

FH AI Risk Management Process

ISO 31000 provides a great foundation for the corporate risk management process. ForHumanity's mission is exclusively focused on downside risks caused by AI, algorithmic or autonomous systems to humans, society and the environment. Using that lens, we build upon ISO 31000 framework comprehensively for AI Risk Management with 3 key focus approaches:

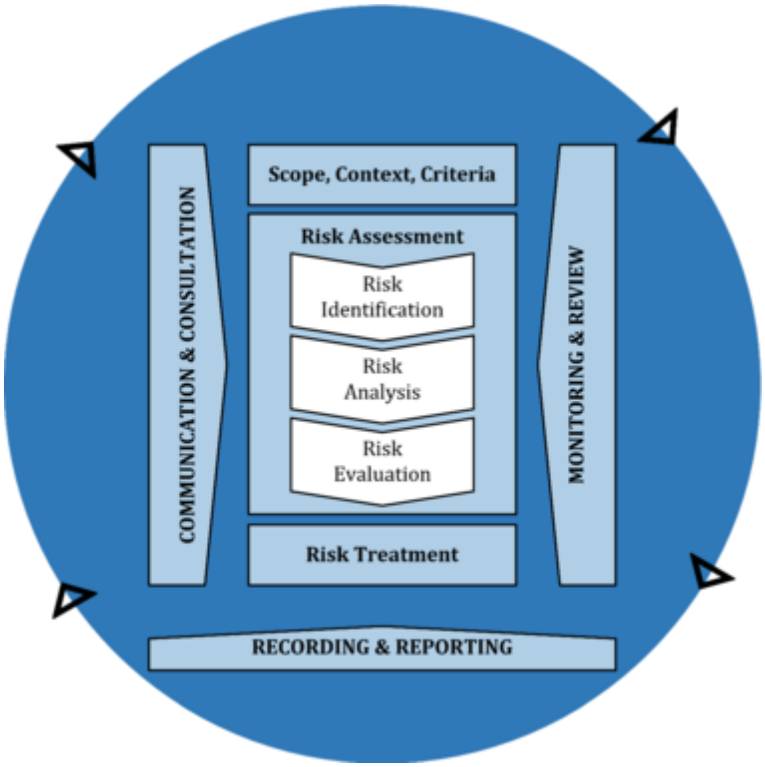
1. Risks are considered from the perspective of impact to humans, society and the environment. These are socio-technical systems where humans are ever present and therefore risks emerging are categorized by Ethics, Bias, Privacy, Trust and Cybersecuirty and subsequently considered.
2. Risks to humans are identified through Diverse Inputs and Multi Stakeholder Feedback (internal and external - including civil society) mechanisms all through the lifecycle (design, development, deployment and decommissioning of AI, algorithmic or autonomous systems)
3. Risks to humans emerging from incidents, adverse incident reporting and/ or post market monitoring of AI, algorithmic or autonomous systems are integrated and feed-back to provide specific insights

The FH Risk management process is operationalized at the Functional Risk Management level managed by respective committees (eg. Algorithmic Risk Committee) or duly

designated individuals. Execution is either integrated with the Organization's Enterprise Risk Management System or documented and fed into the ERM's residual risk repository.

1. **AI Risk Management process**: The Risk Management process involves the systematic application of policies, procedures and practices to the activities of communicating and consulting, establishing the context and assessing, treating, monitoring, reviewing, recording and reporting risk with reference to data processing and AI, algorithmic or autonomous systems. (adapted from ISO guidelines 31000 - https://www.iso.org/obp/ui/#iso:std:iso:31000:ed-2:v1:en)

   AI Risk Management cycle



2. **Scope, Context and Criteria:** FH Risk management process considers Scope, Nature, Context and Purpose as foundational elements for examining risks to humans, society and the environment from AI, algorithmic and autonomous systems.

3. **Risk Categories:** FH Risk Management examines risks from the perspective of areas (Risk Categories) where the organization has a responsibility and opportunity to treat, mitigate and/or manage risks. These Risk Categories permit useful grouping of new AAA Systems related risks to individuals, society, and the environment. FH Risk Management contains 13 key Risk Categories. They are:

1. Privacy	7. Transparency
2. Security	8. Explainability
3. Safety	9. Accountability
4. Bias	10. Accessibility
5. Governance	11. Diversity
6. Ethics capability	12. Human agency
	13. Sustainability

4. **Risk spectrum:** FH Risk spectrum focuses on negative impacts that can result from non existent, inadequate, ineffective or inefficient mitigations for the Risk Categories in AAA systems, with potential to impact people, society, and the environment. Risk Categories are best handled with appropriate mitigations and controls applied to each risk input including policy, process, communication, review mechanism, oversight activities and governance & reporting mechanism.

5. **Risk identification:** Risk input is derived from perception or experience of the organization and humans considered from the perspective of Diverse Inputs and Multi Stakeholder Feedback. Risk can be identified (a) by populating possible known risks (given the industry/ domain / use case) of AI, algorithmic or autonomous system, (b) by conducting secondary research and/ or (c) by enquiring the stakeholders (internal and external) directly or via a survey and/ or a discussion or (d) by expanding perceptions towards emergent and foreseeable risks from innovative technologies deployed within or for the organization.

Risk Input	information, insight or perspective about negative impacts along with their causality/ root cause that provides clarity for mitigating or managing them.
Risk Indicator	information, insight or perspective about negative impacts with unknown causality/ root cause that requires further examination for gaining clarity on the root causes prior to mitigation or management (eg. **Key Risk Indicators** in

	IAAIS audit criteria, **Adverse Incident Reporting System**).
How Risk Indicators become Risk Inputs?	Risk inputs are derived from risk indicators by testing, iteratively learning, understanding causal relationships, understanding interactions between the indicators, conducting experiments and/ or performing any other scientific or time tested method to gain clarity on associable root causes for the given risk indicator.
What happens if Risk Indicators don't become risk inputs?	In instances where the associated causality could not be established, the risk indicator remains as a residual risk.
Residual Risk	Unmitigated risk pertaining to a specific risk input or the aggregation of all risk in an AI, algorithmic or autonomous system. (Refer AVR3)

6. **Perceiving risk input and or indicator**: The following illustrative (non-exhaustive) methods help stakeholders think about and perceive the risk input or indications. They are:
    a. Examining the intended use and reasonably foreseeable misuse.
    b. Examining adequacy of coverage, appropriateness of application and effectiveness of controls with reference to each of the Risk Categories from the perspective of people, process and technology.
    c. Examining what can go wrong in each of the Risk Categories.
    d. Scanning or probing for emerging threats.
    e. Reviewing the adverse incidents gathered from post market monitoring.
    f. Gathering experiences or perceptions on negative impacts from Diverse Inputs and Multi Stakeholder Feedback that maximizes inclusive thought processes and lived-experiences.

7. **Risk analysis**: A process of examining Risk Input impact on people, communities and the environment, and their underlying root causes. A process of examining impact of the Risk Indicators to people, communities and the environment (tracking as residual risk). Risk inputs along with their impact form a Risk Log.

8. **Risk Evaluation**: A process wherein the risks from the Risk Log are ranked in the context of the likelihood and severity of the said risks.Risk level is determined based on Severity and Likelihood and added to the Risk Log.

    a. **Risk severity and likelihood**: The Organization in conjunction with Diverse Inputs and Multi Stakeholder Feedback shall determine the scale for risk severity and likelihood and risk levels at their intersections.
    b. Illustrative Risk Matrix with Risk severity, likelihood and Risk Level matrix is provided below:

Risk Management Matrix (AS/NZS ISO 31000)
Likelihood x Consequence

	Consequences				
Likelihood	1 Insignificant	2 Minor	3 Moderate	4 Major	5 Catastrophic
A	Moderate	High	Extreme	Extreme	Extreme
B	Moderate	Moderate	High	Extreme	Extreme
C	Low	Moderate	High	Extreme	Extreme
D	Low	Low	Moderate	High	Extreme
E	Low	Low	Moderate	High	High

9. **Risk Treatment**: The process of identifying and implementing appropriate measures to modify risk impact. The measures include
    a. Avoidance (activity that gives rise to risk is not undertaken),
    b. Reduce (risk mitigated by deploying internal controls)
    c. Share (risk mitigated in parts by fixing one or more third party accountability to prevent or recover from risk impact)
    d. Transfer (risk mitigated by fixing third party accountability to prevent or recover from risk impact) and
    e. Acceptance (residual risk)

10. FH risk management requires **Residual Risk** to have appropriate disclosure to the users (humans) for enabling informed choices while using AI, algorithmic or autonomous systems.

11. Communication & Consultation and Monitoring & Review are ongoing feedback loops to enable and enhance the risk assessment (including reassessment), evaluation (reevaluation) and treatment process.

12. **Communication and Consultation:** Given the socio-technical and multi-disciplinary nature of the AI, algorithmic and autonomous systems in conjunction with the involvement and impact to humans in the process-communications need to be wider, transparent and benefit from robust disclosure. Teams involved in Design, development, deployment and decommissioning are inadequately equipped to manage multivariate and evolving risk to humans that manifest in AI, algorithmic or autonomous systems.

    In response, such communications and consultations contribute to effective risk management by integrating Diverse Inputs and Multi Stakeholder Feedback to maximize risk inputs and associated mitigations, thereby enabling a virtuous feedback loop for risk management.

    The Algorithmic Risk Committee shall establish a mechanism for enabling Diverse Inputs and Multi Stakeholder Feedback across the risk management process and define criteria for identifying and engaging with this group. The EC shall ensure sufficient diversity and representativeness, consistent with the Code of Ethics and associated diversity and antidiscrimination policies, of Diverse Inputs and Multi Stakeholders pool in line with the Scope, Nature, Context and Purpose of the AI, algorithm or autonomous systems.

13. **Monitoring and review:** The Algorithmic Risk Committee shall ensure that (a) Risk reassessment, (b) Control validation, (c) Adverse Incident Reporting System and other post market monitoring mechanisms that trigger the need for reassessment, reevaluation and/ or incremental or alternative treatment.

14. **Reporting and Governance**: The Algorithmic Risk Committee shall determine the periodicity of preparing & sharing of the Algorithm Risk Assessment as it relates to AI, algorithmic or autonomous systems. The Algorithmic Risk Committee shall be responsible for reporting on material risks to relevant stakeholders based on the risk levels, likelihood and consequences (including adverse incidents).

15. **Evaluating effectiveness of the risk management process:** Effectiveness of risk management requires few considerations including (a) Extent of adverse incidents prevented with the risk management process, (b) Extent of adverse incidents not assessed as part of risk assessment, (c) Efficiency and Effectiveness of mitigation measures adopted to treat the risks, (d) Lead time to get gain awareness of a risk and to respond to a risk event/ adverse incident and (e) Ability to track emergent risks on a consistent basis, based on industry and domain awareness and (f) Critical observation on societal impact contributed by the risks.

16. Maturity Scale

FH risk management adapts CMMI maturity scale for risk maturity to assess the maturity level of the compliance. They are:

- **Initial**: Processes are unpredictable, poorly controlled, and reactive towards dealing with risks. There is an ignorance of the need to proactively assess risks associated with Algorithms / ML / AI.
- **Managed:** Awareness of potential for risk, but failure to define risks in a local context and/or failure to consistently engage with an adequate diversity of viewpoints during risk identification. Action continues to be reactive and not proactive.
- **Defined:** Organization wide standards provide guidance across applications, programs and portfolios. E.g Identifying risks in the local context and engaging an adequate diversity of viewpoints in risk identification, however, there may be failures in consistently applying those insights to an assessment of AI development, AI procurement, AI-related changes, and/or use of AI outputs in decision making.
- **Quantitatively managed:** Organization is driven by quantitative performance driven objectives that are aligned to meet the needs of internal/ external stakeholders. Assessing risks associated with AI-related development, procurement, change, and/or use of AI in decision making, and consistently including adequately diverse insights for risk assessment, but failing to consistently manage and mitigate risks linked to identified gaps in awareness, skills, tools, processes, data sets, models, and/or governance.
- **Optimizing:** Managing identified risks using insights and inputs gained from diverse perspectives during risk identification and assessment. Ensuring risk treatment decisions involve the same diversity of inputs. Ensuring risk information feeds back into plans for future change. Ensuring there are appropriately senior owners for both risks and actions, with the necessary knowledge and influence to effect change. Ensuring risk management decisions are documented. Ensuring processes, benchmarks, stakeholders, and high risks under ongoing management and accepted risks are reviewed regularly (at least annually).

CMMI View	Risk Management	Diversity	Accountability
Initial	Ad hoc	Not considered	Undefined
Managed	Awareness of Risk	Requirement considered, but not defined	Ad hoc
Defined	Identifying Risks	Requirement	Assigned to SMEs

		defined for local context and applied to risk identification	
Measured	Assessing Risks	Diverse inputs to be applied to risk assessment	Demarcation refined via RACI, but not embedded
Optimized	Managing Risks and feeding back into process and solution evolution.	Diverse inputs applied to risk treatment and ongoing risk management	Demarcation refined and formally associated with roles

This process of incorporation, especially in the form of risk mitigation is crucial. Many AAA systems begin as High Risk endeavors and only with sufficient and necessary risk mitigations can the AAA systems proceed into production.

—----------