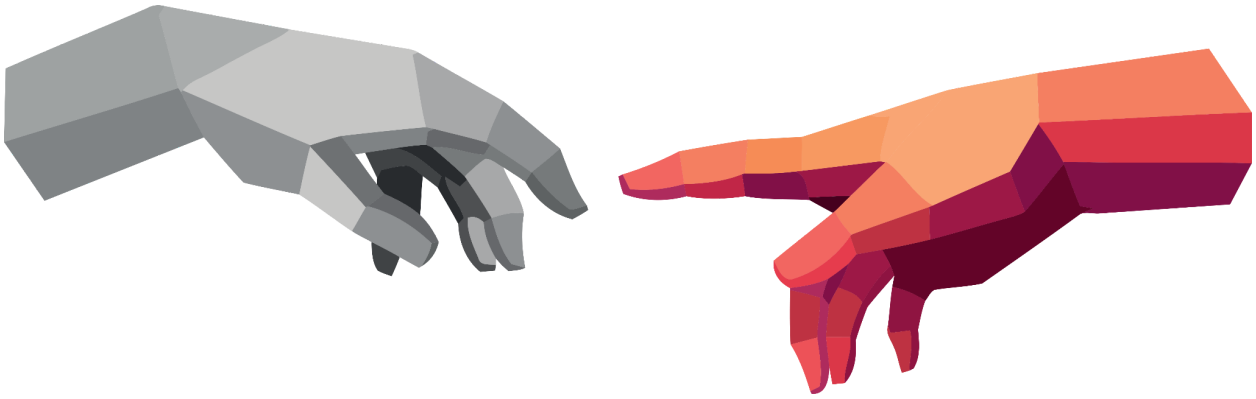


FORHUMANITY

ForHumanity[1]

a non-profit public charity

Government and Regulatory Services for the governance, accountability and oversight of Artificial Intelligence, Algorithmic and Autonomous Systems

(Pre-Production Draft)

[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 900+ 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.

Ryan Carrier
Executive Director
980 Broadway #506
Thornwood, NY 10594

Re: NYC AEDT Bias Audit law - Local Law 144 2021

Dear Chair and Department Members:

It is ForHumanity's pleasure to submit this letter and our Government and Regulatory services in regards to Local law 144 of 2021 related to Automated Employment Decision tools (AEDT). The protection afforded by the law to candidates of AEDT are aligned with ForHumanity's mission to *examine and analyze downside risk associated with the ubiquitous advance of AI, algorithmic and autonomous systems and where possible to engage in risk mitigation to maximize the benefits of these systems… ForHumanity.*

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 1000+ international contributors, from more than 70 countries 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.

In support of the NYC AEDT Bias Audit law, ForHumanity has regularly convened a team of volunteers (all humans are welcome in our transparent, crowdsourced process) to draft ForHumanity's NYC AEDT Bias Audit - a certification scheme aimed to satisfy Local Law 144's term -"bias audit" . It is our belief that a "bias audit" is not a widely understood and accepted term, whereby all auditors know all steps that are required to satisfy such an audit. In our conversations with auditors, AEDT providers, plaintiff-side attorneys and employers, great ambiguity remains on how audit satisfaction will be achieved. In light of this ambiguity most compliance will error on the side of minimum compliance. The ambiguity exists as a result of the law's language copied here, "*Such bias audit shall include but not be limited to the testing of an automated employment decision tool to assess the tool's disparate impact on persons of any component 1 category required to be reported by employers pursuant to subsection (c) of section 2000e-8 of title 42 of the United States code as specified in part 1602.7 of title 29 of the code of federal regulations".* The "*but not limited to*" clause rightly highlights that bias is not only about disparate impact. In fact, bias exists in many forms, such as statistical bias, cognitive bias and non-response bias. Further bias manifests in data, architectural inputs and outcomes from AI, Algorithmic and Autonomous

systems (AAA Systems), like AEDTs. ForHumanity agrees with the Council that we ought to maximize bias mitigation ("*but not limited to*") in AEDTs and our audit criteria already is designed to mitigate a wider array of bias.

The law also did not appear to fully embrace all Protected Categories (the subjects of bias), such as the Disabled. AAA Systems by their very design (seeking "best-fit" conclusions) are often exclusionary, especially in the areas of Disability and neuro-divergence. ForHumanity's audit criteria can help the Council include bias remediation in AEDTs for all New Yorkers

ForHumanity offers to assist the Council in overcoming these challenges with our expertise in drafting audit criteria and our focus on mitigating risk for AAA Systems for all humans. We offer this service for the Council's consideration as a means to establishing uniformity, certainty and an infrastructure of trust for "bias audits" of AEDTs. This offer is not unique for ForHumanity. We have provided the UK's Information Commissioner's Office with a similar submission of audit criteria for the General Data Protection Regulations (GDPR) and we have been retained by CEN/CENELEC JTC 21 as a technical liaison on the conformity assessment called for in the EU's Proposed AI Act. ForHumanity is conducting this work in numerous other jurisdictions globally as law-makers race to place guardrails around these largely ungoverned AAA Systems.

Financial audits have a series of critical elements of infrastructure, including checks and balances leading to successful governance, oversight and accountability. Those key elements are discussed in the attached document laying out a comprehensive framework establishing an infrastructure of trust and are summarized here:

1) Trained bias audit professionals - like CPAs
2) Independent third-party rules (Like Generally Accepted Accounting Principles - GAAP) - accepted and approved by the Council
3) A body to ensure independence, anti-collusion and uniformity of audits prevail.
4) A code of Ethics and Professional Conduct governing auditors and their actions

This set of criteria would dramatically enhance the impact and compliance with the law, providing a leveraged enforcement mechanism of trained auditors abiding by a set of rules the council has approved. ForHumanity provides the services, under the authority of the council for all four elements at no cost to the Council or New York City. As a non-profit, public charity, 501(c)(3) registered, the Council can be assured that our goals are aligned - protecting New Yorkers from bias in Automated Employment Decision Tools.

Thank you to the Council and the City of New York for the opportunity to testify on behalf of all New Yorker's who are the beneficiaries of ForHumanity's work and mission. We hope

you will consider our assistance and would welcome any opportunity to further share our work in support of Local Law 144 2021.

Kind regards,

Ryan Carrier

Executive Director, ForHumanity

Table of Contents

ForHumanity's Mission Statement

ForHumanity is a US 501(c)(3) tax-exempt public charity and our mission is to *examine and analyze downside risk associated with the ubiquitous advance of AI, algorithmic and autonomous systems and where possible to engage in risk mitigation to maximize the benefits of these systems… ForHumanity*

What is ForHumanity (FH)?

ForHumanity is:

1. Mission-driven, non-profit, public charity
2. Consisting of 900+ members from more than 70 countries
3. Only natural persons are permitted to join ForHumanity
4. We accept no corporate funding
5. All works performed have been conducted on an all-volunteer basis
6. All work-projects are executed transparently via crowdsourcing
7. All decisions and governance within ForHumanity are mission-aligned and executed by the Executive Director, or the Board of Directors
8. Majority of Board of Directors are elected from the community of ForHumanity Fellows and by the ForHumanity Fellows

FH Government and Regulatory Services

We offer to facilitate critical portions of the ecosystem under the authority and in cooperation with governments.

ForHumanity has developed a comprehensive ecosystem - an infrastructure of trust for Artificial Intelligence, Algorithmic and Autonomous Systems (AAA Systems) - modeled on the ecosystem of financial accounting and reporting called Independent Audit of AI Systems (IAAIS). All elements of the ecosystem adhere to common accepted practices by many governments and are meant to be accepted, adopted and integrated by government approval.

ForHumanity provides a unique toolkit for the benefit of legislators and regulators offering unprecedented secretariat services:

1. We draft audit criteria to used by third-party, independent auditors on for AAA Systems
2. We adapt law, guidelines, regulations, standards and best-practices in binary (compliant/non-compliant) criteria **submitted for approval by governments and regulators**
3. We educate and train individuals on the audit criteria - accrediting them upon examination as ForHumanity Certified Auditors (FHCAs)
4. We uphold a Code of Ethics and Professional Conduct for FHCAs
5. We maintain an open forum for crowd-sourced, transparent, all-inclusive input on the audit criteria - available to all natural persons without meaningful barriers-to-entry to volunteer contribution
6. We operate a licensing system for approved audit criteria that ensures:
    a. Independence (see below for further definition)
    b. Fair and level playing field
    c. Anti-Collusion principles
    d. Uniformity of certification practices and compliance
    e. Post Audit Compliance Reports
    f. Verification/Trust services (blockchain verifiable)
        i. Verified practitioners
        ii. Verified credentials
        iii. Verified compliance and certifications
7. We provide oversight for certifying bodies (in the absence of a national accreditation service) and for certified individuals, including reviews of past certifications for quality control
8. We advocate for mandatory third-party independent audits on all AAA Systems that impact humans that are not excluded on the basis of low-risk of negative impacts to natural persons
9. ForHumanity provides governments and regulators access to a set of harmonized global best practices tailored and specified to the laws and regulations of your jurisdiction.

Independent Audit of AI Systems

Independent Audit of AI Systems is an all-inclusive term to describe the entire ecosystem of governance, oversight, accountability and trust for AAA Systems. It is highlighted by the following characteristics:

1. IAAIS is applied across the entire lifecycle of the AAA systems including design, development, deployment and decommissioning
2. IAAIS captures and mitigates risk to natural persons across five pillars (ethics, bias, privacy, trust and cybersecurity)

3. IAAIS is designed to identify and mitigate the unique and specialized risks occurring from the very nature of socio-technical systems (e.g. Data Entry Point Attacks[2] and embedded instances of Ethical Choice)
4. IAAIS integrates with a comprehensive risk management framework that operates with four lines of defense for risk mitigation:
    a. Designers, Developers, Product managers and Data Scientists
    b. Managers, Overseers, Human-in-Command, Committees
    c. Internal Audit, Risk Reviews
    d. External, Independent Auditors
5. IAAIS mitigates risk to a wider collection of stakeholders beyond ISO 31000's stakeholder list inclusive of not only organizational risk, but also risks to natural persons, communities and the environment
6. IAAIS establishes binary (compliant/non-compliant rules) and specific documentary evidence required for sufficient proof of compliance

Assuring Trust - IAAIS

Trust is assured when three characteristics come together:

1) Independent, third-party, objective and widely accepted criteria are universally applied
2) Assurance is executed by accredited, well-trained, independent experts with verifiable credentials
3) Independent, accredited certification bodies, acting on behalf of society and not on behalf of the auditee, assure compliance with government approved rules

Independent, external auditors provide a service to the public and the constituency of the government by assuring compliance with the rules and regulations put forward in audit criteria. Certification is a strong signal, annually verified, that compliance with the law is being maintained. The process, like with financial reporting and audit, generates a compliance-by-design approach, whereby best practices are built into the beginning of AAA system development.

Independent, external auditors provide objective certification founded upon their Code of Ethics and Professional Conduct. Additionally, their obligation to receive no other remuneration from auditees, coupled with their risk of false assurance of compliance solidifies their objectivity. The marketplace is assured that an infrastructure of trust is present and may be relied upon by organizations and individuals that need assurance.

No transparent system is foolproof. When the rules are transparent, then companies or persons seeking to commit fraud and malfeasance may succeed. However, a robust

[2] Capitalized terms reflect a ForHumanity defined term in our audit criteria and can be found here https://forhumanity.center/definitions/

ecosystem of transparency and disclosure will eventually identify their misdeeds. Even some of the greatest financial frauds (e.g. Enron and WorldCom) were eventually caught by the system itself as a result of transparency and disclosure. In most cases, IAAIS provides a reliable infrastructure of trust for the willfully compliant.

Independence

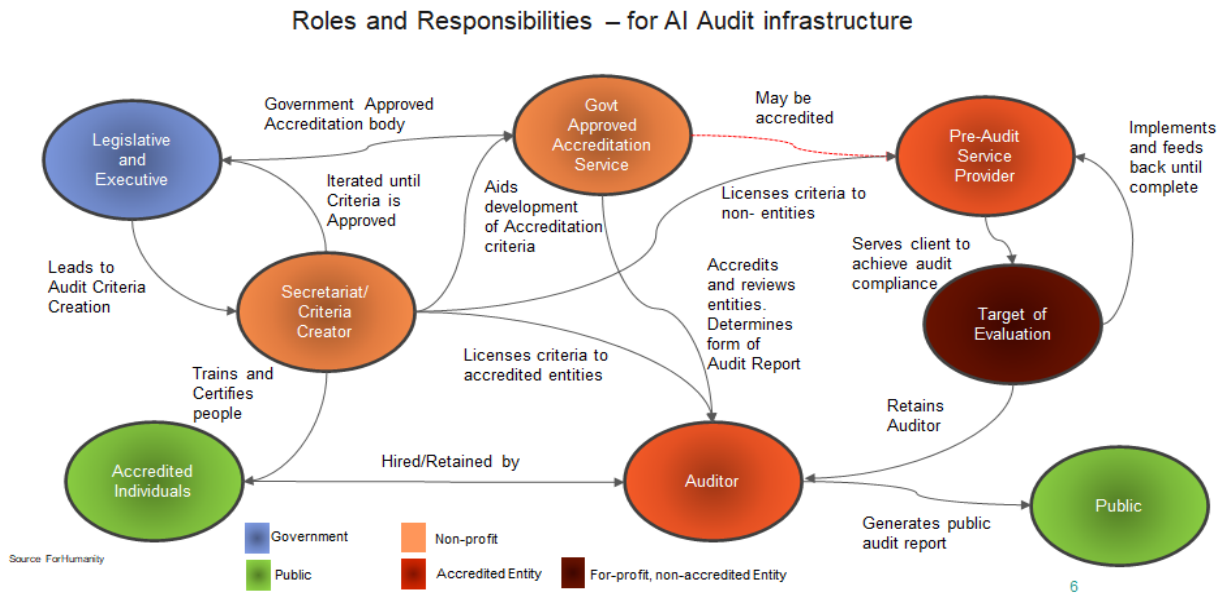
A legal term defined by The Sarbanes-Oxley Act of 2001 that requires a certifying body (an Auditor) to receive no other remuneration from an Auditee beyond reasonable audit fees. In its license agreements, ForHumanity further stipulates that a licensee holder cannot be an Auditor and an Assessor/Consultant (or provide any other form of service) to the same Auditee within a 12-month period. ForHumanity has adopted this rule and determination into the licensing agreements for certification bodies and abided by FHCAs under their Code of Ethics and Professional Conduct.

Independence and independent audit increases compliance with established laws and regulations. Time and again, human nature has proven that self-assessment is useful but insufficient, thus requiring the need for further enforcement mechanisms. However, government and regulatory enforcement requires resources to examine societal compliance. Enforcement bodies can mandate uniform criteria that satisfies compliance (e.g. the Securities and Exchange Commission mandating adherence to Generally Accepted Accounting Principles GAAP for publicly traded companies in 1975). Then, Independent Audit when mandated by governmental enforcement agencies creates a leveraged, overarching compliance mechanism - examining and assuring compliance - accomplished by third-party trained practitioners, accredited robustly (and equally overseen - “watching the watchers”), using uniform rules, regularly assure compliance, at their own risk of false assurance of compliance. Under this ecosystem, conflicts are mitigated, objectivity is maximized, and trust is built.

More details on specific examples of Independence can be found in **ForHumanity’s Certified Auditor Code of Ethics and Professional Conduct v1.0.**

Ecosystem Explained - Roles and Responsibilities

Discussed in detail in ForHumanity’s Infrastructure of Trust for AI - Guide to Entity Roles and Responsibilities, the image below depicts the interactions and segregation of duties in Independent Audit of AI Systems. This ecosystem mirrors that of financial audit and reporting. The graphic includes the role of government, government-approved accreditation bodies (if appropriate), accredited entities, auditors, pre-audit service providers, auditees and the role of individuals and the public in general.



This ecosystem takes a government enforcement role and transfers compliance oversight to the marketplace. Compliance oversight is conducted by accredited entities employing accredited individuals trained in audit compliance and certification. Both ForHumanity and the Accreditation Body "watch the watchers" and have authority to revoke accreditation or licensing rights for insufficient governance, accountability, and oversight. More details are available in the guide.

Taxonomy: AI Audit, Assurance and Assessment

Described and delineated in ForHumanity's Guidance: Taxonomy: AI Audit, Assurance & Assessment. ForHumanity describes differentiated roles amongst third-party services providers. This is critical, because it is the unique characteristics of the auditor that establishes an infrastructure of trust by providing services that are truly independent and conflict-free to provide the public with confidence that compliance has been assured.

	Internal Audit	3rd-Party, Independent Audit	Assurance	Assessment	Consulting
Certified Practitioners Required?	No, Employees	Yes	Yes	No	No
Objective/Subjective	Objective	Objective	Objective	Subjective	Subjective
Independent	Yes	Yes	Yes	No	No
Known 3rd Party transparent Binary, Rules or Laws	Yes	Yes	No	No	No
Known 3rd Party transparent non-binary, Standards, Frameworks or Guidelines	No	No	Yes	No	No
Service provided for?	Management	Users, Society	Users, Society	Contracting Party	Contracting Party
Feedback Loop with the Company, iterative problem solving, teaching, tailoring	No	No	No	Yes	Yes
Consequences for False Compliance assertions	job loss	Liability	Liability	No liability	No liability
Written Report produced for the Public	No	Yes	Yes	No	No

Sources: ForHumanity, COSO, IAASB, Sarbanes-Oxley, IFAC

ForHumanity's Perspective

ForHumanity drafts audit criteria from a specific perspective - what is best for humans. This human-centric, mission-driven focus is intentional to counterbalance the corporate-permissive focus prevailing across most western societies. AI, algorithmic and autonomous systems are socio-technical, meaning the human is "in" the system through the use of Personal Data and is the target of the outcomes of the system as well. This integration necessitates a 360-degree perspective of risk, beyond the corporate and including people, communities and the environment.

When drafting audit criteria, the guidance is simple - "*does this criteria mitigate risk to humans from the AAA System*?" ForHumanity asserts that sustainable profitability for corporations will occur when risks to humans are mitigated.

Audit Criteria - Government Submissions

All humans may contribute to the ForHumanity audit drafting process. It is transparent to all who agree to abide by the community's Code of Conduct to assure decorum. Upon entry into the community, all contributors have full and complete access and transparency to ForHumanity work projects. ForHumanity celebrates Diverse Inputs and Multi Stakeholder feedback in order to maximize risk assessment from a 360-degree perspective of impact. The same holds true for the creation of audit criteria. All may view, all may comment. The filter and adjudication on comments and their inclusion in final audit criteria drafts is simple - "*does a new word, definition or audit criteria mitigate risk to humans*.?" If so, it finds its way into our work.

Unless a certification scheme is deployed by organizations voluntarily, the government always has the final authority as to what audit criteria is approved.

All government approved audit criteria become available publicly under Creative Commons BY-ND-NC[3]. Licenses are offered to all qualified certification bodies and fees are due to ForHumanity upon receipt of revenue. As a non-profit, public charity, ForHumanity is restricted on revenue generation and all revenues must be put towards the operating budget and the mission. Qualified certification bodies are those entities employing FHCAs on the licensed criteria they deploy by contract with clients. The criteria are available for all non-revenue generation applications, such as research and academic study, freely.

Determination of Audit Criteria Development

As jurisdictions enact laws governing AAA Systems, ForHumanity will maintain pace with the law and have audit criteria drafted for and submitted for approval to governments. Additionally, ForHumanity maintains an active engagement process with the market to determine demand for new certification schemes for voluntary adoption or as guidance for policymakers and judicial settlements. ForHumanity intends to produce audit criteria for every AI, algorithmic and autonomous system that impacts a human.

Global Harmonization

In the interest of humanity, ForHumanity drafts audit criteria in accordance with local jurisdiction laws. However, having contributors from more than 70 countries around the world provides diverse inputs and broad perspectives - a lens that no single country can emulate. This provides ForHumanity with a unique perspective on global best practices, ones we offer to each country to uncover solutions and risk mitigations that might not be apparent or readily available to a country. Furthermore, since many organizations are international, a set of criteria with maximum harmonization will minimize the cost of compliance resulting from compliance requirements in multiple jurisdictions.

Relevant Legal Frameworks

ForHumanity tailors each audit criteria to the Relevant Legal Frameworks applicable to the jurisdiction. Auditors shall refer to these local laws for determination of Protected Category Variables, human rights and freedoms afforded to Data Subjects or natural persons especially in the areas of equality and anti-discrimination law, access to goods and services, children's laws, sector, platform- or service-specific law.

[3] https://creativecommons.org/licenses/by-nc-nd/4.0/

Jurisdictional Sensitivity

Under Independent Audit of AI Systems, nation-states retain their authority;in fact have their enforcement capabilities enhanced. Audit criteria are jurisdictionally sensitive, drawing upon local law and regulations to specify such details as, for example, Protected Categories. By focusing on local regulations, the audit avoids "legislating" compliance but instead leaves these governance questions in the hands of elected officials. Furthermore, the audit will occasionally fall back on the legal concept of "reasonable" relying on either past jurisprudence or current examples of possible solutions without being prescriptive.

Under IAAIS, proactive compliance can be achieved through the certification process - an evidentiary based proof-statement, independently verified by an objective, third-party auditor working for the public good.

An example of this jurisdictional sensitivity can be found in Protected Category Variables. Bias, in itself, is a statistical term describing a characteristic of a data set. However, when society then dictates that certain activities shall not be biased in their execution, it becomes something we need to account for in our systems. In the case of Protected Category Variables, each jurisdiction may be different. In Scotland, for example, socioeconomic status is a Protected Characteristic, but that is not true of law in the United States.

Each jurisdiction's laws will be considered in the adaptation of the audit rules.

Criteria Mapping

ForHumanity draft criteria takes time to conduct "map" a specific service related to audit criteria that delineates the difference between two sets of audit criteria at a micro-level:

1. Definition to Definition
2. Legal Term to Legal Term (e.g. meaning of Consent)
3. Authority/Regulator to Authority/Regulator
4. Gaps from one jurisdiction to another

ForHumanity publishes official mapping documents to enable gap analysis services using official ForHumanity criteria.

Ethics-by-Design

The nature of socio-technical systems is to embed the human in the system through the use of Personal Data, while producing outcomes that have impacts on the human. The result of this interaction creates systems with a specific shared moral framework: that of the organization and/or the designers and developers. The ethics of the systems are meaningful to the human and will have significant impact on outcomes.

This interaction of corporate ethics and new/changing law governing AAA Systems often requires ethical/soft law considerations and user's ethics. This intersection requires trained experts to adjudicate the myriad instances of ethical choices embedded in AAA Systems such as:

1. Necessity
2. Proportionality
3. Adjudication of soft law
4. Statistical benchmarks for representativeness
5. KPI design for concept drift
6. Interface design choices
7. Explainability

Responsibility for managing the process ranging from the creation of a public Code of Ethics to implementing controls around instances of Ethical Choice is a standing, trained and empowered Ethics Committee, presided over by experts in algorithm ethics and applied ethics.

However, "ethics washing" and superficial applications of ethics remain a risk without governance, oversight and accountability on instances of ethical choice within organizations. IAAIS criteria require the Ethics Committee to examine and consider all instances of Ethical Choice to be documented and attested by an independent, external auditor to ensure objectivity, oversight and accountability over the design, development, deployment and potential decommissioning of AAA Systems. The implementation of ethics-by-design is discussed in ForHumanity's paper on the Rise of the Ethics Committee.

Privacy-by-Design

Of ForHumanity's five pillars for AAA Systems (ethics, privacy, bias, trust and cybersecurity), privacy is the most well developed because of legal efforts on data privacy and protection, notably advanced by the General Data Protection Regulation (GDPR). ForHumanity has developed a comprehensive set of certification criteria designed to be used by certification bodies to assure compliance with the GDPR (both EU and UK versions).

The criteria provide assurance around:

1. General Governance, Accountability and Oversight
2. Necessity
3. Proportionality
4. Lawful Basis (including Consent)
5. Specified Scope/Nature/Context/Purpose
6. Data Minimization
7. Data Protection
8. Technical and Organizational Controls
9. Security
10. Cybersecurity
11. Data Subject Rights and Freedoms
12. Fairness
13. Transparency and Notice
14. Automated Decision Making (Profiling)
15. Explainability
16. Governance of Data Transfers

Certification of privacy law and regulations provides proactive compliance and fosters compliance-by-design thinking to enable economies of scale and general efficiency that can be leveraged into all AAA Systems.

Bias Mitigation

AAA Systems using Personal Data examine historical data to make inferences about people. Eradicating bias is a statistical impossibility and thus ForHumanity's goal is to maximize bias mitigations. Also, fairness and equity are enforced in different ways through law around the world.

We have identified three stages in AAA Systems where bias can be examined and mitigation rendered:

1) Bias in data
2) Bias in architectural inputs
3) Bias in outcomes

Bias in Data has many potential manifestations as ForHumanity explored in Bias Mitigation in Datasets:

1. Source data
2. Cleaning
3. Labeling

4. Anomaly and outlier treatment
5. Training and testing/validations splits
6. Representativeness
7. Cognitive bias
8. Non-Response Bias

ForHumanity requires documentary evidence of bias mitigations in audit criteria to tackle each of these forms of bias and fight against discrimination results from the data. Many of these mitigations must be documented in a Data Transparency Document and made public allowing for a higher form of discourse, regarding the appropriate mitigations in data sets.

Data however is not the only source of bias, there are two more elements of a AAA System to examine for bias:

1) Architectural Inputs to models
2) Model Outcomes

As designers and developers construct their models, the choices they make and the methods they choose may result in bias. ForHumanity has drafted numerous audit criteria to examine and consider all appropriate mitigations to ensure fair and equitable treatment for people in decisions that designers and developers make across model architecture.

AAA Systems take on many forms with varied degrees of understanding on exactly how conclusions are reached. Some large language models will even have billions of parameters making it virtually impossible to replicate the decision process. For this reason, IAAIS examines outcomes against accepted and tested standards, like the American 4/5th rule[4]. To be certain, there is no magic to the arbitrarily-chosen 80%but at least there is a legislatively-accepted threshold that demands further analysis and justification. As these outcomes manifest as Residual Risk, they are disclosed to the natural persons prior to engagement with the AAA System. Under this disclosed, risk-based system, model designers and developers will encounter market-based feedback on the risk/reward profile of their model.

Trust

Trust is ForHumanity's catch-all category, focused on the many ways in which trust is earned, demonstrated, proved and sometimes forfeit. Below, we describe a series of sub-categories that all are captured under Trust.

[4] https://www.ecfr.gov/current/title-29/subtitle-B/chapter-XIV/part-1607

Transparency/Disclosure

Often feared by corporations, transparency and public documentation (disclosure) are necessary elements of any trustworthy system. However, Intellectual property (IP) and trade secrets are protected under the infrastructure of trust that Independent Audit of AI Systems creates. IP review can be governed by Non-Disclosure Agreement with an Auditor in the rare occasion that it would be necessary for compliance. For example, the IAAIS approach to bias mitigation examines data, architectural inputs and outcomes of AAA Systems; there are no source code audits. The theory of this practice is simple - companies with problems in their models (e.g. bias, discrimination, unethical choices), are accountable for the inputs to the model as well as the outcomes. IP does not need to be examined or disclosed because the auditee is accountable for their outcomes, regardless of the root problem. In most instances, IAAIS is the best compromise between transparency and assurance of compliance for the public.

Typical transparency/disclosure under IAAIS are not intellectual property and trade secrets, instead are proof statements to the public of critical information to allow for users of the system to make informed decisions about their interactions with the AAA Systems. Transparency and disclosure notifications will describe decisions about data around representativeness or confirmation of scope/nature/context/purpose of the AAA Systems including the description and usage of the Personal Data being collected under Consent.

When transparency/disclosure is deployed properly, the intent is to ensure that users are well informed about the scope/nature/context/purpose of the AAA System and the risks that are present for the natural purpose during the interaction with the system. This is a defining characteristic of trust: two parties agreeing to an interaction, knowing the responsibilities and expectations for each party. Transparency/disclosure mitigates an enormous amount of risk for both parties in the interaction.

Finally, transparency/disclosure is one of the last lines of defense. Even in record-setting frauds, like Enron and WorldCom, it was transparency and disclosure that finally allowed the system to catch on to the misdeeds. Under compliance-by-design infrastructure, transparency/disclosure requirements often become systemic productions. Therefore, when they are absent or malformed, their absence or anomalous compliance can become the bread crumbs leading to uncovering non-compliance. Transparency and disclosure are the cornerstones of compliance as they represent absolute accountability.

Accessibility

AAA Systems cannot be considered trustworthy if they cannot be accessed broadly. Systems that exclude groups of persons unfairly because of insufficient accessibility harm two kinds of users, those who do not have access and those who value AAA Systems that respect human dignity.

Accessibility of AAA Systems is rarely a problem of ability, but instead a problem of attention, knowledge and resource allocation. Many countries have equality and anti-discrimination laws requiring accessibility. Such laws designed to protect people in vulnerable situations proved necessary in order to provide added incentives (through legal enforcement) for organizations to ensure their services are available to all persons. AAA System accessibility is rooted in human dignity and organizations should either provide accessibility or the meaningful accommodations to meet the needs of all people.

Control/Safety

Especially in the realm of AAA Systems where physical and psychological harms are possibilities, control and safety are mission critical. For example, ensuring that a AAA System remains true to the intended scope/nature/context/purpose without concept drift is a requirement under GDPR and most informed consent lawful basis. Moreover, it would be unethical to operate a AAA system or any machine designed to benefit humanity without assurance that the system can be turned off and will remain off until the human providers of the system intentionally return the system to service. Such characteristics demonstrate control.

In regards to safety, ForHumanity advocates for a risk-based approach, similar to the US National Transportation Safety Board and other international standards, where systems are rigorously stress-tested in a broad range of conditions that challenge the system in all foreseeable environments ensuring sufficient reliability, robustness and resilience to avoid system failure, resulting in harm to humans.

Control and safety require a robust and comprehensive risk management process centered in a culture that embraces risk management across all three lines of defense (as defined by most traditional risk management frameworks) and includes Diverse Inputs and Multistakeholder Feedback to maximize risk treatment across the entire spectrum of risk inputs.

Explainability

Around the world and across numerous industries, especially when organizations wield power over natural persons, the law requires that corporate decisions are accompanied with explanations. These laws cover decisions rendered from AAA Systems and call for the outcomes to be explained in clear and plain language, however compliance with this rule of law is often a minimalist approach.

The theory behind such laws is based on human dignity. No one who receives a favorable decision is often concerned with an explanation, so this is clearly centered around persons

receiving a negative result. ForHumanity's work ensures that automated decision-making explainability is accomplished commensurate with relevant legal framework requirements. However, ForHumanity recommends that AAA Systems go one step further to *Explainability+*.

The theory of *Explainability+* is simple. Being informed as to "why" an automated decision making system has produced a result, especially when that explanation is perfunctory does not help or empower the person to remedy their situation. *Explainability+* recommends the provider of the AAA System to go one extra step in support of the human and their humanity.

Explainability+ provides the natural person with the education required to achieve a favorable result from the AAA System: steps they might take to improve their situation and thus in a second iteration receive a more favorable outcome. Or, if a favorable outcome proves too challenging, other remediation services from within or outside of the organization designed to help the natural person achieve their desired goals. ForHumanity believes this will lead to great sustainable revenue for the organization, engender positive relations between the parties and celebrate human dignity with care for resolution and opportunity for satisfaction of the person's original goal. Interactions under ForHumanity's *Explainability+* create trust.

Inclusion

Most AAA Systems seek an average - a fitness across the data set - trained to explain all of the data. However, as a result of the very nature of most algorithmic models, fitness and inclusion are frequently at odds. Data scientists seeking greater accuracy often eliminate outliers and anomalies that by their very nature reduce model fitness. In socio-technical systems using Personal data, anomalies and outliers equate to people. The testing and evaluation of the algorithmic modeling process necessarily struggle with the balance between model accuracy and outlier inclusion. This tension can result in discrimination against Protected Categories and inclusion failures.

ForHumanity advocates for "edge-in thinking", a design concept that works to include edge cases, anomalies and outliers from the outset of the design and development process. From this starting point, modeling can consider appropriate and equitable accommodations for persons who would not otherwise be included in the process.

Finally, Diverse Inputs and Multi Stakeholder Feedback maximizes human inclusion in the risk input, analysis, evaluation phases of risk management. Please see the Diverse Input & Multi stakeholder Feedback section below for more details the advancement of inclusivity built into Independent Audit of AI Systems.

Cybersecurity

Cybersecurity is a fairly mature industry by comparison to artificial intelligence, algorithmic and autonomous systems, however, the socio-technical nature of AAA Systems creates new and unique vectors for cyber attacks. IAAIS incorporates existing gold-standard cybersecurity frameworks with tailored controls designed to address the new and unique attack vectors. Notably, data entry point attacks (e.g. data poisoning, model inversion and membership inference attacks) present innovative challenges that the marketplace is still grappling with. ForHumanity's audit criteria for AAA System cybersecurity is built upon the US NIST framework. Organizational solutions to AAA System cybersecurity should be tailored to specific systems and never shared publicly.

ForHumanity's audit criteria represent a foundational governance, accountability and oversight framework for cybersecurity and helps to guide the minimum infrastructure to operate a successful cybersecurity system. However, the downside of transparency is that the rules, processes and criteria are available to bad actors as well, educating them on the methods and procedures that might lead them to find a weakness. Therefore ForHumanity strongly suggests, on behalf of the humans impacted by cyber breaches and data entry point attacks, that entities go above and beyond the foundational ForHumanity cybersecurity criteria. In regards to the specifics of a robust cybersecurity system, opacity is in the best interest of humanity.

Diverse Inputs and Multi Stakeholder Feedback

Diverse Inputs and Multi Stakeholder Feedback (DI&MSF) describes the 360-degree perspective of risk beyond ISO 31000 stakeholders (listed below):

- executive-level stakeholders
- appointment holders in the enterprise risk management group
- risk analysts and management officers
- line managers and project managers
- compliance and internal auditors
- independent practitioners

ForHumanity adds to the list of risk assessors:

- external domain experts
- natural persons (users and impacted)
- communities
- employee organizations (e.g. Unions)
- environmental representatives

Additionally, Independent Audit of AI Systems ensures diversity in the risk assessors by relying upon the Ethics Committee to determine the definition of diversity for the

organization. ForHumanity recommends such a definition includes diversity of thought and lived experience in addition to the inclusion of Protected Categories and intersections thereof. DI&MSF risk assessors are trained in assessing risk in AAA Systems and provide valuable diversity in the risk input, analysis and evaluation process. For more detailed information, see ForHumanity's paper on Diverse Inputs and Multi Stakeholder Feedback as well as the guidance in the Risk management Framework below.

Special Committee Structure

AAA Systems exist in a myriad of forms impacting numerous specific groups of people (e.g children, persons with disabilities or persons in vulnerable situations). ForHumanity recognizes the special needs of each of these groups and requires teams of trained experts to provide governance, accountability and oversight over such systems.

Each special committee expert requires deep and specific knowledge.For example, a member of the Children's Data Oversight Committee should have expertise in understanding and interpreting the UN's Rights of the Child declarations including the ability to evaluate and adjudicate impacts and outcomes that affect their health, well-being, safety, avoidance of sexual abuse or exploitation, avoidance of economic exploitation, rights to privacy and exercising their own data privacy, supportive structures for their family relationships, elements that support their physical, psychological and emotional development, support of their right to develop their own identity, views and have their perspectives heard. These expertise are deep and specific and are necessary to ensure the Child is well represented in the design, development and deployment of AAA Systems.

Biometric Data

Biometric data is Personal Data and sometimes Special Category Data or Sensitive Personal Data, however, ForHumanity argues that it is more dangerous and sensitive than most types of data relating to humans:

1. **The immutable nature of many Biometric Data items** makes them more intrinsic to people's being than most other data, and so results in enormous risk to people from breach, theft, misuse and misappropriation
2. **The richness of information extractable from biometric data** makes drifts in scope, nature, and purpose extremely common and potentially widespread
3. **The conspicuous nature of many biometrically-scannable identifiers** (say, our faces or gaits) makes such data potentially very public, and has already led to a significant loss of privacy whenever people are in public places, either physically or virtually

There are many uses of Biometric Data already in commercial implementation. While this is usually considered benign, there is an extreme asymmetry with respect to the benefits of the use of such data given the risks outlined above. Such risks are rarely characterized with transparency and disclosure. If the risk/reward profile was more widely recognised, it is our belief that uses of Biometric Systems would be held to higher standards, and fewer usages of frivolous Biometric Systems would occur. Further, if the law required sufficient levels of governance, protection and oversight be applied to uses of Biometric Data, then the "cheap and easy" solution would begin to lose its luster.

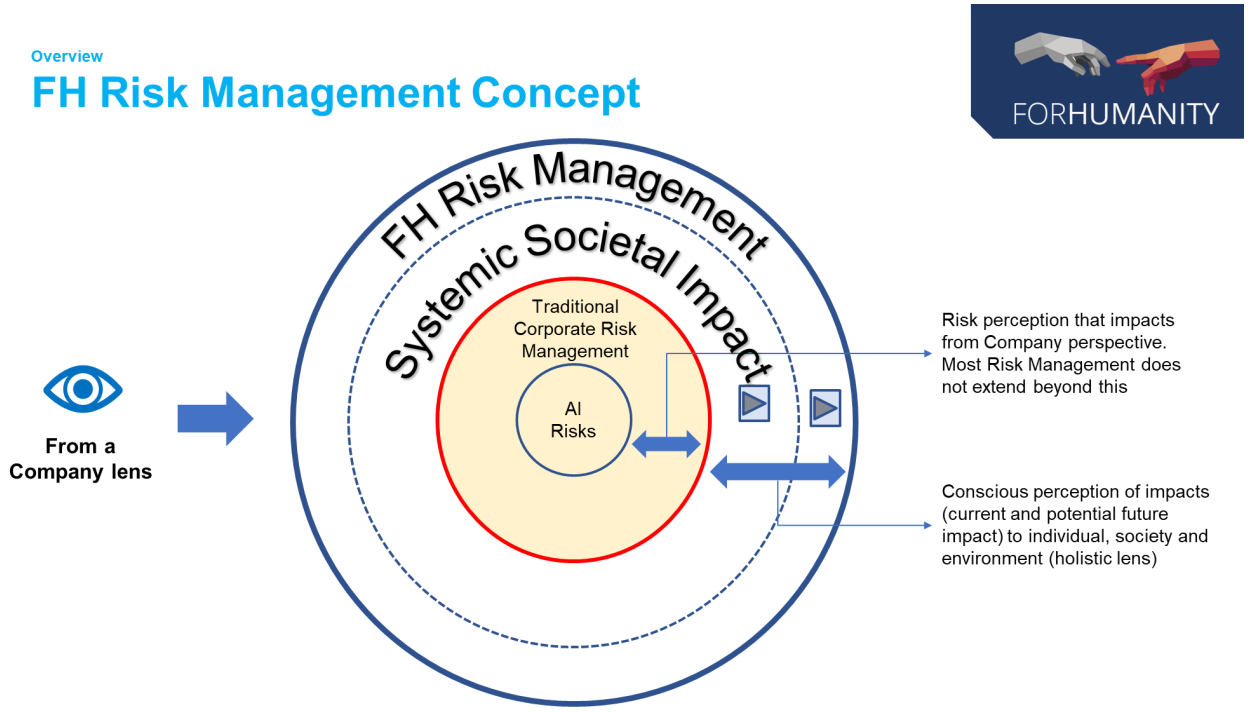
Current adoption of Biometric Systems is built on a flawed risk/reward profile. The reward is skewed to the system operator and the risk is almost exclusively held by individuals subject to these systems. As of now, many biometric systems are being tailored to maximize profitability or efficiency, without enough regard to the possible - and often probable - harms that humans may incur, including often unproven and potentially unprovable conclusions reached by these systems about us.

ForHumanity has deep concerns regarding the accuracy, validity, and assumed causality of many Biometric Systems. Many of the inferences (e.g race, mood, personality characteristics, mental state) are very hard to test against a ground truth, as they are often "fuzzy" even when being defined by humans and are frequently based on scant scientific evidence. Furthermore, most of these systems are poorly adapted to edge cases, like the disabled, neuro-divergent or other at-risk Protected Categories and intersections thereof. In consequence, ForHumanity requires a risk-based approach to accepting Accuracy, Validity, Reliability, Robustness, Resilience (AVR3). Biometric Systems call for especially high transparency, high disclosure, and an assumption of high risk. For further insight on our treatment of Biometric Data, please read our published work.

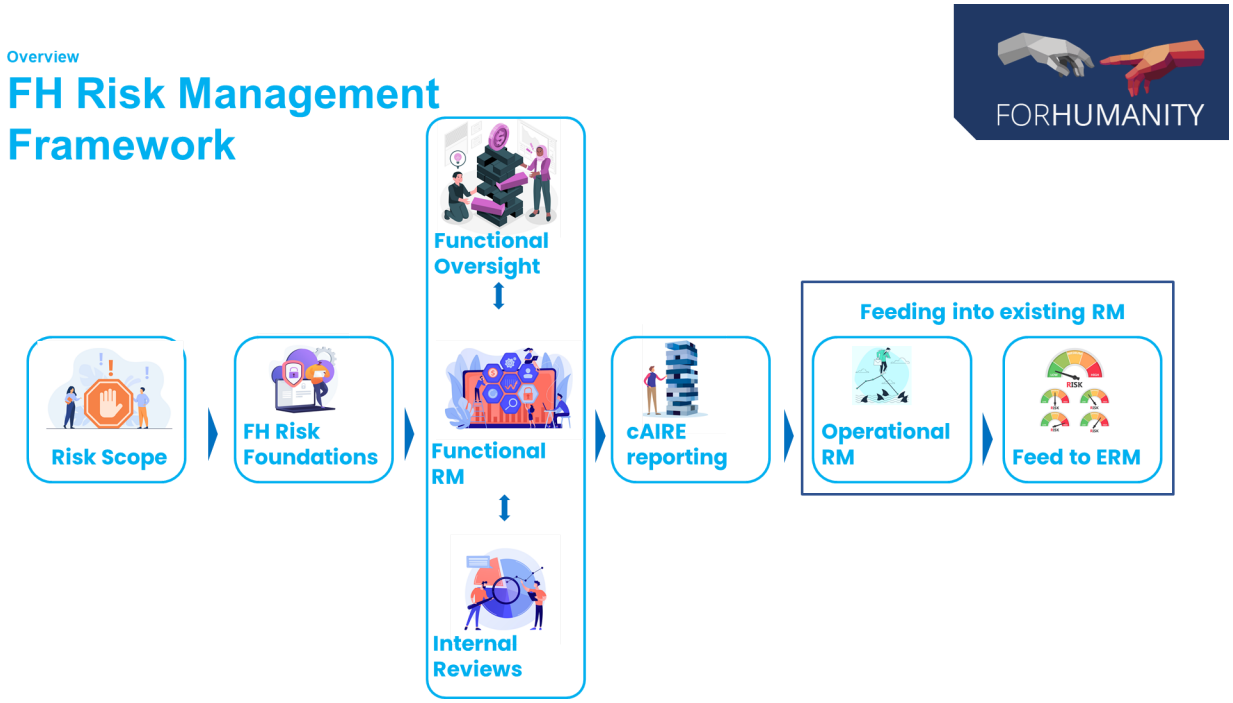
FH AI Risk Management Framework

ForHumanity's Risk Management Framework is designed for AAA Systems specifically and for integration into ISO 31000 and COSO ORM and ERM applications. The information below can also be found here.

Concept

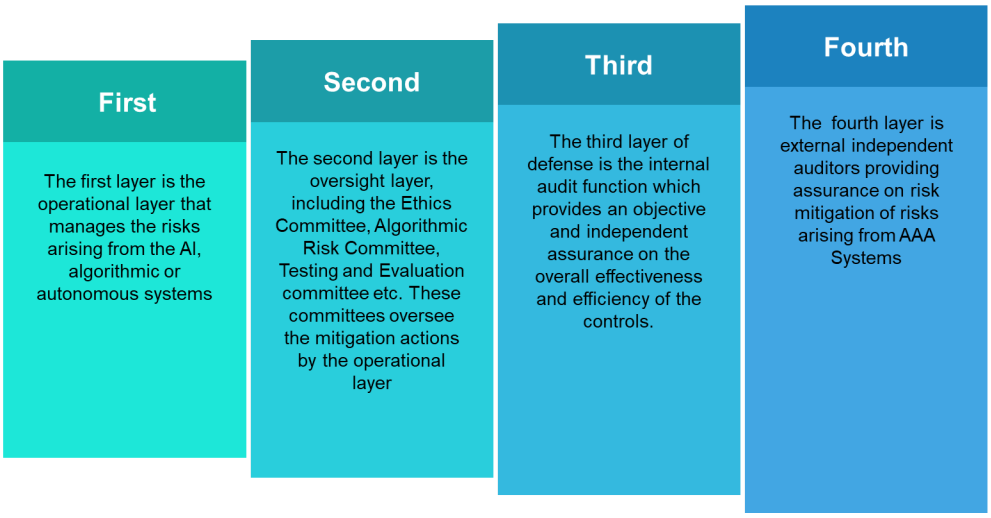


Overall Framework



Overview

Four Layers of Defense



ForHumanity Risk Foundations

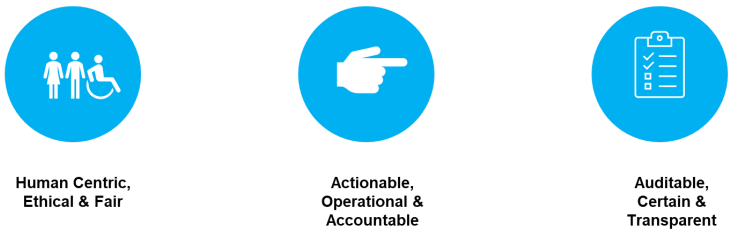
ForHumanity foundational reading on risk management (Principles)

ForHumanity's mission is to mitigate downside risks posed by AI, algorithmic and autonomous systems. One of the clear ways to mitigate risk is to implement and operationalize a robust & agile Risk Management framework.

ForHumanity's approach to risk management is centered on Ethics, Bias Privacy, Trust and Cybersecurity. Considered from a 360-degree multi stakeholder perspective, these pillars encapsulate the range of negative impacts and risk from socio-technical systems. ForHumanity wraps those pillars with a risk management framework that ensures compliance, mitigation and operability including characteristics such as: ethical, human-centric, accountable, governable, overseeable, transparent, documentable, proveable, evidence-based, and independently auditable.

Foundational Principles

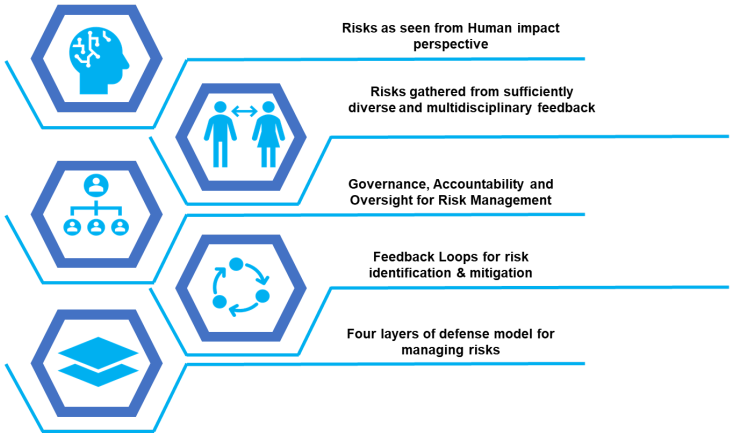
FH Risk Perspective



From a ForHumanity context, Risk management is an essential component to not just enable compliance with the criteria, but also sustainably prevent, detect and respond to emergent risks. ForHumanity advocates for a risk management framework that is omni-directional and multivariate. Multivariate in that the framework considers corporate risk (which damages employees and shareholders), risk to humans (which damages users/clients/prospects and unwitting participants), societal risk (which damages our systems, groups, communities, markets and collectives) and environmental risks (which damages nature and sustainability considerations). As risk is never wholly removed, residual risk will always remain. These residual risks, well disclosed and considered, will empower an increased ability to identify emerging risks, support concentrated research on novel mitigations and encourage informed acceptance of consequences when residual risk manifests itself.

Foundational Principles

FH Baseline Risk Principles



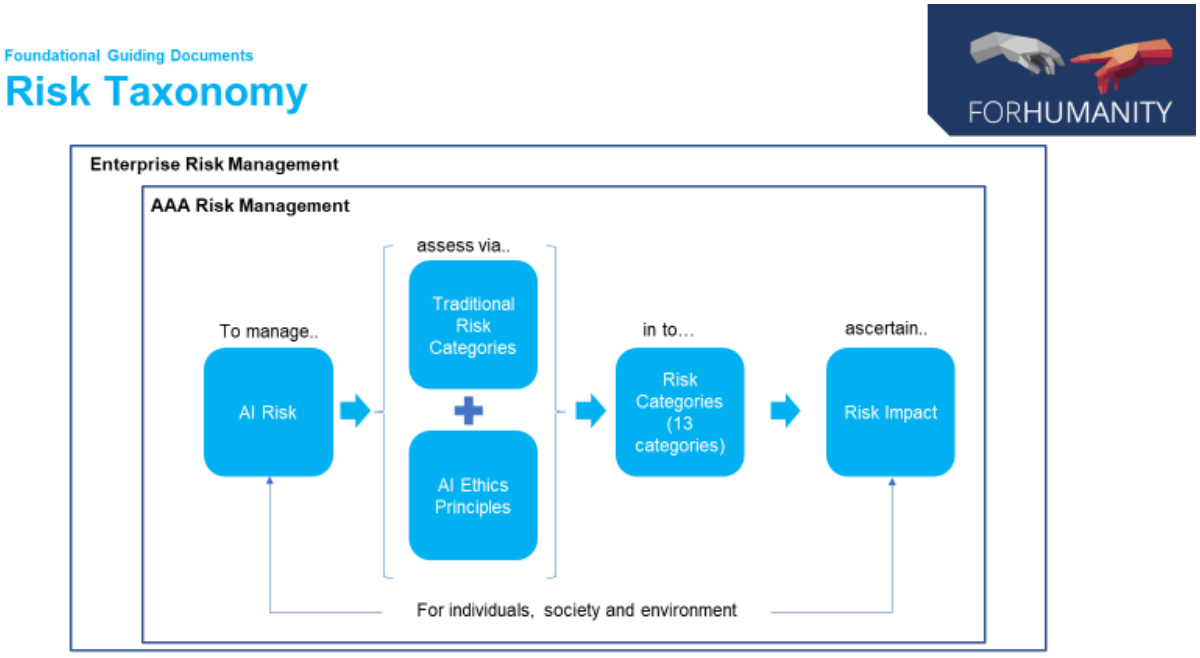
Read the complete brief here: FH foundational reading on Risk Management

Also read about Low risk AAA system identification process here
Identifying Low Risk AI, Algorithmic and Autonomous Systems

FH Risk Guiding Documents

Risk Taxonomy

Risk Taxonomy v2.docx



Risk Management Policy

Risk Management Policy - Guidance

Maximizing Risk Mitigation for Humans

Maximizing risk mitigation for humans

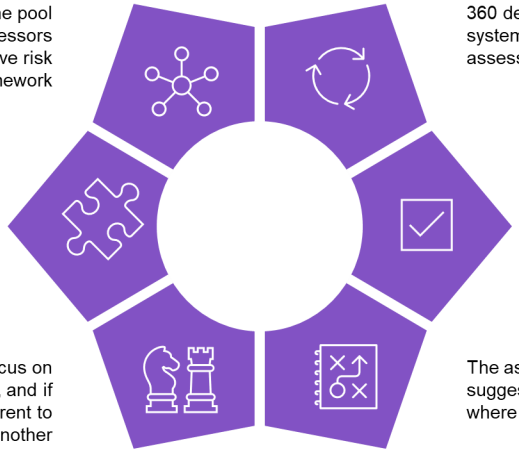
Diverse Inputs & Multi-stakeholder feedback

Diverse Inputs and Multi-Stakeholder Feedback & associated guidance

DIMSF - Guideline and template

Operationalizing Risk Management

Diverse input & Multi-stakeholder feedback



DI & MSF describes the pool of risk assessors comprehensive risk management framework

360 degree perspective of the system and risks being assessed

Enabling to gather diverse thinking and lived experiences

The Ethics Committee determines if/when the pool of DIMSF assessors are sufficient

The assessors focus on impacts of the risks, and if those impacts are different to one stakeholder over another

The assessor offer their suggestions for risk treatment where relevant

Need for committees

Need for committees

Functional Risk Management

Committees contribute to

Multi-disciplinary with varied skills and experience from relevant domains

Risks & impact- ability to assess risks and evaluate their impacts.

Establish and fairly value **Risk Tolerance and Risk Appetite**

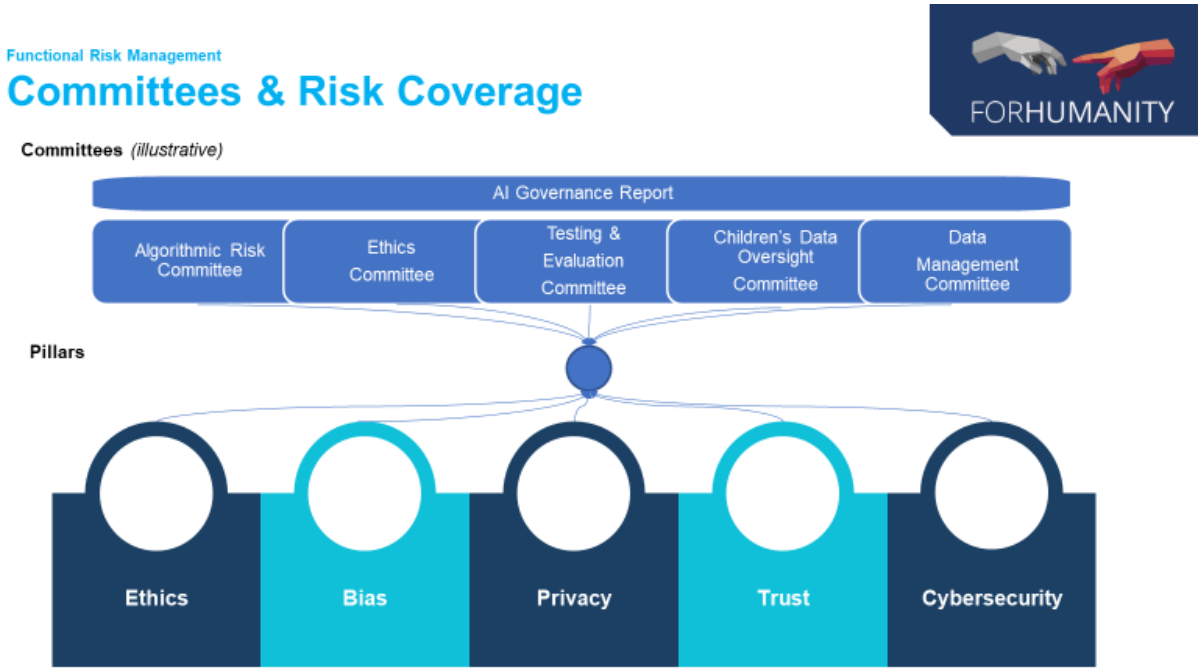
SOD & Accountability have clear segregation of duties & function as a Second Line of Defense (2LOD

Increase **Diverse Input and Multi Stakeholder Feedback** and risk input process

Manage roles and responsibilities **risk** from departures, role changes, and institutional knowledge

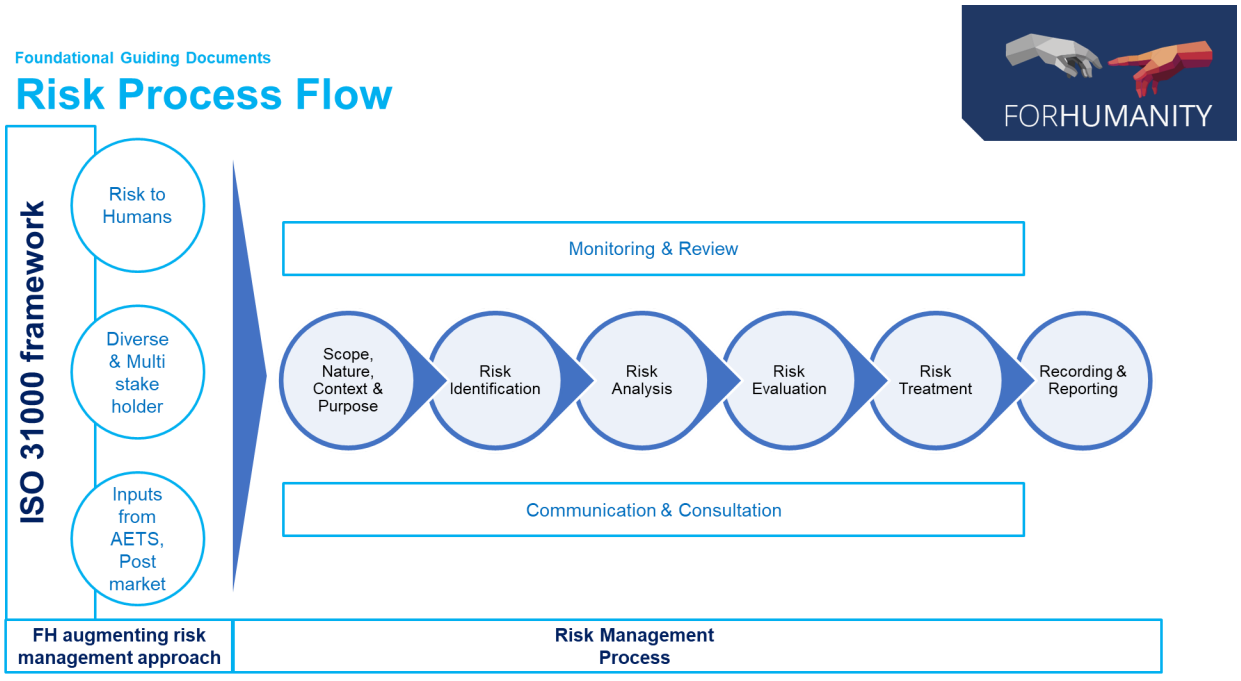
Reduce risk of autonomy, and bias (including cognitive biases)

Manage residual risk by treating the risks and disclose the residual risks



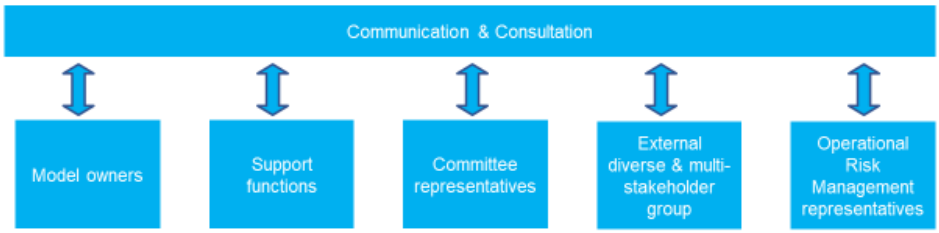
FH Risk Management Process

FH - AI Risk Management Process



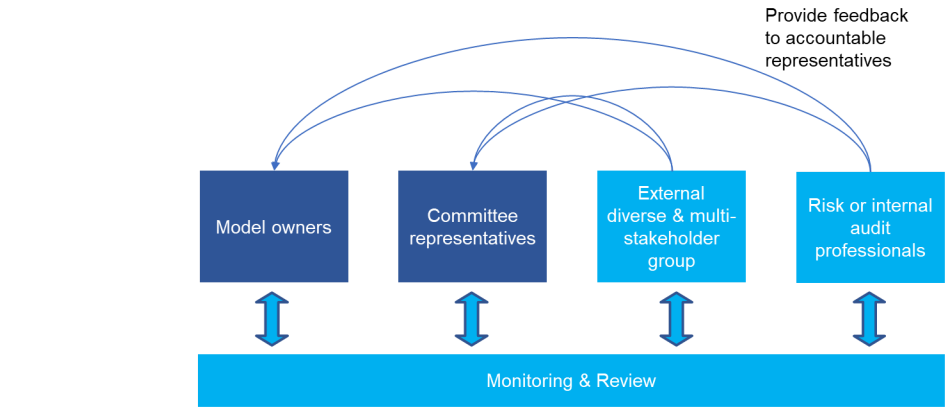
Foundational Guiding Documents

Risk Process Flow



Foundational Guiding Documents

Risk Process Flow



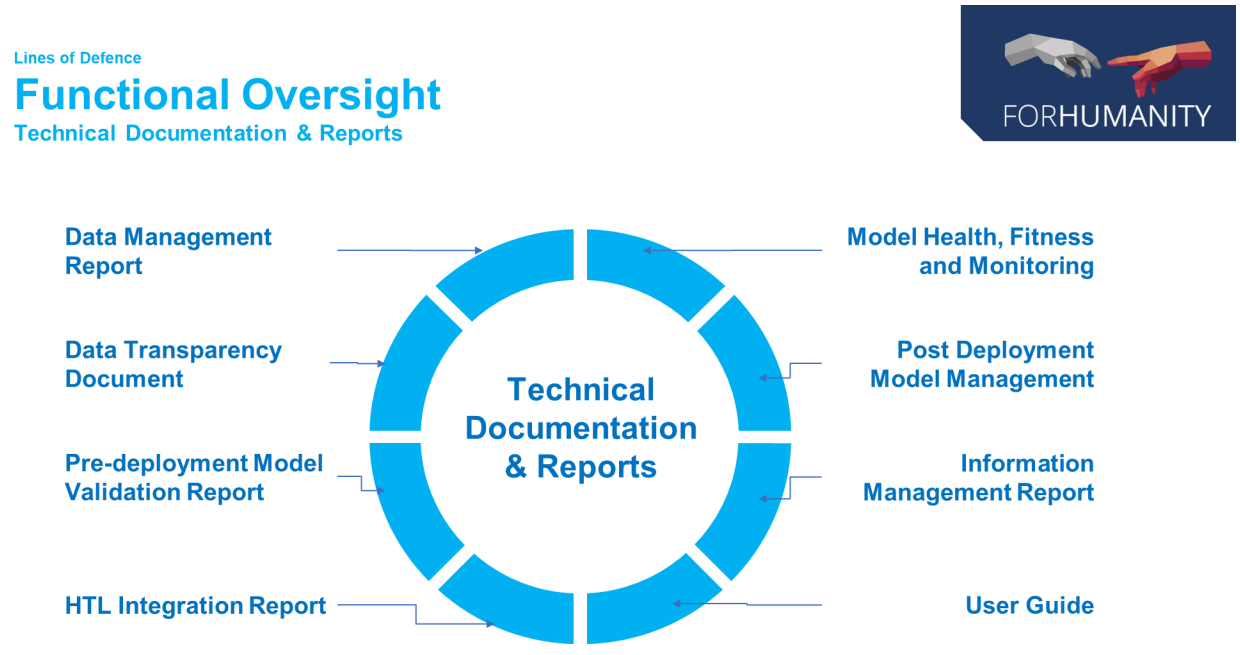
Accountable

Input providers

Guidance on operationalizing risk categories from human impact perspective

Guidance on operationalizing risk categories from human impact perspective

Functional Oversight



Functional Risk Management

Responsibilities of Committees

Responsibilities of committes in AI lifecycle

Explaining the role of committees across the lifecycle of the AAA systems

Role of product, business and other stakeholders in risk management in AI lifecycle

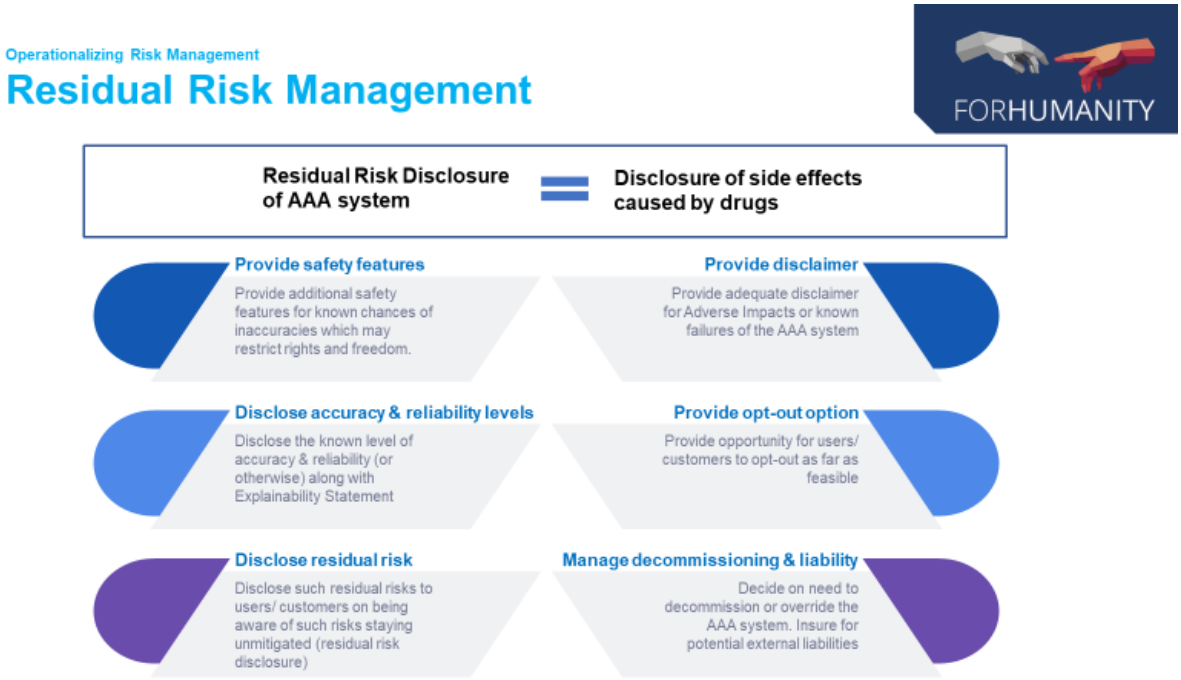
Functional Risk Management reports

Committee	Subject	Guidance & Templates
Algorithmic Risk Committee (ARC)	ARC Structure and Governance	BoK on ARC and ARA
	Algorithmic Risk Assessment Components	ARA Components and Guidance
	Algorithmic Risk Assessment Template	ARA-Risk template
Ethics Committee (EC)	EC Structure and Governance	BoK on EC Structure and Resp…
	Ethical Risk Assessment Components	ERA Component and guidance
	Ethical Risk Assessment Template	ERA-Risk templates
Testing & Evaluation	TEC Structure and Governance	
	TEC Components	
	T&E At-Risk Report Template	
Children's Data Oversight Committee	CDOC Structure and Governance	BoK on CDOC Structure and R…
Data Management Committee (DMC)	DMC Structure and Governance	
	Data Management Report Components	
	Data Management Report Template	
AI Governance	AI Governance Structure and Governance	
	AI Governance Components	

	AI Governance Assessment Template	

Residual Risk Management

Residual Risk Management and Response action



Internal Reviews

Note: Brief Intro Incident Management essentials in the context of AAA systems

cAIRE Reporting

Understanding cAIRE report -

cAiRE report

Risk and Control Scope template-

Risk and Control Scope

cAIRE residual risk log -

Residual risk log

Threat and Risk (Emergent & Horizon scanning) & Systemic Societal

a. Threat and Risk Template - to be created
b. Systemic Societal Risks - an Introduction: Systemic Societal Impact Analysis
c. Systemic Societal Risk template - to be created

Feeding into Operational Risk Management at an Organizational Level

Enterprise Risk Management - Guidance for integrating AI risk with ERM
COSO ERM: AI Risk Management integration

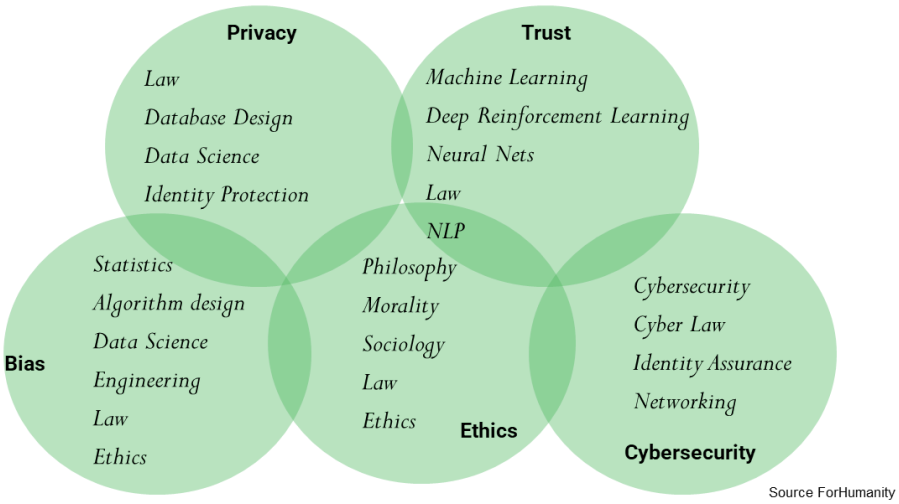
ForHumanity University

ForHumanity University is an online teaching environment designed to educate individuals on the precise skills required to execute in the ecosystem of Independent Audit of AI Systems. Coursework such as *Foundations of IAAIS* prepares the student with a broad understanding of terminology, theoretical and historical underpinnings, checks and balances, segregation of duties, the nature of audit criteria and IAAIS itself.

Building upon *Foundations of IAAIS*, ForHumanity University takes two pathways, the first is the ForHumanity Certified Auditor (FHCA) program discussed in more detail in the next section. This curriculum assures the student has a deep understanding of the audit criteria for which they are becoming accredited. Coursework covers terms and definitions, individual criteria, and associated documentary evidence. Upon completion, the student will have demonstrated a comprehensive understanding of the audit criteria through examination.

The second pathway is the Accredited Expert program such as the certifications for ForHumanity's Risk Management Framework and Ethics Committee. These curricula are designed for practical application by students in the field of AAA System risk management, or as a dedicated Ethics Officer on the Ethics Committee responsible for algorithm ethics. These expert certifications will provide job seekers and employers with a proven credential by which job requirements may be partially fulfilled in an area of expertise where training is still in its infancy.

These studies cannot replace the enormous amount of multi disciplinary study necessary for understanding the interplay between Ethics, Bias, Privacy, Trust and Cybersecurity. Below is a chart to illustrate some of the multidisciplinary studies that universities and institutes of higher learning can provide their students, with either generalist knowledge or deep specialist knowledge in any specific subset. ForHumanity aims to coordinate our accreditations with universities and institutes of higher learning around the world.



Examples of multi-disciplinary study in IAAIS

Trained and Accredited Workforce

The ForHumanity Certified Auditor (FHCA) accreditation process is exclusively for individuals and not organizations. This rigorous training teaches an individual about the Foundations of Independent Audit of AI Systems and the specific elements of audit criteria, documentary evidence for compliance and the critical elements of the Code of Ethics and Professional Conduct by which all FHCAs abide upon completion of their accreditation.

The accreditation process assures uniformity across all audits under the Independent Audit of AI System. Auditees can expect the same process and evaluation of compliance, regardless of the auditor or certification body.

FHCAs are required to maintain continuing education on the ever-changing landscape of laws, regulations, and best practices in order to maintain their good standing. ForHumanity administers the continuing education and certification of good standing and provides the marketplace with verifiable credentials.

Compliance-by-Design

ForHumanity's Certification Schemes are designed with the intention that most of the elements in the audit rules will promote and allow for compliance-by-design. It is anticipated that designers and developers will build the audit backup satisfaction into the system, plugging into the COSO system of internal risk control, governance and audit compliance as well as operational risk management and enterprise risk management.

Board of Directors Audit

There are audit criteria dedicated to Boards of Directors and which must be answered according to the required audit documentary evidence. It is not expected that the Board of Directors will have day-to-day responsibilities associated with audit compliance, however, the Board should have accountability for systemic failures of governance and accountability systems. Audit criteria are designed to ensure culpability, designed to ensure that the Board has adequate knowledge and oversight of key elements of the audit process. Notably, the requirement to establish Algorithmic Risk Committee, the Ethics Committee and the Children's Data Oversight Committee. All audit criteria referencing the Board are the sole responsibility of the Board of Directors. Auditors should ensure that the Board is the respondent of record to relevant audit questions.

Evaluation Methods

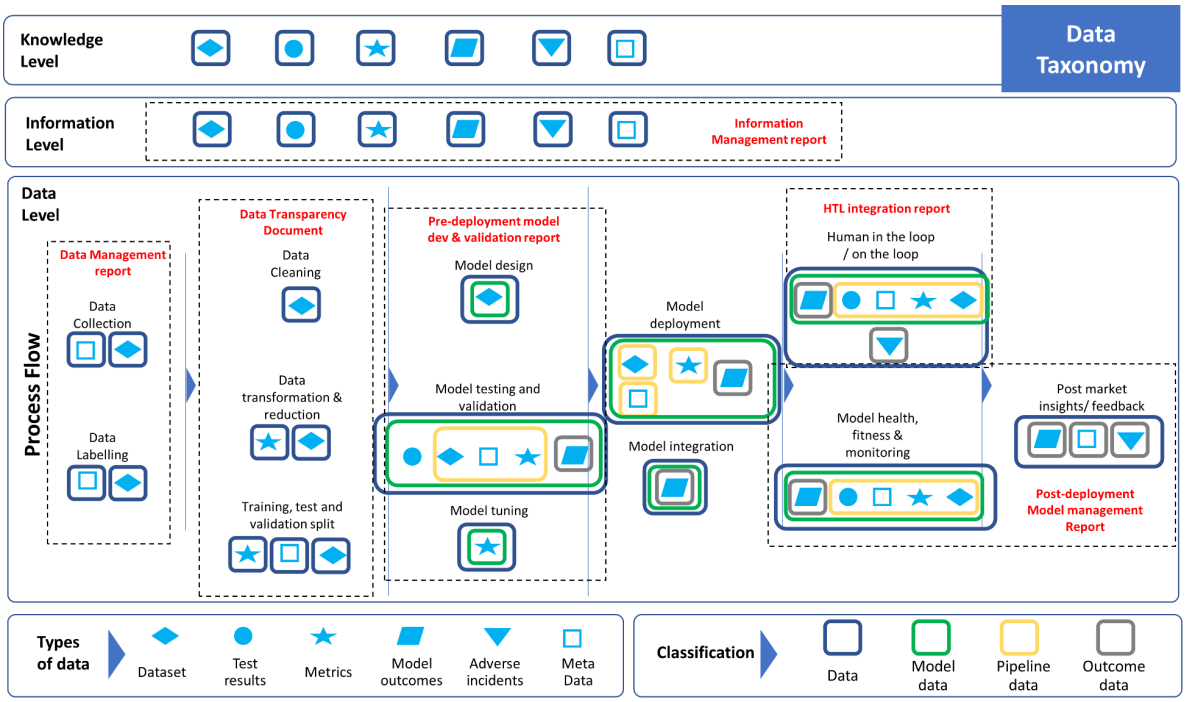
Each of the scheme criteria identifies a typ[e of evaluation method. The auditor may vary the evaluation method type where it provides additional assurance, but not so that it provides less. The following types are listed:

1. *Contract*. An executed contract can be examined and demonstrates compliance with the criteria.
2. *Correspondence*. Historical correspondence is available that demonstrates compliance with the criteria.
3. *Internal log, register or database*. Internal records and reports or systems with proof of authenticity can be examined by the auditor, and demonstrate compliance.
4. *Internal procedure manual*. Internal procedural documentation can be shown to the auditor that demonstrates compliance with the criteria. Note that these procedures should be of sufficient detail to show that they are up-to-date, implemented, operational and complete. High-level policies are not sufficient to demonstrate implementation.
5. *Public disclosure document*. A publicly disclosed document will demonstrate compliance. This may include comparison to other evaluation types.
6. *Physical testing*. This can refer to any of the following, at the auditors' discretion:
    a. *Records of previous events that can be examined*. For example, if there is a clear audit trail demonstrating the response to prior Data Subject Access Request, the auditor can review this audit trail to gain confidence that the organization can comply with the criteria.
    b. *Witnessing current events*. For example, to ensure that an organization can restore from backup, the organization can demonstrate its ability to do so to the auditor.

c. *Technical testing.* For example, to demonstrate that network traffic is encrypted, the auditor may inspect the traffic.

Copies of all evidence obtained during the evaluation should be stored in encrypted form by the auditor, except where this includes personal data and does not comply with the principle of data minimisation.

Data Taxonomy and Technique Documentation of AAA Systems



ForHumanity has established a data taxonomy to enable increased precision and specification of the overused word "data." The EU Artificial Intelligence Act requires significant amounts of new documentation systems to achieve compliance with the Act. The required documentation is listed on the diagram above in red font, with the exception of the AAA Systems User Guide, a document presented to Users by providers of AAA Systems:

1) Data Management Report
2) Data Transparency Report
3) Information Management Report
4) Pre-Deployment model development and validation report
5) Post-deployment model management report
6) HTL Integration Report

This diagram is from a forthcoming paper on Data Taxonomy. Beyond satisfying legal requirements for documentation, the Data Taxonomy paper provides clarity on the word "data" in the following ways:

1) Process Flow - from Left to Right - Sourcing Data to Design to Development to Deployment to Integration to Decommissioning
2) Data Hierarchy - from bottom to top - Data to Information to Knowledge
3) Data Type - keyed in from shapes
    a) Dataset
    b) Test Results
    c) Metrics
    d) Model Outcomes
    e) Adverse Events
    f) Metadata
4) Data Classifications
    a) Data
    b) Model Data
    c) Pipeline Data
    d) Outcomes Data
5) Data Task - described at each step of the Process Flow, to each Data Type

Process Flow

1) Design
    a) Source Data
    b) Data Management and Preparation
2) Development
    a) Training, testing/validation and transformation
    b) Pre-deployment, Development, Validation and Functional Correctness
3) Deployment
    a) Pipeline
    b) HTL Integration (Human-in/on-the-Loop including Human-in-command)
    c) Downstream Integration
    d) Model health, fitness and monitoring
    e) Post Market monitoring and Adverse Impact Reporting Systems (AIRS)
4) Decommission

The paper provides a detailed explanation of each term, each step and each differentiation. These definitions already exist in our GDPR audit criteria or EU Artificial Intelligence Act criteria, but are behind combined in the paper for clarity.

Body of Knowledge - Knowledge Stores

The Body of Knowledge and its specific Knowledge Stores are guidance notes for Auditors, to be applied when examining items of compliance sufficiency and maturity. They do not represent normative criteria. Instead they reflect measures, tools and thresholds that help an Auditor understand if the documentary evidence is sufficient, or even a mature level of compliance. Further, the Knowledge Stores highlight frequent insufficiencies related to documentary compliance evidence designed to draw attention to common mistakes with sufficiency. The Body of Knowledge - Knowledge Stores can be found here.

Certifications

Certification Merits

Independent certification of conformity to ForHumanity's Certification scheme for AAA systems represents the highest form of compliance.

1. Certification demonstrates the organization's willingness to transparently and objectively document controls, governance, and accountability with respect to the approved certification schemes.
2. Certification helps your organization demonstrate compliance to the regulator, the public and in your business-to-business, supply chain relationships.
3. Mitigation of risks associated with AAA Systems.
4. The Certification mark received by the organization upon compliance is an outward, public expression of a commitment to uphold the law to the highest standard and to indicate to clients, customers, prospects and employees that their Personal data/Personal Information is well protected, used ethically and fairly and that interactions, interfaces and outcomes will be fair and according to the Code of Ethics and Code of Data Ethics.
5. Certification enables an opportunity to have an independent examination of the data supply chain, the AAA System life cycle, including the associated supply chain and includes robust documentation and disclosures.
6. It enables organizations to prove monitoring of ethical, bias, privacy, trust and cybersecurity risks, to implement proportional controls, and encourage a company culture upholding privacy.

Limitations of Certification

Certification does not provide immunity from regulatory scrutiny or action, and is not a guarantee of continuous compliance with the law and/or certification schemes.

Certification marks must be used in association with the pre-agreed disclaimer language to disclose the data processing purpose that is covered by the certification mark and any limitations.

Engagement with a Certification Body

The Auditee shall engage with an Auditor (Certification Body) by executing an Audit Engagement Letter. This letter will explicitly identify a Target of Evaluation (TOE) upon which the certification will be conducted. This letter shall stipulate the following:

1. Scope of the Data Processing Purpose including beginnings and ends (ToE)
2. Disclaimer for the certification mark
3. Rules and guidelines for use of the certification mark
4. Expectations from an auditee to provide documentary evidence
5. Expectations regarding ongoing and post market monitoring
6. Certification Plan, including, if applicable:
    a. Opening meeting where the scope is verified and the names of organizations and individuals participating, and their roles
    b. Confirmation of the authorisation of the auditors to award the certification, and their impartiality
    c. The Target of Evaluation (ToE, as documented in the contract)
    d. The legal basis of the AAA System and the role of the Auditee
    e. Expected documentary evidence
    f. Physical testing scheduling
    g. Any site or network access required, and any special requirements for that access (e.g. permission to conduct intrusive network scanning)
    h. Closing meeting for presentation of Certification Report, issuance of Certification or issuance of Non-Compliance Letter
7. Certification Report, including:
    a. Clear explanation of the scope agreed in the Audit Engagement Letter and the beginnings and ends, also expressed in the disclaimer
    b. Any deviations from the certification plan
    c. Process narratives, walkthroughs, flowcharts, diagrams, control descriptions, codes, policies
    d. The specific software and hardware versions and assets inspected including third-party assets, as applicable
    e. The actual dates of inspection(s)
    f. A list of documentation and assets that will be retained as audit evidence, and explanation of deviations
    g. A duly authorized signatory
    h. A list of deficiencies, if certification will not be issued
    i. A determination of sufficient/mature levels of compliance

j. Whether a certification is awarded, and its duration
k. Sufficient deliverable for disclosure requirements
l. Sufficient, robust and resilient ongoing monitoring systems and explicit statement that systemic failures of ongoing monitoring systems will preclude future certification

Auditor - Auditee agreement on Scope

See section Target of Evaluation Determination Process

Certification Warning/Certification At-risk

The Certifying body (Auditor) may issue a written warning to an auditee that they are not compliant with the terms of the Audit Engagement Letter. This written warning shall include a timestamp, remediation period, and the expected remedy. Failure to satisfy may result in the withdrawal of certification. Potential warnings could include:

1. Misuse or misrepresentations in use of certification mark and their stated purpose
2. Contravention to any of the contractual clauses for certification
3. Failure to maintain documentary evidence related to the certification
4. Failure to maintain post market, robust and ongoing monitoring on the data process
5. Failure to uphold agreed and documented thresholds, Key Performance Indicators on the data process
6. Concept drift and deviations from scope, nature, context or purpose of the data processing
7. At the launch of an investigation based upon a report or complaint by the FH certified auditor highlighting potential misrepresentation, falsification or fraud associated with information provided for audit
8. Reported data privacy breaches

Warnings and at-risk certification may or may not lead to revocation of certification based upon this guidance and failures to remediate in a timely fashion, at the discretion of the Auditor.

Withdrawal of Certification

Certification may be withdrawn for any of the following reasons:

1. Regulatory action related to the data process
2. Successful civil litigation of a case directly pertaining to the data process certified
3. Failure to maintain documentary evidence related to the certification

4. Failure to maintain post market, robust and ongoing monitoring on the data process
5. Failure to uphold agreed and documented thresholds, Key Performance Indicators on the data process
6. Concept drift and deviations from scope, nature, context or purpose of the data processing
7. Material change in organizational governance, accountability, oversights or controls related to the data process
8. Reported data privacy breach
9. Fraud, misrepresentation or malfeasance associated with material information related to the certification

The Auditor will notify the Auditee that certification has been withdrawn with a Letter of Withdrawn Certification and will be required to provide the auditee with the associated reason for the withdrawn certification from the list above. This may be done at their sole discretion according to the Audit Engagement Letter for any reasons listed above.

Certification mark use standards and guidelines

See the ForHumanity license agreement and website for standards and use guidelines for certification marks.

Certification Steps

Define Scope

ForHumanity designs certification schemes for specific AAA Systems. The scheme may only be applied to systems that fall within parameters and scope as defined in the certification scheme.

Target of Evaluation Determination Process

The Auditee determines the data process(es) to which the Auditor will apply the scheme and documents this agreement in a contract. The Target of Evaluation (ToE) shall be defined by an Audit Engagement Letter between the auditor and the organization (the Auditee).

The Audit Engagement Letter shall document all information required by the Auditor for a sufficient Certification Plan and shall include all of the following:

1) Name/identifier of the ToE, specifically noting the boundaries of the data process

2) Beginnings and Ends of the Data Process(es) where Personal Data is processed (including a visual representation)
3) Systems or organizations expected to be "in" or "out" of scope (including data Processors under contract), including a visual representation as appropriate
4) Description of the lawful basis for processing, as well as its scope, nature, context and purpose
5) Description of the data deployed in the system, specifically noting the Personal Data/Personal Information and Special Category Data/Sensitive Personal Data/Biometric Data that may be present (including Inferences and/or potential Proxy Variables

The Auditor will only perform an audit of the documented scope. The Auditee bears the responsibility of ensuring that all relevant aspects, infrastructure, data, data processes, storage, interfaces, software, service providers and output system necessary for the proper function of the AAA System identified as a ToE undergo an Audit.

The Auditor and Auditee shall document in Audit Engagement Letter the wording of an associated disclaimer from ForHumanity to be published alongside the Certification mark to provide clarity on what has been certified.

Conduct Pre-assessment/ Pre-audit

Certification requires extensive preparation and work to ensure that requirements will be met completely. During a certification audit, there is limited ability to remedy meaningful shortcomings. Therefore, the organization is strongly advised to invest time in pre-audit compliance, designed to meet the requirements of the audit in advance, while preparing the organization for ongoing maintenance needed to maintain certification. Pre-audit compliance should establish the following key components for certification compliance:

1. Establish infrastructure for governance, accountability and oversight as specified in the certification criteria
2. Identification of certification criteria requiring documentary evidence
3. Establishing process for compiling and storing documentary evidence
4. Identification of training needs and sourcing auditable solutions
5. Drafting codes and procedure manuals
6. Identification of system requirements (hardware and software)
7. Verifying operational risk management and control processes
8. Preparation for disclosure requirements

Identify Certification Body

ForHumanity relies upon local, government-approved accreditation services (e.g United Kingdom Accreditation Service (UKAS)) when one exists. ForHumanity will provide

information and resources to the accreditation service in support of their mission. In the event that a government-approved accreditation service does not exist, ForHumanity has a process for evaluating sufficiency of accreditation bodies and is willing to provide that service.

Identify Auditors for Certification

Auditors must be trained and certified in the scheme that they intend to provide. Not all persons engaged in the provision of certification services must be individually certified, however the issuance of a certification may only be provided by a named ForHumanity Certified Auditor (FHCA).

FCHAs are well-trained in the certification scheme criteria and the audit process that leads to certification. They are required to maintain their knowledge through continuing education and their current status can be checked on the ForHumanity Certified Auditor website found here.

FHCAs must abide by the principle of Independence.

Independence Enforced via License

As a legal term defined in America by The Sarbanes-Oxley Act of 2001, a certifying body (an Auditor) must receive no other remuneration from an Auditee beyond reasonable audit fees. ForHumanity further stipulates, in its license agreements, that a licensee cannot be an Auditor and an Assessor/Consultant (or provide any other form of service) to the same Auditee in a 12-month period.

More details on specific examples of Independence can be found in **ForHumanity's Certified Auditor Code of Ethics and Professional Conduct v1.0.**

Anti-Collusion

Independence is further enforced through licensing requirements enforcing anti-collusion amongst auditors and pre-auditors. As the market for data auditing matures and grows, it is impermissible for pre-audit service providers and auditors to regularly guide clients to each other excessively. This prevents pre-auditors and auditors from becoming overly comfortable with each other's processes/expectations and failing to deliver the maximum diligence and objectivity owed to the client and the public. Anti-collusion requirements ensure maximum mitigation of risk to humans and implement complete compliance.

Certification Issuance

Upon completion of the audit process, an accredited auditor in their sole discretion will either issue certification or explain why certification has been denied. This occurs after all attempts at remediation have been made by the auditee within a reasonable time period as determined by the auditor.

ForHumanity and Accreditation Service Examinations

Both government-appointed accreditation services and ForHumanity have the right and responsibility to periodically review certification reports to ensure that accredited certification bodies are conducting their work in a responsible and proper manner, consistent with the requirements of the approved certification scheme. The organization receiving the certification for a AAA System would only be notified if there were a discrepancy or shortcoming of the certification process. The auditee would have a reasonable opportunity to rectify and meet the requirements of certification.

Audit Period of Validity

A certification is good for one year. Compliance should be renewed annually and an auditee is expected to maintain compliance with the current version of the audit. In any areas where the audit has changed, the auditee will have until the next annual audit to bring their systems into compliance.

Significant changes in the nature, scope, and purpose of an AAA System should require updated certification. Significant changes to an algorithmic system may jeopardize the certification status. Some examples which may require recertification to maintain status are:

1. Acquisition/Change in Control
2. Complaint
3. Regulatory intervention
4. ForHumanity's Cause for Concern

Recertification

It is expected that organizations will want to maintain their certified status for the data processing purpose. The organization will be welcome to recertify against the current version of the certification scheme, and it is expected that recertification will be substantially easier resulting from the investment in compliance from the original certification.

