Identifying Low Risk AI, Algorithmic and Autonomous Systems

ForHumanity's risk management framework requires the identification and exclusion of systems that represent a low risk to humans and their rights and freedoms. This process will allow for efficient allocation of resources to systems that present a higher risk—systems that require governance, oversight and accountability to ensure the safety of people and the maximization of benefit from AI, Algorithmic and Autonomous Systems (**AAA Systems**).

This paper is about the operationalization of Step 2 in the framework below.

ForHumanity's Risk Management Framework

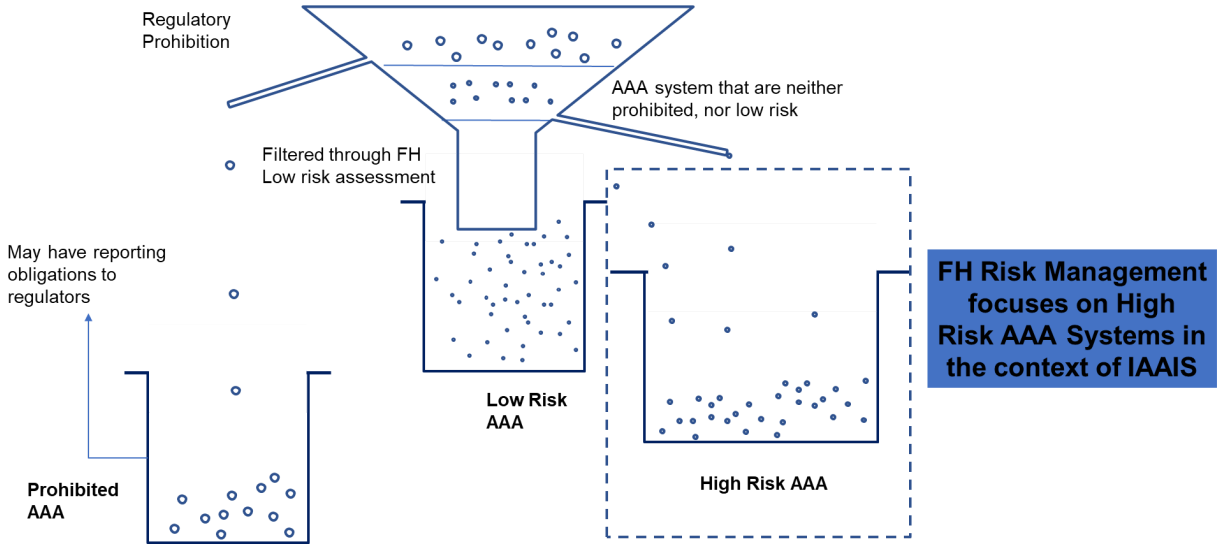
The risks we are measuring are the impacts to people from all **AAA Systems**. In terms of Risk Inputs, we want internal and external risk assessors to rank or consider the severity and likelihood of each risk identified in a given system. However, before that step, there are many systems that we can eliminate prior to requiring a comprehensive compliance audit. This "low risk" assessment must be monitored and reassured continuously.

—---

Step 1: Is it prohibited by law?
Step 2: Is it low risk and excluded from independent audit?
Step 3: Risk management consideration for all the remaining AAA systems.
Step 4: If required, report to regulators after risk mitigation or as per relevant legal framework(s).

Operationalizing Risk Management

Risk Classification of AAA systems



The primary step towards examining **AAA Systems** is to understand if the systems or their nature or components have a high negative impact on health, wellbeing, safety and/ or fundamental rights and freedom of natural persons and is prohibited by regulation. For instance, the EU AI Act prohibits certain **AAA Systems**.

Step 1 refers to the **Relevant Legal Framework** applicable to each system. Many system owners will have to consider multiple jurisdictions and relative prohibitions for each jurisdiction as applicable to their **AAA System**.

Step 2 determines if the **AAA System** may be excluded from required compliance with ForHumanity's **Independent Audit of AI Systems (IAAIS)**. This step defines a filter for **AAA Systems** considered "low risk." This filter is a self-assessment that, if passed through successfully, should serve as an adequate mechanism for documenting and assuring the system as low risk. If an entity is interested in verification, low risk determinations could also be assured by an Independent Auditor.

ForHumanity's intent in creating the following low risk AAA System assessment is to help organizations alleviate compliance burdens through effective allocation of risk mitigation and assurance resources.

In addition, the independent assurance of an AAA System as low risk may be used as social proof of organizational values and used toward increasing perceptions of trustworthiness and public transparency.

In most cases, a determination of low risk for an AAA System will not relieve entities of the potential risk of non-compliance with the law. Accordingly, entities are encouraged to seek legal counsel. ForHumanity will endeavor to have the following low risk exclusions added to applicable legal frameworks. Until then, the best we can do is advise.

Step 2: Low Risk AAA System Determination

Below is a series of criteria for low risk AAA Systems determination. Failure of any one of these criteria may result in the AAA System being classified as other than low risk, requiring Independent Audit of AI Systems.

1. **Privacy.**
    a. The **Data Subject** does not have an expectation of data protection or privacy associated with the specific items provided as **Personal Data**. Under this criteria, **Personal Data** are typically public already. The risk of data breach is well signaled, deemed non-consequential (eg. Personal Data is limited to one or two variables, primarily used for contact or to facilitate end-to-end connection); and
    b. Where the lawful basis of data collection is constrained to consent or contract performance; and
    c. Where there is no ability to sell, share, transfer or trade the data, whether internally (eg. within a single organization) or externally (eg. third-parties.)
2. **Trust.**
    a. Where the **AAA System** is non-complex (eg. limited upstream, downstream interoperability, limited Human-in-the-loop interaction and interpretation), with limited need for Accountability, Governance and Oversight - and the technology may not be innovative
    b. Systems where Personal and non-personal, synthetic data data usage is not monopolistic and transparent
    c. Where the **AAA System** outcomes (a) are known not to cause or have had any adverse impacts to the safety or health of humans, or (b) are known not to cause or have any adverse impacts to the environment or (c) do not create any legal implications
    d. When the **AAA System** is not widely or completely accessible in the context of its deployment
3. **Bias.**
    a. When **AAA System** decisions are not being applied to a Data Subject differently than any other Data Subject, except by the Data Subjects' opt-in choices; and

b. Where the **AAA System's** primary purpose is to "inform" on pre-chosen information; and
c. Where the **AAA System** has no autonomy except for functions based on the **Data Subject's** stated preferences.

4. **Ethics.**
    a. Risk to rights and freedoms (risks to human values) is non-existent (Where models do not discriminate people based on their profiles/ preferences) including processes where there is no potential for influencing behavior or democratic beliefs of individuals or society.

Why is the identification of Low Risk Processing necessary?

Our answer is that this system of risk assessment provides future proofing to the overall function of risk assessment for AAA Systems. The EU AI Act is poised to accomplish a robust mechanism of governance, accountability and oversight, but it tied its own hands on determining what systems would be governed by it, likely because our political leaders did not yet have a robust mechanism for differentiating risk. Under the current risk assessment mechanism, the EU has forced itself into a regular, likely politically charged debate as to when a system is sufficiently risky enough to be labeled "high-risk" and added to Annex III (a threshold that is yet undefined in the proposed law). Whereas, the ForHumanity approach labels all systems that are not prohibited and not low risk as worthy of Independent Audits based upon system specific audit criteria. The ForHumanity Risk Management framework and notably the exclusionary mechanism for Low Risk provides a flexible mechanism for the pre-assessment of all systems avoiding the complexity of political debates around thresholds of risk conducted without consideration of specific or appropriate mitigations.

First, lawmakers will always have the power and the authority to determine prohibited systems - again, something beyond ForHumanity's remit. Once eliminated we have everything else - all systems. Beginning with this remainder, if we can eliminate the systems that are not sufficiently risky to humans or our humanity (Low Risk), then we are left with everything else that must exist in an equilibrium between balancing the benefits from these innovative systems and the trade-offs that come with each system.

We enter this phase having eliminated systems that are too risky and others that are essentially not risky. ForHumanity believes that "everything else" should have governance, accountability and oversight tailored to each individual system's riskiness - literally all other systems. Each system clearly has sufficient risk to humans and humanity to not be excluded from governance, oversight and accountability, otherwise they would be Low Risk. However, we would be obtuse to believe that all remaining systems have the same amount of risk or require the same risk mitigations. A recommendation system will not have the same impact on humanity as content moderation on a major social media platform (even if

they do largely the same thing - recommend). Risk and risk mitigation must be considered in each individual system.

This is where Independent Audit of AI Systems (IAAIS) does its work. Demanding risk assessments across the five pillars of Ethics, Bias, Privacy, Trust and Cybersecurity that are tailored to the risk of each individual system and its impact on humans and our humanity. IAAIS brings tailored compliance requirements designed to address the level of risk present in AAA systems with precise metrics and criteria for compliance - assured by Independent Auditor acting as a proxy for society to verify compliance. Criteria are designed in a crowdsourced manner, considering Diverse Inputs and Multi Stakeholder Feedback to broadly encapsulate risk to humans and humanity. These criteria are flexible and specific and can be altered or expanded to meet each new challenge without disrupting the rest of the system.

