

## **Guidance on selection, engagement and oversight mechanisms for gathering Diverse Inputs and Multi-Stakeholder Feedbacks**

### **Setting the context**

ForHumanity advocates for a risk management framework that is omni-directional and multivariate. Multivariate in that the risk framework considers corporate risk (damage to employees, business reputation and shareholder wealth), risk to humans (damage to users/clients/prospects and unwitting participants), societal risk (damages to systems, groups, communities, markets and collectives) and environmental risks (damages to nature and sustainability considerations).

### **Concept of Diverse Inputs and MultiStakeholder Feedback**

Diverse Inputs & MultiStakeholder Feedback is the process of injecting human agency into AI, Algorithmic and autonomous systems to maximize risk mitigation in socio-technical context. The word “Multi Stakeholder” is a 360 degree perspective of the system and risks being assessed. Providers of data, providers of inputs, providers of systems (software and hardware), providers of networks, data processors, partners, clients, prospects and most importantly the human users, humans impacted, civil society and researchers from many different backgrounds, cultures and user experiences.

This concept is not new. For instance, a bug bounty program attempts to gather insights from experts or practitioners to understand previously unconsidered or unknown risks. This essentially is a form of multi stakeholder feedback mechanism. However, these mechanisms fall short in the context of socio-technical systems when they do not include human diversity and associated lived experiences consistently across the value chain of AAA system.

### **Why Diverse Inputs and Multi-Stakeholder Feedback matters?**

DI&MSF matters because it enables systematic capturing of such diverse inputs in the evaluation of risk. This thereby brings in previously unknown or unfamiliar perspectives of risks that are not considered by the organization in the current scale of operations. In addition, it provides an opportunity to understand if the risk impacts are different for different groups of people.

### **Considerations for using engaging Diverse and Multi Stakeholder Groups (Human Impact Assessor (HIA)) and collecting Input and Feedback**

- **Nature of diversity:** Diversity needs to be considered contextually. For instance, if a platform is at the risk of discriminating against people of color in few geographies, diversity shall have consideration to have representations of people of colour from those geographies.

- **Contractual arrangements:** Human Impact Assessor (HIA) shall be engaged only on contract (with or without a fee associated with such activity). The contract shall provide specific clarity on the role, responsibility, and confidentiality. It shall explicitly define the expected conduct of a Human Impact Assessor (HIA) (in the form of a Code of Conduct/ Code of Ethics as the case may be).
- **Code of Conduct for HIA:** The Code of Conduct / Code of Ethics for the Human Impact Assessor (HIA) shall have clear expectations on the conduct of Human Impact Assessor (HIA) during the course of engagement.
- The organization may determine the process of gathering inputs. This may include approaches which guide specific inputs and the ones which gather open inputs on an AAA system. This may involve training (even at micro-levels for input processes)
- While seeking guided inputs, great care should be taken to balance the Trade-off between diversity of thought and lived experience and the conformity that manifests through uniform training. Teach the essentials of execution. Teach the essentials about the tool. Forego guidance on “how” or “what” to access allowing the assessors the freedom to explore negative impacts, harms and potential mitigations.
- **Confidentiality:** The organization shall set clear boundaries on the nature of information that will be shared to avoid potential intellectual property being exposed during the process of collecting DI&MSF. Further, the organization shall adopt alternative approaches in nature of surveys, workshops, case study deliberation or discussion on an illustrative scenario. The approach should be to ensure that adequate information is shared with the Human Impact Assessor (HIA) to gather DI&MSF. Confidentiality should be required and Human Impact Assessor (HIA) should grant and abide by the confidentiality agreements.
- The organization shall also guide the Human Impact Assessor (HIA) to determine an appropriate structure for participating, examining, consolidating and ranking the risk inputs. These need not be hierarchical structure within the Human Impact Assessor (HIA)., however, it shall have clear guidance on decision making, quorum or dispute resolution etc. The organization shall revisit the guidelines as may be required from time to time.
- **Independence declaration:** Independence declaration is one of the critical aspects in engaging a Human Impact Assessor (HIA). The Organization and the individual representative participating in the process of DI&MSF shall essentially declare that they are free from any conflicts (including investment or similar monetary) associated with the Organization that is seeking inputs, the AAA system (including its creators/ owners) and associated individuals including Designated Personnel who play a key role in the organizational process.
- **Avoiding conflict of interest:** A conflict of interest occurs when an entity or individual becomes unreliable because of a clash between personal (or self-serving) interests and professional duties. This will include scenarios where the conflict will be perceived due to the role or position that the person or organization plays in the relevant business context, though there are no demonstrable direct conflicts. For

instance if the researcher (person) participating as a Human Impact Assessor (HIA) is also receiving funding for their research from the same business unit, we may have a perceived conflict in such a scenario. These examples will be found in the Code of Ethics and Professional Conduct for HIAs.

### **Process of diverse and Multi Stakeholder groups**

- **Setting diversity expectations:** The Ethics Committee, using the Code of Ethics and/or the diversity policy, shall determine the specifics around the diversity that is required for Human Impact Assessor (HIA) in the relevant business context. This shall include specifications around the gender, color, race, religion, age, national origin, demographic, domain, skill, experience and relevance for the said business context. Further the Ethics Committee shall ensure that HIA are aware of the 5 pillars of FH (Bias, Ethics, Privacy, Trust and Cybersecurity).
- **Balanced Diversity:** Human Impact Assessor (HIA) group shall be balanced with representations from all relevant classes of protected categories as may be contextual to the scope, nature, context and purpose of the AAA System. Care shall be taken to avoid significant concentration of Human Impact Assessor (HIA) for a class of protected category (including demographic, role and capability).
- **Designate personnel for managing HIA process:** The organization shall also designate personnel via a Duty Designation letter who will have the responsibility of managing the selection, engagement and oversight process along with the support of committees (Ethics Committee, Algorithmic Risk Committee, Childrens Data Oversight Committee as required by criteria or otherwise necessary).
- **Funding and resources:** The organization shall ensure adequate funding and resources (including engaging personnel with sufficient authority) are available for gathering DI&MSF commensurate with the risks posed by AAA systems.
- **Representativeness in HIA selection:** Ethics Committee shall ensure that the organization takes adequate considerations to include representatives from Civil society organizations, Non profits, individual experts, Community representatives, Users, Outsourced Human-in-the-loop and foreseen at-risk or vulnerable groups.
- **Diversity in spirit:** Further, the organization shall ensure that they do not assume diversity for groups that do not bring alternative perspectives. Such groups shall be evaluated as part of the Human Impact Assessor (HIA) evaluation to determine the effectiveness of inputs received from them and diversity of their perspectives on a need basis.
- **Ineligibility to join as a HIA:** The Ethics Committee shall also detail the type of organization or individuals who are ineligible to participate in the Human Impact Assessor (HIA) and provide DI&MSF.
- **Reasonable efforts to attract HIA:** The organization shall provide reasonable efforts to retain relevant and trained Human Impact Assessor (HIA) for gathering risk inputs.
- **Type of engagement of HIA:** Human Impact Assessor (HIA) shall be engaged at identified stages of the AAA system and can also be engaged in an ad hoc manner

based on relevant business circumstances. Illustratively, from a Risk consideration perspective, Human Impact Assessor (HIA) shall be engaged in the following stages: Concept , Design, Development (including redevelopment or overlays), Deployment and Post market monitoring.

- There could be different groups of Human Impact Assessor (HIA) who are engaged in different aspects associated with each of these stages. For example, different Human Impact Assessors (HIA) could be engaged for examining ethical considerations and for accessibility considerations. While there could be common members between both groups, they shall have representative capability and experience to contribute to both areas.
- **Frequency of HIA intervention:** The frequency of intervention of Human Impact Assessor (HIA) shall be determined by the Ethics Committee based on need considering the nature and complexity of the AAA system.
- **Minimum requirements of HIA:** In addition, the Ethics Committee shall detail the minimum requirements including the number of participants, aspects to seek feedback on and considerations for evaluation of the feedback, in consultation with the other committees (Algorithmic Risk Committee, Testing & Evaluation Committee and Childrens Data Oversight Committee to name a few).
- **Due diligence of HIA:** The organization shall provide adequate guidance on the due diligence requirement and associated process for selection of Human Impact Assessor (HIA) and make the designated personnel responsible for it. The designated personnel shall document the process, outcomes and exceptions associated with due diligence.
- **Review of HIA due diligence process:** The Ethics committee shall examine the due diligence reports to identify if there are any gaps in the selection process including risk of not including relevant Human Impact Assessor (HIA) and inadequate considerations for due diligence outcomes. The Ethics Committee shall report the unresolved issues as residual risks in the Ethical Risk Analysis report. In addition, the Ethics Committee shall produce a report to the management detailing the potential impact associated with the risks which were resolved after a delay and risks which remain unresolved.
- **Nature of engagement with HIA:** Process of engagement with the DI&MSF needs to be documented. It shall clarify the expectations and the approach towards gathering feedback. Besides specific feedback sessions as part of the process, there could also be such a process on an adhoc basis to gather feedback (including based on insights from AETS) post production via polls or other electronic information gathering.
- **Process and guidelines for HIA:** The organization shall publish a process for DI&MSF including the process relating to selection, engagement and oversight of the Human Impact Assessor (HIA). The process of engaging with the Human Impact Assessor (HIA) shall include details of process (reasonably explained) to seek risk indicators/ risk inputs, platform or medium to seek risk inputs, guidance on how to

resolve conflicting opinions within or between groups (essentially measures to deal with tensions and tradeoffs) and establish mechanism to streamline inputs into risk log. In addition, the organization shall encourage the Human Impact Assessor (HIA) to report concerns associated with any instance of oppression, mistreatment or other forms of disregarding key risk highlighted by Human Impact Assessor (HIA).

- **Onboarding HIA:** The organization shall ensure adequate process in place to provide onboarding support for the Human Impact Assessor (HIA) including awareness sessions and accessible guidance documentation.
- **Involving committee representatives in DI&MSF sessions:** The organization shall require the designated personnel to include representatives (to the extent relevant) from the Ethics Committee, Algorithmic Risk Committee, Childrens Data Oversight Committee and Testing and Evaluation Committee for DI&MSF sessions.
- **Documentation of HIA outcomes and treatments:** The organization shall require the designated personnel to document the sources, inputs and feedback gathered during the process. This shall include the risks, concerns or adverse events highlighted by the Human Impact Assessor (HIA).. In case, DI&MSF resulted in a number of risk indicators, the designated personnel shall ensure that adequate efforts are undertaken to understand the root causes associated with these risk indicators.
- **Mechanism to evaluate effectiveness:** There should be a well documented mechanism to evaluate the effectiveness of the Human Impact Assessor (HIA) and assess the need for de-empanelling or re-empanelling a select representative.
- **Review of HIA outcomes and subsequent treatments:** The Ethics Committee in consultation with the Algorithmic Risk Committee and other committees (as may be relevant) shall examine the documentation (including HIA evaluations) to understand the risks gathered, subsequent actions to treat the risks and weightage provided to the risk inputs gathered through the DI&MSF process. The Ethics Committee along with the other committees shall include the unresolved risks as residual risks as part of their respective reports (Ethical Risk Analysis, Testing and Evaluation Committee At-Risk Report, Algorithmic Risk Analysis etc).
- **HAI's as an data poisoning attack vector:** The Algorithmic Risk Committee shall ensure that DI&MSF is not used as an attack vector for model poisoning or exposing intellectual property information.
- **Risk Management:** The organization shall deploy adequate measures to manage the risks associated with selection, management and oversight of Diverse and Multi Stakeholder Group (Human Impact Assessor (HIA)). This includes minimum threshold for selection of HIA, minimum expectations on outputs from HIA (including risk log creation align with risk inputs, risk indicators, adverse event references, risk evaluation and identified risk mitigations), audit compliance ensuring appropriate treatment of risks identified through the process and contractual compliance measures to address potential deviations.

- Template to collect these inputs - [+ Template for Risk inputs from DI&MSF](#)
- Need to document disclosure requirements
  - Disclosure requirement on the demographic or other characteristics of the diverse group for each system