FH AI Risk Taxonomy

Enterprise Risk (Business model / impact context)

This level is representative of the overarching strategic risk environment for an organisation. We would foresee risks being divided up between the following kinds of areas.

If this does not fit the executive risk profile of a target organisation, it can be amended as much as desired, but organisations need to ensure there is a mapping to operational risk types within each operational vertical - see the Risk Taxonomy below this.

As long as mappings are attached to new top level categories, it should not impact the taxonomy if changed.

1. Economic
2. Political
3. Social
4. Technology
5. Legal & Regulatory
6. Environmental
7. People
8. Third party

AI, algorithmic and autonomous systems (AAA) Risk Taxonomy

This level is to enable connection into the top level Enterprise Risk Management strategic risk landscape. The risk categories listed below are typical for larger companies. They are likely the same across all operational areas. These are scaled based upon metrics for more specific contributory risk categories under management.

This contributes context for operational risk management, but we would recommend, for the purpose of human-centric risk management those risk categories are either linked to preferred AI principles, or replaced by AI principles to roll up to a more meaningful measure of socio-technical AI related risk to pass up for reporting to the board.

Traditional Risk Categories	**AI Principles**	**FH AI Ethics Principles**
These are typical organisational risk categories, which cannot cater for the full socio-technical risk picture. These are likely static across specialist operational risk areas e.g. Strategic / Financial IT risk,	One generally well accepted list of ethics-focused principles is included below. It can be mapped to AI control capabilities / domains and impact types. Then to criteria required to effectively manage	Principles that are applied within FH for minimizing downside risks to humans. Augments to the commonly accepted AI Ethics principles like HLEG

Strategic / Financial Marketing risk. Categories may usefully map to control domains/systems, but careful consideration should be given to moving towards a more human-centric list Requires evolution to incorporate risk categories associated with impacts to individuals, society, and the environment as opposed to typical historical focus: the organisation	socio-technical systems risk. As noted above, we would recommend that these are mapped to or replace the more traditional risk categories. OECD and UNICEF are just two other organisations that have outlined such principles. Most have at least some reference back to the EU AI High Level Expert Group list below.	
1. Strategic 2. Financial 3. Reputational 4. Operational 5. ESG 6. Business **Risk universe in AI, algorithm and/ or autonomous systems that has a potential to impact people, people groups, society, and environment.**	• Human agency and oversight • Technical robustness and safety • Privacy and data governance • Transparency • Diversity, Non-discrimination, and fairness • Societal and environmental well being • Accountability Source: High level expert group	● Human Centric ● Ethical ● Fair ● Actionable ● Operational ● Accountable ● Auditable ● Certain ● Transparent
AI Risk Categories (socio-technical) Replicating structure in standards such as ISO27002. Identifying socio-technical risk categories that permit useful grouping of new AI related risks to individuals, society, and the environment. In addition, these risk categories support in creating appropriate control capability that were required, but frequently missing from pre-existing risk models. For instance, diversity and accessibility.		

These sit at a layer above the typical control capabilities for specialist disciplines such as security or privacy. This is recognising that AI governance and risk management brings together many such pre-existing expertise, each with their own subsets of specialized skills and controls.

Control capabilities represent reasonably independent subsets of control that have already been identified as necessary to avoid exposure to one or more adverse outcomes that might result from use of AI, ML, Autonomous systems.

An iterative process of feedback will enable change to the list, where risks or incidents do not usefully map to a pre-existing control domain for mitigation, or criteria do not usefully map up to control domains on a one to one, one to many, or many to many basis for reporting.

Level 1 (Risk Categories)	Level 2 (Activities/ measures)	Level 3 (root causes)
1. Privacy 2. Security 3. Safety 4. Bias 5. Governance 6. Ethics capability 7. Transparency 8. Explainability 9. Accountability 10. Accessibility 11. Diversity 12. Human agency 13. Sustainability	Accuracy Validity Reliability Robustness Resilience Interpretability Performance Ethics Assessment	Data Quality Information Quality Pipeline and Infra quality Model quality Policy or process Training and communication Data ethics

Risk Impact Types

- Environmental impact
  - Preserve, Conserve, Limit and Enhance/ enrich – Environment (Air, Forest, Water, Animal)
    - Global warming
    - Curtailing climate crisis response
    - Deforestation
    - Animal Extinction
    - Water, Soil and Air Quality
    - Overcrowding/ Gravitational pull of urban centers
    - Extreme weather
- Societal impact
  - Impact to democracy

    - Impact to rights and freedom
    - Impact on social policies
    - Impact on behaviours/ beliefs
- Impact to individuals/ groups (subset of larger scale)
    - Life impact
    - Physical, Mental and Psychological impact
    - Damage to reputation and/ or identity
    - Privacy exposure and associated harassment
    - Resource or Monetary or time loss
    - Limits to access or opportunity
    - Petty disturbance

Writing a good risk statement

[Adverse Outcome/s that has an effect on people / society / the environment] caused by [missing controls, insufficient control/s] compromised by [inside or outside threat actor/s, or harmful when operating as expected], that may result in [impacts/s]

How to enhance the risk taxonomy:

1. Link each adverse outcome to the control domains (preferably level 1), impact types, principles and risk categories. If there are any adverse outcomes that do not fit with any of the control domains or impact types or risk categories or vice versa, please do notify to enable updating the taxonomy.

Next steps

1. Create a table to populate with existing lists of risks or control failings, with drop down lists to add taxonomy labels.
2. To map each FH audit criteria to one or more of the risk categories/control domains. Potentially using the pre-existing FH pillars and just adding other applicable capabilities / domains. Noting exceptions where there is a missing or poor fit. Enabling different audiences to filter different ways
3. Attempt risk statement for each listed risk to link the control failing or potential adverse outcome to the contributory components. This will give us means to communicate any gaps.
4. When logging and labelling risks / adverse outcomes, or matching FH criteria to risk categories/control domains, note any that don’t fit well or have no fit and flag to potentially revised taxonomy.