AI Impact Identification Framework

Type: Framework

Target Audience: AI Ethics Teams, Risk Managers, Product Owners

This framework provides a comprehensive categorization system for identifying potential AI harms across multiple domains. Use during AI impact assessments to ensure all harm categories are considered.

Assessment Information

Field	Details
AI System Name	
Assessment Date	
Assessor(s)	
Use Case Description	

1. Individual Impacts

Harms that affect specific persons directly.

Impact Category	Examples	Present?	Severity
Rights Violations	Privacy invasion, denial of due process, discrimination	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Economic Harm	Job loss, credit denial, unfair pricing, financial fraud	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Physical Harm	Injury from autonomous systems, medical misdiagnosis	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Psychological Harm	Manipulation, addiction, anxiety, loss of autonomy	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Reputational Harm	False accusations, defamation, deepfake victimization	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Loss of Opportunity	Unfair hiring decisions, educational gatekeeping	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C

2. Group Impacts

Harms that affect communities or demographic groups.

Impact Category	Examples	Present?	Severity

Discrimination	Disparate impact on protected groups, proxy discrimination	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Stereotype Reinforcement	Biased representations, harmful associations	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Cultural Harm	Erasure, misrepresentation, appropriation	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Digital Divide	Exclusion of groups lacking technology access	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Collective Surveillance	Targeting of communities, chilling effects	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C

3. Societal Impacts

Harms that affect society at large.

Impact Category	Examples	Present?	Severity
Democracy Threats	Election manipulation, misinformation at scale, erosion of trust	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Social Cohesion	Polarization, filter bubbles, breakdown of shared reality	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Economic Inequality	Wealth concentration, labor displacement, market manipulation	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Power Concentration	Monopolistic control, surveillance capitalism	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Security Risks	Autonomous weapons, critical infrastructure vulnerabilities	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C

4. Environmental Impacts

Harms to the natural environment.

Impact Category	Examples	Present?	Severity
Energy Consumption	Training compute, inference at scale, data center power	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Carbon Footprint	CO2 emissions from compute and cooling	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
Water Usage	Data center cooling water consumption	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C
E-Waste	Hardware obsolescence, disposal of AI chips	[ ]Y [ ]N	[ ]L [ ]M [ ]H [ ]C

5. Risk Scoring Matrix

Calculate overall risk score using Likelihood × Severity for each identified impact.

Likelihood Scale

Score	Likelihood	Description
1	Rare	Unlikely to occur (<10% probability)
2	Unlikely	Could occur occasionally (10-30%)
3	Possible	Might occur sometimes (30-60%)
4	Likely	Will probably occur (60-90%)
5	Almost Certain	Expected to occur (>90%)

Severity Scale

Score	Severity	Description
1	Negligible	Minor inconvenience, easily remedied
2	Minor	Limited harm, short-term impact
3	Moderate	Significant harm, requires intervention
4	Major	Serious harm, difficult to reverse
5	Catastrophic	Severe/irreversible harm, rights violation

Risk Matrix (Likelihood × Severity)

	Sev 1	Sev 2	Sev 3	Sev 4	Sev 5
Likelihood 5	5	10	15	20	25
Likelihood 4	4	8	12	16	20
Likelihood 3	3	6	9	12	15
Likelihood 2	2	4	6	8	10
Likelihood 1	1	2	3	4	5

- **1-4 (Green):** Low risk - Monitor | **5-9 (Yellow):** Medium risk - Mitigate
- **10-14 (Orange):** High risk - Significant controls | **15-25 (Red):** Critical - Avoid/transform