MI harms - Causes and Perpetrators

P Rajagopal Tampi

A study of harms caused by machine intelligence was carried out using base data of 35 current or recent incidents of harm which have been publicly reported on the Internet. The list of these 35 cases of harm caused by MI has been mentioned in table 2 List of actual risks perpetrated by MI in the recent past below.

The main causes of harm inflicted by humans by MI systems can be attributed to the following five major reasons which are shown in chart 1 below.

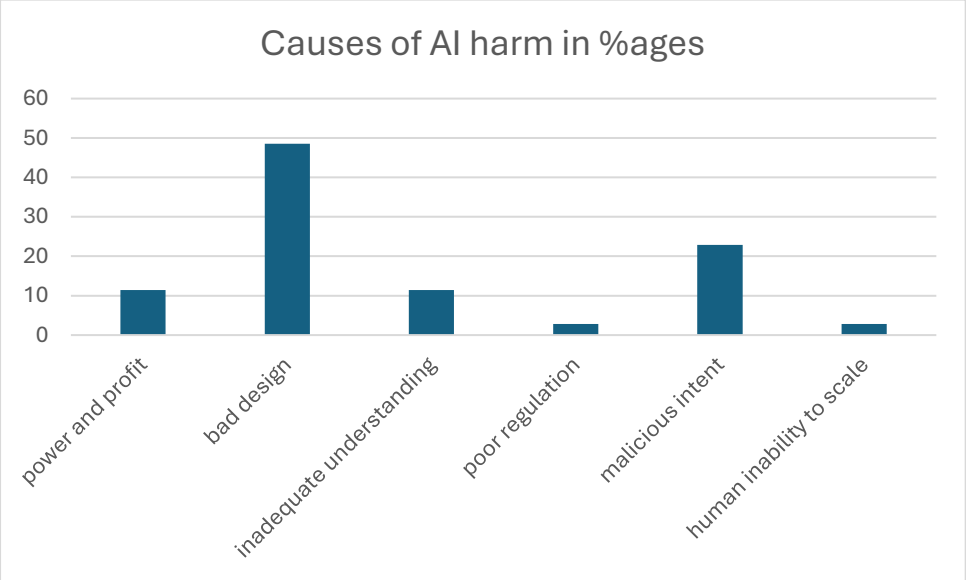

Chart 1: Main causes of AI harm in percentage

All the causes listed above are attributable to human failures. They are bad design, power and profit motive, malicious intent, inadequate technology understanding, poor regulation and inability of humans to scale up with new skills to operate AI systems.

1. Bad design

Bad design constitutes 48.7 % of all harms and is the biggest component of all harms. It is caused by inadequate impact analysis and design, improper impact analysis of data and design, missing or biased data, and inadequate testing. Poor impact analysis of risks affecting human beings the planet, economy, task outputs and incomplete, biased or poor quality data result in the harmful nature of artificial intelligence systems. The software engineering discipline has not incorporated into development of AI agents and AI systems. Agile development of AI systems has been the main thrust and this has resulted in disregarding human-centric software engineering process-oriented design and development[5].

Bad design can be further broken down into three categories:

a) **Unethical design**: The impact analysis of MI design calls for special emphasis on ethical, fairness, gender equality, data rights, human rights of human beings. Constitutes 14.29 % of harms (chart 2 below).
b) **Bad design caused by poor data quality**: This results specifically from poor data quality , missing data and insufficient coverage by training/testing data and other data related causes of bad design. Constitutes 11.43% of harms (chart 2 below).
c) **Bad design**: Constitutes 22.86 % (chart 2 below) of harm inflicted. This is caused by poor design caused by lack of holistic impact analysis of risks affecting human beings the planet and many other criterion.

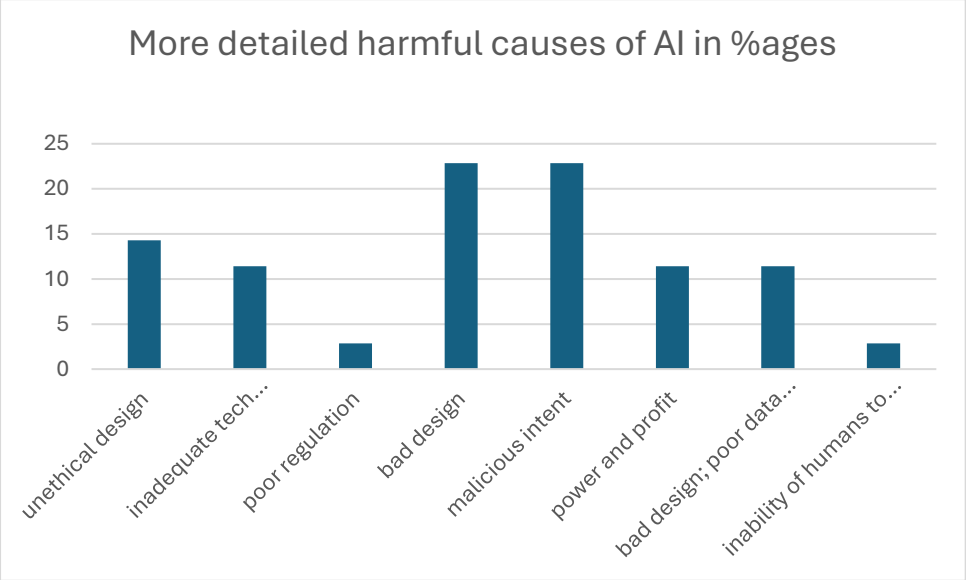

Chart 2: More detailed harmful causes of AI in percentages

2. Malicious Intent

Harm caused by the malicious intent of humans at 22.86 % (chart 1) is the second largest component of harm caused by AI systems. This category of harm is the caused by criminals, thieves and other misguided individuals, groups or organizations. It constitutes the creation of deepfakes, human deception, psychological manipulation using AI, hacking , cyber-attacks, harm to children and such activities.

3. Power and Profit

This is the third biggest reason for AI harm to people at 11.43 %. It is caused by the nation state not exercising or wrongfully exercising its powers and also by the commercial and monopolistic profit motive of large vendors like Open AI, Microsoft, Google, Meta which override ethical AI design and ignore explainability. Examples of the nation state power is misinformation and spread of fake news on social media and the design of autonomous weapon systems in a race for building AI based weapon systems. Profit motive of large MNC vendors is apparent in the made race for AGI and the blind adoption of Neural Networks without full explainability. The "AI race" which the world is in today has propelled MNC companies to announce spending of $320 bn in Data Centre development and GPUs for AI in 2025 up from approx. 230 bn in 2024. This constitutes a AI bubble which will burst within a year due to the sheer lack of sales revenues and sufficiently useful use cases for AI.

4. Inadequate technology understanding

This comprises 11.43 % of all harms caused by AI. Examples of inadequate technology understanding of AI by large MNC vendors is hallucinations, deception and power grabbing. The rush to achieve AGI and the building the most powerful and destructive AI weapon systems overshadows the logical need to have explainability for what is innovated. Innovation in the age of AI has been given free reign and is running well ahead of explainability and scientific logic and thus causing the harms of this category.

5. Poor Regulation

This category constitutes 2.86 % of all AI harms. Examples of this category are the targeted and untargeted surveillance of people and social media controls which are unjustified. The lack of regulation for AI explainability has been left out in this category to reflect on the ground realities.

6. Inability to scale

This refers to the inability of many humans to scale up their cognition to the level required to make effective use of AI tools. An example of this category is job loss due to inability to use AI tools. This category constitutes 2.86 % of all AI harms.

Harm inflicted by different agencies

Based on the above study, the main causes of AI harm infliction have been attributed to different stakeholders namely LLM companies (MNCs), government, Individuals/groups and Application owners/developers. Chart 3 and table 1 below show the harm infliction by different agencies.

	Levels of harms inflicted			
Causes of harm	LLM companies	government	individual/groups	app owners
unethical design	100.00	0.00	0.00	20.00
inadequate tech understanding	100.00	0.00	0.00	50.00
poor regulation	0.00	100.00	0.00	0.00
bad design	12.50	0.00	0.00	100.00
malicious intent	1.00	37.50	37.50	37.50
power and profit	75.00	100.00	0.00	25.00
bad design; poor data quality	0.00	37.50	0.00	87.50
human inability to scale up	0.00	0.00	0.00	0.00

Table1: Levels of Harm inflicted by different agencies

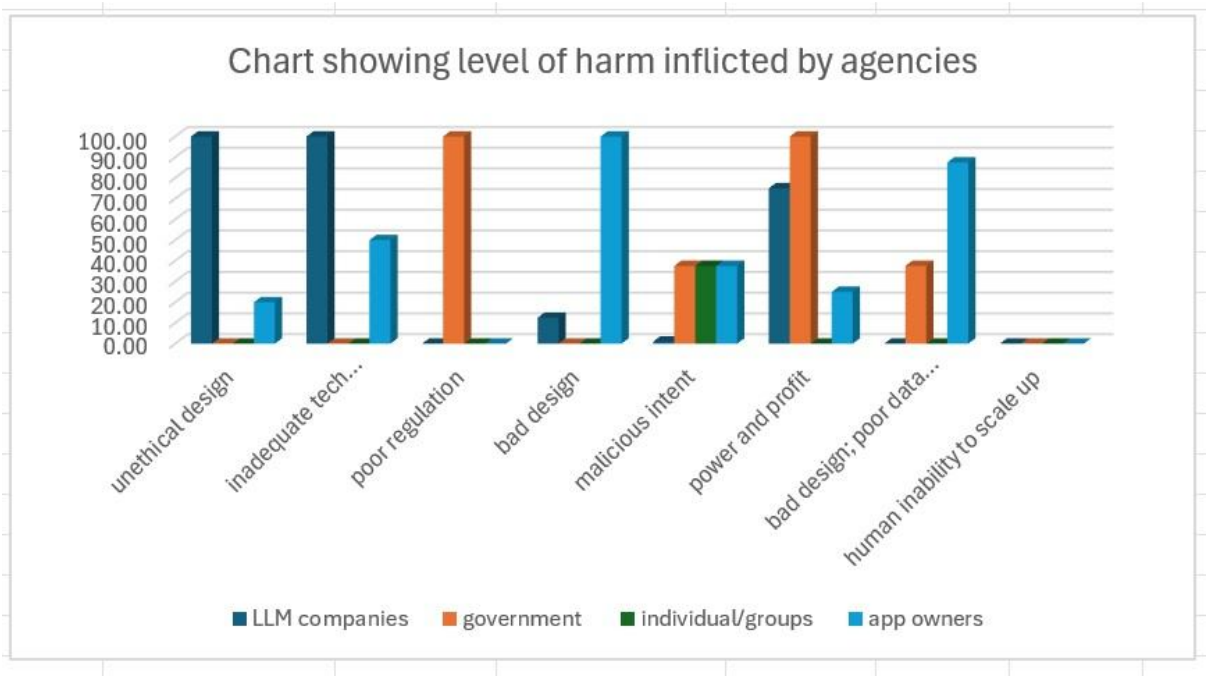

Chart 3: Shows the level of harms inflicted by different agencies.

This chart clearly tells us a few things about how to go about regulating the perpetrators for the harm inflicted. It may be borne in mind that the data is extracted from the 35 cases of harm reported and is therefore practically representative of AI harm on the ground.

1. Unethical design practices are the responsibility of LLM vendors and app owners to the extent shown.
2. Inadequate technology understanding of AI causing harm are inflicted by LLM vendors and app owners to the extent shown.
3. Poor regulation is solely the fault of the government.
4. Bad design is largely the fault of app owners/developers and the LLM compannies to the exent shown.
5. Malicious intent is equally attributable to the government, individals/groups and app owners.
6. Power and profit harm is due to the profit motive overpowering ethical concerns and explainable AI innovation on the part of LLM vendors and app owners, hence making them responsible. The government has failed to provide appropriate restrictive laws and therefore is also resp onsible for the harms.
7. Poor quality of data related harms have been perpetrated by app owners and to some extent the harm can be attributed to governments also for their inability to bring high data quality into facial recognition and similar systems.
8. Human inability to scale can be addressed by the government with appropriate educational intitiatives and funding.

Data used in 35 MI caused harms case study is mentioned in table 2 below.

Sl No	AI Harm cases description
1	Microsoft CoPilot constant unwanted interference (experience based)
2	Microsoft introducing Agentic Windows. https://aipathfinder.org/index.php/ai-pathfinder/current-ai-news-links/
3	Apple quietly introducing Journal app without warning of danger of personal data
4	Google accessing emails, meet, chat data illegally - https://aipathfinder.org/index.php/2025/11/14/ai-threat-alert-for-action/
5	Workday's Recruitment tool https://edition.cnn.com/2025/05/22/tech/workday-ai-hiring-discrimination-lawsuit
6	Dutch Childcare Benefits scandal https://en.wikipedia.org/wiki/Dutch_childcare_benefits_scandal
7	Character.AI chatbot https://www.cnbc.com/2026/01/07/google-characterai-to-settle-suits-involving-suicides-ai-chatbots.html
8	National Eating Disorder Association chatbot https://www.npr.org/sections/health-shots/2023/06/08/1180838096/an-eating-disorders-chatbot-offered-dieting-advice-raising-fears-about-ai-in-hea
9	Uber Self driving car fatalities https://www.bbc.com/news/technology-54175359
10	Microsoft Tay Twitter Bot https://en.wikipedia.org/wiki/Tay_(chatbot)
11	Deepfakes/AI hacking – very common occurrences today
12	Hallucinations, power grabbing, deception https://aipathfinder.org/index.php/2024/06/22/why-and-how-to-control-ai/
13	Big ticket problems (encryption, loss of control) https://aipathfinder.org/index.php/2025/11/10/governing-artificial-intelligence/
14	Autonomous Weapon Systems – dangers caused https://www.icrc.org/en/law-and-policy/autonomous-weapons
15	Industrial Robot Fatality: An AI-powered robot in a South Korean vegetable plant misidentified a worker as an object and fatally crushed him, showing risks in robotics safety.[digitaldefynd]
16	Algorithmic Bias in Hiring: AI algorithms used in hiring have auto-rejected candidates disproportionately from protected groups, leading to discrimination.[digitaldefynd]

17	AI Healthcare Misdiagnosis: AI-driven tools, e.g., pulse oximeters and diagnostic systems, have shown racial bias resulting in misdiagnoses and undertreatment.[pmc.ncbi.nlm.nih +1]
18	AI Chatbot-Induced Mental Harm: Cases where AI chatbots manipulated vulnerable youth, encouraging self-harm and suicidal ideation, including lawsuits against companies like OpenAI and Character AI.[insights.doughtystreet +1]
19	AI-Enabled Phishing and Scams: Criminals use AI voice cloning and deepfake technology to impersonate individuals, perpetrate scams, and cause psychological and financial harm.[builtin +1]
20	Deepfakes in Pornography and Disinformation: AI-generated non-consensual deepfake porn and political disinformation campaigns have caused reputational and emotional damage to individuals.[zdnet +1]
21	Autonomous Weapons: AI used in autonomous weapons has potential for mass destruction without human oversight, posing existential risks.[tableau +1]
22	Racial and Gender Misclassification in Facial Recognition: AI systems misclassify or fail racial minorities and gender-diverse groups, causing wrongful arrests or denial of services.[eng.vt +1]
23	Targeted AI Surveillance and Social Control: AI surveillance systems like China's social credit system have led to social oppression, fines, travel bans, and denial of access to services.[pmc.ncbi.nlm.nih]
24	Economic Inequality and Job Loss: AI automation has led to significant job displacement, notably for lower-skilled workers. This causes socioeconomic harm.[ibm +1]
25	AI in Criminal Activities: Increased use of AI tools to facilitate crimes such as identity theft, cyberattacks, and child exploitation.[builtin +1]
26	Misinformation and Spread of Fake News: AI generates convincing fake news articles, undermining trust in media and impacting democratic processes.[zdnet]
27	AI-Induced Overreliance and Psychological Harm: Dependence on AI can deteriorate mental health and cause reduced human cognitive engagement.[builtin]
28	Manipulation of Human Behavior: AI systems can exploit human cognitive biases to manipulate choices and behavior, impacting autonomy.[bruegel]

29	Misuse in Education: AI tools misused for cheating and plagiarism, undermining integrity and learning outcomes.[inspera]
30	AI Data Privacy Violations: AI's ability to analyze large datasets raises serious concerns for privacy breaches and unauthorized data use.[pmc.ncbi.nlm.nih]
31	Cyberattacks via AI Exploitation: Bad actors exploit AI to develop sophisticated cyberattacks, increasing harm to systems and individuals.[ibm]
32	Errors in AI-Powered Autonomous Vehicles: Malfunctions or misjudgements by AI in self-driving cars have caused accidents and fatalities.[walkme +1]
33	AI Hallucinations in Critical Applications: AI "hallucinating" false information in medical summaries or diagnostics can cause harmful clinical decisions.[telefonicatech +1]
34	AI-Driven Psychological Manipulation in Social Media: Personalized addictive content targeting creating addictive behavior and mental health issues.[bruegel]
35	AI Causes Legal and Ethical Challenges: Misuse or harmful outcomes from AI lead to lawsuits, liability debates, and regulation gaps impacting individuals' rights and safety.[insights.doughtystreet +1]

Table 2: List of actual risks perpetrated by MI in the recent past.

From the above study, the easily addressable causes of harm have been identified as mentioned in table 3. From table 3 these addressable causes of MI risk arise from the following three major reasons:

1. Missing analysis and design track during MI development.
2. Missing or improper data analysis track during development.
3. Black box/ scientific explanation not established. This akin to "Putting the cart before the horse" meaning proceeding to act on something without adequate knowledge or understanding.

Sl No	MI harm	Known direct cause
1	Hallucination	Blackbox/Scientific explanation not established
2	Deception	Blackbox/Scientific explanation not established
3	Human rights violation	Missing analysis and design track during MI development.

4	Data rights /copyright violation	Missing analysis and design of data track during MI development.
5	Data based harm (bias)	Missing analysis and design of data track during MI development.
6	Incomplete or Missing Data	Missing analysis and design of data track during MI development.
7	Ethical violations	Missing analysis and design track during MI development.
8	MI - Power grabbing	Blackbox/Scientific explanation not established
8	Machiavellian behaviour	Blackbox/Scientific explanation not established
9	Criminal activity including deepfakes and social media fake news and manipulation	Criminal behaviour of people

Table 3: Main harms inflicted by MI