Reducing Harmful AI Development through better Analysis and Design

Rajagopal Tampi

Unethical and enshittification technology practices have been on the rise ever since November 2022 when Chat GPT debuted. The rate of growth of such undesirable practices has increased ever since all the data on the internet has been consumed for training LLMs. It seems that Open AI, Microsoft, Google and Meta see themselves in an existential AI race triggering all sorts of undesirable and illegal practices. These practices are being forced upon users sometimes blatantly and at other times surreptitiously much against their will and without their consent.

Examples of Harmful AI

Case 1: Microsoft's Copilot is the epitome of intrusiveness into user space. Popping up and trying to insert itself when we do not want it at all. Sometimes it jumps in anyway and reads everything we write when we do not want it to. If that's not intrusion and undesirable behaviour what is? There are indeed times when I would want to use Copilot but that's my prerogative to decide when and not Microsoft's.

Case 2: Pavan Davuluri, the Head of Microsoft's Windows announced on X that Windows is evolving into an Agentic OS. This meant that it would perform the role of a user assistant performing tasks for users in addition to running apps. The outrage on the internet was immediate and universal essentially saying "We don't want this". A new term has been coined for this sort of technology company behaviour "enshittified" meaning something filled with unwanted changes and advertisements.

Case 3: A few years ago, Apple slipped in the "Journal" app into iPhone without a specific alert or information about how the intimate personal data captured in the app will be used. We enter our most confidential information in our journals. It is not something most users would like to share due to privacy concerns. It was ethically incumbent on Apple to specifically highlight the default sharing of captured data in the case of Journal app. Unsuspecting and gullible users have surely used the app and exposed their most personal data to significant risk and at their own peril.

Case 4: In a surreptitious move to gather private data using Gemini AI Assistant without user consent, Google accessed the emails, chats and google meet content of millions of its users including the author's by using "ON" as the default choice of AI settings during a major upgrade of its Gemini AI tools installation in October 2025. **Google's wants to capture as much data legally or illegally to train AI models even at the cost of modelling its own customers without their consent or knowledge.** A major lawsuit has recently been filed in the US District Court for Northern District of California (San Jose) for this transgression against Google.

Case 5: Workday's AI based Recruitment product rejected all applicants over 40 years of age. Derek Mobley applied for over 100 jobs through the Workday system and was rejected within

minutes each time. The case achieved nationwide class action certification in the US in May 2025.

Case 6: The Dutch Childcare Benefits Scandal (Toeslagenaffaire) was caused by an AI algorithm used by Dutch tax authorities which wrongly flagged thousands of families for fraud related to childcare benefits based on biased criteria like dual nationality and low income. This resulted in families being forced to repay large sums of money they did not owe, causing severe financial and emotional distress. Over 20,000 families were harmed, with more than 1,000 children placed in foster care due to these wrongful accusations.

Case 7: Another particularly serious case that resulted in the suicide of a minor was led by a Character.AI chatbot. The plaintiff claimed the vendor knowingly used highly toxic datasets for training, that the chatbot manipulated vulnerable users emotionally, and that they intentionally allowed minors to use the chatbot without adequate warnings or protections. The vendor was accused of breaching their duty to warn users of inherent dangers, contributing to a tragic outcome.

Analysis of AI Harm

Large MNCs (LLM vendors) are accountable for the lapses in first 4 cases above. These are clearly cases where the companies are at fault and have not played by acceptable standards of ethical conduct impacting their own users and customers. They have prioritised innovation and growth over more important human risks like human rights violations and data privacy.

In the 5th case, Workday took a shortcut in analysis and design and will pay the price. In the 6th and 7th cases above the fault is again with the Dutch tax authorities and Character.AI respectively for bad analysis and design of their AI based application.

The above cases and another additional 7 harmful AI cases (details in Annexure) were analysed for cause and the agency that was responsible for the lapse. The pivot chart on the agency responsible is shown in Fig 1 below. The pivot chart on the causes is shown in Fig 2 below.

Agency responsible for AI harm	No of cases	%age
Companies	9	64
Individuals/groups	1	7
Govt/Rules lapses	2	14
Organizations - AI app owners	2	14
Total	14	100

Fig 1: Agencies causing AI errors

Root cause of AI harm	Numbers	%age
Analysis and design lacking	5	36
Deliberate human decision	5	36
Lack of technology understanding	1	7
Poor regulation	1	7
Power and profit motive	2	14
Total	14	100

Fig 2: Causes of AI errors

Take aways

Leading or forcing user behaviour and new functionality adoption without alerting users of the dangers involved and without their consent is condemnable behaviour by technology companies which should be penalized heavily. It is not enough to bury technical details of new

functionalities in legalese within Terms and conditions, companies should be mandated to publish new feature explanations and impacts in non-technical language for the easy understanding of users.

They should provide ways of opting in for new AI features rather than opting out. This means that all new AI features must be by default switched off. This is simple and can be enforced through regulation giving a level of user protection that also encourages users to become more knowledgeable about AI use and themselves get used to making choices for their own good. This will help develop learning skills in people which are essential in the age of AI. Users also have to scale up in their knowledge and cognitive capabilities to use AI tools in an optimal manner.

From the above analysis, Analysis and Design has been a consistent weakness in most of the cases analysed and constitutes 36 percent of the cases of AI harm inflicted (Fig 2 above). This can easily be rectified if the software engineering practices in my book "Applied Human-Centric AI" are adopted during analysis, design and development of AI systems. It may be hard to do so but is the only way to design human-centric AI. Adoption of the processes suggested in my book will result in tools being developed to automate the processes.

Conclusion

The short-cutting of the analysis and design stages in the development of AI systems, the need for "agility" of development in vogue today are responsible for a major part of the harms caused by AI to people. This must be rectified by introducing software engineering processes in the AI project life-cycle as explained above.

Technology is meant to complement the individuals' efforts when the person chooses to use it. It is not meant to lead the way individuals think and act. The moment we allow AI take over the role of leading our thoughts and actions, we lose control of AI in a significant way. That is surely a recipe for disaster for us and not for AI.Download pdf copy

Disclaimer: The opinions expressed in this article are personal opinions and futuristic thoughts of the author. No comment or opinion expressed in this article is made with any intent to discredit, malign, cause damage, loss to or criticize or in any other way disadvantage any person, company, governments or global and regional agencies.

Author

Annexure

AI harms to humans – other cases considered in the study

1. National Eating Disorder Association chatbot - https://www.evidentlyai.com/blog/ai-failures-examples
2. Uber Self driving car fatalities - https://research.aimultiple.com/ai-ethics/
3. Microsoft Tay Twitter Bot - https://en.wikipedia.org/wiki/Tay_(chatbot)
4. Deepfakes https://www.amazon.in/dp/B0CW1D7B88
5. Hallucinations, power grabbing, deception https://aipathfinder.org/index.php/2025/11/10/governing-artificial-intelligence/
6. Big ticket problems (encryption, loss of control) https://aipathfinder.org/index.php/2025/11/10/governing-artificial-intelligence/
7. Autonomous weapons systems - https://www.amazon.in/dp/B0CW1D7B88

