Responsible and Ethical AI development

10892938, 10864307, 10842878, 10882588

Introduction

AI is fast becoming integral to our everyday life. With each passing day, new technologies are developing that require the use of AI and thus increase its importance (Schraffenberger et al. 2019). With the increasing implementation of AI, it may also be necessary to increase the ethical oversight in its implementations due to the dangers that may be involved(Vandberg and Mott 2023). The responsible use of AI is an attempt to create an ethical framework governing the development, implementation and oversight of Artificially Intelligent systems(Liu 2024).

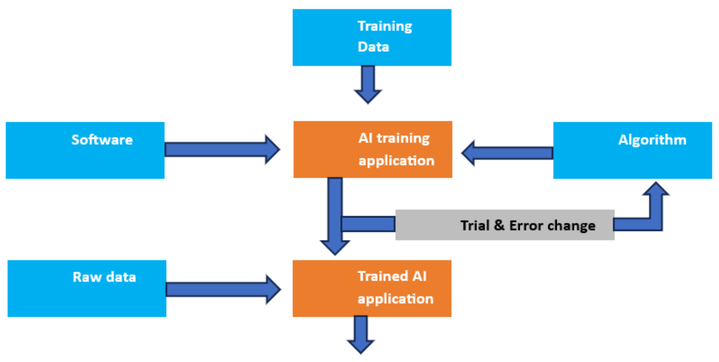


Figure 1: Artificial Intelligence Components.(Jones Day 2022)

Dangers Of AI

Long Tail Situations: i.e AI is not good at adapting to situations that have lots of special cases. Examples of this may include autonomous self driving. AI functions that require massive amounts of data in order to deal with special cases that occur rarely in the data to prevent misgeneralisation (Langasco et al. 2022).

Discriminatory Practices i.e AI follows statistically correct stereotypes, which may cause group discrimination.

Uncontrollable self-aware AI i.e. AI that gains too much autonomy to the point where it becomes impossible to shut down.

Lack of data privacy using AI tools.

What is Responsible AI Development

The responsible use of AI according to the Model AI governance framework(World Economic Forum 2020) is:

1. Internal governance structures and measures.
2. Determining the level of human involvement and decision making.
3. Operations management
4. Stakeholder interaction and communication.

Internal Governance Structures and Measures

Internal governance structures tries to integrate AI danger considerations into the risk management system of the organization. Creating a decentralized structure to manage such risks may be preferable to using a centralized structure(World Economic Forum 2020).

Determining the level of Human Involvement and Decision Making

The organisation involved could try to predict the probability and severity of harm associated with implementation of AI use.

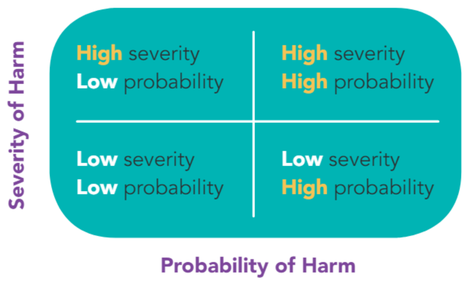


Figure 2: AI risk quadrant. (World Economic Forum 2020)

As the type of decision moves from the bottom right to the top left, the amount of executive oversight over Artificial intelligence applications should also increase(World Economic Forum 2020).

Operations management

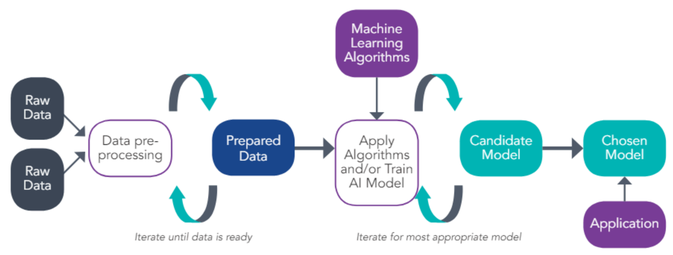


Figure 3: AI development Cycle.(World Economic Forum 2020)

This framework aims to mitigate the problems that come about in AI systems due to errors in the gathering of data, choice of machine learning algorithm or the application of a model or any other factor in the AI development cycle by exercising more judicious management of the choice of each operational element(World Economic Forum 2020).

Stakeholder Interaction and Communication

This framework aims to mitigate the cost of AI that comes about due to a lack of transparency. Organisations should disclose whether or not AI is used in their services, how such AI use may affect their customers and third parties, and should try to avoid decisions that may erode the trust of these stakeholders for the organisations in question(World Economic Forum 2020).

Objectives of Responsible and Ethical AI Development(Gillis 2023)

Fairness: Datasets used for training the AI system must be given careful consideration to avoid discrimination.

Transparency: AI systems should be designed in a way that allows users to understand how the algorithms work.

Non-maleficence: AI systems should avoid harming individuals, society, or the environment.

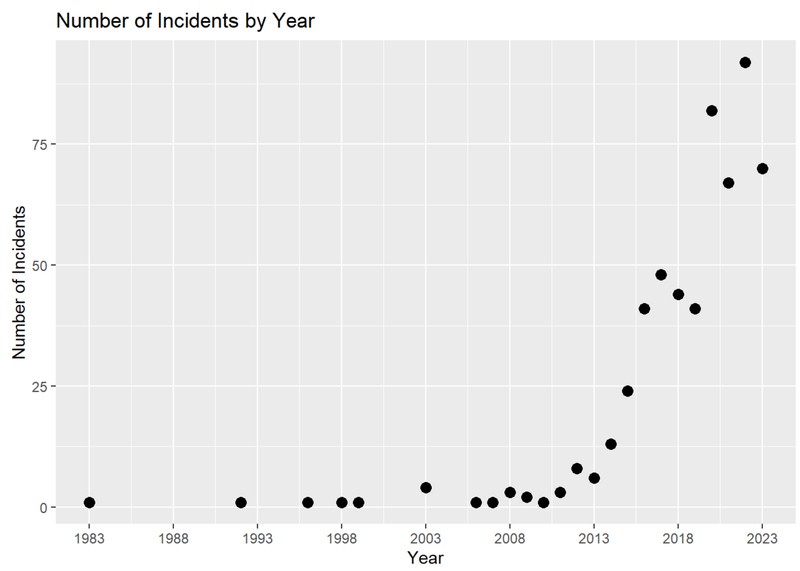
Responsibility: Developers, organizations, and policymakers must ensure AI is developed and used responsibly.

Privacy: AI must protect people's personal data, which involves developing mechanisms for individuals to control how their data is collected and used.

Inclusiveness: Engaging with diverse perspectives helps identify potential ethical concerns of AI and ensures a collective effort to address them.

Analysis of the Importance of AI Ethics(Bhalerao 2023)

The graph below show the frequency of AI incidents which may justify a greater awareness of its risks from 1983 to 2023.



The result of the above plot shows a growing frequency of incidents over time, creating a need for a greater awareness of AI and its risks.

Conclusion

It is important that before developing any AI technology, it should be human-centered. This thinking should be the ideal method for producing responsible AI technology. It will ensure that the future of AI remains hopeful from century to century.

The dangers associated with artificial intelligence should always be discussed, so leaders and stakeholders can continuously decide ways to use the technology for noble purposes.

References

Bhalerao, Nikhil. 2023. "AI Incidents." Kaggle. https://doi.org/10.34740/KAGGLE/DSV/6655184.
Gillis, Alwxander. 2023. "A Guide to Artificial Intelligence in the Enterprise." *EnterpriseAI*. https://doi.org/DOI Number.
Jones Day. 2022. "Rising Global Regulation for Artificial Intelligence." *One Firm Worldwide*.
Langasco, Jack, Lauroamd Koch, Lee Sharkey, Jacob Pfau, Laurent Orseau, and David Kruegar. 2022. "Goal Misgeneralization in Deep Reinforcement Learning." *International Conference on Machine Learning*. https://proceedings.mlr.press/v162/langosco22a/langosco22a.pdf.
Liu, Kai. 2024. "Artificial Intelligence and Ethical Frameworks in Pediatrics." *JAMA Pediatrics*.
Schraffenberger, Hanna, Yana Sande, Gabi Schaap, and Tibor Bosse. 2019. "Investigating People's Attitudes Towards AI with a Smart Photo Booth." *BNAIC/BENELEARN*. https://web.archive.org/web/20220420201132id_/http://ceur-ws.org/Vol-2491/abstract113.pdf.
Vandberg, Jessica, and Bradfort Mott. 2023. "'AI Teaches Itself': Exploring Young Learners' Perspectives on Artificial Intelligence for Instrument Development." *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education* 1: 485–90. https://doi.org/10.1145.
World Economic Forum. 2020. *Model Artificial Intelligence Governance Framework*. Davos, Switzerland.: World Economic Forum Annual Meeting. https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf.