Ethics in AI Assignment

Discuss the importance of developing and deploying transparent AI systems with reference to the reading by Mittelstadt et.al. (2016) on “The Ethics of Algorithms”.

Over the last 20 years, Artificial intelligence has transformed from a novel idea to an essential part of the modern digital world, AI technologies are shaping critical decisions in fields like criminal justice,healthcare,hiring and finance. While these systems offer efficiency, scalability their lack of transparency(opacity) in how the particular decision was made remains an ethical challenge. It remains a crucial question to answer who should be held accountable for the harm caused by the AI person who deployed the system or AI. Philosophers like Immanuel Kant and John Rawls have proposed strong frameworks for effectively addressing these issues:Kant focuses on individual autonomy and the right to make free, independent choices without unwanted influence, whereas Rawls stresses fairness, equal opportunities for the marginalized communities. Together ,their viewpoints stress that transparency in the AI systems is crucial for responsible, ethical use. (Kant, 1785; Rawls, 1971).

It is the most common perception that AI systems are objective and neutral. However, they can unintentionally introduce, magnify, or reinforce bias if not carefully administered. The root cause of bias in AI systems lies in the data on which they are trained data that may carry implicit cultural, social, or historical biases. In Coded Bias, MIT researcher Joy Buolamwini discovered that facial recognition systems failed to detect her large dark skinned face unless and until she wore a white mask.This indicates how AI systems can reinforce bias related to culture, society and history, reflecting the values and assumptions of creators.(Kantayya, 2020). Big tech companies like IBM,Microsoft have been criticized for similar issues in their facial recognition systems, perpetuating bias based on racial discrimination, gender and ,stereotypes. Transparency in the AI systems is very crucial to ensure that AI systems are being implemented responsibly and ethically. But, in modern days most of the AI systems are ‘black box’ by nature, which creates difficulty in understanding how a particular decision is made by it, complicating accountability during post-deployment. This marks the crucial urge to make AI systems transparnt in their working in development as well deployment stage.Considering the case of COMPASS , which is a widely used tool in US justice system to predict the rate of reoffending of criminal person , was biased against black defendants and assigned them higher sentencing scores as compared to white , producing scores even judges could not comprehend highlighting the urgency of transparency in life threatening cases.

Similar concerns exist in employment, healthcare, and finance, where AI systems are trained on biased historical data. For example, Amazon's AI-powered hiring tool discriminated against female applicants because it was developed using male-dominated data (Mittelstadt et al., 2016). Without transparency, the system reinforced gender bias for years. Transparent systems, in contrast, make it easier to audit, identify, and correct such flaws early on. Privacy and identity concerns are closely tied to transparency. Mittelstadt et al. (2016) explain how algorithms categorize people into behavioral profiles without their consent, which are then used for decisions like dynamic pricing or content filtering. Luciano Floridi's theory of informational privacy argues that individuals must control how their data is used (Floridi, 2011). Transparency in AI development ensures ethical data handling, while transparency during deployment ensures that individuals can consent, contest, or opt out of manipulative profiling. Transparency is essential for supporting individual autonomy. If AI systems make opaque decisions whether in healthcare, sentencing, or loan approval people cannot exercise meaningful agency. For instance, an AI system recommending treatment in healthcare must clearly explain its rationale so patients can give informed consent. As Kant argues, true autonomy requires understanding, which only transparency can ensure. Furthermore, transparency fosters public trust in AI technologies.Being open about how these systems work helps build confidence and allows others to check for mistakes or unfairness. Transparent systems, developed with explainability and deployed with oversight, encourage external review, ethical auditing, and peer correction. In diagnostic AI, transparency boosts clinician adoption and leads to better healthcare outcomes. Importantly, trust is earned not just during development but through responsible deployment practices that demonstrate accountability.

Developing and deploying transparent ai systems is essential to ensure fairness ,accountability,public trust.AI systems should be fair and unbiased, avoid discrimination, and ensure equal treatment to everyone and this can be achieved by intense testing of AI models in various domains and mitigate biases wherever identified, Also there should be a mechanism to decide who is responsible for for the outcomes produced by AI tools which ensures the appropriate solution to any harm caused by AI tools.To preserve the user's privacy by analyzing how an AI tool is handling the private data of the user and whether any sensitive information is safeguarded or not is important.There are seven essential requirements for an AI system to be trustworthy. These include human agency and oversight, technical robustness, transparency, diversity, privacy and data governance, non-discrimination, and environmental well-being(European Commission, 2019). These regulations are provided by the High-Level Expert Group on Artificial Intelligence which is established by the European Commission.

References:

- Floridi, L. (2011). *The philosophy of information*. Oxford University Press.

- Kant, I. (1785). *Groundwork for the metaphysics of morals* (H. J. Paton, Trans.). Harper & Row. (Original work published 1785)

- Kantayya, S. (Director). (2020). *Coded Bias* [Film]. 7th Empire Media.

- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 2053951716679679. https://doi.org/10.1177/2053951716679679

- Rawls, J. (1971). *A theory of justice*. Harvard University Press.

- European Commission. (2019). *Ethics guidelines for trustworthy AI*. High-Level Expert Group on Artificial Intelligence.

.