Ethics and Bias in AI

# Introduction

**(15 marks)**

## Overview

As artificial intelligence becomes increasingly integral to various fields, its widespread adoption raises pressing ethical concern, notably bias. This paper looks at the importance of ethical impact assessments in software development. Through an ethical impact assessment, I’ll show practical steps for addressing potential risks and harms. The paper concludes with concise recommendations for both technical and non-technical aspects of software development.

## Importance of Ethical Impact Assessments

## The Value Sensitive design approach

The Value Sensitive Design approach is approach based on theoretical principles to the design of technologies, while accounting for human values that are deemed important [1 (insert the link) ].

# Ethical Impact Assessment

**(35 marks)**

*TABLE 1*

*Ethical Impact Assessment, informed by Value Sensitive Design*

|  |  |  |
| --- | --- | --- |
| **Stakeholder** | **Values** | **Potential Risk/Harm** |
| **Employer/ Employee**  (System Administrator/ Project Team Member/ Project Manager)  Direct | **Privacy** – Protection of personal data and information. | **Privacy** – Data leaks or unauthorised use of employee data, leading to violation of privacy. |
| **Affected Individuals (the people the software is used on)**  (Subjects of Analysis)  Direct | **Privacy** – Protection of personal data and information.  **Fairness and Discrimination** – Ensuring the software isn’t biased (on non-relevant attributes such as race, gender, age etc…). Training dataset should have bias mitigated against.  **Respect for Human Rights** – upholding human rights. | **Privacy** – Data leaks or unauthorised use of user data, leading to violation of privacy.  **Fairness and Discrimination** – Biased software for the emotion recognition could reinforce or encourage stereotypes and discrimination.  **Respect for Human Rights** –  Right to privacy could be breached. Furthermore, the client may not be protected against unwanted surveillance. |
| **External Data Provider (e.g. social media platform)**  (Data Provider)  Indirect | **Accountability** – The data provider should accept responsibility for managing user data ethically, ensuing privacy standards and regulations are met.  **Transparency** – The data providers should explicitly communicate to clients/users that their data is being shared to external entities. | **Accountability** – The data provider may not take responsibility if harm falls onto a user.  **Transparency** – The data provider may not openly disclose how user’s images of faces are used. |
| **Professional Customer/ Client (e.g. the public or psychologist)**  (System End User)  Direct | **Accuracy/ Reliability** – Professional integrity will be impacted by the system’s performance | **Accuracy/ Reliability –** The software could be utilized by professionals (e.g. psychologist) for research or diagnosis. Faults in the software could detrimentally impact the user’s credibility |
|  |  |  |

# Recommendations and Considerations

**(45 marks)**

# References