*A blue and black sign with black text

Description automatically generated*A picture containing font, logo, graphics, symbol

Description automatically generated

|  |  |
| --- | --- |
| **Student ID Number** | HE30667 |
| **Programme Title** | Research method |
| **Module Code** |  |
| **Module Title** |  |
| **Assessment Number** | [e.g. 1] |
| **Assessment Title** |  |
| **Date of Submission** | *dd/mm/yy* |

***MSc Software Engineering***

Research Methods

COM7302

***Ethical Considerations in Machine Learning***

Presented by

**Student ID#: HE30667**

Supervised by

**Dr. Liaqat Ali**

**Submission Date**

December 2024

Contents

[**Introduction** 2](#_Toc180394726)

[**Background of Research** 2](#_Toc180394727)

[**Problem Statement** 2](#_Toc180394728)

[**Aims and Objectives** 3](#_Toc180394729)

[**Research Questions** 3](#_Toc180394730)

[**Importance and Significance** 3](#_Toc180394731)

[**Rationale for Research** 3](#_Toc180394732)

[**Time Frame** 4](#_Toc180394733)

[**Methodologies** 4](#_Toc180394734)

[**Expected Outcomes** 4](#_Toc180394735)

[**Challenges and Limitations** 5](#_Toc180394736)

[**Summary** 5](#_Toc180394737)

[**References** 5](#_Toc180394738)

### 

# **Introduction**

Machine learning has been behind the revolution of many industries, including healthcare, finance, and policing, through insight delivery from data and task automation. Though this has far gained a wide range of applications, it has raised issues of fair treatment, transparency, privacy, and accountability in critical applications like granting loans, employment opportunities, and court orders.  
  
The research proposal is aimed to address the machine learning ethical issues in the algorithms’ bias, which can entrench the discrimination and inequality that already plague many marginalized groups. This research will cover the ways through which biases find themselves in algorithms and opacity of decision-making to provide solutions toward more ethical and fair systems.

# **Background of Research**

### Machine learning in decision-making has advantages and disadvantages. Precisely, while it might be able to process huge data for decision accuracy, data and algorithm bias might be prejudicing, eventually leading to unfair results. Evidence that algorithms are biased comes from a study done in 2016 about the US criminal risk assessment algorithm, the researchers found racial bias in the algorithm raising the risks of reoffending that the defendants would score for. Such a case indicates how much loss a biased algorithm could expose when developed for a sensitive application.

### Moreover, opaque reasoning in most machine learning models brings about what is commonly referred to as a black box, tough explaining decisions and raising questions about transparency and accountability. Such should be the focus of this research: the ethical issues in machine learning with particular attention to bias, transparency, and accountability so that appropriate strategies can be advanced to counter these challenges for fair systems.

# **Problem Statement**

Despite the numerous benefits of ML, ethical concerns, particularly regarding bias and transparency, continue to plague its widespread adoption. This research seeks to address these concerns by exploring how bias is introduced into ML algorithms and how transparency and accountability can be enhanced in ML decision-making processes.

# **Aims and Objectives**

* To examine the ethical concerns associated with ML, particularly in relation to bias, transparency, and accountability.
* To analyze the ways in which bias is introduced into ML algorithms and its impact on decision-making.
* To propose strategies for reducing bias and ensuring more equitable outcomes in ML systems.
* To explore methods for improving transparency and accountability in ML decision-making processes.

# **Research Questions**

1. How does bias enter ML algorithms, and what results does it bring about?
2. What is the connection between transparency and the proper, ethical use of ML systems?
3. How might one design ML systems so that they incorporate a minimal amount of bias and are more fair?
4. What are the concrete actions that would further accountability in ML systems?

# **Importance and Significance**

This research is highly relevant because it will help to contribute to the ethical development and use of machine learning systems. Given that machine learning is very rapidly becoming a core constituent of decision-making processes, getting it right the fairness and transparency of systems is critical. Because fairness and transparency breed trust, an added benefit in pursuing fairness in ML systems is that doing so fosters efforts to use transparency to build trust with users and stakeholders.

It is important to the extent that it relates to, in fact has meaning for, any policy maker or developer or an organization which uses ML systems for making decisions. This research will make those in charge more aware of ethics and be able to give well-reasoned responses, thus leading to more responsible and ethical ML implementations.

# **Rationale for Research**

The motivation for this research lies in the increasing reliance on ML for organizational decision-making and its impacts on people and society. Accompanying such reliance is the need to make sure these systems do not perpetuate harm and exacerbate existing forms of inequality. Bias is an urgent issue of concern in ML systems that is already causing wrong, unprincipled decisions across sectors. By concentrating on the diminution of bias and increasing levels of transparency, this research can be angled to try and meet along these lines a more ethical creation of ML systems.

# **Time Frame**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Activities | Week 6 | Week 7 | Week 8 | Week 9 | Week 10 | Week 11 |
| Literature Review |  |  |  |  |  |  |
| Identifying Ethical Issues |  |  |  |  |  |  |
| Data Collection |  |  |  |  |  |  |
| Developing Ethical Guidelines |  |  |  |  |  |  |
| Editing and Proofreading |  |  |  |  |  |  |
| Final Report Submission |  |  |  |  |  |  |

# **Methodologies**

It would use a mixed-method approach, balancing both the qualitative and quantitative approaches. The qualitative aspect would assess existing studies on bias, transparency, and accountability in ML systems. From these notes, critical ethical concerns would be drawn. The quantitative part would access case studies of biased ML, answering how, when, and why bias made the decisions going into the process of decision-making.  
  
Data for this will be culled from academic journals, case studies, and interviews with ML experts. An analysis will be done to identify the patterns of bias in ML systems and how to solve them.

# **Expected Outcomes**

The expected outcomes of this research include:

* An increased understanding of how bias is introduced into ML algorithms.
* Proposed solutions for reducing bias and increasing the level of fairness in ML systems.
* Ways to achieve transparency and accountability in making decisions with ML technologies.

# **Challenges and Limitations**

One of the key challenges for this research will be in accessing sufficient data on ML algorithms since many companies and organizations treat their algorithms as a trade secret. Other complexities of the ML system can sometimes afford it hard to understand the exact decision-making process in it especially when some algorithms are a black box. This bestows the limitation that results from an individual case study may not be generalizable to all ML systems.

# **Summary**

This paper will attempt to discuss the ethical concerns that revolve around bias-reduction and efforts towards more transparent and accountable ML systems. It is in this view that this paper will explore how bias is introduced into ML algorithms, and strategies capable of reducing this kind of bias with an aim to have more ethical and equitable ML systems.

# **References**

1. Lo Piano, S. (2020) *Ethical principles in machine learning and artificial intelligence: Cases from the field and possible ways forward*, *Nature News*. Available at: https://www.nature.com/articles/s41599-020-0501-9 (Accessed: 23 October 2024).
2. Venkateswaran, A. (2017) *Ethics in machine learning*, *Medium*. Available at: https://towardsdatascience.com/ethics-in-machine-learning-9fa5b1aadc12 (Accessed: 12 October 2024).
3. Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L., 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), pp.1-21. Available at: https://doi.org/10.1177/2053951716679679 [Accessed 23 Oct. 2024].
4. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning: Limitations and Opportunities. MIT Press.
5. Dastin, J. (2018). "Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women." Reuters.
6. Binns, R. (2018). "Fairness in Machine Learning: Lessons from Political Philosophy." Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency.

Top of Form