Individual Assignment

Ethics, Regulation & Law in Advanced Digital Information Processing & Decision Making
(ECS7025P)

Student name	:	Pranav Narendra Gopalkrishna
Student number	:	231052045
Programme	:	Master’s in Artificial Intelligence
Submission date	:	10 May 2024
Word count	:	3642

Table of Contents

Abbreviations

Abbreviation	Full form
AI	Artificial Intelligence
ML	Machine Learning
TEA	Transparency, Explainability, Accountability
US	United States
UK	United Kingdom
EU	European Union
EIA	Ethical Impact Assessment

Introduction

"*The field of AI ethics has largely emerged as a response to the range of individual and societal harms that the misuse, abuse, poor design, or negative unintended consequences of AI systems may cause*." (Leslie, 2019)

Ethics is the study of the values by which humans must guide their choices at the broadest levels of decision-making; this includes not just the identification of the values but also of the means of achieving them, namely virtues and principles. As a result, ethics forms the bedrock of long-term human existence and flourishing. Due to AI's impact and potential impact, issues in AI use and abuse become important ethical issues that can have (and often do have) a large impact both on individual users, organisations as well as the society at large. In particular, I shall focus on the issues of transparency, explainability and accountability, which are vital in making informed, rational and safe AI-based decision-making, thereby protecting individual rights and interests from the misuse or neglectful use of AI systems.

The ethical concerns of privacy, consent and justice can be tied to the technical-business concerns of producing high-quality solutions that create the greatest value in terms of profits and long-term viability, but there may be a divergence between these concerns with respect to (1) some ethical, regulatory or legal frameworks, or (2) some socio-economic conditions. This report focuses primarily on transparency, explainability and accountability, since these are both key technical/business requirements and key requirements in ensuring ethical practices and legal/regulatory compliance.

This report aims to address some relevant ethical ideas in AI, primarily transparency, explainability and accountability. Further, it aims to integrate these ideas with technical/business requirements, explore AI ethics using a case study involving an ethical and legal/regulatory breach in AI use and finally, explore the application of AI ethics in a hypothetical case involving the development of an ethical framework for a particular technical/business context.

Executive Summary

Section A gives an overview about AI ethics and its relationship to technical/business requirements, particularly with respect to transparency, explainability and accountability (TEA) while also touching upon relevant ethical issues in AI that are addressed through TEA. Section A also touches upon the value of ethical frameworks for organisations, indicating a few relevant frameworks that may be referenced later in the report.

Section B explores a breach in ethical principles with respect to AI use, particularly AI use in recruitment and the unfair gender-discrimination that resulted from deficiencies in TEA. Section B also explores the breach with respect to regulatory frameworks and their implementations. Finally, it expands on the ethical responsibilities of developers in AI and their role in ensuring ethical practices with respect to AI.

Section C applies AI ethics to a hypothetical medical organisation, with a focus on integrating TEA as a primary consideration. Here, an ethical framework as well as an ethical impact assessment procedure is defined for the organisation and its particular needs. Section C also aims to integrate business concerns with ethical concerns, showing how they work hand-in-hand and why ethical concerns are vital for business. Finally, it expands on the role and importance of training, monitoring and continuous improvement in both the application of the ethical framework as well as the framework itself.

NOTE: *Definitions of key terms used in the report are given in the* ***appendix****. The appendix also contains additional content that could not fit in the report.*

Section A

Transparency, explainability & accountability (TEA)

Definitions, technical/business value and ethical relevance

NOTE: *Transparency, explainability and accountability are defined in the appendix*.

Transparency can be considered as an impactful non-functional requirement (NFR) that facilitates (among other things) other NFRs such as the *accessibility*, *usability*, *informativeness*, *privacy* and *security* of a software (Horkoff, 2019). Furthermore, explainability as an NFR is a means to improve *transparency*, *trust*, *scrutability* and *satisfaction* (Chazette et al., 2019). In this way, we see how the ethical dimension of transparency and explainability are tied to the technical-business dimensions; in the long run, it is in the self-interest of developers to achieve ethical goals (e.g. privacy, scrutability or accessibility) through technical or business-driven goals (e.g. security, usability or satisfaction). Hence, transparency and explainability improve our ability to examine and assess a system, which in turn improves the accountability of the system (i.e. the ability to examine and assess the system for potential biases, errors and guidelines). This module provides the ethical and legal overview of the above concepts, which – when combined with the technical-business perspective – can help make informed, effective decisions in the design, development and deployment of AI solutions.

Accountability is a key requirement in ensuring that a system can be made to comply with ethical and legal constraints; this ties transparency, explainability and accountability in both an ethical sense and a technical-business sense. Leslie, 2019, identifies the two subcomponents of accountability as **answerability** (i.e. explainability, but with the responsibility of explanation on the human authorities behind the AI system) and **auditability** (i.e. the ability of the AI system to be monitored by a third-party via records and accessible information on every stage of the system's development and deployment) (Leslie, 2019). He also distinguishes between **anticipatory accountability** (i.e. accountability before deployment, during development stages) and **remedial accountability** (i.e. accountability after deployment), noting the importance and necessity of both (Leslie, 2019). The relevance of accountability lies in the fact that, "*automated decisions are not self-justifiable. Whereas human agents can be called to account for their judgements and decisions in instances where those judgments and decisions affect the interests of others, the statistical models and underlying hardware that compose AI systems are not responsible in the same morally relevant sense*." (Leslie, 2019)

Use of AI and the relevance of TEA

Consider some wide-spread or significant AI use-cases. (1) Large language models (such as ChatGPT) used for analysing, interpreting and summarising content from technical topics. (2) Large language models (such as ChatGPT) used to generate code to solve programming problems. (3) Advanced generative AI models (such as Stable Diffusion) used to create art for personal and commercial use (e.g. book covers, advertisements, visual aids in videos, etc.). (4) AI tools for decision-making, such as reviewing job applications[1] and business risk and legal/regulatory compliance assessment[2]. Apart from current use, AI also has significant potential uses, such as in healthcare (e.g. digital twin, treatment planning, simulating drug trials and medical tests, etc.), assessing insurance claims, reviewing loan and mortgage applications, and media content creation (e.g. animation, visual effects, script-writing, etc.) and governance. The actual and potential uses of AI indicate its actual and potential significance in human lives on a both personal and societal level, which makes the validation of AI processes and decision-making all the more relevant, which in turn leads to the relevance of TEA in AI.

Ethical frameworks for TEA

Why bother with ethical frameworks? Firstly, existing frameworks can help guide ethical considerations, particularly around maintaining or promoting TEA. Secondly, knowing about existing frameworks, to the extent they are implemented, can inform decisions to ensure smooth development and deployment. Lastly, existing ethical and legal/regulatory frameworks and guidelines can help indicate areas of ethical deficiency in the functioning and product of an organisation (public or private). Examples of such frameworks and guidelines are:

- European Commission's ethical guidelines for trustworthy AI
- *Understanding artificial intelligence ethics and safety*
  — by Dr. David Leslie (The Alan Turing Institute)
- *AI Ethics Guidelines: European and Global Perspectives*
  — by Marcello Ienca and Effy Vayena (part of the Ad Hoc Committee on AI, CAHAI)

A key feature of these frameworks is that they are globally applicable, i.e. they are not based on country-specific or region-specific factors but AI development and usage as such. The frameworks outline considerations that would be relevant for any law-makers when considering legal and regulatory measures for AI development and usage. As for private organisations, the frameworks can give insights into the potential legal and possibly regulatory concerns regarding AI that could become relevant in the future, thereby "future-proofing" the organisations'

[1] Inferring from the existence of AI-based recruiting tools such as Manatal and tools used in Breezy HR: https://breezy.hr/resources/breezy-updates/candidate-match-score

[2] Moody's AI-review tool for third-party risk-assessment: https://www.moodys.com/web/en/us/kyc/products/review.html

practices and ensuring long-term stability to some extent; such insights are particularly relevant with respect to TEA, which are important non-functional considerations in software in general and are likely to be relevant factors in any ethical, legal or regulatory considerations regarding AI. This module's content has given a greater exposure to and awareness of such frameworks.

Other relevant ethical issues in AI that TEA can address

Unfair discrimination in AI use is often due to a lack of understanding about the training data quality or algorithmic processes that lead to the discriminatory decisions. Hence, TEA can help address unfair discrimination (e.g. gender or age discrimination for employment, racial discrimination for medical or insurance-related decisions, etc.).

Privacy and the misuse of personal data is a key concern in data-driven AI. TEA can ensure that how data is used in the AI system is always apparent and auditable, thereby helping users and outsiders to detect privacy breaches and potential for such breaches. In this way, TEA ensures that AI is used in accordance with human rights and not against them (*the right to privacy is vital to safeguarding the right to the freedom of expression and the right to autonomy in many contexts*).

This module's content gives a basis for understanding the use of AI with respect to human rights and dignity (which involves fair, ethical treatment of individuals).

Section B

Introducing the case study

The chosen case study is Amazon's now-scrapped ML recruitment engine, which may have been also used between the years 2014-2015 by Amazon's recruiters, perhaps not as the sole basis for recruitment but was at least used for looking at recommendations. In 2015, Amazon realised that its engine was rating candidates for technical posts such that it discriminated against female applicants despite gender not being a relevant factor in an applicant's competence. A key reason was that Amazon's computer models were trained on résumés submitted to Amazon over a 10-year period wherein most submissions for technical roles — including top submissions — were men (Dastin, 2018). Hence, the key points of failure were: (1) biassed training data, (2) the lack of transparency in the system's training and decision-making and (3) the lack of explainability of the decision-making process. The lack of transparency and explainability means the usage of such a system in practice would be a failure in accountability, since human decision-makers would be not validating their decisions using a logical, human-interpretable basis.

Discussing the case with respect to ethical AI decision-making

Breach of ethical guidelines

A relevant set of guidelines for such a case is given by the European Commission's guidelines for trustworthy AI ('Ethics guidelines for trustworthy AI', 2019), which concerns AI in decision-making. The relevant requirements for trustworthy AI among the ones listed in the guidelines are: (1) human agency and oversight, (2) transparency, (3) diversity, non-discrimination and fairness and (4) accountability. In Amazon's case, human recruiters were practically the primary decision-makers, since they had the discretion to decide whether or how to act upon the AI system's results; hence, we see a human-in-command system, in line with requirement (1). However, the AI system's decisions were not explainable and it is not indicated whether the system's capacities and limitations were communicated to the recruiters who used it.

Furthermore, there was no indication that the rejected applicants were informed of the fact that they were rejected by an AI system. Moreover, if they were rejected by the AI system, the reason behind their rejection by the AI system could not be explained to them. According to people familiar with the project, the recruiters did not rely only on the AI system and used it only for recommendation; nonetheless, the recommendations may have led to unexplained, non-transparent rejections. Given that the system was found to be unfairly discriminatory, both requirements (2) and (3) were not fulfilled. Lastly, the accounts on the case indicate that the AI system was not sufficiently audited before deployment and that the system was not auditable[3] to begin with, since the discriminatory nature of the algorithm used and the bias in the training data were not caught until after it was already put to use for a while. Hence, requirement (4) was also not met.

Breach of legal/regulatory frameworks

There are regulations as well as ethical and regulatory frameworks proposed or in place with respect to which Amazon's actions in the above case would be legal or regulatory violation. In particular, the Ad Hoc Committee on Artificial Intelligence (CAHAI) proposes that, "*a legally binding transversal instrument should contain provisions on ensuring that gender equality and rights related to vulnerable groups … are being upheld throughout the lifecycle of artificial intelligence systems.*" (CAHAI, 2021).

An implementation of this proposition is by the EU's Artificial Intelligence Act (passed in the European Parliament in 2024). The AI system in the case study would fall under the category of "high-risk AI applications" and would thus be subject to certain legal requirements. In particular, as per the act, the training data for the AI system would have to undergo examination in view of possible biases that are likely to have a negative impact on fundamental rights or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future operations (Paragraph 2, Article 10, EU Artificial Intelligence Act)[4].

Another such implementation can be found in the Equal Employment Opportunity Commission[5], a US federal agency that handles cases involving unfair discrimination in private-sector employment. For example, they sued the iTutorGroup for age discrimination in their software-driven recruitment tool ('EEOC Sues iTutorGroup for Age Discrimination', 2022). Amazon in the above case study would be in violation of such regulations due to the demonstrable gender-based discrimination.

Breach in terms of stakeholder perspectives

Stakeholder: *Applicant*
Potential unfair gender-discrimination against the applicant by the AI system can lead to a loss of legitimate employment opportunities and unjustified uncertainty about the strength of their application. Ethically, it breaches the principle of fair, rational treatment based on their character, skills and experience rather than unchosen, irrelevant attributes.

Stakeholder: *Recruiters*

[3] Auditability enables the assessment of algorithms, data and design processes.
[4] Direct link to article 10 of the EU AI Act: https://artificialintelligenceact.eu/article/10/
[5] Website available at: https://www.eeoc.gov/

The AI system's irrational bias — which is likely to be unknown to the users, i.e. the recruiters — would breach the recruiters' trust in the system's reliability in choosing objectively good potential candidates. Furthermore, it can make the recruiters legally liable for discrimination that they did not intend to commit.

Stakeholder: *Amazon*
Unethical discrimination is harmful to the company's business interests as it arbitrarily excludes potentially good candidates. Such discrimination also harms its reputation among potential applicants, who may seek opportunities in enterprises with more ethical hiring practices.

Ethical responsibilities in development

There, the responsibility of AI developers is to ensure that the capacities and limitations of their AI system are clearly explained to the users of the system; this ensures transparency and accountability. Furthermore, where AI is used for critical decisions, it is the responsibility of both the developers and the concerned authorities to ensure that (a) the AI's development involves quality-testing at critical stages (e.g. training, validation and testing of the ML algorithm's performance), (b) the AI's decisions are traceable to explainable causes or reasons, (c) the use of AI is communicated to the affected parties (e.g. the applicants), and (d) the AI's decisions are validated by one or more humans who can be held accountable for them (because only humans can and should take ethical responsibility).

Section C

Consider a hypothetical medical practice called Feigenbaum ENT Clinic (FEC) that handles general health checkups and medical diagnoses for ENT-related concerns. To improve the accuracy and cost-effectiveness of their diagnoses and treatment decisions, they decide to use an AI-based clinical decision support system (CDSS).

Importance of an AI-ethics committee and framework

Since medicine deals directly with a patient's well-being, medical decisions can pose critical risks to a patient which must be mitigated as far as possible or — if not — which must be made clear to the patient, since doing so is the only way in which a voluntary, rights-respecting agreement can be formed that places the patient's current or future well-being and/or life in the hands of the medical practitioners. Hence, when an AI CDSS is used, establishing the principles — hence, an ethical framework — by which its usage and results must be evaluated, processed and conveyed to the relevant stakeholders is crucial in ensuring that actions and outcomes using the AI system are consistently ethical.

From a business perspective, consistently ethical practice of medicine is crucial in building trust between practitioners and patients. Moreover, ensuring ethical practice is also necessary to mitigate avoidable harm to patients and thereby to the clinic's own revenue (affected by legal fees in fighting lawsuits), reputation and long-term business interests (due to the loss of clients or legal actions). From a legal perspective, explainability is important in a medical context for (1) informed consent, (2) certification and approval as medical devices and (3) liability for errors or failures in usage (Kiseleva et al., 2022).

AI-ethics committee objective

The key objective of the committee is to outline a framework to (1) uphold the conditions for informed consent by transparency in the use of AI, (2) ensure that AI-based decisions are validated and can be validated (i.e. that are explainable) by humans, and (3) maintain a record of the use of AI and AI-decisions whose responsibility can be tracked to individuals within FEC, thereby maintaining accountability via liability for errors or failures in usage. Note that objective (2), i.e. explainability, is a key factor facilitating the certification and approval of the AI CDSS as a medical device for more extensive use.

Key values and guidelines

Values

1. **Information-transparency**[6]
2. **Accountability**
3. **Privacy**

Information-transparency here is catch-all term meaning that the development, testing, deployment, use and decisions of the AI should be communicated without fail to the stakeholders (chiefly the patient and the practitioner), the communication should be explainable, the AI system should be auditable and the AI system's use should be traceable (definitions for each are given in the appendix).

In essence, information-transparency ensures that (1) the conditions for informed consent are always met, (2) the AI's decision-making can be validated by the practitioners for every decision, and (3) problems in the AI's functions or outputs can be traced as far back as needed (e.g. to the development, algorithm or training data). This ensures ethical communication of information and a robust AI-system that can be easily managed, updated or corrected.

Accountability here refers specifically to the idea of ensuring that the AI's decision-making is the responsibility of humans to validate or invalidate. It also implies that any errors or misuse of the system should always be prevented and preventable by human intervention so that failure to prevent such errors or misuse would essentially not be an AI error that cannot be reasoned with but a human default that can be recognised and dealt with ethically and legally.

Finally, **privacy** is an extension of the patient's own right to his own personal data, which the patient agrees to share with the clinic for medical use only. Hence, any leakage of his personal data by the clinic is a violation of his rights, and thus, unethical and illegal. Hence, privacy is a key value in the proposed ethical framework.

Guidelines to ensure the fulfilment of the above values

1. Maintaining documentation of the AI-system's development and testing
2. Informing patients of the use of AI
3. Keeping records of decisions the practitioners made using AI
4. Assigning the responsibility of medical decision-making to the lead doctor in the case
5. Securely encrypting patient data
6. Limiting access to patient data based on legitimate need and access level[7]

[6] Information-transparency as used here is different from transparency in the usual AI context.

[7] Access level means the level of security up to which a user of a database can access its data.

Guidelines 1, 2 and 3 ensure that information-transparency is maintained. Guidelines 3 and 4 ensure that accountability is maintained. Guidelines 5 and 6 ensure that privacy is maintained.

Ethical impact assessment (EIA)

An EIA is a procedure (e.g. a sequence of questions) by which the impact of an AI system can be evaluated with respect to a set of ethical objectives. The EIA developed for FEC consists of the following questionnaire:

1. **Does each training dataset have enough data diversity in its relevant features to prevent overfitting and bias?** *Low diversity can lead to overfitting, i.e. poor generalisation and a potential bias with respect to irrelevant characteristics*.

2. **Is there a mechanism to ensure that the usage of AI is always communicated clearly to the patient?** *This is especially relevant for online/electronic applications, forms and reports.*

3. **Does each responsible practitioner have a clear means of knowing when a decision was made by AI?** *Especially when sharing decision-making responsibility for a case to multiple practitioners.*

4. **Can the AI be queried for explanations of its decisions or at least the availability of explanations?** *If no explanations exist, the practitioner will understand that the AI system is a black-box and act accordingly.*

5. **Is patient data accessible by anyone other than the responsible practitioners?** *Specifically, can the AI system expose patient information to any querying third-parties?*

Role of training, monitoring and continuous improvement

Since the medical practitioners must bear the responsibilities of AI use, and since they are — in general — not familiar with the intricacies of AI and its proper use, training them is essential in ensuring that the ethical framework is put into practice. Monitoring helps track the actions of the stakeholders — particularly the practitioners — with respect to AI use, which helps verify whether, to what extent and in what way is the ethical framework being followed or violated. The results of monitoring can help guide and improve the actions of the relevant stakeholders. Furthermore, practical complications in the implementation of the ethical framework and overlooked or new ethical/legal considerations must be incorporated into the framework or the method of applying the framework; this requires a continuous monitoring and improvement process of both the framework and its practical application.

References

Leslie, D. (2019). ‘Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector’. *The Alan Turing Institute.* Available at: https://doi.org/10.5281/zenodo.3240529 (Accessed: 9 May 2024).

Horkoff, J. (2019). ‘Non-Functional Requirements for Machine Learning: Challenges and New Directions’. *2019 IEEE 27th International Requirements Engineering Conference (RE)*. Jeju, Korea (South). pp. 386-391. doi: 10.1109/RE.2019.00050.

Chazette, L., Karras, O. and Schneider, K. (2019). ‘Do End-Users Want Explanations? Analyzing the Role of Explainability as an Emerging Aspect of Non-Functional Requirements’. *2019 IEEE 27th International Requirements Engineering Conference (RE)*. Jeju, Korea (South). pp. 223-233. doi: 10.1109/RE.2019.00032.

Dastin, J. (2018). ‘Insight - Amazon scraps secret AI recruiting tool that showed bias against women’. *Reuters*. Available at: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/ (Accessed: 9 May 2024).

‘Ethics guidelines for trustworthy AI’ (2019). *European Commision*. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (Accessed: 9 May 2024).

Ad Hoc Committee on Artificial Intelligence (CAHAI) (2021). ‘Possible elements of a legal framework on artificial intelligence, based on the Council of Europe’s standards on human rights, democracy and the rule of law’. *Council of Europe*. Available at: https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d (Accessed: 9 May 2024).

‘EEOC Sues iTutorGroup for Age Discrimination’ (2022). *U.S. Equal Economic Opportunity Commission*. Available at: https://www.eeoc.gov/newsroom/eeoc-sues-itutorgroup-age-discrimination (Accessed: 9 May 2024).

Kiseleva, A., Kotzinos, D. and De Hert. P. (2022). ‘Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations’. *Frontiers in Artificial Intelligence*. Available at: https://doi.org/10.3389/frai.2022.879603 (Accessed: 9 May 2024).

Rouse, M. (2023). ‘What is Transparency?’. *Technopedia*. Available at: https://www.techopedia.com/definition/30368/transparency-data (Accessed: 9 May 2024).

O. Boluwatife, V. (2023). ‘Explainability in AI and Machine Learning Systems: An Overview’. *Comet*. Available at: https://www.comet.com/site/blog/explainability-in-ai-and-machine-learning-systems-an-overview/ (Accessed: 9 May 2024).

Appendix

Generative AI usage declaration

No generative AI was used in the creation of the report.

Section	**Gen. AI system**	**Version**	**Publisher**	**AI system URL**	**Text description**	**The rationale for using the Gen AI tool**
Introduction	None	NA	NA	NA	NA	NA
Executive Summary	None	NA	NA	NA	NA	NA
Section A	None	NA	NA	NA	NA	NA
Section B	None	NA	NA	NA	NA	NA
Section C	None	NA	NA	NA	NA	NA
Appendix	None	NA	NA	NA	NA	NA

Definitions (in alphabetical order)

Accountability (***algorithmic*** **accountability)**:
A quality that enables the examination and assessment of the decision-making processes of the models, i.e. examination of potential biases or errors and assessment of the software's compliance with ethical guidelines and legal requirements (Boluwatife, 2023).

Auditability:
In finance, auditability is the ease with which an auditor can obtain sufficient, appropriate evidence to evaluate financial statements or internal controls. When used for AI development and use, auditability refers to the quality of an AI system that ensures the system is capable of demonstrating both the responsibility of design and use practices as well as the justifiability of outcomes to an independent third-party. (Leslie, 2019).

Equality (moral/legal equality):
The ethical/legal position that each individual must have the same fundamental moral and legal rights and responsibilities.

Equity:
The view that the socio-economic benefits of any action or policy must be evenly distributed among the population.

Explainability:
A quality that provides human-interpretable explanations, i.e. reasoning about the processes of a software system. It may or may not correspond to the actual processes, but it does work to consistently justify the results of the process with reason. It is not simply the explanation of the inner workings of the system but the reasoning that supports the system's processes (Boluwatife, 2023).

Data quality:
The representativeness of a dataset with respect to the population or random process being studied.

Transparency (system transparency):
The quality that makes the inner workings of ICT systems and networks open to observation to users, administrators, developers and software engineers (Rouse, 2023).

Traceability:
The quality that allows the results of the usage of a system to be traced back to a definitive request or query to the system.

Section B: Additional content

Why was ML used at all in this case?
The hiring process for companies is a time-consuming and expensive task, involving background checks, résumé reviews and one or more rounds of interviews. AI tools — using NLP and ML — can automate more routine or impersonal parts of the task, namely background checks and résumé reviews.

Why does ML pose such a challenge?
In ML, the system learns decision-making based on the relationships between the input's features (which may be automatically extracted or pre-defined). In essence, ML algorithms aim to approximate a function that maps the inputs (e.g. résumé) to an output (e.g. score). For high-dimensional data — such as natural-language texts — the function that is approximated is too complex to define in terms that can be feasibly interpreted by humans. Hence, such algorithms are black-boxed wherein the relationship between the input's features and between the inputs and outputs are opaque and unexplainable. Furthermore, the quality of the ML algorithm's performance depends in large part on the quality of the data used to train it; bias in

the training data leads to bias in performance, since the algorithm works to fit its outputs according to what it can observe or extract from the inputs.