

Original Article

# Machine Learning in Automated Decision Systems with Fairness Awareness

Steven Henry

Ladoke Akintola University of Technology

## Abstract

Machine learning nowadays is being used in more and more important places, such as jobs, health care, criminal justice, and banks. Some people believe these systems will be good for justice because they might worsen or perpetuate social biases. This paper discusses FAML, which is a new area that aims at improving the fairness of machine learning models by incorporating factors and constraints that guarantee fairness. We discuss key notions of fairness regarding equal opportunities and equalized odds. Further, we mention the different ways in which algorithms can be used to reduce bias in pre-, in-, and post-processing. We also discuss the positive and negative aspects of honesty, fairness, and transparency. By case studies and theoretical analysis, we show that incorporating fairness into machine learning helps to reduce bias, builds societal trust, and makes the system accountable. Finally, we identify some major issues and provide novel ideas for investigation. These will pave the way for building automated decision-making systems that are not only fair but also moral.

## Keywords

Fairness-aware machine learning, Algorithmic bias, Automated decision systems, Discrimination mitigation, Ethical AI, Fairness metrics, Socio-technical systems, Responsible AI.

Article  
History

Received:  
22.12.2025

Accepted:  
09.01.2026

Published:  
23.01.2026

## 1. Introduction

### A. Motivation: Automated Decision Systems' Ascent

Within the last couple of years, machine learning automated decision systems have made many things very important in society. The systems help people find jobs, get loans, get out of jail, or get medical care. People want to use such systems because they promise to make things easier, to enable people to save money, and to make decisions faster and more easily compared to earlier when decisions were hard and took a very long time. However, when these very important apps make use of machine learning algorithms, the choices made by people have a greater bearing on their lives and chances. Their rapid diffusion indicates a need for reflection on how ADS impact technology and society from a moral point of view.

### B. Hazards of Inequitable Results and Their Effect on Society

On one hand, some of the ways that automated decision systems can help are very positive, but there is also the potential for them to make biases that exist even worse or new. This may happen in those instances when biases, common in society, find their way into data being utilized for training algorithms, and from there, biases are taught to models that learn from data. For instance, an algorithm that learns from resumes of a predominantly men workforce might always pick the men over the women. The famous COMPAS case well illustrates that some people can be wrong in court when using risk assessment tools. It can also include exacerbation by making structural inequalities worse, corroding trust in government, and treating people inequitably based on their race, age, gender, or income level. These risks show the importance of machine learning algorithms being fair.

### ***C. The Necessity of Machine Learning That Considers Fairness***

Fairness-aware machine learning is thus gaining increasing importance to fix the problems of algorithmic bias and hence to make sure that machine learning systems will work in a fair and socially responsible way. "FAML" is an abbreviation for tools and methods that help find, measure, and fix algorithm decisions that are not fair. While most traditional machine learning models care about nothing but how well or quickly, they do their job, models that understand what fairness is also think about how fair it can be. These methods try to ensure at least that the model's guesses don't always hurt people or groups that aren't being treated fairly. Now it is necessary to devise and employ FAML methods in order to make AI systems fair that do not violate any law or moral code.

### ***D. This Paper's Contributions and Scope***

This paper performs a deep analysis of fairness-aware learning in automated decision-making systems. First of all, it considers the more general social and technological motives that underpin the desire for equity. A very important contribution of this work is the analysis of formal definitions of fairness in the machine learning literature, giving a critical evaluation of their advantages and disadvantages. Then, the three main ways in which algorithms can decrease bias are discussed: pre-processing, in-processing, and post-processing. Afterwards, this work will show how such techniques can be applied to realistic settings. Finally, moral, legal, and technical issues related to fairness arising when trying to achieve equity are discussed, together with novel ideas for constructing schemes that will always be fair, easy to use, and functional.

## **2. Background and Motivation**

### ***A. Automated Decision Systems Overview***

An automated decision system is a computer program designed to support or make decisions for humans by looking for patterns in data and analysing them. Most use machine learning models, often trained with historic data to determine how to act on inputs or predict outputs. Consider the following applications: employers use tools that automatically screen applicants' resumes; police use face recognition systems; systems that predict how well someone will repay a loan. These can now more widely assist human choice in the physical world because they have access to more and better data and greater processing power. Yet, if you rely too heavily on data to inform your choices, then you are likely to be wrong. This is especially true when the information upon which your decisions are based indicates that people have made mistakes in the past or that society is unfair. Even though ADSs function properly, we shouldn't use them in real life until we consider how they impact humans in a way that's both fair and moral.

### ***B. Examples of Algorithmic Unfairness in the Real World***

There are several well-known cases of automated decision systems making unfair or biased choices. COMPAS helps the US criminal justice system decide how likely it is that a person who has just been arrested will commit another illegal act. This study found that the system was ranking Black defendants as "high risk," despite their criminal records being roughly the same as those of white defendants. Because of this, racial bias was suspected. Most of the resumes Amazon received came from men, so for women, it would be unfair to apply there. These examples very clearly show that algorithms, which seem to be fair, actually make matters worse for everybody. They show that making machine learning fair is difficult—a problem deeply rooted in history, law, and society.

### ***C. Definitions of Justice in Legal and Human Contexts***

Fairness means various things in various contexts. It emanates from law, philosophy, and the struggle for equity in the world. Most people immediately think of civil rights laws that prohibit the differential treatment of individuals based on their race, gender, or sexual orientation when they think of the concept of fairness. Another example is the EU's General Data Protection Regulation, while the Equal Employment Opportunity Commission is a US example. Utilitarianism posits that being fair means doing what is in the best interest of everyone, while Rawlsian justice says that what is fair is giving everyone an equal opportunity. These diverging concepts cause algorithmic systems to design different ways to determine what is fair. Apart from ensuring your ML models are working correctly, you need to understand what the law and society consider "fair."

### 3. Fairness Definitions in Machine Learning

#### A. Individual vs. Collective Fairness

The two key notions that bring fairness to machine learning, therefore, take up the form of individual fairness and group fairness. What this means is that under ordinary contexts, individual fairness refers to a situation where an algorithm will have to treat everybody just the same. Of course, the complexity of telling individuals apart is not easy to establish and is dependent on the context. In contrast, group fairness focuses on individuals of different ages, gender, or race receiving the same things or results. Most times this simply implies that every group possesses the same amount of information or the same number of errors. It is easier to identify if a group is fair than when one person alone would be fair. Fairness to everyone does present an interesting idea; however, this is tough to pursue because it needs a really strict similarity function. Yet another challenging task of FAML is the fact that it must cope with both kinds. Below table shows the difference between Individual Fairness and Group Fairness:

**Table 1: Individual Fairness vs. Group Fairness**

Aspect	Individual Fairness	Group (Collective) Fairness
Main Idea	Treat similar individuals similarly.	Ensure different demographic groups receive similar outcomes.
Focus Level	Person-level fairness.	Group-level fairness (e.g., gender, race, age groups).
Requirement	A strong "similarity function" to identify how similar two individuals are.	Statistical comparison between groups (e.g., accuracy, error rates).
Difficulty Level	Harder: It is complex to define what makes two individuals "similar."	Comparatively easier: Group metrics are straightforward to measure.
Strength	Provides fairness at a very fine-grained, personal level.	Ensures fairness across protected groups and avoids discrimination.
Limitation	Hard to implement because defining similarity is subjective and context-dependent.	May miss unfairness happening to specific individuals.
Example	Two loan applicants with similar credit profiles should get similar decisions.	Loan approval rates for men and women should be roughly equal.
Use Case	When individual-level decision equity is important.	When preventing group discrimination is the priority.

#### B. Parity in Statistics

Statistical parity, or demographic parity, is a rule about fairness that says all groups should have the same opportunity to receive a positive outcome. A statistically fair loan approval algorithm, for instance, would dole out the same number of loans to men as to women. But this scheme may now be blind to real differences in how quickly people qualify, thus creating "reverse discrimination" or loss of benefits. Statistical parity can lead you to the presence of bias in algorithmic systems but, generally speaking, it is not an accurate method.

#### C. Equalized Odds and Opportunity

If the odds are equal, then the true positive rate and false positive rate of the algorithm are equal across all protected groups. The rule for equal opportunity is not as demanding since it requires only equality of true positives. These notions of fairness are important in situations where stakes are high, like in healthcare or criminal justice, where false negatives or positives have a very significant cost. On the other hand, equalized odds avoid unjust reward or punishment for group membership while still allowing proper differences in baseline outcome rates.

#### D. Forecasted Parity

This is called calibration or predictive parity when all groups have the same chance of receiving the same outcome. If a system says there is a 70% chance some-one will default on a loan, then the actual number of people defaulting should be roughly the same across different groups of people who receive that score. Predictive parity

ensures the predictions are fair by guaranteeing their proximity to accuracy. But it may not work with other rules, such as when everyone has the same chances but the base rates are different across groups. This trade-off is one of many that illustrates how difficult it can be to be fair when one must adhere to more than one rule at once.

#### ***E. Fairness in Counterfactual Situations***

Counterfactual fairness, on the other hand, is a more modern and philosophically sophisticated view of fairness derived from causal inference. Roughly speaking, it considers the decision to be fair if it would have been fair to have the opposite decision if the person had belonged to a different group of people. In order to apply this approach, one has to develop cause-effect models to investigate what the outcome would be in various scenarios. This may be a challenge, but once more, understanding why a particular person is racist will be facilitated too. The above definition of counterfactual fairness follows from moral theories and examines both the objective as well as the result. Now it is clear how crucial knowing what happened is when making a decision if something was fair or not.

#### ***F. Trade-offs and Definitional Incompatibilities***

One big problem in machine learning including fairness is that many definitions of fairness are not mathematically sound. If the base rates across different groups are different, it is impossible to achieve statistical parity, predictive parity, and equalized odds all at once. This has been formalized by Kleinberg et al. as their "impossibility theorem." That means professionals have to select one form of fairness measure based on the goals and purpose of the app. You have to be very honest and careful with those trade-offs when you use and make machine learning systems in sensitive situations.

## **4. Fairness-Aware ML Approaches**

### ***A. Procedures for Pre-Processing***

Yourselves Methods for preprocessing attempt to remove any bias in the data before you train a machine learning model. The premise of these methods is that issues exist in historical data, which must be corrected at their origin. When data transformation changes the features, they become less related with protected attributes. Reweighting assigns a weight to each group, which ensures equal representation from all groups. You can ensure the balance of your dataset by taking another sample from the dataset. It does this by taking fewer samples in case some groups are too large, or it takes more samples in case of small groups. This is because these methods can be used with any type of learning algorithm, and no particular model should necessarily be used. If not used correctly, they can make either the model or the data worse.

### ***B. Methods Used During the Process***

Because of how in-processing works, it is harder for the model training process to be fair from the beginning. Now that these changes have been made, the learning algorithm works better and more fairly. One way to stop people being unfair is to add penalties for unfairness to the loss function. Adversarial debiasing is another way to teach an enemy how to use the model's results to find the trait that is protected. If the other person loses, the model is fair. When you learn, constrained optimisation makes sure that rules like "equalised odds" are followed. In-processing methods are usually better and more flexible, but they are more difficult to understand because you need to be able to view how the model works from the inside.

### ***C. After-Processing Methods***

Post-processing methods remove bias after training the model by making all the outputs the same. Since sometimes it is impossible to change the model itself-for example, some old or proprietary systems-these methods work great. For example, calibration. It makes predictions that are fair for all. For example, it uses the same numbers. Another example is modifying the threshold. By using different decision thresholds, it keeps the number of mistakes the same for all groups. If the predictions are not clear, you can also change your mind, using reject option classification. You should only do this if it would make things fairer. If you have to change the model, people may think things aren't fair or straightforward but, usually, changing after the processing is much easier.

## 5. Case Studies and Applications

### A. Criminal Justice (such as Tools for Risk Assessment)

Machine learning aware of fairness is very much in use, almost on a daily basis, in the criminal justice system. Judges and others study things like COMPAS: Correctional Offender Management Profiling for Alternative Sanctions for guidance on whether someone should get out on bail or go to prison. Studies have since demonstrated that these methods make racially biased predictions-assigning Black defendants to a higher-risk category than white defendants who have committed similar crimes. The difference in outcomes is because the historical data is wrong, the police aren't always applying the same rules, and there were no just rules on how to train the model in the first place. One case revealing possible biases in algorithms, even when they appear to treat people with equity, is that of COMPAS. There are two ways to make the law treat everyone alike: by making the odds equal for all people and by making things fair in a way that doesn't actually happen.

**Table 2: Fairness Issues in Criminal Justice Risk Assessment (COMPAS Case Study)**

Key Element	Explanation
Purpose of Tools	Tools like COMPAS help judges predict whether a person may re-offend or should get bail.
Observed Problem	Research found COMPAS gave higher risk scores to Black defendants compared to white defendants with similar records.
Reason for Bias	The historical data used was already biased, police practices were unequal, and the model was not trained with fairness rules.
Type of Unfairness	Although the tool looks neutral, it produced racially biased outcomes, affecting certain groups more harshly.
Impact on Justice	Biased scores can lead to unfair bail decisions, longer sentences, or denial of release, harming individuals and communities.
Lesson Learned	Fairness in machine learning must include clean data, proper fairness metrics, and transparent training methods.
Why This Case Is Important	COMPAS shows that even advanced algorithms can look fair but act unfairly, reminding developers to check for hidden bias.

### B. Credit Scoring and Financial Lending

Banks and other financial institutions employ different machine learning models to make decisions on who to give a loan. Most of the models review the historic lending data for unfair lending practices, such as gender-based lending or redlining. This may make the credit issue even worse since those who already face difficulties in receiving credit will face more problems. More and more people are using methods that are fair so they don't make bad financial choices. Some of these are pre-processing methods which get rid of the effects of sensitive attributes and others are in-processing algorithms which make sure demographic parity is maintained. Regulators also say financial tools which work on their own should be fair and easy to use. This is because bad credit decisions can take a great toll on people's lives and the economy.

### C. Recruitment and Resume Screening

A lot of businesses use automated hiring tools to bring in new people. The tools include algorithms that rank candidates and resume parsers. Yet, these very tools have been shown to be prejudiced against race and gender, especially when they are trained on biased past hiring data. Amazon, for example, turned off its test hiring system after finding that resumes with the word "women's" or the names of women who had applied were less likely to get hired. Fair hiring means removal of bias in data, use of adversarial methods to prevent guessing gender, and model modifications that make sure candidates from groups not well represented get a fair deal. It's not only the right thing to do, but in many places, it's also illegal to hire people who are like you.

#### ***D. Medical Diagnostics***

Health is increasingly leveraging machine learning models to help with diagnosis, suggested treatment pathways, and resource allocation decisions. However, predictions can go astray if the medical data is not consistent, possibly a situation where there is not a sufficient number of members of a certain race or gender. Some algorithms that detect skin cancer have been found to perform very poorly on dark skin since these algorithms have not been sufficiently exposed to various skin tones. The biased health models might cause misdiagnosis, imbalanced access to care, and poor outcomes for the already vulnerable groups. Other critical fairness-aware healthcare practices include training models on diverse subgroups, monitoring subgroup performances, and counterfactual fairness, providing insight into how model predictions change when subject demographics are varied. Ensuring AI in health is fair is extremely crucial in order not to exacerbate health disparities.

### **6. Challenges and Trade-Offs**

#### ***A. Equitable versus Accurate***

The immediate effect of machine learning in trying to be fair would mean finding a delicate balance between being fair and being correct. In other words, it may be more challenging for a model once there is a set of rules about fairness to achieve its initial objective. This is also true, considering the data it has to work on is not quite even. If one is taking a racial bias perspective, such as in loan approvals, the model may fail to estimate how much money it will make. Because of this trade-off, it becomes the decision maker's job to find a middle ground between what he considers best for the business and what he wants it to do, or how well it should do. The key goal here would be to find a middle ground between making things more equitable but sacrificing some accuracy.

#### ***B. Complexity vs. Interpretability***

Some fairness-aware models require complicated procedures that are difficult to understand, such as adversarial debiasing or constrained optimisation. This is a problem in domains like law, health, and finance, where it is particularly important that things be explicable. The more complex models may be impenetrable with regards to what happened, while the simpler models, which are more interpretable, may not be able to find complex patterns. The tension between fairness and interpretability suggests that transparency regarding design choices is very important, and also explaining inner workings of models is important, especially when automated decisions have a direct impact on people.

#### ***C. Transparency versus Privacy***

Governance has to be careful about the balance between privacy and being open. Federated learning, differential privacy, and safe multi-party computation are ways of letting people decide if something is fair without putting privacy in danger. In the assurance of machine learning not being biased and fair, you may need to learn about race, gender, or even income level. However, it may be inappropriate or even unlawful to ask for such information and make use of it. Moreover, if you make the model more open by showing the features' importance or how decisions are made, it would be easier for people to get around the rules, even cheat.

#### ***D. Stakeholder Alignment and Contextual***

Not everyone-for example, stakeholders, regulators, developers, and communities who are affected by the situation-may agree on what fairness is or how to measure it. For instance, a medical diagnostic tool may look for equalised error rates, while a college admission algorithm looks for demographic parity. You really need to use participatory design processes, and bring these different points of view to work together in order to align all these different opinions. We need to know what each of the situation's values, is aiming for, and its limits to build systems that are fair and really help people.

## **7. Regulatory and Ethical Considerations**

### ***A. Legal Frameworks (e.g., GDPR, EEOC, AI Act)***

Laws increasingly influence how people use machine learning systems aware of fairness. The GDPR by the EU requires that decisions which machines make are transparent and people can be held accountable. People can ask for an explanation and give reasons for the choices made. EEOC in the United States ensures that hiring tools utilizing computers do not violate laws that protect people against discrimination. The proposed EU Artificial Intelligence Act places strict limits on the use of AI systems that might potentially cause serious harm; for example, they should be tested for fairness and human-moderated. These laws demand that businesses fairly treat people and their systems adhere to rules related to privacy and civil rights.

### ***B. Fairness's Place in Ethical AI Guidelines***

Fairness has been heavily emphasized in the ethical AI frameworks developed by a lot of businesses and governments. Great examples include IEEE's Ethically Aligned Design, Google's AI Principles, and the OECD's Principles on AI. These ideas express the notion that AI systems should not be biased against any person, harm any group, or degrade human society, but rather improve it. These moral promises may not be legally binding, yet they do alter ways of thinking and institutional practices. They consist not only of rules but also of the basis for checklists on fairness, value-based processes for AI development, and internal audits.

### ***C. Accountability and Auditing***

But fair machine learning requires more than just technical fixes. We also need systems that hold businesses accountable for what they do. Increasingly, people are using algorithmic audits to make sure their models are fair, function well, and don't give one group an unfair advantage. Audits happen all the time. Audits may be performed by people inside or outside the company, may be voluntary, or may be required. They often review lists of stakeholders and do impact analyses and independent reviews of their findings by third parties outside their company. Companies are also keeping track with "model cards" and "datasheets for datasets" of the fairness, limitations, and intended uses of AI systems. These actions provide much more transparency and help build a culture of accountability around algorithm development and deployment.

## **8. Future Directions**

### ***A. Equitable Feedback Loops and Dynamic Systems***

Fair systems and ways for people to give feedback Because of that Shift Dynamic, many automated decision-making systems evolve over time and impact the things surrounding them. For example, predictive policing systems can alter how police officers do their jobs. This means they will have to learn new things and get new kinds of information in the future. These kinds of feedback loops have the potential to make biases worse. To ensure equity in dynamic systems, it is crucial to foresee long-term outcomes using causal modelling, simulations, and longitudinal analysis. For long-term fairness, future studies will have to investigate how fairness-aware algorithms work operationally over time and ways of intervening in complex socio-technical systems.

### ***B. Participatory and Multi-Stakeholder Design***

Developers cannot guarantee fairness of ML systems just by bug fixing. Design also needs to involve the people who will be impacted. Stakeholders who care about a system, though not well-represented, are asked to set goals about fairness, help create indicators, and use participatory design methods to discover how that system impacts them. Opening up AI design to anyone imbues the design with reality, morality, and awareness of what is happening in the world. This requires future research to develop appropriate tools and frameworks which enable participatory fairness evaluation and governance involving a variety of stakeholders

### ***C. Equity in Generative and Multimodal Models***

The emergence of multimodal models, for instance, vision-language models, and large generative models, such as GPT and DALL-E, has brought out new fairness issues. The models have the capability to learn and amplify societal

biases from large training datasets to produce results that reflect and perpetuate prejudices. Given the wide range of issues, such as the generation of toxic content, biased linguistic phenomena, and inconsistent performance across different groups, this calls for complex architectures in fairness research. Bias audits, fairness fine-tuning, and debiasing are on the rise in frequency and popularity, even though they are still considered novel methods.

#### **D. Tools and Standards for Assessing Fairness**

For fairness to work in the real world, we need benchmarks and standard toolkits. If-Then from Google, Fairlearn from Microsoft, and AI Fairness 360 from IBM are open-source libraries enabling developers to take a closer look at the algorithms, measurements, and visualizations around fairness. What is still needed is more domain-specific benchmarks for each field, reflecting realistic difficulty. Standardized datasets and challenge tasks streamline collaboration among diverse groups in the study of fairness. There are advantages to this, making replication of results and reproduction of experiments far easier.

## **9. Conclusion**

### **A. A Summary of the Need for Equity in Machine Learning**

As systems of machine learning become increasingly vital in providing access to key resources and opportunities, it also means that it is our moral obligation, together with our social one, to ensure that the systems are nothing short of fair. If biases of the past do not get fixed, they can transfer to other groups and be applied in algorithms. This will also be part of an effort at making AI open, responsible, and fair for all.

### **B. Overview of Key Findings**

This paper has looked into fairness from several dimensions: conceptual, technical, legal, and practical. We have looked at key definitions of fairness and the associated trade-offs. We have discussed three key types of fairness-aware approaches: pre-processing, in-processing, and post-processing. We have talked also about how they can be helpful in real life. We have discussed the new challenges and possible directions going ahead, along with the legal and moral norms that have shaped this area.

### **C. An Appeal to Scholars, Decision-Making Officials, and Professionals**

Not only computer workers but also technology students, legal experts, politicians, ethicists, and social scientists should all be brought together. Researchers must find approaches that are fair and can work under a wide range of circumstances. Lawmakers should make sure that the laws which guarantee online rights for people are enforced. AI developers should pledge to use it judiciously, be in collaboration with others to create it, and be transparent about what they do. In constructing machine learning systems that are intelligent but just, everybody has something to contribute.

## **10. References**

- [1] Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*. fairmlbook.org.
- [2] Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153–163.
- [3] Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. *Proceedings of Innovations in Theoretical Computer Science*.
- [4] Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*.
- [5] Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency (FAT)\**.
- [6] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6), 1–35.
- [7] Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *AAAI/ACM Conference on AI Ethics and Society*.

- [8] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. *ProPublica*.
- [9] Ustun, B., & Gebru, T. (2019). Fairness audits and audits of fairness. *ACM FAccT*.
- [10] Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 33(1), 1–33.
- [11] Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap. *FAT Conference*.
- [12] Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? *CHI Conference on Human Factors in Computing Systems*.
- [13] Binns, R. (2020). On the apparent conflict between individual and group fairness. *Proceedings of ACM FAccT*.
- [14] Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and Discrimination*.
- [15] European Commission. (2021). Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act).