BIGMUN 2025

ECOSOC 4: Commission on Science and Technology for Development (CSTD)

Research Report

Topic 1: Establishing policies on bias in machine learning as a result of human involvement.

DENIZ TARHAN & EVA DUZINSCHI

Table of Contents

Introduction

With technology and mechanical research evolving every day, the importance of the reliability of machines have increased. AI and other technologies have been getting integrated in various areas from healthcare to law enforcement. Efficiency in the modern world is playing a larger part in life than ever, and autonomous systems are developing to solve this need. Even though AI provides many gateways for easier decision making, faster and more efficient research, system-based recruitment, diagnosis, organisation yet it also carries a large error that is still unfixed today: ML bias.

Since Machine Learning Principles were formed by humans, models like AI act with bias in many cases. In the process of machine learning, the program is given a set of data and based on this data, the machine acts.[1] Bias enables the recognition of patterns which is a highly efficient aspect for machine learning however this recognition of patterns is in many cases impractical. In the modern world, the cruciality of human rights have been emphasised yet the faults in machine learning contradicts with the rest of actions against inequality. When societal and personal biases are integrated in machine learning, then the machines and programs work with a discriminative mindset, enforcing inequality with unfair decisions and providing less accurate outcomes.

[1] Schwartz, Oscar. "Untold history of AI: How Amazon's mechanical turkers got squeezed inside the machine." *IEEE Spectrum*, 30 Sept. 2021, spectrum.ieee.org/untold-history-of-ai-mechanical-turk-revisited-tktkt.

Definition of Key Terms

Accountability: is the state of being responsible to a certain action. During the detection and revolution for machine learning with bias, accountability must increase.

Artificial Intelligence Bias: Refers to the discriminative systems inside AI programmes that lead to inequality and chaos. This term also refers to machine learning bias, an essential term in the debate. During the process of machine learning, if at any point, bias has also been integrated, then the system itself acts upon bias. [2]

Artificial Intelligence: is the developing technological system where the computer programme can act on given tasks and provide responses mimicking human intelligence with the relevant data is supplied.[3]

Human Bias: Bias is found to be a phycological tendency towards ideas whether it is in favour or against. Bias can develop from anything whether it is experiences, social backgrounds, economic conditions, race or education. It is a way for the brain to adapt and survive in the environment, developing positive or negative ideas. While it is common, it is highly disruptive for the person and the surroundings, effecting societal development worldwide and enforcing inequalities.

Machine Learning (ML): Refers to the development of technological systems where with the analysis of the provided data, the computer systems learn and act. ML is the basis of this issue since the given data effects the learning process highly potentially creating biases on several aspects. ML is an essential part in modern technology which makes AI possible with the algorithms that know and understand data like humans. [4]

Transparency: refers to the clarity and explainability of a certain action or system. In this

---

[2] SAP. "What Is AI Bias? Causes, Effects, and Mitigation Strategies." *SAP*, 2025, https://www.sap.com/resources/what-is-ai-bias.

[3] Oxford Dictionary. "Artificial Intelligence", https://www.oed.com/dictionary/artificial-intelligence_n?tl=true , Access date: 15.12.2025

[4] "Machine Learning Pipeline." IBM, www.ibm.com/think/topics/machine-learning-pipeline.

context, transparency is a must for machine learning since to be able to find the root cause of bias in machines, every decision and step must be transparent for researchers. It is also important to add that the result given by the system must be clear, giving the clear reasons on why that decision was made.

Background Information

While developing machines, engineers struggle to keep machines in a lower bias rate aiming to reveal accurate patterns and detection. The reason for this struggle comes from the fact that most of the data received are highly biased. With the rise of trend of reaching high efficiency and automation for profit has increased the frequency of bias in machines. The way of leading a machine learning characterised by bias can happen for many different reasons.

One of the first instances where bias was seen was in 1979. At St. George Hospital Medical School, it was decided that hiring processes were taking too long. A more efficient way of conducting this process was by developing a computer programme that aped the previous choices and applied to the criteria. This programme was finalised in 1979 and in 1982 all the applicants were evaluated by this system. Soon after it was noticed that there was a lack of diversity between the people that were hired. After thorough research, it was found that the machine was weighing down applicants by unnecessary criteria such as name and place of birth, in addition, classifying the people by Caucasian and non-Caucasian, male and female. Unfortunately, this incident was just the start of many others.

Machine learning has a clear system that provides a structured plan for achieving machine learning smoothly. The process that involves the design, development and finalisation is named "The Machine Learning Pipeline".[5] First, the plan starts with deciding on the problem and setting a clear goal. This stage is referred to as stage 0 since there is no clear action present just the first step for a more efficient model. Stage 1 is considered the most critical especially for the issue of bias. In this part, data processing occurs. All the necessary data is collected and inserted into the model. While exploring the pipeline, it is beneficial to consider where some of the types of biases occur. During data processing, since the data reflect understanding of the past it can clearly reflect

[5] "Machine Learning Pipeline." IBM, www.ibm.com/think/topics/machine-learning-pipeline.

historical bias to the model. For example, if the data shows that males have been favoured against women for a certain task, then from the data, the model will act the same, repeating and enforcing inequalities.

Another type that is seen often is exclusion bias. While developers collect data sometimes data that is deemed unnecessary may be critical. If this seemingly useless data is not included in the model, the machine will not act upon enough knowledge and prepare faulty outputs. Example for this is if data regarding a minority is excluded, in this case, underrepresentation will be strengthened. During data labelling, an important step is data labelling. Here, sets of data are prepared and presented to the model. If labelling is inconsistent in any way, recall bias occurs. Since the data that the model receives might have small differences yet labelled the same, the model will not be sensitive enough to produce truthful, reliable, and fair outcomes. Finally, if the data that is provided does not reveal the truth, this type of bias is named sample bias. Each of these biases have similar starting points will dangerously results. The next stage is known as model development. The implementation and development of algorithm take place. When the algorithm reflects bias, then the type of bias is algorithmic bias. The final stage of the ML pipeline is model deployment and here the model is finalised and prepared for use. As it could be understood, at any and every stage of the machine learning process, the machine could become biased and for this reason it is critical to assess every stage of the process of machine learning while searching for the root cause of bias.[6]

An emphasis for the fact that, the most common types of bias are seen during the first stage of machine learning also known as data processing is because throughout history people have acted on bias and the data reflects all of it, is crucial while exploring a way for preparing solutions.
The consequences of ML bias are severe and must be acknowledged by the member states. Bias results in wrongful outcomes, reinforcing inequalities for race and gender. Bias is an alarming risk for human rights since it can easily undermine anyone with just a single fault. Since the model will provide results with low accuracy it will affect corporates that use it. This will lead to many financial and legal problems. For text and image generation, study has shown that light skinned males are portrayed better than others. This amplifies gender stereotypes and increases the wrong depictions of minorities or races. While encouraging discrimination, these models

[6] "Machine Learning Pipeline." IBM, www.ibm.com/think/topics/machine-learning-pipeline.

negatively affect education, hiring processes as it has become common to use AI in these areas. The presence of errored conclusions disturbs policing and management as well. All in all, machine learning bias poses great threats and must be reduced significantly with clear functional policies.

Major Countries and Organisations Involved

United States of America - The U.S. has taken important national steps for ensuring safety and transparency in machine learning. The White House Office of Science and Technology Policy have been critical for they have prepared the Blueprint for an AI Bill of Rights. This framework is unfortunately now nonbinding, yet it lays important rules and principles to follow for the development and use of AI. This action has accelerated the work towards establishing AI issues that are being faced which includes bias. An organisation under the United States called National Institute of Standards and Technology is worth mentioning. The organisation has assembled the AI Risk Management Framework, a standard for the design and development of reliable AI. The work emphasises that the root of problems must be dealt with to prevent more inaccurate outcomes. National policy has been evolving with respect to the initiatives. Finally, the Executive Order 14110, released on October 23rd of 2023 is a clear sign that the government has been working on principled ways for fair and safe AI.

China - The People's Republic of China has had the role of being one of the first states to develop legitimate rules for algorithmic recommendation programs and generative AI systems. The prevention of discrimination thus bias has been a clear point of focus. The draft of the rules for AI has a part entirely dedicated to discrimination. The issue has been taken as a risk for governance and social balance. The draft emphasises that clear measures must be taken against discrimination and acknowledges the different reasons for this issue. The problem has been very clearly found and stated however there could be improvement for implementation. There still is not a clear standard for the assessment of data and identification of bias. It has been underlined that transparency plays an important role for the issue. The Cyber Space Administration of China has guided the preparation of the AI Safety Governance Framework 2.0, clearly demonstrating interest and commitment to the issue faced by AI.[7]

[7] "Overview of Draft Measures on Generative Artificial Intelligence." China Law Translate, www.chinalawtranslate.com/overview-of-draft-measures-on-generative-ai/.

European Union - The EU has been the leader for understanding and regulating AI. The European Union AI Act by the EU AI Office is the first administrative and inclusive framework in the world. The assembly of requirements and guidelines include the development and safe use of AI aiming for minimised bias and transparency. With the goal of increasing compliance, a fine of 35,000,000 € has been included. The AI act is an exceptional material to take reference from for member states as it outlines many of the possible solutions to be found. One of the requirements are the impacts analyses for more ethical development.[8] ,

Canada - When speaking of Canada, the Artificial Intelligence and Data Act must be mentioned. This act aims to enforce human rights laws and target violations for constraining discrimination. AI bias has remained implicit in many of the solutions however discrimination has clearly been targeted. AIDA, though a wholesome act has not been approved by the Advisory Council on AI thus has not yet been implemented. Canada's efforts against AI bias must be considered while drafting resolutions.

United Kingdom - One of the eye-catching initiatives by the U.K. has been the Fairness Innovation Challenge. This competition, by the Department of Science and Technology, has played ground for scholars and researchers to collaborate in finding solutions regarding AI bias and other relevant issues. The different areas where AI plays a role such as healthcare and education has been emphasised. One project promotes fairness in higher education while the other is about early warning systems placed in health care and how it puts lives at risk when providing biased and discriminative conclusions especially for underrepresented people. The government of U.K. has also prepared the White Paper, namely The UK Science and Technology Framework, that highlights the necessity of monitoring and improving AI systems. It has been a guideline that encourages innovation and new solutions. The UK has hosted the first ever AI Safety Summit.[9]

International Organisation for Standardisation- An organisation that is worth mentioning is

---

[8] Kosinski, Matt, and Mark Scapicchio. "What Is the Artificial Intelligence Act of the European Union (EU AI Act)?" IBM Think, www.ibm.com/think/topics/eu-ai-act "EU Artificial Intelligence Act." Artificial Intelligence Act, artificialintelligenceact.eu.

[9] The AI Safety Summit will be discussed more during the part about Past Attempts.

the International Organisation for Standardisation (ISO). The aim of this organisation is to compose global principles on mechanical systems for enhancing reliable AI models. Valuing systematic solutions for handling and thus reducing complications about AI, they have provided a framework, the International Organisation for Standardisation 420001, that includes the first ever Artificial Intelligence guidelines with a clear cycle-based plan for improvement. In addition, a report for AI bias has been published and touched upon core principles and techniques, an advantageous guideline for improving solutions on ML bias. Their attempts for standardising technical language in the world have been inspirational for other organisations.

Relevant UN Resolutions

1. *A/HRC/RES/47/23*

The Human Right Council Resolution 47/23 was issued on 13th of July in 2021 for the topic of "New and emerging digital technologies and human rights". This foundational resolution introduces the idea that technology can be used for inclusivity and for protecting human rights. The preambulatory clause below states:

"*Recognizing also* the risks that new and emerging digital technologies may have for the protection, promotion and enjoyment of human rights, including but not limited to the right to equality and non-discrimination, the right to freedom of opinion and expression, the rights to freedom of peaceful assembly and freedom of association, the right to an effective remedy and the right to privacy, in accordance with States' obligations under international human rights law,"[10]

Even though the focus of the resolution is not machine learning bias, its suggestion for organising impact assessments for developing technologies is highly related. This resolution has been inspiration to many other resolutions making it one of the first concrete examples of the UN's actions against discriminative technologies.

2. *A/RES/78/311*

The focus and attention have newly been turning into AI as the incidents have demonstrated the

[10] "Resolution 47/23." United Nations Human Rights Council, digitallibrary.un.org/nanna/record/3936036/files/A_HRC_RES_47_23-EN.pdf.

risks it poses. Adopted on the first of July 2025 by the UN General Assembly during the seventy eighth session for the issue of *Enhancing International Cooperation on Capacity Building of Artificial Intelligence.* This resolution does not have its explicit focus on AI bias however reaffirms the resolution below A/RES/78/265 for its solutions. Bias is managed through the efforts on capacity building and cooperation. Operative clause 7 includes:

"*To adopt proactive measures to counteract racism, discrimination and other forms of algorithmic bias".*[11]

The resolution holds clear attention in developing the systems that are respectful to human rights. Operative Clause 1 states that technological divides must be closed especially in the case of AI. This clause highlights the importance building capacity and strengthening defences as a means for fighting against bias and unfair circumstances.

3. A/RES/78/265

Issued on 21 March 2024 by the UN General Assembly, this resolution holds a clearer focus on AI bias. The agenda item is *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development*. The resolution, in this issue, is extremely important as it defines a core standard for secure AI increasing the ethical aspect of ML development against bias while drawing attention to human rights. The preambulatory clause is clear evidence of the papers' focus on developing technological systems that will not harm society and instead protect the sustainable development goals.

"*Recognizing also that the improper or malicious design, development, deployment and use of artificial intelligence systems, … could hinder progress towards the achievement of the 2030 Agenda for Sustainable Development and its Sustainable Development Goals and undermine sustainable development in its three dimensions – economic, social and environmental; widen digital divides between and within countries; reinforce structural inequalities and biases; lead to discrimination; undermine information integrity and access to information; undercut the protection, promotion and enjoyment of human rights and fundamental freedoms, including the right not to be subject to unlawful or arbitrary interference with one's privacy; and increase the potential risk for accidents and compound threats from malicious actors…*

[11] "Resolution 78/311." United Nations General Assembly, docs.un.org/en/A/res/78/311.

The resolution overall shows that bias and discrimination can affect all areas in life and must be dealt with immediately. Reminding that this resolution was constructed nearly a year before 78/311 reflects that the problem and its importance have become a baseline for active solutions included in the more recent documents.[12]"

4. *Recommendation on the Ethics of AI*[13]

Though technically this is not a resolution, it is the most influential paper and the groundwork written about AI principles. Prepared by the United Nations Educational, Scientific and Cultural Organisation in 2021, it highlights equity, mentions the way of ensuring implementation and proposes clear principles. The recommendation includes ethical assessments, data controls, transparency measures and many more. Precisely the issue of discrimination is mentioned in the 29th paragraph as given below:

"AI actors should make all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of the AI system to ensure fairness of such systems. Effective remedy should be available against discrimination and biased algorithmic determination."

Previous Attempts to Address the Issue

Principles for the Ethical Use of Artificial Intelligence in the United Nations System:[14]

Aside from independent national solutions and UN resolutions, the work that includes principles towards AI is highly influential today. Published on September 20, 2022, by the High-Level Committee on Programmes, the document starts with a clear identification of the current problem faced by AI explicitly referencing the recommendation made by UNESCO. The principles include a clear section dedicated to fairness and equity as well as transparency. Though the

[12] "Resolution 78/265." United Nations General Assembly, docs.un.org/en/A/res/78/265.

[13] Delegates are specifically advised to read the whole document as this will be a beneficial resource for resolution drafting and the debate.

[14] "Principles for the Ethical Use of Artificial Intelligence in the United Nations System." United Nations System Chief Executives Board for Coordination, Sept. 2022, unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf.

document clearly acknowledges that discrimination must come to an end when speaking of AI, lacks implementational aspects, Member States are advised to read the paper beforehand and comply to the principles provided by the UN.

ACM Conference on Fairness

This conference focuses solely on Algorithmic bias. Researchers and scholars come together from all around the world and examine the issues of ML while debating on solutions. The conference is known to be one of the sources of debiasing research and solutions. Admirably, the conference hosts people from many nations increasing inclusivity and equality. The conference has become an environment for researching and discussing new ways to solve the root causes of bias that is damaging equality and challenging transparency.[15]

AI Safety Summit:

In the summit, influential Artificial Intelligence companies, NGOs, scholars and governments collaborate on understanding AI, ensuring its safety by analysing the risks it poses, the root causes of the risks and most importantly how, collaboratively, these issues can be solved. The summit provides ground for enhanced cooperation gathering many representatives with unique knowledge. In 2023, from the summit The Bletchley Declaration was prepared which focuses on Artificial Intelligence risks that affect many different areas. Bias is a deep concern, according to the declaration, that promises great damage. It emphasises the importance of collaboration, accountability, and regulation.[16]

Proposed Solutions

Analytical Assessment:

This idea has been introduced by many different bodies yet failed to reach the positive outcome that was expected. The assessments must be detailed clearly ensuring that discriminative conclusions from the past never repeat again. There could be several groups that investigate different stages within the Machine Learning Pipeline. This way the source of bias whether it is the data, or some other variable can be detected easily before any tragedy and widespread

[15] "ACM Conference on Fairness, Accountability, and Transparency." FAccT Conference, facctconference.org.

[16] "AI Safety Summit 2023." UK Government, www.gov.uk/government/topical-events/ai-safety-summit-2023.

misinformation. These analytical assessments must be clearly stated, with the consideration of every possible action for maximum efficiency in the proposed resolutions.

Impact Research:

This solution is not directed to the algorithms or data of machine learning. In this case, no matter the efficient purpose of the clause, bias that has already been done cannot be erased, and the damage will not disappear when the solutions only focus on what will happen in the future. There have been countless instances of bias hurting people, and the impacts of bias must be further elaborated. Research must be organised with the consideration of all member states for aiding the societal concerns bias poses. By clarifying the evidence on how bias effects people, the entirety of solutions will be more inclusive.

Capacity Building Measures:

As mentioned before in the UN resolutions section, Capacity building is another very important aspect while handling machine learning development. There are member states that are developing and that do not have the same number of resources for rapid change. This vulnerability is a great risk for the technical divide between states. There must be found a mutual way to bridge the gaps between technologies and determine a standard that all nations can follow.

Bibliography

"ACM Conference on Fairness, Accountability, and Transparency." FAccT Conference, facctconference.org.

"Addressing AI Hallucinations and Bias." MIT Sloan Educational Technology, Massachusetts Institute of Technology, mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/.

"AI Safety Summit 2023." UK Government, www.gov.uk/government/topical-events/ai-safety-summit-2023.

"Algorithmic Fairness in Machine Learning Systems." ACM Digital Library, doi.org/10.1145/3468264.3468537.

"Artificial Intelligence and the Law: Bias and Accountability." International Association for Artificial Intelligence and Law, iacajournal.org/articles/598/files/68107ee94c966.pdf.

"Artificial Intelligence." United Nations, www.un.org/en/global-issues/artificial-intelligence.

"Bias and Fairness in Artificial Intelligence Systems." Applied Sciences, vol. 14, no. 19, 2024, article 8860, www.mdpi.com/2076-3417/14/19/8860.

"EU Artificial Intelligence Act." Artificial Intelligence Act, artificialintelligenceact.eu.

"Fairness: Types of Bias." Google for Developers, developers.google.com/machine-learning/crash-course/fairness/types-of-bias.

"How China Views AI Risks and What to Do about Them." Carnegie Endowment for International Peace, Oct. 2025, carnegieendowment.org/research/2025/10/how-china-views-ai-risks-and-what-to-do-about-them.

"ISO/IEC 42001: Artificial Intelligence Management System." International Organization for Standardization, www.iso.org/standard/42001

"Machine Learning Bias (Algorithm Bias or AI Bias)." TechTarget, www.techtarget.com/searchenterpriseai/definition/machine-learning-bias-algorithm-bias-or-AI-bias

"Machine Learning Bias and Ethical Risk." ACM Digital Library, doi.org/10.1145/3492853.

"Machine Learning Bias." Deepchecks, www.deepchecks.com/glossary/machine-learning-bias/.

"Machine Learning Pipeline." IBM, www.ibm.com/think/topics/machine-learning-pipeline.

"Measures for the Management of Generative Artificial Intelligence Services." Cyberspace Administration of China, 15 Sept. 2025, www.cac.gov.cn/2025-09/15/c_1759653448369123.htm.

"Overview of Draft Measures on Generative Artificial Intelligence." China Law Translate, www.chinalawtranslate.com/overview-of-draft-measures-on-generative-ai/.

"Principles for the Ethical Use of Artificial Intelligence in the United Nations System." United Nations System Chief Executives Board for Coordination, Sept. 2022, unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf.

"Recommendation on the Ethics of Artificial Intelligence." UNESCO, www.unesco.org/en/artificial-intelligence/recommendation-ethics.

"Regulation of Artificial Intelligence in Canada." Osler, www.osler.com/en/insights/reports/ai-in-canada/regulation-of-ai-in-canada/.

"Resolution 47/23." United Nations Human Rights Council, digitallibrary.un.org/nanna/record/3936036/files/A_HRC_RES_47_23-EN.pdf.

"Resolution 78/265." United Nations General Assembly, docs.un.org/en/A/res/78/265.

"Resolution 78/311." United Nations General Assembly, docs.un.org/en/A/res/78/311.

"Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." Federal Register, 1 Nov. 2023, www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

"The Fairness Innovation Challenge: Key Findings." UK Government, www.gov.uk/government/publications/the-fairness-innovation-challenge-key-findings/the-fairness-innovation-challenge-key-findings.

"What About Fairness, Bias, and Discrimination?" Information Commissioner's Office, ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-about-fairness-bias-and-discrimination/.

Coursera. "What Is Bias in Machine Learning?" Coursera, 7 Apr. 2025, www.coursera.org/articles/what-is-bias-in-machine-learning.

GeeksforGeeks. "Bias and Variance in Machine Learning." GeeksforGeeks, 4 Dec. 2025, www.geeksforgeeks.org/machine-learning/bias-vs-variance-in-machine-learning.

Holdsworth, James. "AI Bias." IBM, 17 Nov. 2025, www.ibm.com/think/topics/ai-bias.

SAP. "What Is AI Bias? Causes, Effects, and Mitigation Strategies." *SAP*, 2025, https://www.sap.com/resources/what-is-ai-bias.

Schwartz, Oscar. "Untold History of AI: The Birth of Machine Bias." IEEE Spectrum, 30 Sept.

2021, spectrum.ieee.org/untold-history-of-ai-the-birth-of-machine-bias.

Schwartz, Oscar. “Untold history of AI: How Amazon’s mechanical turkers got squeezed inside the machine.” *IEEE Spectrum*, 30 Sept. 2021, spectrum.ieee.org/untold-history-of-ai-mechanical-turk-revisited-tktkt.

Kosinski, Matt, and Mark Scapicchio. “What Is the Artificial Intelligence Act of the European Union (EU AI Act)?” IBM Think, www.ibm.com/think/topics/eu-ai-act