NIMUN' 26

UNHRC

United Nations Human Rights Council

STUDYGUIDE



United Nations Human Rights Council

Topic: AI Bias and Algorithmic Discrimination: Human Rights in Automated Decision-Making

Table of Contents

1. Committee Overview 3

1.1 Mandate and Responsibilities of the Council 3

1.2 Structure and Functioning of the Council 3

2. Introduction to Agenda 4

2.1 Definition and Scope of AI Bias 4

2.2 Context of Automated Decision-Making 5

2.3 Scale and Severity of the Issue 5

3. Glossary and Key Terms 6

4. Background 7

4.1 Evolution of AI in Decision-Making Process 7

4.2 Historical Context of Discrimination 7

4.3 Emergence of Algorithmic Discrimination as a Human Rights Concern 8

5. Case Studies 10

5.1 Past Implications 10

5.1.1 Case Study 1: Gender Bias in Employment (Amazon Recruitment Tool, 2014-2018) 10

5.1.2 Case Study 2: Racial Bias in Criminal Justice (COMPAS Algorithm and state v Loomis, 2016) 11

5.2 Present Implications 11

6. Past UN Actions and Resolutions 13

6.1 Actionable Policy Options 15

7. Questions A Resolution Must Answer (QARMA) 16

8. Bibliography 17

Authors & Contributions

Editorial Oversight: Zainab Fawad - Director Publications

Editors

Safa Nasar - Deputy Director Publications

Saifullah Shamim - Deputy Director Publications

Coordinators

Mahad Saeed-Director Secretariat Affairs

Contributing Authors

Team	Authors
Publications	Muhammad Hussain Sair
Secretariat Affairs	Muhammad Basim Manzoor

1. Committee Overview

1.1 Mandate and Responsibilities of the Council

The United Nations Human Rights Council (UNHRC), created in 2006, works to promote and protect human rights worldwide. It addresses violations, encourages prevention through dialogue and cooperation, and issues policy recommendations. With the rise of digital technologies, its mandate now includes human rights in the digital age—covering AI, data protection, algorithmic transparency, and corporate accountability. Through resolutions like A/HRC/42/15 and A/HRC/59/32, the Council ensures that emerging technologies respect international human rights law, protect privacy, and promote responsible innovation.

The principal objectives of the UN Human Rights Council's mandate comprise:

1. The Council's mandate comes mainly from **Resolution 60/251** which includes promotion and protection of human rights, addressing violations of human rights, Universal Periodic Review (UPR), work with special procedures (that may include special rapporteurs, independent experts and working groups) as well as cooperation with states.
2. The Council is entrusted with the responsibility to champion the respect for and protection of the rights and fundamental freedoms of every individual, without distinction or discrimination.
3. It is imperative to respond effectively to instances of widespread human rights violations and to proffer viable suggestions for their resolution.
4. Fostering collaboration constitutes a core focus of the UN Human Rights Council (UNHRC), which aims to facilitate substantive discussions and strategic partnerships on human rights issues to enhance the coherence of human rights initiatives within the broader United Nations system.
5. The Council provides essential support and training, including capacity building and technical assistance, to member states to aid them in fulfilling their international human rights obligations.

1.2 Structure and Functioning of the Council

The Council has 47 member states. Members are elected by the UN General Assembly through a secret ballot. Seats are distributed by regional groups:

a. Africa (13)
b. Asia-Pacific (13)
c. Latin America & Caribbean (8)
d. Western Europe & Others (7)

e. Eastern Europe (6)

Members serve 3-year terms and may not serve more than two consecutive terms. UNHRC holds three regular sessions each year (March, June, September). Special sessions may be called out at any time upon request of 1/3 of member states (in the event of emergencies such as conflicts, genocides, etc.). Certain subsidiary bodies may work under the Council providing technical support and the Council may refer any said matter to any of its subsidiary bodies for further evaluation. These may include Universal Periodic Review (UPR) Working Group (reviews human rights performance of all countries), Complaint Procedure (allows individuals and groups to submit complaints), Advisory Committee (think-tank providing expertise), Fact-finding missions and commissions of inquiry (created for specific country situations). The Council is entitled to also appoint independent human rights experts to monitor and report on certain conflicts, or thematic violence of any sort.

2. Introduction to Agenda

2.1 Definition and Scope of AI Bias

AI bias, alongside algorithmic discrimination poses a major threat to human rights in the modern era, impacting countless people worldwide as automated decision-making machines become more prevalent in important aspects of life [1]. AI bias emerges if the machine models unfairly favour a certain group while it disadvantages another one, caused by skewed training data or an incorrectly designed learning system for the AI [2]. This bias becomes discrimination when this leads to AI systems being able to restrict or deny an individual's access to opportunities, goods and services solely due to their characteristics such as race, ethnicity, gender, disability or income level.

The scope of AI bias and algorithmic discrimination is wide and yet, still growing. Criminal Justice systems increasingly make use of AI in order to show chances of the offender making the same mistake, an appropriate bail amount and make jail sentence decisions. Research demonstrated that these AI's frequently displayed race bias, leading to people of color receiving harsher punishment than others [3]. Healthcare systems utilize AI for diagnosis, recommended treatments and management of supplies. Data revealed that the AI provided less accurate results for ethnic minorities leading to disparities in the care they received [4]. Employment systems use AI as well to filter CV's, evaluate interviews and pick candidates, however evidence shows that a lot of times AI puts women and specific ethnicities at a disadvantage when choosing who gets hired [5]. Banking systems reported having smarter models to verify a person's credit scoring and loan approvals, though patterns showed that the approval rates were inexplicably high among disadvantaged populations [6].

2.2 Context of Automated Decision-Making

Automated decision-making systems have spread rapidly in recent years, largely due to recent advancements in computer processing, availability of large data collections and business incentives to automate decision making processes [7]. These systems bring advantages like speed, consistency alongside being able to process large volumes of data efficiently [8]. However, when these systems are biased, using them can violate people's basic rights more widely and quicker than ever before, in turn affecting millions of individuals at once through decisions that would be impractical for a human to make [9].

The context in which AI's make decisions differs significantly from how people usually land on their choices, especially when it's relevant to the protection of a human's rights [10]. And because automated systems can operate at a larger scale and speed beyond the capabilities of a human, it means that if a system contains bias, the bias will affect an uncountable number of decisions before it is identified and fixed [11]. Algorithmic decisions lack clarity, not stating their reasons for the final decision made, which makes it difficult for affected individuals or overseers to understand whether a decision was fair or due to a bias [12]. Additionally, algorithmic decision-making lacks the sense of human responsibility, as the prosecutor often refuses to claim the decision and state it was merely the choice of the algorithm [13].

2.3 Scale and Severity of the Issue

The speed at which AI is being used in key areas such as healthcare and judicial cases makes tackling this bias more important especially since the rights of humans are at stake [14][15]. While previous forms of discrimination could be fought using established laws, automated decision-making introduces new problems which the legal framework may not be able to address [16]. Due to the complexity of the AI models and how hard it is to interpret its thought process, people may not even realise they've been discriminated against. Without clear insight into how these choices are made, proving harmful intentions or even contesting the AI's decisions becomes immensely difficult [17].

The risk that's faced by marginalized groups from biased algorithms is immensely high. Evidence consistently demonstrates that AI systems in use in criminal justice cases, healthcare, employment, housing and welfare systems significantly harm ethnic minorities, women, people with disabilities and people of lower income [18]. These communities often don't have the power to contest the system and might also struggle to obtain legal support after being persecuted [19]. The urgency is further heightened because the algorithmic systems tend to repeat and worsen their responses to certain situations, which could possibly increase societal division and undermine democratic values [20].

3. Glossary and Key Terms

1. **Artificial Intelligence (AI)** is defined as technologies that allow machines to perform tasks that normally require human intelligence, such as decision-making, recognition, or prediction.
2. **Algorithms** are a set of instructions that a computer follows to solve a problem or make a decision.
3. **AI Bias** is systematic unfairness in AI outputs caused by biased data, flawed design, or societal inequalities embedded in the system.
4. **Algorithmic Discrimination** is defined as the procedure when automated systems treat individuals unfairly based on characteristics like race, gender, ethnicity, age, socioeconomic status, disability, or religion.
5. **Algorithmic Impact Assessment** is defined as formal evaluation done before deploying an AI system to identify potential risks, discrimination, or human rights violations.
6. **Automated Decision Making** are processes where algorithms or AI models make decisions without, or with very limited human involvement.
7. **Black Box Decision Making** are AI-generated decisions that cannot be explained due to complexity or lack of transparency.
8. **Criminal Justice Algorithms are** AI tools used to predict re-offending risk, assign bail, or influence sentencing often criticized for racial bias.
9. **Civil Society Organizations are** NGOs and advocacy groups that research and expose harm caused by AI systems.
10. **Data Bias** is defined as data skewed, imbalanced, or unrepresentative, leading the AI to favor or discriminate against certain groups.
11. **Digital Profiling** is the use of automated tools to analyze personal data and profile individuals, often used in immigration, advertising, or policing.
12. **Ethical AI** is the usage of AI in light of guidelines emphasizing fairness, transparency, accountability, and human rights in AI development.
13. **Human Rights Framework** is the set of international laws and principles (e.g., UDHR, ICCPR) that protect individuals from discrimination including algorithmic forms.
14. **Fairness Metrics** are defined as technical measures used to evaluate whether an AI model treats different demographic groups equally.
15. **Machine Learning Models** are AI systems that learn patterns from data instead of following fixed, programmed rules.
16. **Risk Assessment Software** are tools used in criminal justice to evaluate the likelihood of an individual reoffending shown to produce racial biases.
17. **Systemic Bias** are defined as biases that arise from underlying social, economic, or historical inequalities embedded in data.
18. **Skewed Data** is unbalanced or unrepresentative datasets that lead to unfair outcomes in AI decision-making.

19. **Surveillance Technologies** are AI-driven monitoring tools that may include facial recognition, biometric databases that raise data breach concerns.

4. Background

4.1 Evolution of AI in Decision-Making Process

AI has undergone immense development over the past two decades, evolving from a tool to test the limitations of computer intelligence to being present in some of the most important systems, making decisions which affect individuals lives [21]. Early AI systems in the 1980s and 1990s only worked in limited areas and a fixed set of instructions [22]. The emergence of machine learning in the 2000s brought forth more flexible AI systems which could adapt to new data and perform across diverse fields [23].

The rapid development of machine learning, combined with the growing availability of large data sets and higher computational power has enabled AI systems to be used in key fields like law enforcement, criminal justice, healthcare, banking and government [24]. Many organizations adopted AI systems aiming for faster processes and lower expenses, often without taking into consideration the potential impact it might have on human rights [25]. Early adopters of algorithmic decision-making in legal cases and social welfare systems often lacked checks to discover bias' or finding actual impact of the decisions [26]. Over time, research talking about the bias in AI's deployed in systems has led to an increased amount of attention to human rights risk caused by AI.

4.2 Historical Context of Discrimination

To understand algorithmic discrimination, it is important to recognize that it does not appear erratically, rather it emerges from and reflects historical patterns of discrimination rooted in societies, organizations and data [27]. Historical discrimination against racial and ethnic minorities, women, people with disabilities, and other marginalized groups has created long standing inequalities in wealth, health, employment, criminal justice outcomes and educational opportunities. Since artificial intelligence often learns from records shaped by these imbalances, its decisions tend to copy old biases present in the datasets [28].

The U.S. provides particularly clear examples of how historical discrimination gets built into AI systems [29]. Due to past cases of criminal injustice which led to arrests, the conviction rates for African Americans compared to white Americans is significantly higher [30]. When AI systems are trained on these past cases in order to predict future offenders or decide the risk of setting a bail, the system ends up copying the discriminatory pattern which was present in the data [31]. Similarly, past discrimination

in healthcare including the under treatment of African Americans and other minorities, means that healthcare datasets underrepresent the prevalence and severity of these diseases [32]. As a result, AI's used in healthcare which were trained on such data provide less accurate diagnosis for minority patients [33].

One other significant area is Child Welfare and Family Services. Historically, low income and minority families have been monitored an immense amount leading to a higher rate of family separation and foster care placement for Black and Indigenous children [35]. Modern "predictive risk modeling" (PRM) tools which are utilized by social services are trained on this specific administrative data. This is due to wealthy families being more likely to use private services like private therapy. This does not generate public records thus the algorithm flags poor families as a "high risk" based on their interaction with the public support systems [36]. As a result, the algorithm has reproduced a class and racial bias, pushing vulnerable families to automated suspicion and investigation at a rate far higher than their wealthier counterparts [37].

Another critical area is Migration and Border Control. In the past, immigration policies in many nations have been shaped by xenophobia and racial exclusion, favouring people from some places while treating people from other places as threats [38]. Today, this history is encoded into biometric systems and automated visa screening tools. For example, facial recognition technologies used at borders often perform poorly on darker-skinned individuals because they were trained on datasets dominated by white faces, leading to false matches and wrongful detentions [39]. Furthermore, automated risk assessment algorithms for visa applicants often use "country of origin" as a weighted factor, effectively digitizing historical national stereotypes and systematically rejecting applicants regardless of their individual merit [40]

4.3 Emergence of Algorithmic Discrimination as a Human Rights Concern

Recognition of algorithmic discrimination as a human rights concern is slowly growing as evidence accumulates documenting the bias developed into AI systems, and as researchers, civil society organizations and lawmakers began analyzing how this technology affected human rights [41]. The first cases of recognition came from criminal justice lawyers documenting bias in the widely used risk assessment software used for bailing and sentencing decisions [42]. Researchers analyzing these systems found a significant racial bias, with the algorithms assigning higher risk scores to African American defendants compared to white defendants with similar criminal histories, leading to a higher bail amount and harsher sentences for African Americans [43]. Recognition of algorithmic discrimination can harm people's rights accelerated after 2018, with multiple international human rights organizations and UN bodies releasing reports detailing bias in AI systems and its human rights implications [44]. The UN

Human Rights Council adopted Resolution 42/15 in 2019, requesting a detailed study on AI and human rights. This marked an important step in formally recognizing that algorithmic discrimination is a human rights concern requiring international attention as well as action [45]. Later UN decisions, findings by independent experts and international human rights declarations have solidified racial discrimination due to algorithmic means to be a serious concern requiring urgent responses from governments, international organizations and the private sectors who develop these systems.

The deployment of biased AI systems is not just a technical failure, it is an active assault on the most essential human rights protected by international law. The right to life and physical integrity is directly threatened when healthcare algorithms systematically deprioritize minority patients for critical care, effectively deciding who lives and who dies based on flawed data [46]. Simultaneously, the right to privacy is ruined by mass surveillance tools that harvest biometric data without a person's consent, creating permanent digital footprints which specifically target marginalized groups [47]. And worst of all is the violation of freedom of opinions and expressions, as automated content moderation tools silence legitimate political speech from specific cultures while amplifying hate speech against others [48].

The urgency to address this issue is critical because unlike human bias, algorithmic bias scales instantaneously and invisibly across entire populations. A single flawed algorithm used by a national healthcare provider or a global hiring agency can discriminate against millions of people in a fraction of minutes, causing widespread harm that is nearly impossible to reverse once done [49]. As these technologies become more autonomous and integrated into critical infrastructure, from border control to child welfare, the window to establish meaningful oversight is closing [50]. Without intervention, we risk cementing a “digital caste system” where automated discrimination becomes a permanent, uncontestable feature of governance, leaving vulnerable groups no legal recourse and undermining the very foundation of international human rights law [51].

5. Case Studies

5.1 Past Implications

5.1.1 Case Study 1: Gender Bias in Employment (Amazon Recruitment Tool, 2014-2018)

Between 2014 to 2018, Amazon developed an experimental automated recruitment system which aimed to speed up the hiring process. The company’s engineering hub in Edinburgh, Scotland, created 500 computer models designed to scour through the internet and private records to locate strong candidates to

hire. The system's objective was to assign job applicants a rating from one star to five stars, allowing recruiters to focus only on those applicants with high ratings. This tool was intended to process up to 100 resumes and output the top five candidates for hiring [52][53][54].

The main issue in the system arose from its training data. The models were trained on resumes submitted to Amazon over a ten-year period. Due to the tech industry historically being male dominated, a vast majority of these resumes came from men. Consequently, the machine learning models assumed a strong statistical correlation between male applicants and positive hiring decisions. The algorithm effectively taught itself that male candidates were preferable and began to penalize resumes that included the word "Woman", one instance having a resume mentioning "Women's Chess Club Captain." Additionally, it ranked graduates of women-only colleges lower, effectively preventing them from being within the top ranks [52][53][54]..

Technically, the system didn't penalize resumes due to explicit instructions to discriminate, rather it prioritized keywords found in past successful male resumes such as "executed" and "captured," which were statistically less common in female applicants resumes. Amazon engineers attempted to patch the algorithm by creating rules to ignore specific terms, but the models continued to find other gender-based differences such as linguistic patterns which were associated with male speech. By 2018, Amazon shut down the project due to persistent cases of skewed results [52][53][54].

5.1.2 Case Study 2: Racial Bias in Criminal Justice (COMPAS Algorithm and state v Loomis, 2016)

The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), developed by Northpointe (Now Equitant), is a risk assessment tool which is widely used in the U.S. legal system to predict the likelihood of an offender re-offending (recidivism). It draws data from a 137-question survey, covering criminal history as well as socioeconomic factors like education, family background and their place of stay to generate a "risk score" spanning from 1 to 10 [55][56].

In 2016, ProPublica analyzed the risk scores of over 7000 defendants in Broward County, Florida. The analysis revealed a significant racial disparity in the number of incorrect decisions made by the algorithm. While the systems were set to predict recidivism for both Blacks and Whites defendants accurately, implying that a score of 7 carried the same risk probability for both groups, it failed

differently for each group. Black defendants who did not reoffend were nearly twice as likely to be mislabeled as “high risk” (45%) compared to their White counterparts (23%). To contrast, White defendants who did reoffend were more likely to be mislabeled as “low risk" (48%) as compared to Black reoffenders (28%) [55][56].

A fallout from this bias showed up in State v. Loomis (2016), a ruling by the Wisconsin Supreme court. Eric Loomis challenged the use of COMPAS in his sentencing, arguing that it violated his due process rights because the proprietary nature of the algorithm prevented him from inspecting how his risk score was calculated. The Court ruled against Loomis, allowing the continued use of COMPAS, provided that judges were given a written warning about the tool's limitations and that the score was not the "determinative" factor in sentencing. This case highlights the "black box" problem in judicial AI, where proprietary trade secrets conflict with a defendant's right to a fair and transparent trial [56][57][58].

5.2 Present Implications

Current data from 2024-2025 reveals that algorithmic discrimination has become a systemic crisis across all major sectors of life. In criminal justice, predictive policing tools seemingly designed to be race neutral continue to subject Black communities to a disproportionate amount of surveillance. A 2024 analysis found these systems target minority neighborhoods at significantly higher rates than white areas with similar crime levels, creating a self-fulfilling feedback loop of over-policing and arrest [59]. This inequality is mirrored in the healthcare sector, where a 2025 Yale study confirmed that widely used clinical risk algorithms systematically underestimate the medical needs of minority patients, predicting them to be "healthier" than equally sick white patients because the models rely on biased data proxies [60]. Financial systems further entrench these divides through “digital redlining,” with 2024 research showing that fintech AI models consistently deny loans to Black applicants at rates that would require credit score 120 points higher to equalize, through which it effectively would lock vulnerable populations out of homeownership [61].

Beyond these domestic issues, the humanitarian and migration sector is witnessing the rise of “automated borders” that disproportionately harm refugees from the Global South. Recent reports indicate that AI-driven asylum processing tools, now being tested in multiple European jurisdictions, commit serious “epistemic violence,” in which it erased the testimonies of asylum seekers and replaced it with its own threat ratings for the individuals [62][63]. Studies show these algorithms, trained on historical rejection data,

misclassify applicants from sub-Saharan Africa and conflict zones as "fraudulent" at rates up to 25% higher than other groups effectively automating the rejection of valid asylum claims under the disguise of efficiency [64].

In the private sector, a silent crisis is emerging in automated insurance claim denials, where AI is increasingly used to aggressively reject valid medical and life insurance. A 2024 investigation revealed that major insurers are using algorithms to deny claims en masse, one documented case showed that a single algorithm denies 300,000 claims within the span of two months with an accuracy rate of less than 10% for certain conditions [65]. These “denials-by-default” systems disproportionately affect patients with complex, chronic conditions who cannot afford the legal resources to appeal. The burden is particularly high for low-income families [66]. When life insurance claims are flagged by AI for “suspicious patterns,” which happens often due to new policies being set in place, families are left without support during grieving, forced to fight opaque “black box” decisions that insurers refuse to explain [67]

Finally, the gig economy and housing sectors have converged to create a precarious “surveillance trap” for low-income workers. In 2025, “algorithmic wage discrimination” has become standard practice, platforms like Uber and Lyft use AI to calculate the lowest possible wage a driver is likely to accept based on their personal data and location [68][69]. This system extracts maximum value from workers in poorer neighborhoods who have fewer employment alternatives. This economic instability grows worse because of AI-based rental checks used by property owners. These algorithms, which scrape eviction and criminal records, are riddled with errors [70]. A 2025 report highlighted that they frequently misidentify minority applicants as having eviction histories due to name mismatches, leading to automatic housing denials that are virtually impossible to correct, further deepening the cycle of homelessness and exclusion [71].

5.2.1 Case Study 3: Systemic Discrimination in Social Welfare (Dutch Childcare Benefits Scandal, 2013–Present)

The Childcare Benefits Scandal in the Netherlands represents one of the most severe instances of algorithmic discrimination in a modern democracy. Beginning in 2013, the Dutch Tax and Customs Administration utilized a self-learning algorithm to create risk profiles of applicants for childcare benefits in an effort to detect fraud. The system used "risk indicators" such as dual nationality and low income to flag individuals for investigation. Over 50,000 parents, disproportionately those with dual citizenship or ethnic minority backgrounds, were wrongly accused of fraud.

The algorithm's "risk classification model" operated on a feedback loop: once a group was flagged and investigated, any minor administrative error (such as a missing signature) was treated as "fraud" and resulted in a demand for the repayment of all benefits ever received, often totaling tens of thousands of euros. This confirmed the algorithm's bias that these groups were "high risk," leading to further targeting. The Dutch Data Protection Authority later found that the use of dual nationality as a risk indicator was a direct violation of anti-discrimination laws and the General Data Protection Regulation (GDPR)

The implications of this scandal remain acute in 2025. The Dutch government resigned en masse in 2021 accepting responsibility, but the remediation process has been catastrophic. As of late 2024, the recovery operation is estimated to cost €14 billion, yet many families have still not received full compensation. Over 1,600 children were removed from their homes into state care as a direct result of the poverty caused by these algorithmic debts. In May 2025, children of the victims launched a new civil lawsuit against the state for the damage they suffered, keeping this issue at the forefront of European human rights law

6. Past UN Actions and Resolutions

The United Nations Human Rights Council (UNHRC) has approved resolutions and declarations focusing on facets of human rights issues over time. One key area the council has tackled is to ensure data safety for every individual, by passing resolutions that primarily focus on tackling AI bias as well protecting the privacy and data of every individual.

1. **A/HRC/RES/42/15** adopted by the United Nations Human Rights Council (UNHRC), 26 Sept 2019. This resolution mandated that the right to privacy must be respected in the digital age; it laid the groundwork for scrutiny of AI, profiling and automated decision-making (ADM) in light of human rights.
2. **A/HRC/59/32** submitted in 2025 by the Office of the United Nations High Commissioner for Human Rights (OHCHR) under UNHRC mandate. The resolution focuses on how technology firms (especially those developing or using AI) should implement human rights safeguards and are directly relevant for accountability in algorithmic decision-making and corporate responsibility.
3. **A/78/ L.49** adopted by the United Nations General Assembly (UNGA) on 27 March 2024. This was the first major global resolution on AI by UNGA. It emphasizes respect for human rights in design, deployment, and use of AI by calling on states and stakeholders to avoid AI systems that can't comply with international human rights law.
4. **UNESCO Recommendation on the Ethics of Artificial Intelligence** adopted Nov 2021 by United Nations Educational, Scientific and Cultural Organization (UNESCO) which applies to all UN member states. It provided a global normative standard for AI ethics rooted in human rights, fairness, transparency, accountability, non-discrimination, data

protection, human dignity, and human oversight. Helps translate abstract rights into policy and governance frameworks.

5. **The UN Guiding Principles on Business and Human Rights (UNGPs)** have been endorsed by the UN Human Rights Council in 2011, with strong support by global business and civil society alike, and have emerged as the authoritative global normative framework for responsible business conduct. The UNGPs are structured into three pillars that provide a framework for governments to protect and businesses to respect human rights.
    1. The State duty to protect human rights.
    2. The corporate responsibility to respect human rights
    3. Access to remedy.

The UN Human Rights B-Tech Project was launched in 2019 with the goal of promoting the uptake of the UNGPs in the technology sector. Tech companies increasingly use the UNGPs to identify, assess and mitigate downstream human rights impacts, also with regard to digital technologies including a range of maturity stages of AI.

6.1 Actionable Policy Options

One of the key challenges emerging from the UNGPs, the UNESCO AI Ethics Recommendation, and UNHRC resolutions such as A/HRC/42/15 is the absence of mandatory human rights scrutiny before AI systems are deployed in sensitive public sectors. As states increasingly rely on algorithms for hiring, welfare distribution, policing, and border control, these systems continue to be treated as neutral tools rather than power-bearing decision-makers. This creates a regulatory gap where harm is addressed only after violations occur. Delegates must consider whether international law should shift from a reactive to a preventive model, who would oversee such scrutiny, and whether developing states would face disproportionate compliance burdens.

Moreover, A/78/L.49 and the UNESCO Recommendation highlight the growing use of black-box AI systems that neither users nor governments fully understand. These models challenge due process when applied to criminal justice, financial access, and public surveillance. While transparency is framed as a human rights necessity, it often clashes with corporate intellectual property claims and national security concerns. Delegates must therefore assess whether full transparency is essential in high-risk contexts or whether limited disclosure can meaningfully protect rights without stifling innovation.

The UNGPs and the 2025 OHCHR submission A/HRC/59/32 underscore that many powerful AI systems are developed by multinational corporations but deployed by states across borders. This separation enables companies to distance themselves from downstream human rights harm, particularly in countries with weaker legal systems. Delegates must evaluate how far corporate liability should extend across jurisdictions,

whether parent companies should be responsible for foreign deployments, and how victims can access justice when responsible actors lie beyond their legal reach.

A/HRC/42/15 establishes privacy as a core right in the digital age, yet the rapid expansion of biometric surveillance, facial recognition, and predictive monitoring shows the normalization of constant observation. With states adopting sharply different models of digital governance, global consensus remains elusive. Delegates must therefore examine how international law can regulate surveillance without undermining national security, and where the boundary between legitimate protection and digital authoritarianism lies.

The UNESCO Recommendation and UNGA Resolution A/78/L.49 emphasize ethics and human dignity, yet global AI development is increasingly driven by geopolitical competition. States are racing for dominance in military AI, economic automation, and data control. This creates a paradox in which governments endorse ethical commitments while accelerating high-risk deployments. Delegates must analyze whether ethical governance can realistically survive under conditions of strategic rivalry, and whether binding global norms are feasible in a power-driven international system.

Lastly, AI systems trained on historically biased data risk reinforcing existing patterns of discrimination based on race, gender, class, and nationality. Although international frameworks condemn discrimination in principle, enforcement within algorithmic systems remains legally and technically weak. Delegates must therefore question whether existing anti-discrimination laws are sufficient for machine-based decision-making, and how equality before the law can be preserved when bias is embedded in data rather than explicitly written into policy.

7. Questions A Resolution Must Answer (QARMA)

1. How can the United Nations ensure equality, and remove partial bias with regards to the usage of AI across the continents? Should international human rights law require mandatory scrutiny of AI systems before their deployment in sensitive public sectors?
2. Can full transparency of high-risk AI systems coexist with national security and corporate intellectual property protections?
3. Should local governments impose stricter restrictions on the usage of AI and technology?
4. To avoid algorithmic discrimination, should the United Nations appoint subsidiary groups to avoid bias against certain factions or individuals?
5. Without disruption in innovations and technology, how can states ensure safety and transparency for each individual with regards to the usage of AI?
6. How can developing countries provide adequate help and support to underdeveloped countries to ensure fair share of distribution of technological resources?
7. Should tech companies (operating in the Silicon Valley or across the board) be entitled to algorithmic impact assessments to ensure transparency and safety with regards to AI usage?
8. Can the right to privacy realistically be protected in an era of expanding biometric surveillance and predictive monitoring?

8. Bibliography

1. Wischmeyer, T., & Rademacher, T. (2022). "AI Bias: Exploring Discriminatory Algorithmic Decision-Making in the Public Sector
2. European Union Agency for Fundamental Rights (FRA). (2022). "Bias in Algorithms – Artificial Intelligence and Discrimination
3. ProPublica. (2016, updated 2025). *Machine Bias: Risk Assessments in Criminal Sentencing.*
4. Ghafur, S. et al. / Imperial College London. (2022). *AI Could Worsen Health Inequities for the UK's Minority Ethnic Groups*
5. Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). *Does AI Debias Recruitment? Race, Gender, and AI's "Eradication of Bias" Claims*
6. Stanford Institute for Human-Centered AI (HAI). (2025). *How Flawed Data Aggravates Inequality in Credit*
7. Hadjitchoneva, J. (2019). *Efficient Automation of Decision-Making Processes in the Digital Economy*
8. Symbio6. (n.d.). *Automated Decision-Making Benefits: Efficiency & More*
9. Council of Europe. (2018). *Discrimination, Artificial Intelligence, and Algorithmic Decision-Making*
10. Council of Europe. (2019). *Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*
11. UK Government. (2020). *Review into Bias in Algorithmic Decision-Making*
12. Grimmelikhuijsen, S. et al. (2022). *Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making*
13. European Judicial Training Network (JuLIA Project). (2025). *Artificial Intelligence, Judicial Decision-Making and the Rule of Law*
14. Khosravi, M. et al. (2024). *Artificial Intelligence and Decision-Making in Healthcare*
15. OECD. (2025). *AI in Justice Administration and Access to Justice: Governing with Artificial Intelligence*
16. Rodrigues, R. (2020). *Legal and Human Rights Issues of AI: Gaps, Challenges and the Way Forward*
17. Burrell, J. (2016). *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*
18. Zuiderveen Borgesius, F. et al. (2024). *Algorithmic Discrimination: Examining Its Types and Regulatory Responses*
19. Dencik, L. et al. (2025). *Civil Society and the (Im)Possibilities of Algorithmic Contestation*
20. Greenlining Institute. (2021). *Algorithmic Bias Explained*
21. Russell, S., & Norvig, P. (latest edition). *Artificial Intelligence: A Modern Approach*
22. Buchanan, B. G. (2005). *A (Very) Brief History of Artificial Intelligence*

23. Jordan, M. I., & Mitchell, T. M. (2015). *Machine Learning: Trends, Perspectives, and Prospects*
24. Hickok, M., & Hu, H. (2021). *Beyond State v. Loomis: Artificial Intelligence, Government Decision-Making, and the Law.*
25. Chen, J. (2021). *The Impact of AI/ML on Human Rights*
26. UK Centre for Data Ethics and Innovation. (2020). *Review into Bias in Algorithmic Decision-Making*
27. Barocas, S., & Selbst, A. D. (2016). *Big Data's Disparate Impact*
28. Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*
29. Office of the UN High Commissioner for Human Rights (OHCHR). (2024). *Racism and AI: "Bias from the Past Leads to Bias in the Future".*
30. Tonry, M. (2009). *Are Blacks and Hispanics Disproportionately Incarcerated Relative to Their Arrests?*
31. U.S. Department of Justice. (2022). *Artificial Intelligence and Criminal Justice: Final Report*
32. Obermeyer, Z. et al. (2019). *Racial Bias in Algorithms Used to Manage the Health of Populations*
33. Leslie, M. et al. (2023). *Fairness of Artificial Intelligence in Healthcare: A Review and Evaluation of Bias Mitigation Strategies*
34. UN Guiding Principles on Business and Human Rights (UNGPs) for tech companies
35. U.S. Department of Health and Human Services, ASPE. (2022). Avoiding Racial Bias in Child Welfare Agencies' Use of Predictive Risk Modeling
36. Oxford Insights. (2024). Predictive Analytics, Public Services and Poverty
37. Berkeley Public Policy Journal. (2021). Algorithms and Child Welfare: The Disparate Impact of Family Surveillance and Risk Assessment
38. Access Now. (2022). Uses of AI in Migration and Border Control
39. National Institute of Standards and Technology (NIST). (2019). Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects
40. Foxglove & Joint Council for the Welfare of Immigrants. (2020). Home Office to Stop Using Racist Visa Algorithm
41. Ma, E. (2025). Algorithmic Discrimination and the Existing International Law Framework
42. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, updated 2023). How We Analyzed the COMPAS Recidivism Algorithm.
43. Hamilton, M. (2023). Justice Served? Discrimination in Algorithmic Risk Assessment
44. Kaye, D. (UN Special Rapporteur). (2018). Report on Artificial Intelligence Technologies and Implications for Freedom of Expression and the Information Environment
45. UN Human Rights Council. (2019). Resolution 42/15: The Right to Privacy in the Digital Age.

46. Obermeyer, Z. et al. (2019). *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*
47. ARTICLE 19. (2021). *When Bodies Become Data: Biometric Technologies and Free Expression*
48. Gameiro, S. (2023). *Artificial Intelligence and Online Hate Speech Moderation*
49. European Union Agency for Fundamental Rights (FRA). (2022). *Bias in Algorithms – Artificial Intelligence and Discrimination*
50. Rodrigues, R. (2020). *Legal and Human Rights Issues of AI: Gaps, Challenges and the Way Forward*
51. Australian Human Rights Commission. (2021). *Addressing the Problem of Algorithmic Bias*
52. Dastin, J. (2018, October 10). *Insight: Amazon scraps secret AI recruiting tool that showed bias against women*
53. CNBC. (2018, October 10). *Amazon scraps a secret A.I. recruiting tool that showed bias against women*
54. BBC News. (2018, October 10). *Amazon scrapped 'sexist AI' tool*
55. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23; updated 2023). *How We Analyzed the COMPAS Recidivism Algorithm.*
56. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks*
57. State v. Loomis, 881 N.W.2d 749 (Wis. 2016).
58. Kehl, D., Guo, P., & Kessler, S. (2017). *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments.*
59. The Markup. (2021). *The Disparate Impact of Predictive Policing Software*
60. Obermeyer, Z. et al. (2019). *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*
61. Lehigh University (2024). *AI Exhibits Racial Bias in Mortgage Underwriting Decisions*
62. European Parliamentary Research Service (EPRS). (2025). *Artificial Intelligence in Asylum Procedures in the EU*
63. Ahmed, S. (2025). *Asylum Processing Algorithms and Epistemic Violence.*
64. Global Governance of Algorithmic Refugee Status Determination Systems." (2025).
65. Thomas, P., & Herman, B. (2024). *How Cigna Saves Millions by Having Its Doctors Reject Claims in Bulk.*
66. The Guardian. (2025). *New AI Tool Counters Health Insurance Denials Decided by Automated Algorithms*
67. Life Insurance Attorney LLC. (2025). *AI's Role in Denying Life Insurance Claims*
68. Dubal, V. (2023). *On Algorithmic Wage Discrimination*
69. Human Rights Watch. (2025). *The Gig Trap: Algorithmic Wage and Labor Exploitation in Platform Work in the US.*

70. TechEquity Collaborative. (2025). *How AI-Powered Tenant Screening Hurts Renters*
71. Urban Institute. (2025). *Oversight of Tenant Screening Companies May Change: Here's What Landlords and Renters Should Know*