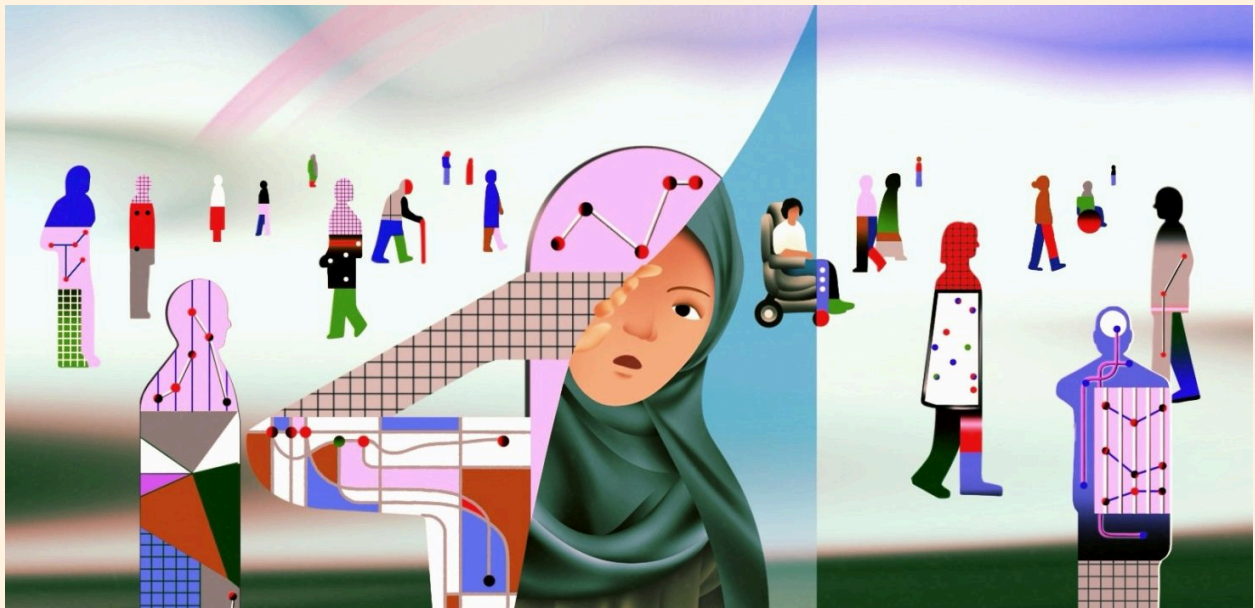


SOCHUM

Discrimination in Social Welfare Systems



**39th Annual Model United Nations
Conference at St. John's High School**

By David Wang and Vismay Prasad

Table of Contents

<i>Letters From the Chairs:</i>	2
<i>History of the problem</i>	4
<i>Statement of the Problem</i>	7
<i>Possible Solutions</i>	10
<i>Bloc Positions</i>	13
<i>Bibliography</i>	18

Letters From the Chairs:

Dear Delegates,

My name is David Wang, and I am a Junior at St. John's High School. I sincerely welcome you to SJMUNC XXXIV. It is my pleasure to chair this committee along with Vismay Prasad.

This is my third year as a member of Model UN at St. John's and at this conference. This is my first time as a chair (I was co-chair last year). I have attended national conferences twice, once at Brown University and once at the University of Chicago. I hope my skills can help us have a good time at this conference!

We figured this is an interesting committee topic, as it is highly relevant in this age of Artificial Intelligence. However, we figured that by stripping our typical understanding of "ChatGPT"/Language Models away, and looking at AI as just simple prediction/classification algorithms. We can approach such topics with a more "human" perspective, considering how this technology might affect the poor and the marginalized. I would strongly recommend, in addition to reading this briefing paper, doing some additional research outside. The rules and regulations are constantly being updated, and new policies might emerge every few weeks.

If you have any questions about this committee or MUN in general (particularly if this is your first committee), please feel free to contact me. I will try my best to answer them.

May your models never overfit,

David Wang

wangh28@stjohnshigh.org

Dear Delegates,

My name is Vismay Prasad, and I am a senior at Saint John's. Surprisingly, this is only my second year participating in MUN, but regardless, I am excited to co-chair this committee with David, who is far more experienced than I. I initially joined MUN because I had a lot of friends who did it, and I thought it would be good to practice public speaking, but I realized how much fun and enriching it can really be to converse and argue with people you have never met before. If you put your best foot forward, I doubt any of you would finish off a year without learning at least one valuable lesson or making a new friend.

While AI might seem like a repetitive buzzword to many, it remains influential as it brings about changes that are becoming inevitable with its growth. This led David and me towards this topic, as the tools that artificial intelligence brings lend themselves to boosting the use of algorithmic and automated systems, and we have a responsibility to question that use. We hope that this unique topic encourages everyone to explore views on technology that you have rarely considered, especially when it comes to realizing the responsibilities surrounding those technologies.

Model UN is consistently more fun when you aren't in fear of being unprepared, so try your best, and if you find yourself with questions or are seeking clarification, please reach out to either of us, and we will try our best to be of help.

Good luck,

Vismay Prasad

prasav27@stjohnhigh.org

History of the Problem:

Over the past two decades, developed countries worldwide have increasingly shifted the administration of social benefits to automated systems rather than traditional human evaluation. Eligibility assessment, payment calculation, and most importantly, fraud detection have increasingly been handed over to algorithms and data-matching pipelines. The motives are consistent: greater efficiency, shorter response time, and reduced spending through the detection of overpayments and fraud.

The human rights implications of this transformation were brought to global attention in October 2019, when Philip Alston, the UN Special Rapporteur on extreme poverty and human rights, presented a thematic report to SOCHUM: *Digital welfare status and human rights* (A/74/493). In it, he warned that societies risk stumbling, in his words, “zombie-like, into a digital welfare dystopia,” in which data and technology are being used to surveil, target, and punish the poor rather than to support them. He also argued that the private companies behind those “welfare systems” frequently function in what he called a “human rights-free zone”. Additionally, he argued that these automations tended to fall hardest on the most marginalized, those people who often lacked the digital access or skills required to navigate the very systems that were imposed on them. The report drew on 60 submissions from civil society organizations across 34 countries and effectively names the phenomena this committee hopes to address.

Ever since the release of this report, a number of national scandals have occurred, proving that this issue is not an abstract warning but rather documented cases. The following three cases are some examples.

The Netherlands produced two of the earliest and probably most consequential cases. The first concerned the SyRI (Systeem Risico Indicatie, or “System Risk Indication”), a fraud-detection tool that pooled personal data previously held in separate government databases, such as employment, benefits, taxes, debt, and more, then applied a hidden algorithmic risk model to flag individuals deemed likely to commit benefit fraud. It was introduced through the welfare reforms around 2014. SyRI was developed almost exclusively in low-income neighborhoods with very high proportions of minority and immigrant residents. In 2018, a coalition of civil liberties NGOs and a trade union federation sued the Dutch government. On February 5th, 2020, the District Court of the Hague ruled that the SyRI legislation was unlawful because it violated Article 8 of the European Convention on Human Rights, the right to respect for private life. The court held that SyRI was not transparent in thought and warned that it could discriminate on the basis of socio-economic or immigration status. This was one of the first times a court anywhere stopped a digital welfare tool on human rights grounds.

Australia’s Robodebt scheme is another classic example. It was introduced in 2015 and fully automated from July 2016. The program sought to recover supposed welfare overpayments by replacing manual review with an automated data-matching process. It compared income reported to Centrelink with averaged annual income data from the tax office (income averaging) and used the result to automatically raise debts against recipients. At this point, the issue should be clear: the algorithm treats people as if they had a steady income across the entire year when many do not, which causes a large number of false or inflated debts and shifts the burden of disproving them to the recipients, while many already struggle to maintain basic survival.

The scheme ultimately affected more than 500,000 Australians. The federal court found it unlawful in late 2019, the program shut down in 2020, and a class action led to a settlement of roughly 1.8 billion Australian dollars. In August 2022, the incoming government established a Royal Commission, Australia's most powerful form of public inquiry. Its report, delivered on July 7th, 2023, described Robodebt as a “crude and cruel mechanism” that was neither fair nor lawful and made recipients feel like criminals. The commission heard evidence from families of young people who had died by suicide after debts were raised against them, issued 57 recommendations, and referred several officials for further investigation. Robodebt is one of the clearest examples of how an automated system can cause mass harm with minimal government accountability, as it is merely a misdesigned algorithm.

These cases have begun to generate a body of legal precedents and regulations. The 2020 SyRI judgment established that human rights law, particularly the right to privacy, can be used to challenge opaque welfare algorithms. National data protection authorities have found welfare systems unlawful and imposed fines. Most importantly, the EU’s Artificial Intelligence Act, which entered into force on 1 August 2024, classifies systems used to evaluate eligibility for public benefits as “high risk,” imposing requirements for transparency, human oversight, and non-discrimination testing. However, issues persist: there are no universal laws addressing them, the definition of an “automated system” remains in question, and those affected by unfair systems have received limited redress.

France offers a third example of proxy-based discrimination that predates and outlasts the EU's newest regulations. Since the early 2010s, the Caisse Nationale des Allocations Familiales (CNAF), France's family and social security agency, has used a machine-learning model to

assign nearly every welfare recipient a fraud "risk score" between zero and one. A 2023 investigation by Lighthouse Reports and Le Monde, based on documents obtained under freedom-of-information law, found the model weighted single parenthood, low income, and disability status heavily, despite no established link between those traits and fraud. People with the highest scores faced invasive home investigations. In 2024, a coalition of 15 human rights groups filed legal action against the French government over the algorithm, arguing it unlawfully discriminates against single mothers and disabled recipients.



Statement of the Problem:

The transition to an automated welfare system described above has not simply digitized a set of bureaucratic tasks for the government; it has quietly altered the relationship between the state and its most vulnerable citizens. Those systems, introduced in the name of efficiency and fraud prevention, had increasingly trained poverty itself as the source of suspicion, subjecting the very people that social protection is meant to shield to surveillance, investigation, and exclusion. The core problem that this committee must address is not only about the deployment of digital social welfare systems, but also the way these systems embed and amplify discrimination against already marginalized groups – the poor, the disabled, women, ethnic and religious minorities, immigrants, and refugees. Digital social welfare systems led to those groups facing discrimination at a much faster and far greater scale than human bureaucracies.

One central mechanism of discrimination in these systems is that they don't need to name *one* specific characteristic of the client to punish them. Since the algorithms rely on “proxy” variables indicating the client's status, such as income level, employment status, neighborhood, family structure, or disabilities, these variables correlate strongly with marginalized groups. As such, it would often indirectly discriminate against groups that tend to have these traits.

Compounding the problem of indirect discrimination is the issue of transparency. Because the internal logic of these systems is hidden, recipients flagged as “high-risk” often cannot see, understand, or challenge the system's results, which effectively hollows out their right to appeal. The burden is now on recipients to disprove decisions or flags made by the system, as in Australia's Robodebt scheme. Additionally, such automation can strip away human

discretion that often absorbs the messiness of real-life situations, reducing them to an algorithm that often ignores the unique circumstances individuals may face.

Underlying all of this is the issue of accountability. Many of these systems are built or operated by third parties not in the government's control, and they can behave as a "human rights-free zone," as Philip Alston wanted. Leaving the affected citizens confused about who was even responsible for making that decision. The scale of benefit fraud is often exaggerated to justify intrusive data collection from citizens, even though the systems are often ineffective at detecting the fraud they were built to find. Furthermore, the increased reliance on digital technology widens the "digital divide," which makes those with the least digital access and literacy, such as the elderly, the poor, and the marginalized, least able to navigate and correct the systems and algorithms that stand between them and a truly well-lived life.

Crucially, these issues are not confined to the wealthy Western developed countries; they are rapidly spreading across the globe. Investigations by Lighthouse Reports and its partners have uncovered similar risk-scoring systems across the Netherlands, Spain, Denmark, and Serbia. International leaders are actively financing the export of such welfare systems to lower-income countries with weaker safeguards. As these tools spread, so does the danger that algorithmic discrimination will be scaled across the globe.

These social welfare systems, designed to accelerate government workflows and assist the marginalized communities, have been proven to discriminate against the very people they were designed to serve. It is up to this committee, SOCHUM, to decide exactly what to do about the problem of discrimination in social welfare systems that is expanding worldwide, and to

ensure that the digitization of welfare systems does not plunge the world into a digital welfare dystopia.



Possible Solutions:

A good resolution will need rules that advise governments to be careful, ways to check whether those rules are being followed, and options for individuals to fix mistakes when the system gets it wrong. The ideas below are grouped into four categories. Delegates do not need to pick just one; most real resolutions combine several. Keep in mind that not every country has the same level of wealth or technical capacity, so a solution that works in a wealthy country may not work everywhere.

Create rigorous, thorough methods to validate the system before use. Before a government uses an algorithm to decide who receives benefits, it may be required to assess whether the system treats certain groups unfairly. For example, it denies benefits to immigrants or people in poor neighborhoods more often than others. This is sometimes called an "impact assessment." The European Union already requires something similar for welfare algorithms.

Give people the right to a real explanation. If someone is denied benefits or accused of fraud by an algorithm, they should have the right to know why and to have an actual person review the decision, not just click "confirm" on what the computer said.

Put the burden on the government, not the individual. In Australia's Robodebt case, people had to prove the algorithm was *wrong* about them, which is backward. A resolution could state that when a government accuses someone of fraud using an algorithm, it is the government's responsibility to prove the accusation, not the recipient's to disprove it.

2. Ways to Check the System Is Working Fairly

Create an independent watchdog group. This would be a team that does not work for the same government agency running the welfare system, so it can honestly check the algorithm for bias without pressure to protect the agency. In the Netherlands, courts ended up playing this role for the SyRI system, but it would be better to have a watchdog in place before harm happens, not after a lawsuit.

Hold private companies responsible, too. Many of these algorithms are built by private tech companies, not the government itself. A resolution could require that these companies also be held accountable if their systems discriminate, rather than allowing them to hide behind "trade secret" excuses (claiming they can't reveal how the algorithm works because it's private business information).

3. How the Technology Itself Should Be Designed

Check for hidden bias in the data. An algorithm doesn't need to directly ask about someone's race or immigration status to discriminate against them; it can use something else, like zip code or family size, as a stand-in that ends up targeting the same group. These stand-in factors are sometimes called "proxy variables." Governments could be required to test for this kind of hidden bias before using a system.

Keep a human involved in tough cases. Instead of letting the algorithm automatically deny borderline cases, the system could be designed to send difficult or unusual cases to a real person for review, so that context and fairness aren't lost.

4. Making Sure No One Gets Left Behind

Keep a non-digital option available. Not everyone has reliable internet access or knows how to use complicated government websites, often the same people who need help the most. Countries could be required to keep an in-person or paper option available for applying and appealing.

Set a minimum global standard. Right now, a country with weak regulations can simply buy an unfair algorithm that a stricter country wouldn't allow. Poorer countries also often can't afford to properly test these systems before adopting them. A resolution could set basic fairness standards that all countries are expected to follow, along with support (funding or training) to help lower-income countries meet them. The hardest question for this committee is how far the UN can go. Rules that are too strict may simply be ignored by countries that don't want to follow them. Rules that are too soft may not protect anyone. A strong resolution should seek a middle ground: real, enforceable protections that are still realistic for every country to implement.

Bloc Positions:

Delegations are divided into seven categories: regulators, fraud-control, public infrastructure, sovereigns, targeting veterans, excluded, and crisis-affected.

The regulators, as the name suggests, already regulate welfare algorithms. They are all members of the EU and, as such, implement the EU's regulations to flag AI used to decide public benefits as high risk and to prohibit any harmful social scoring. Additionally, many of them have been sued over these algorithms (such as the Netherlands' SyRI, mentioned above). Recognizing the benefits as well as the potential issues posed by these algorithms, this bloc would most likely align with existing EU regulations and push for impact assessments, rigorous human review, and independent watchdog monitoring. They might not agree to a global ban on such algorithms, since their own governments still run them.

The fraud-control bloc mostly uses digital systems to find incorrect payments, which can include human mistakes or fraud claims (hence the name), and they do not intend to stop. It is important to note that they do not use algorithms to determine eligibility for social benefits, unlike some countries in the regulator bloc. Therefore, those countries in the fraud-control block will likely take a stance similar to regulators, supporting impact assessments and advocating for human-in-the-loop. However, they will oppose any attempts to ban automated data-matching or transaction error monitoring.

The public infrastructure countries take a different approach. They implement digital IDs, biometrics, and social registries as the means by which people obtain their social benefits. In these counties, not having a digital ID or being outside of the registry usually means no benefits. These countries rarely use AI-based algorithms to determine eligibility; they simply rely on

digital identification to qualify and then use statistical formulas to determine eligibility. This bloc will likely defend digital public infrastructure as inclusive and accept additional steps to include those who are more marginalized and lack proper digital access. Their stance towards AI algorithms is unclear, as most don't run EU-style automated eligibility assessments.

The sovereigns place greater emphasis on where their data flows than on anything else. They do have digital IDs and welfare systems (China's Social Insurance, for example), but the core similarity is how they treat their country's data. As a result, sovereign countries will oppose any suggestions for a third-party independent regulator or UN bodies to govern their systems.

Targeting veterans is similar to the public infrastructure bloc. They maintain large national social registries to decide who gets the benefits. However, they do not use any digital identification as a gate to eligibility. The state already holds the information via other means. As for disability itself, it is determined via a statistical algorithm, not an AI-based one. They are likely to support advances in statistical formulas and national registries for benefits, as well as improvements in the registries themselves.

The excluded bloc is trying to set up digital IDs and social registries. However, many of their citizens, women, older people, and those in remote areas in particular, cannot sign up for such services, as they lack the cellphones or internet connection to even begin the registration step. These countries will support fallback plans when possible and may request funding to expand their core infrastructure.

The crisis-affected countries are dealing with war, displacement, or economic collapse, so a stale state registry or a strict targeting algorithm can cut people off from survival assistance.

The challenge for these countries is navigating social welfare amid turbulence, since the usual solutions do not apply to them. As a result, they are more likely to support more adaptive systems that can be quickly updated after a crisis. They will resist those who try to impose a universal system that ignores their unique situation.

- **Regulators (7)**

- Netherlands
- France
- Germany
- Denmark
- Sweden
- Spain
- Poland

- **Fraud-Control (5)**

- United Kingdom
- Australia
- Canada
- Japan
- Republic of Korea

- **Public Infrastructure (7)**

- India
- Brazil

- Indonesia
- Philippines
- Kenya
- Rwanda
- Estonia

- **Sovereigns (7)**

- China
- Russia
- Iran
- Egypt
- Saudi Arabia
- Vietnam
- Cuba

- **Targeting Veterans (6)**

- Chile
- Colombia
- Mexico
- Argentina
- Peru
- Costa Rica

- **Excluded (7)**

- Nigeria
- South Africa
- Ghana
- Uganda
- Ethiopia
- Senegal
- Togo
- **Crisis-Affected (5)**
 - Jordan
 - Serbia
 - Lebanon
 - Pakistan
 - Ukraine

Bibliography:

- Alston, Philip. “Digital Welfare States and Human Rights.” United Nations General Assembly, A/74/493. October 11, 2019. <https://docs.un.org/en/A/74/493>
- Amnesty International. “Trapped by Automation: Poverty and Discrimination in Serbia’s Welfare State.” December 4, 2023. <https://www.amnesty.org/en/latest/research/2023/12/trapped-by-automation-poverty-and-discrimination-in-serbias-welfare-state/>
- Australian Government. “Government Response: Royal Commission into the Robodebt Scheme.” November 2023. <https://www.pmc.gov.au/sites/default/files/resource/download/gov-response-royal-commission-robodebt-scheme.pdf>
- Caisse nationale des allocations familiales. “Contrôles Caf : un nouvel algorithme plus transparent.” January 15, 2026. <https://www.caf.fr/allocataires/actualites/actualites-nationales/controles-caf-un-nouvel-algorithme-plus-transparent>
- DigiChina. “Personal Information Protection Law of the People’s Republic of China.” November 1, 2021. <https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021/>
- District Court of The Hague. “NJCM et al. v. The State of the Netherlands (SyRI).” ECLI:NL:RBDHA:2020:1878. February 5, 2020. <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBDHA:2020:1878>
- Economic Commission for Latin America and the Caribbean. “Social Information Systems and Registries of Recipients of Non-contributory Social Protection in Latin America in Response to COVID-19.” LC/TS.2021/56. Accessed September 7, 2026. https://repositorio.cepal.org/bitstream/handle/11362/46868/S2100029_en.pdf
- European Parliament and Council. “Regulation (EU) 2024/1689 (Artificial Intelligence Act).” June 13, 2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

- Lighthouse Reports. “How We Investigated France’s Mass Profiling Machine.” December 4, 2023.
<https://www.lighthousereports.com/methodology/how-we-investigated-frances-mass-profiling-machine/>
- Lighthouse Reports. “Suspicion Machines.” March 6, 2023.
<https://www.lighthousereports.com/investigation/suspicion-machines/>
- Office of the United Nations High Commissioner for Human Rights. “World Stumbling Zombie-like into a Digital Welfare Dystopia, Warns UN Human Rights Expert.” October 17, 2019.
<https://www.ohchr.org/en/press-releases/2019/10/world-stumbling-zombie-digital-welfare-dystopia-warns-un-human-rights-expert>
- Royal Commission into the Robodebt Scheme. Report. July 7, 2023.
<https://robodebt.royalcommission.gov.au/publications/report>
- Treasury Board of Canada. “Directive on Automated Decision-Making.” Accessed September 8, 2026. <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>
- UK Government. “Fraudsters Face Tougher Action as Government Gains New Powers to Tackle Benefit Fraud.” December 2, 2025.
<https://www.gov.uk/government/news/fraudsters-face-tougher-action-as-government-gain-s-new-powers-to-tackle-benefit-fraud>
- Unwanted Witness. “Chased Away and Left to Die.” 2021.
<https://www.unwantedwitness.org/chased-away-and-left-to-die-new-human-rights-report-finds-that-ugandas-national-digital-id-system-leads-to-mass-exclusion/>
- World Bank. “Global Digital Public Infrastructure Program.” May 6, 2026.
<https://www.worldbank.org/en/results/2026/05/06/global-digital-public-infrastructure-program>
- World Bank. “Mind the Gap: Addressing Exclusion in Pakistan’s Social Protection System.” November 4, 2025.
<https://www.worldbank.org/en/news/feature/2025/11/04/-mind-the-gap-addressing-exclusion-in-pakistan-s-social-protection-system>
- World Food Programme Innovation. “Building Blocks.” Accessed September 7, 2026.
<https://innovation.wfp.org/project/building-blocks>