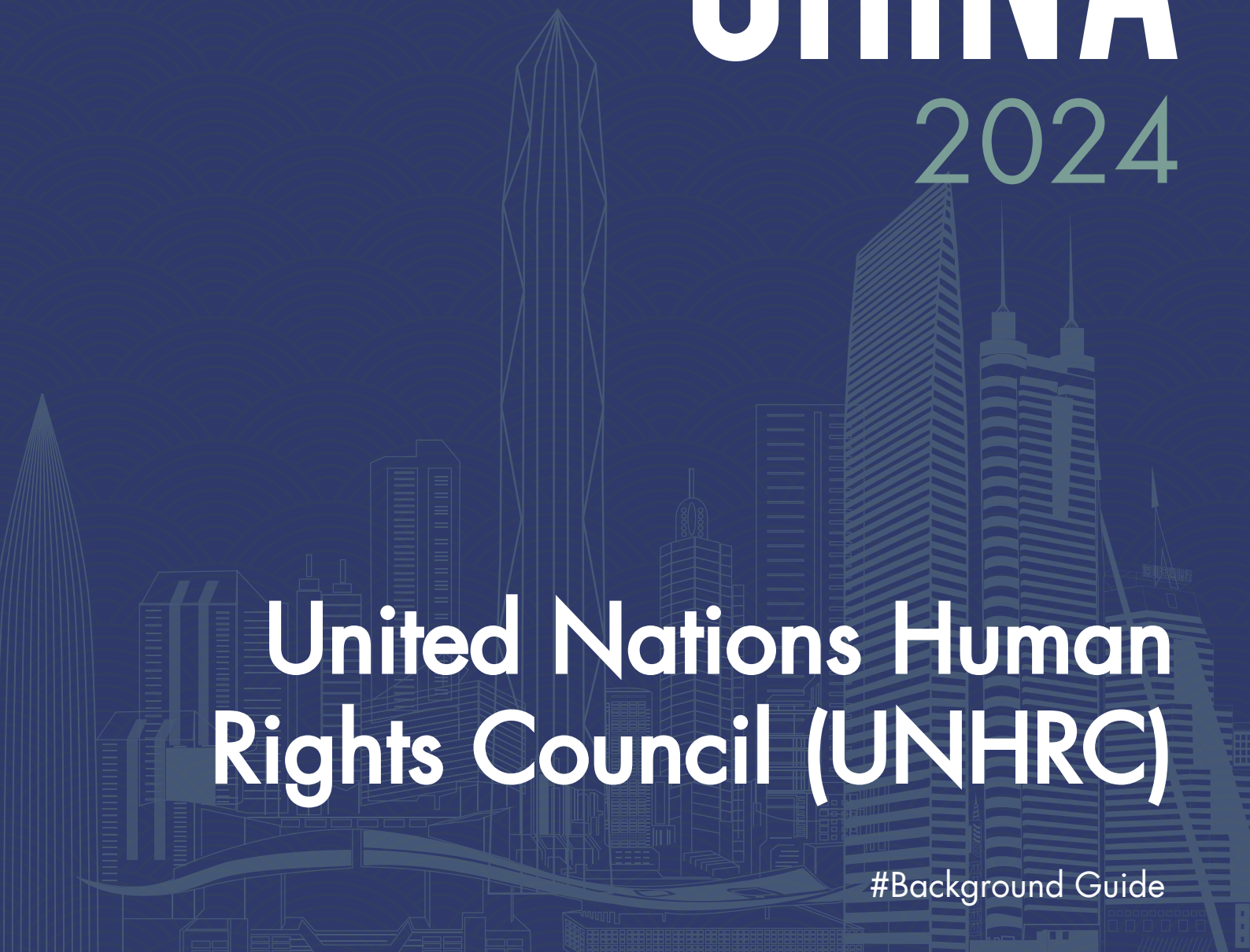
YALE MODEL UNITED NATIONS CHINA

YALE

MODEL
UNITED
NATIONS
CONFERENCE

CHINA

2024



United Nations Human Rights Council (UNHRC)

#Background Guide

UN HUMAN RIGHTS COUNCIL (UNHRC)

Director: Ivana Ramirez

TABLE OF CONTENTS

Letter from the Dais

Dear Delegates,

Hello! My name is Ivana Ramirez (she/her) and I am so excited to see y'all at YMUN China.

For a little introduction, I am a junior at Yale from Greenville, South Carolina and the Dominican Republic. I am a Global Affairs major pursuing a Certificate of Advanced Language Study in Chinese. Outside of school, I am a Tour Guide for the Yale Admissions Office, a research assistant at Yale Law School, and am a proud member of Mixed Company of Yale acapella. When I have free time, I enjoy film photography, reading fiction, and going to the beach.

I am very passionate about the UN Human Rights Council — the research that I do at the law school is primarily focused on human rights, and I hope to continue advocating for human rights after graduating. My work involves creating a human rights education project to incorporate human rights history to curricula around the world. I have worked extensively on human rights education, specifically with high schoolers, so I am very excited to work with you all! Analyzing human rights through the lens of AI is especially important, because it is an emerging market that you may very well become the world's leading experts on someday! Please feel free to reach out to me if you have any questions or need any guidance at ivana.ramirez@yale.edu. We are going to have so much fun!

Warmly,
Ivana

Committee History

The UN Human Rights Council (UNHRC) was created by the United Nations General Assembly on March 15, 2006, with its first session held in June 2006. Throughout each year, the Council enables about 260 civil society organizations to deliver more than 900 statements. One significant resolution passed by the council focused on Syria in 2011, expressing deep concern about the human rights situation amid government responses to protests. Another noteworthy resolution, from 2009, dealt with Sri Lanka's civil war, emphasizing the need for accountability, reconciliation, and reconstruction. The UNHRC has also tackled human rights abuses in North Korea through multiple resolutions, calling for investigations and improvements. In 2017, a resolution regarding Myanmar addressed the Rohingya crisis, advocating for an end to violence and the safe return of displaced individuals. More recently, a 2020 resolution focused on Belarus, expressing concerns over human rights violations during protests and calling for an independent investigation. These examples highlight the council's commitment to addressing global human rights issues and fostering accountability and justice.

While the Council has several involvements in many different aspects of human rights, its commentary in the area of Artificial Intelligence (AI) is relatively new.

In a historic declaration issued just this past July 2023, UN High Commissioner on Human Rights Volker Türk, the head of the UNHRC, emphasized the need for human rights to form the foundation for all AI use worldwide. The High Commissioner highlighted the potential advantages of AI in this remark, including how it may advance scientific research, increase capacity for processing massive volumes of data, improve strategic foresight and forecasting, and democratize knowledge access. But he also listed the possible drawbacks of AI, such as its capacity to operate deadly autonomous weaponry, bolster authoritarian rule, and be used to anticipate criminal conduct in the future in a discriminating manner.

While Türk properly emphasized the private sector's pivotal role in ensuring the regulation of AI, he also outlined UNHRC's involvement as critical in providing human rights-centered guidance to the private sector. Particularly, the High Commissioner proposed creating an international advisory body for very high-risk technologies to provide insights into how regulatory frameworks could be in line with frameworks for universal human rights and the rule of law. The group, as he explained, may make suggestions for AI governance and make the results of its discussions publicly available. The UN Secretary-General also proposed a similar idea as a component of the Global Digital Compact for the Summit of the Future in September 2024.

Because AI as a concept was introduced to the United Nations just in 2018, these debates will be consequential to future policy in the topic area. Our iteration of UNHRC will act as this High-Level Advisory Body on Artificial Intelligence to debate intersectional issues of artificial intelligence and human rights, including the rights of minorities, people with disabilities, and globally impoverished.

Committee Structure

The Human Rights Council is responsible for strengthening the promotion and protection of human rights around the globe and for addressing situations of human rights violations to make recommendations on them. The Council is made up of 47 United Nations Member States which are elected by the UN General Assembly. These candidacies are divided by region, with 13 for the African Group, 13 for the Asia-Pacific Group, 6 for the Eastern European Group, 8 for the Latin American and Caribbean Group, and 7 for the Western European and Others Group. Of the 193 member states of the United Nations, 117 have served as council members for the Human Rights Council.

The Council meets regularly three times per year, totalling to roughly ten weeks or more in a year. Each regular meeting includes a briefing by the UN High Commissioner for Human Rights, the presentation of UN reports on human rights promotion and protection, the holding of interactive dialogues with UN Special Procedures (based on their reports to the Council) and various other civil discourses. Outside of these regular meetings, if any country wishes to hold a special meeting, they must first obtain the support of one-third of members. These special sessions can either have a thematic or country-specific focus. The main output of a given Council session is a series of Human Rights Council resolutions drafted and negotiated by members and designed to address a particular human rights issue or agenda. Action is taken on such resolutions at the end of each Council session. Special sessions last for one day and allow members and other stakeholders to hear updates (for example, from the UN High Commissioner) and present their views on the issue at hand. The session usually concludes with the adoption of a resolution.

1

Safeguarding Privacy in the Age of AI Surveillance

Safeguarding Privacy in the Age of AI Surveillance

Introduction

The widespread use of AI-powered monitoring tools is changing how we think about privacy. These technological advancements, which range from facial recognition to predictive policing, have the potential to violate civil liberties and individual rights. This topic seeks to strike a balance between the necessity of security and the protection of the fundamental right to privacy.

Glossary

Algorithm: a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

Anonymization: a type of information sanitization whose intent is privacy protection; the process of removing personally identifiable information from data sets so that the people whom the data describe remain anonymous.

Artificial intelligence (AI): the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

Data matching: the process of comparing two different sets of data and matching them against each other. The purpose of the process is to find the data that refer to the same entity.

Machine learning: the use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.

Privacy engineering: an emerging field of engineering which aims to provide methodologies, tools, and techniques to ensure systems provide acceptable levels of privacy.

Pseudonymisation: a data management and de-identification procedure by which personally identifiable information fields within a data record are replaced by one or more artificial identifiers, or pseudonyms

Topic History

Artificial intelligence as a concept has several interesting historical milestones. In 2014, the inception of Generative Adversarial Networks (GANs) sparked a transformative era in AI, paving the way for the creation of remarkably realistic images and videos. Harnessing neural networks, a model inspired by the functioning of the human brain, GANs generated examples indistinguishable from authentic data. The year 2016 witnessed the activation of Sophia, the inaugural non-human robot, blending vision and speech algorithms for a remarkably human-like impression. Concurrently, Google DeepMind unveiled WaveNet, an AI voice generator employing an algorithm akin to GANs. The practical applications of AI-generated voices manifested, aiding individuals with neurological disorders in reclaiming their voices while also being exploited by criminals in deceptive scams. Google introduced the groundbreaking Transformer in 2017, revolutionizing machine translation by processing entire sentences at once, discerning meaning within a contextual framework. The year 2018 witnessed OpenAI's launch of GPT-1, employing a Transformer architecture to respond to queries and generate text. Despite its capabilities, GPT-1 struggled with coherence in generating longer passages and exhibited tendencies toward repetition. It wasn't until 2022, however, that AI truly became mainstream with the release of OpenAI's ChatGPT.

August 29, 2018

First documented instance of artificial intelligence in the United Nations. This document, the Promotion and Protection of the Right to Freedom of Opinion and Expression, sought to outline the consequences of artificial intelligence technology on human rights in the information age, with a focus on the freedoms of speech, privacy, and nondiscrimination. This was before the topic was picked up by the UN Human Rights Council as an official topic, so the research was carried out by a Special Rapporteur on the issue. In this report, the Special Rapporteur outlined how AI will continue to advance in the future. This includes voice help, chatbots, online translation tools, self-driving cars, search engine results, and mapping services. He explained that narrow AI techniques, under machine-learning, train algorithms to identify patterns in datasets and provide solutions. For example, AI-enabled smart home appliances learn from common speech and language patterns to interpret user inquiries.

The report highlighted several key areas of concern with privacy and AI, including:

1. Search platforms: Large-scale datasets that incorporate user demographics, browsing histories, sentiment and semantic analysis, among many other variables, are fed into increasingly individualized algorithmic models to rank and filter information — that is, to display people certain information or implicitly remove it.
2. Social media: Content that is sponsored, paid for, or includes hashtags may be promoted above or to the detriment of other content. People may have little to no exposure to some critical social or political topics and content uploaded on social media platforms since these news

feeds provide content based on subjective evaluations of how engaging or intriguing it may be for a user.

3. Advertising: Developments in artificial intelligence have profited from and strengthened the Internet's data-driven business model, which requires users to pay for free goods and services with their personal information. Businesses are able to provide AI systems with rich datasets to create prediction and targeting models that are increasingly more accurate due to the enormous data resources gathered over years of Internet monitoring and profiling. Individualized advertising is now possible for both public and commercial entities; voters and consumers are the focus of "microtargeting," which is tailored to take advantage of and react to unique quirks.
4. Misinformation: AI-driven targeting raises the possibility of manipulating specific users through the dissemination of false information and encourages the extensive acquisition and use of personal data.

The Special Rapporteur concluded that the majority of data used to feed AI systems is either personal data that has been extracted or inferred from personal data, or personal data that has been anonymized (often imperfectly). AI-driven decision-making systems rely on the collection and exploitation of data, ranging from ambient, non-personal data to personally identifiable information. Businesses obtain datasets from third parties like data brokers, use data from digital fingerprinting and online profiling, and create fresh data from large aggregated datasets to feed AI systems.

At the end of the report, the Rapporteur called for a private-sector response to human rights concerns surrounding AI, including auditing for transparency and conducting human rights impacts assessments. At this time, there was little discussion about the role of AI in sustainable development or privacy concerns in terms of international foreign policy.

August 15, 2019

The publication of the Report of the Working Group of Experts on People of African Descent on using data to protect privacy concerns in light of racial justice. The working group emphasized the recent introduction of AI automated systems, which were primarily intended for white people and therefore negatively affect individuals of African origin. For instance, individuals of African origin were 10 percent less likely by automated cars to be recognized as human, putting their lives in jeopardy and exposing them to unnecessary risk. The group also emphasized how facial recognition technology fails to identify persons of African origin based on their gender, further dehumanizing them. During the discussion of this report, the UK delegate made comments about algorithmic prejudice and artificial intelligence, highlighting the dangers these technologies pose to the criminal justice system. Furthermore, research in the paper indicates that the infant formula industry's aggressive targeting of black moms and the misuse of cesarean sections may be factors in the higher infant death rate among black infants. In a more positive light, the report also highlighted the successful German measures in

removing racial bias from algorithms in the criminal justice system. Generalized algorithms for training were developed in Germany so that bias could be fixed and negative outcomes could be avoided.

Current Situation

Since the two initial reports that spurred the conversation about human rights in the context of AI, there have been several follow-up reports and committees in various capacities regarding the very broad topic of privacy and AI.

For example, the UN Secretary-General requested that UN Human Rights Council create UN System-Wide Guidance on Human Rights Due Diligence for Digital Technology Use (HRDD) in order to assist all UN bodies in putting into place and bolstering HRDD policies, procedures, and practices for the use of digital technologies, including development, acquisition, and sharing. The guidelines provide steps to get started and build HRDD throughout time, as well as a useful introduction to HRDD to help each entity establish, implement, and strengthen its HRDD for the use of digital technology. This guidance culminated in the establishment of the UN Hub for Human Rights and Digital Technology to serve as a central repository for authoritative guidelines on the application of human rights norms to the use and governance of digital technologies, including artificial intelligence, from various UN human rights mechanisms.

In 2020, UNHRC hosted a high-level online expert seminar in order to determine how artificial intelligence—which includes machine learning, automated decision-making, and profiling—may, in the absence of appropriate protections, impair the enjoyment of the right to privacy. A primary emphasis of this seminar was the lack of regulatory bodies in this topic area, specifically that around one third of UN member states don't have privacy laws and only 65 to 70 states have independent data protection authorities. The Permanent Mission of Switzerland's delegate brought out the significance of human rights and the users-based approach to AI and voiced concerns that invasions of privacy had an impact on other human rights. For instance, the freedoms of expression, peaceful assembly, and association were all impacted by social media surveillance and the use of facial recognition technology.
AI use for developing countries and impoverished people

In the online expert seminar, the Permanent Mission of India's delegate stressed the significance of the right to privacy in the context of AI technology and the influence that AI has on developing nations like India. For example, though AI technologies provide immense potential for public goods, they also pose human rights issues, particularly with regard to data mining techniques. Specifically, the delegate referred to the 2017 ruling by the Indian Supreme Court that recognized the right to privacy as guaranteed by the Indian Constitution.

Furthermore, many Latin American countries rely on privately owned infrastructures or services by transferring personal data to commercial entities. Data sharing tactics between corporations and nations should fall under the purview of the international human rights framework and advocate for strong regulation of such practices.

Artificial intelligence and other digital technologies have enormous potential to further the numerous advantages that their supporters frequently list. In order to harness this potential, governments will need to take the lead in any such effort through the implementation of suitable fiscal policies and incentives, regulatory initiatives, and a sincere commitment to creating the digital welfare state as a means of guaranteeing a respectable standard of living for all people.

For example, a major public health concern in Kenya is antimicrobial resistance, or AMR. However, there is not enough being done on a worldwide scale to address AMR due to a knowledge gap. AI-collected citizen-generated data (CGD) can contribute to closing this global knowledge and policy gap. Enhancing CGD is essential to raising community awareness of AMR and assisting communities and governments in making decisions that will lead to coordinated action on the issue.

Persons with Disabilities

People with disabilities have a plethora of opportunities thanks to artificial intelligence. Artificial intelligence-powered systems are improving technology, especially assistive technology. One example is the ability for people with impairments to find accessible routes in their neighborhood using accessible phone apps. Other applications that use artificial intelligence-powered navigability tools can improve the personal mobility rights of people who are blind or have low vision. Still others include software for voice recognition and eye tracking that enable persons with disabilities to access information and education.

However, AI has posed significant challenges for people with disabilities as well. For instance, studies have shown that these AIs are acting as gatekeepers for social protection benefits. Artificial intelligence can impose restrictions on eligibility for benefits, such as unemployment aid, by utilizing discriminatory algorithms and biased data sets. AI systems must ensure that people with disabilities are not unfairly denied access to services and support, as there are a lot of people with disabilities living in poverty worldwide, many of them in extreme poverty. Furthermore, people with impairments are beginning to feel the effects of artificial intelligence in their role as customers. Inaccurate increases in health insurance premiums and the denial of access to private health insurance might result from flawed artificial intelligence risk assessments based on specific handicap categories. When artificial intelligence is used to make decisions that result in higher insurance premiums for consumers without clear or understandable reasoning, there have been concerns expressed that this could lead to higher premiums. The general public faces serious problems as a result of the lack of transparency surrounding the underlying reasoning behind automated decision-making and machine learning, and

people with disabilities, who are already excluded in the health and life insurance markets, may be particularly affected. Finally, an applicant with a disability may find resume-screening methods problematic if, for example, their resume does not highlight any internships they have completed or options for remote work. A human almost seldom reviews explanatory information regarding comparable experience. It is common practice to build artificial intelligence-enabled video screening systems without subjecting people with disabilities to experimentation. This makes it possible for a candidate to be excluded from consideration for a job based on unusual qualities before a human interviewer even gets to meet them. An artificial intelligence program has been known to assign a negative weighting to candidates with impairments during the scoring process. This has led to the employment agency providing a disabled candidate with fewer resources and less support during their job hunt.

Racial Justice

People of color are misidentified and misclassified by AI facial recognition technologies. Black men in the United States have been wrongfully imprisoned and falsely arrested at a disproportionate rate as a result of the usage of this technology. In the private sector, businesses use facial recognition technology for hiring as well as tracking the whereabouts and output of their employees.

Consumer data is used by online platforms and other service providers to discriminate against people of color and prevent them from accessing housing, jobs, commerce, and other services. Due to the usage of AI technologies by lenders, people of color seeking loans to buy homes or refinance have been overcharged by millions. Additionally, a lot of companies now interview and screen job applicants using AI-driven technologies, which increases the potential of discrimination against individuals with disabilities and other protected groups. AI has exacerbated discriminatory behaviors rather than helping to end them, endangering the financial stability of underprivileged communities who have long faced systemic discrimination. However, businesses are opaque about how they get data, how they use other data points and personally identifiable information in automated decision-making tools, and how they use these things.

However, creating a more informed AI system may help combat systemic racism. By providing chances for information access and a more dynamic understanding of AI and its effects, we will be able to create a society that is more knowledgeable about AI.

COVID-19

In light of the COVID-19 pandemic and the use of contact tracing applications, there exists a tension between a right to privacy and protection of a right to health. In the past, supporters claimed that in order for states to fulfill their duty to safeguard citizens' health and life, the right to privacy had to give way. Depending on a number of variables (such as what data is gathered, where it is stored (locally or

centrally), for how long, and which organizations or agencies may have access to the data now or in the future), contact tracking apps may cause major violations of the right to privacy.

Questions to Consider

1. To ensure inclusivity and equal opportunities for individuals with disabilities, what specific measures can the UNHRC implement to prevent discrimination by AI systems and provide accessible pathways to information, education, and employment?
2. In the complex scenario of the COVID-19 pandemic, how can the UNHRC propose to strike a balance between the right to privacy and the right to health, specifically in the deployment of contact tracing applications and other AI-driven responses to public health crises?
3. Regarding the impact of AI on racial justice, how can the UNHRC plan to actively address issues related to misidentification and misclassification, particularly within facial recognition technologies, and what initiatives can be explored to promote fairness and equity?
4. How does your country's position land on the tension between AI's power for good and its infringement on privacy rights?
5. How can AI be used for good to promote development in your country without infringing on privacy concerns?
6. Besides the existing frameworks, what additional frameworks can UNHRC create to protect humanity's right to privacy?

Additional Resources

Promotion and Protection of the Right to Freedom of Opinion and Expression

Report of the Working Group of Experts on People of African Descent on using data to protect privacy concerns in light of racial justice

UN Human Rights High Commissioner Statement on AI

Description of Activities on AI of UNHRC

Report of the proceedings of the online expert seminar with the purpose of identifying how artificial intelligence, including profiling, automated decision-making and machine learning technologies may, without proper safeguards, affect the enjoyment of the right to privacy

UN HRC Resolutions Offer Crucial Safeguards for Civil Society in AI-Driven Digital Age

UN Hub for Human Rights and Digital Technology

The right to privacy in the Digital Age

Extreme poverty and human rights

Rights of persons with disabilities

Multistakeholder Advisory Body on Artificial Intelligence

High-Level Advisory Body on Artificial Intelligence

Privacy & Racial Justice

How AI Can Help Combat Systemic Racism

Citizen-Generated Data on Antimicrobial Resistance in Kenya

2 Ensuring Ethical AI Use in Security and War Measures

Ensuring Ethical AI Use in Security and War Measures

Introduction

The human rights implications of AI technology are becoming more prominent as countries integrate them into their security protocols. Concerns of responsibility, prejudice, and the right to due process are brought up by drones, automatic weapon systems, and prediction algorithms during times of conflict.

Glossary

Algorithmic bias: the presence of unfair or discriminatory outcomes in algorithms due to biased data or flawed design.

Asymmetric warfare: a situation where two belligerents possess different levels of military capabilities and strategies. In the context of AI in conflict, asymmetric warfare may refer to the use of advanced AI technologies by one party, creating challenges for the less technologically advanced opponent.

Autonomous weapons: weapons that can operate without direct human control

Biometric Surveillance: the use of biological and behavioral characteristics, such as facial recognition or fingerprint analysis, for identification and tracking purposes.

Cyber warfare: the use of digital techniques to attack and disrupt the computer systems and networks of adversaries. In the realm of AI in conflict, cyber warfare involves leveraging AI technologies for offensive and defensive purposes in the digital domain.

Lethal autonomous weapons (LAWs): a type of autonomous military system that can independently search for and engage targets based on programmed constraints and descriptions.

Robotic warfare: the use of robots or autonomous machines in military operations.

Topic History

Since May 2014, discussions on lethal autonomous weapons systems have taken place at the Convention on Conventional Weapons (CCW) in Geneva, although no significant progress has been made. The primary cause of the CCW's stagnation is that its member nations rely on a consensus-based decision-making process, which allows one nation to reject a proposal even if all other nations support it. This has been used by a few big military powers to consistently thwart attempts to negotiate a legally binding instrument.

In 2021, the UN Security Council (UNSC) recorded one of the first instances of autonomous weapons used in combat to take the lives of human beings during the Libyan civil war. In this incident, logistics convoys were tracked down and attacked in remote battle by lethal autonomous weapons systems. With a real "fire, forget, and find" capacity, the deadly autonomous weapons systems were designed to strike targets without requiring data contact between the operator and the munition. Electronic jamming from the electronic warfare system rendered small drone intelligence, surveillance, and reconnaissance capacity as well as the unmanned combat aerial vehicles useless. After retreating, the affected group faced constant harassment from the deadly autonomous weapons systems and unmanned combat aerial vehicles.

Shortly thereafter, the UN Convention on Certain Conventional Weapons debated whether or not to ban autonomous weapons in warfare, ultimately failing to reach a consensus. Globally, militaries are making significant investments in the development of autonomous weaponry. Between 2016 and 2020, the United States alone allotted US $18 billion on autonomous weapons. Human rights and humanitarian groups are working feverishly to enact laws and bans on the development of such weapons. Foreign policy experts warn that in the absence of such checks, disruptive autonomous weapons technologies could combine with conventional nuclear, biological, and radiological weapons, changing perceptions of strategic dominance and raising the risk of preemptive attacks. This could lead to a dangerous destabilization of current nuclear strategies.

In 2021, then-UN Human Rights High Commissioner Michelle Bachelet called for a moratorium on artificial intelligence systems that could put human rights at risk — at least until stronger safeguards are in place internationally. The statements were made in conjunction with the release of a study by the U.N. Human Rights Council that examined the threats to human rights presented by various AI-powered technologies, such as machine learning, profiling, and automated decision-making.

Current Situation

In November 2023, the UN passed its first revised draft resolution on lethal autonomous weapons systems. Egypt's representative voted in favor of the resolution and stated that an algorithm should not

have complete authority over decisions that involve killing or injuring individuals. The resolution expressed concern about the potential drawbacks and effects of autonomous weapons systems on international stability, regional security, and global security, including the possibility of a new arms race and a lowering of the bar for proliferation and conflict, including with non-State actors. The draft resolution was approved overall by a recorded vote of 164 in favor to 5 against (Belarus, India, Mali, Niger, and Russian Federation), with 8 abstentions (China, Democratic People's Republic of Korea, Iran, Israel, Saudi Arabia, Syria, Turkey, and United Arab Emirates), following 11 separate recorded votes on its individual provisions. Speaking against the resolution, the delegate from Russia stated that the resolution was unbalanced in that it mainly addressed dangers and difficulties, despite the fact that these weapons can be quite useful for defense and counterterrorism and can lower the chance of error and be more efficient than a human operator.

Shortly after in December, the General Assembly resolution on the risks of deadly autonomous weapons systems received votes from 152 nations, 4 against, and 11 abstained. The resolution acknowledged serious challenges and concerns created by new technological applications in the military domain, including those related to artificial intelligence and autonomy in weapons systems. By September 2024, the General Assembly will receive a report from UN Secretary-General António Guterres detailing the countries' and other stakeholders' perspectives on how to address the issues and problems brought forth by autonomous weapons systems from humanitarian, legal, security, technological, and ethical perspectives. The resolution also gives states a platform to take action to address this issue by adding a resolution on lethal autonomous weapons systems to the UN General Assembly's provisional agenda for 2024.

Case Study: Determining Village Destruction from Conflict in Sudan

In 2018, Amnesty International collaborated with academics from University College London and the University of Amsterdam to measure the devastation of villages in Darfur, Sudan. The project, spearheaded by an NGO collaboration between civil society and technical specialists, was marketed as the first application of machine learning for human rights. The partnership developed an AI algorithm trained to scan satellite photos for the remainder of the villages in order to find new human habitats and detect destruction using a multi-task binary classification, with the assistance of a team of AI professionals. The intention of the partnership was to increase public awareness of the conflict on a broader scale, elicit indignation, and encourage participation from the public through an online petition. Through the use of open-source data gathered by Amnesty International during their Eyes on Darfur Campaign, the program "learned." Crowdsourced photos of ongoing events served as the data source (which was later used to train the AI). The model performed exceptionally well in terms of lowering the amount of time and money that humans would have needed to undertake the research as well as increasing accuracy when measured against the performance of human experts. The AI correctly identified 81 percent of the same tiles as the human experts did when identifying the ones with destroyed buildings. In contrast, 85 percent of those that the AI identified were also identified by

human specialists. This case study is an example of AI's capabilities to be used for good in the area of global conflict.

Case Study: Using AI To Forecast Internal Refugee Displacement

In 2021, the Danish Refugee Council (DRC) utilized AI and machine learning technology to predict displacement trends. The tool, initially developed in 2021, collected and analyzed data on 148 indicators related to conflicts, governance, economy, climate, human rights, and societal trends. Based on this information, the model accurately forecasted how many people would be displaced annually over the following three years.

The model used open access data from sources like the World Bank and NGOs to predict forced displacement in a given country. In 2021, the DRC used the tool to make predictions about 2022 and 2023, covering twenty-six countries and predicting an increase of 8.7 million people displaced between 2021 and 2023.

The average margin of error for the 2021 displacement forecasts was 14 percent, with the highest margin of error being 45 percent for Libya. The DRC found that the margin of error across all 188 forecasts was 19 percent.

Predictive modeling can be helpful in advocating for increased humanitarian aid in places where the models predict high levels of displacement and low levels of humanitarian aid. The DRC has been able to compile the findings into a report for the European Union, providing specific recommendations for how displacement should be addressed, including where resources should be allocated. However, one of the limitations of the AI model is that the most current data it uses is still from the previous year, making it largely unable to take into account unexpected developments or major changes in a country's geopolitical realities in real time.

Case Study: U.S. Drone Strike in Afghanistan

In September 2021, up to ten civilians, including seven children, were mistakenly killed by a drone strike in Kabul, Afghanistan. The strike occurred three days after the Taliban swept to power in the capital and U.S.-trained Afghan forces melted away, killing 13 U.S. troops and dozens of Afghan civilians who had gathered outside the airport gates. The chaos was due to a suicide bomber from the Islamic State.

The U.S. military claimed to have killed two Islamic State members in an eastward drone strike that it had carried out in response to the airport suicide bombing. That strike was not investigated. As the Pentagon completed its mission, officials warned they expected more attacks on the airport, including from rockets and vehicle-borne explosive devices. This coincided with the second, unintentional strike.

While artificial intelligence has potential for good, it also has many misuses during wartime that infringes on individuals' human rights. It is critical to analyze how to best strike a balance between the two to ensure that human rights is at the center of every debate related to artificial intelligence.

Questions to Consider

1. To what extent should algorithmic authority over life and death decisions be granted in the context of lethal autonomous weapons, and how can international agreements uphold ethical standards in the use of AI in warfare?
2. What lessons can be drawn from the U.S. drone strike case in Afghanistan regarding the unintended consequences of AI in military operations, and how can these lessons inform responsible AI use in conflict zones?
3. How can the positive applications of AI, as demonstrated in the case studies of monitoring village destruction in Sudan and tracking violence against minorities in Myanmar, be expanded to enhance human rights protection during armed conflicts?
4. How can international collaboration address the challenges posed by asymmetric warfare, ensuring that less technologically advanced nations have access to AI technologies without compromising global security?
5. In the context of cyber warfare and the use of AI for offensive and defensive purposes, how can states strike a balance between the necessity for security and the protection of individual privacy in the digital domain?

Additional Resources

UN fails to agree on 'killer robot' ban as nations pour billions into autonomous weapons research

U.S. says Kabul drone strike killed 10 civilians, including children, in 'tragic mistake'

U.N. official calls for moratorium on artificial intelligence tools that breach human rights

AI for humanitarian action: Human rights and ethics

First Committee Approves New Resolution on Lethal Autonomous Weapons, as Speaker Warns 'An Algorithm Must Not Be in Full Control of Decisions Involving Killing'

Without Adequate Guardrails, Artificial Intelligence Threatens Global Security in Evolution from Algorithms to Armaments, Speaker Tells First Committee

Killer Robots: UN Vote Should Spur Action on Treaty

Note to Correspondents: Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for States to establish new prohibitions and restrictions on Autonomous Weapon Systems

The Future of Artificial Intelligence

International Community Must Urgently Confront New Reality of Generative, Artificial Intelligence, Speakers Stress as Security Council Debates Risks, Rewards

UN and Red Cross call for restrictions on autonomous weapon systems to protect humanity

Human Rights Impacts of Autonomous Weaponry

The Use of Artificial Intelligence in International Human Rights Law

Yale Model United Nations China 2024

May 17-19, 2024
Shenzhen, China

ymunchina.org
@ymunchina

YALE MODEL UNITED NATIONS
CHINA