JejuMUN XII

BACKGROUND GUIDE

UNHRC

Ensuring the Protection of Human Rights in the Age of Artificial Intelligence Against Surveillance and Deepfakes

SDG: 4: Quality Education | 9: Industry, Innovation, and Infrastructure| 10: Reduced Inequalities | 16: Peace, Justice, and Strong Institutions

Authored by Kyle (Shinho) Lee, Yunseong (Sean) Kim, Jiyoul (Yuri) Kim

November 28th-29th, 2025

Last updated on July 26, 2025

Table of Contents

Table of Contents	2
Committee Introduction	3
Agenda Introduction	4
Letter from the Chairs	5
Key Terms	6
Historical Background	6
Current State of Affairs	8
Stances of Parties	9
Possible Solutions	10
Questions to Consider	11
Bibliography	12

Committee Introduction

Following the establishment of the United Nations Human Rights Council in 2006, this committee has stood steadfastly as a multinational forum that strives to address human rights violations to improve the world. Currently, the Human Rights Council meets continuously at the United Nations Office at Geneva (UNOG).

It is also within the responsibilities of the HRC to address a diverse range of issues, from the protection of human dignity to the affirmation of freedom. Thus, the HRC's agendas naturally include tackling ethnic and racial discrimination, reaffirming the rights of individuals from all demographics and genders, ensuring the freedom of speech and expression, and the choice of religion. Furthermore, the council also focuses on the human rights impact caused by indirect conflicts (i.e., Forced displacement and environmental crises).

The Human Rights Council has a number of ways and programs it uses to deal with these issues. For example, the HRC hired the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia, and Related Intolerance to watch over and write down problems with discrimination. It also set up the Universal Periodic Review (UPR) system, which looks at the human rights records of all UN member states. The HRC also sends people on fact-finding missions and sets up advisory committees to look into and fix certain human rights abuses. It also supports giving each country the power to solve problems that are still going on, like those involving gender rights, religious freedom, and the rights of native people.

In simpler terms, the Human Rights Council upholds and protects the rights of anyone around the world. This takes place through the gathering of information following the HRC's rapporteur and Universal Periodic Review(UPR) system.

Furthermore, the Human Rights Council has also investigated human rights abuses in various areas of the world. For example, the HRC established the Independent International Commission of Inquiry in Syria to look into alleged violations during the civil war; the HRC established the Independent International Fact-Finding Mission in Myanmar to document abuses against the Rohingya and other minorities; and the HRC established an Independent International Fact-Finding Mission in Sudan to look into violations in Darfur. Thus, the HRC seeks to condemn abuses and thus offer aspects such as advisory capacity in order to improve a country's human rights record.

Agenda Introduction

As technology continues to develop at a rapid pace, marginal historical action and, in many cases, emerging deep fakes are changing societies, economies, and systems of governance.

With the entry of the information age and striving to go for the 4th technological revolution, various sectors of technology are developing at exponential and unprecedented rates. Alongside other cutting-edge tech such as robotics, big data, and blockchain programs, AI and its following generative products are changing society universally.

Unfortunately, these developments present formidable challenges to human rights, especially in terms of mass surveillance and deepfakes. Both mass surveillance and deep fakes threaten human rights in ways that will be unprecedented regarding threats to privacy, freedom of expression, the right to information, and arguably even the integrity of democracy.

The AI and facial recognition algorithms were increasingly adopted by governments and private entities, reporting by the Carnegie Endowment for International Peace indicated that in 2021, more than 109 countries had adopted facial recognition systems or were considering implementing them and on occasions, these same instruments have had the ultimate effect, surveilling protestors, suppressing dissent, and discriminating against ethnic minorities(Carnegie Endowment for International Peace, VoiceOfAmerica.com). For example, reports of the use of AI surveillance are documented in Xinjiang, China, and the UN Special Rapporteur on the right to privacy has repeatedly warned that the initial stages of mass surveillance systems can and do erode civil liberties and pave the way for authoritarian behaviour.

At the same time, the rapid increase of deep fakes—hyper-realistic synthetic media produced with AI—has exacerbated the dissemination of misinformation, modelled political discourse, and identity theft. A Diptrace report noted that deep fake content on the internet increased by 900% from 2018 to 2020, the vast majority (96%) of which was either aimed at manipulating elections or modelling political leaders. Deep Fakes are now being used to impersonate political leaders, assist election manipulation, and support disinformation campaigns, thereby eroding trust in institutions and media.

These developments contribute to a pace of development that will soon outstrip frameworks of law and ethics. Without concerted international cooperation, if AI is misused, the core rights of the Universal Declaration of Human Rights (UDHR), such as “right to life, liberty, and security of person” of article 3 or “No one shall be subjected to arbitrary interference with his privacy…” of article 12, will be significantly compromised. The role of the United Nations is therefore critical to both regulate the use of AI and to establish ethical standards, and to prevent violations of dignity. This agenda ought to provide a basis for ensuring technology innovation does not discount or erase the principles of accountability, transparency, and human rights that belong to everyone.

Letter from the Chairs

Hello, this is Kyle Shinho Lee from Korea International School, Jeju Campus, who will be serving as your head chair, and welcome to UNHRC! With the drastic advent of AI technology and its side effects, our committee will be deeply discussing preliminary measures, online implementations, and countless strategies to effectively prevent and secure cyberspace from unwanted and coercive surveillance and deep fakes. Do not fear making mistakes and jumping into dog water situations, as the committee's sole purpose is to protect not only the privileged but the minorities, too. Most importantly, prepare to have fun, drive some madness, and understand MUN.

Hello Delegates, I write to you in my capacity as Deputy Chair of the UNHRC at JejuMUN XII, and I'm truly honored to welcome each of you to this year's conference. My name is Jennifer Park from Korea International School Jeju. This is my second year of MUN and third time chairing. Outside of MUN, I play varsity volleyball and basketball. My hobbies include reading and watching Formula 1. Given how this conference provides a great opportunity to step out of your comfort zone, I truly encourage you to challenge yourself. I will do my best as your chair to respond to that effort. Above all, remember to enjoy the whole experience!

Hello, this is Jiyoul Kim from Korea International School, Jeju Campus, who will be serving as your associate chair. I am more than happy to welcome you to UNHRC! Even though I don't have as much experience as our other two chairs, my dedication and passion brought me up here, and I hope to give all of you the same valuable experiences that I had. Our committee will mainly be focusing on the effects of deep fakes and surveillance on human rights. I hope this rising issue will allow you to think deeper and come up with amazing solutions. Don't be afraid to jump into tough situations and have fun!! Enjoy the moment, and we hope this could be a meaningful time for you all.

Best regards,

Shinho (Kyle) Lee | Head Chair | shlee28@kis.ac

Seohyun (Jennifer) Park | Deputy Chair | shyun28@kis.ac

Jiyoul (Yuri) Kim | Associate Chair | jyoulkim29@kis.ac

Key Terms

Artificial Intelligence

A simulation of human intelligence processes and logics by machines or computer systems. AI attempts to mimic human-level to superhuman levels of intellectual services in the contemporary state. AI in today's definition reaches far beyond LLMs(defined below), but is a substitute for a variety of jobs, manual labor, and systems that require human intelligence.

Generative AI

A type of technology that applies AI to create content, usually text, video, code, or images, based on its database and learned ideas from a variety of sources. Generative AI's accuracy and credibility highly depend on the accumulation of data for that specific topic or genre that is asked to answer, and these fluctuations are causing social unrest due to the mistrust caused by distorted forms of generated content.

Artificial General Intelligence

A hypothetical point of AI technology that engages and brings far more human-like cognitive abilities. Much more than a human, AGIs can be adapted and be adept in many sectors without needing to diversify each AI per function or topic, making an all-adroit digital being that mimics human intelligence as a singularity. AGIs are currently at the starting stage of being developed in many tech giants, as they attempt to harness wide databases and strengths in an enriched digital world.

Algorithm

A set of instructions or rules to follow to complete a specific task. These complex calculations help sustain and operate statistical processes. On par with the agenda, certain algorithms assist AI in thinking like a human to form a generalized and apt response, or perhaps a function that derives from those conclusions. These algorithms, while crucial to a multifaceted digital world, bring cybersecurity and privacy concerns upon extensive analysis and critical derivations that might cause unrest, conflict, or adverse effects for users.

Machine Learning

A subset of AI in which algorithms mimic human learning while processing data. In conventional terms, this stage is usually referred to as a prerequisite to deep learning, which simulates human logic more precisely. AI undergoes this process to make decisions based on likely patterns and the derivation of situational data.

Deep Learning

A much advanced and independent form of subset of AI that allows AI to be self-subsistent in terms of decision-making, route resourcing, and overall analysis of data, assimilating itself much closer to a human being. A literal difference between machine learning and deep learning of AI would be the engagement of neural networks(stated below).

Neural Network

A deep learning technique designed to resemble the structure of a human brain. It requires large sets of data to mimic sensory data that is obscured for an AI to understand. At the current state, neural networks are still a prototype, for the processes do not flexibly perceive situations like a neuron alerting the brain.

Weak AI

A type of AI that usually refers to an AI system that is allocated for a specific purpose or function, not a universal tool. These models excel at what they are given, which makes them an AI that is suitable for dedicated functions such as surveillance, service, organization, etc. To date, these AIs are widely used, and many sectors of the world are integrating AI into their systems.

Cybersecurity

The state of being protected against the criminal or unauthorized use of electronic data, or the measures taken to achieve this. Cybersecurity usually refers to the sum of safety measures, a secure environment, and all protected states of privacy and data.

Deep Fakes

A video of a person in which their face or body has been digitally altered so that they appear to be someone else, typically used maliciously or to spread false information. These altered videos are seemingly boosted with AI technology that generates content that could amalgamate a variety of things. In popular social media, videos, and posts, AI-generated deep fakes are becoming a problem, as some of them are acting as if the content were legitimate or was verified.

Threat Detection

A procedure and system that analyzes user behaviors, traffic patterns, and flow of data to presume a threat in the digital space and prevent the discerned situation from spreading to more sectors of the online world. Threat detection systems are crucial to safeguarding servers and the valuable data of individuals, private firms, and many others. However, due to over-surveillance and excessive threat prevention, these procedures are also being questioned for potential breach of privacy in the name of security.

Data Mining

A process of searching and analyzing a large batch of raw data to identify patterns and extract useful information. Firms use data mining software to learn more about their customers, usually through cookies(stated below) or algorithms that suggest particular interests. These processes can be tailored to individuals, yet they raise concerns similar to those surrounding private-sector threat detection.

Cookies

Small pieces of text sent to the browsers of users to indicate a visit, acceptance of a certain collection of data, and miscellaneous functions requiring third-party apps. These small information packets function to assist the users' experiences to be more personalized and user-friendly. However, these *tags* are also a concern as some firms attempt to collect excessive information for their advantage.

Cyber Surveillance

An action of monitoring and moderating online activities that serves multiple functions: identifying threats, gathering data, and observing user behavior. Cyber surveillance, even though it is a prevalent measure to enhance security in the digital world, is a component of restriction and restraint across the web in the contemporary world, as cyber surveillance can also clash with privacy.

Historical Background

Artificial Intelligence (AI) being called a significant contributor to society hasn't been a title that has persisted for a long time. AI's exponential and unprecedented growth in the 2020s has made it possible for AI to be assimilated into modern society as a substantial factor. Despite being somewhat called a "bubble", the AI industry's profits across multiple appliances have septupled since 2020, which showed significant growth compared to other sectors in stagnant technological sectors during the 2010s.

Particularly, telecommunications, semiconductor manufacturing, and information technology have profited extensively from the integration of AI technology or the development of powerful models. Nvidia, which has held a long-lasting throne in GPUs, entered a 4 trillion market cap due to the significant rise in demand for high-end GPUs, acting as a key component in running sophisticated AI. Microsoft (MS), a forever runner-up in market caps, has surpassed Apple, the challenged market cap champion, because of its successful investment in OpenAI's ChatGPT, which crucially supplemented the revitalization of MS products: Bing, Windows, Edge, etc. Internationally, other massive companies such as Samsung, ASML (Advanced Semiconductor Materials Lithography), TSMC (Taiwan Semiconductor Manufacturing Company), etc, have experienced a cornerstone benefit to their business, illustrating how AI has established a global hegemony for profit.

Alongside the technological sectors, conventional sectors of video platforms, social media, and advertisements also took advantage of AI as it stimulated hastened content creation, reduced expenditure, and overall an efficient process than employing human resources. Platforms such as YouTube, TikTok, and Instagram have shown exponential growth of AI-generated content from a varied range of consumers and suppliers and have catalyzed more activity on their platforms with shortened, AI-driven media. Fast food companies such as Subway and McDonald's, though controversial, have widely accepted AI generative tools as trend-setting and effective methods of advertisement. These conventional and everyday niches with AI show the lingering impact of AI as a positive factor. Ambivalently, AI exploitations across surveillance and media have also risen to be concerning in modern-day society, in a simultaneous flow with positive elements.

As early as March 2023, newly introduced AI technologies started deceiving modern-day security protocols. Widely known as ReCAPTCHA, or *Completely Automated Public Turing test to tell Computers and Humans Apart*, was victimized by the acts of GPT-4. Even though ReCAPTCHA worked effectively against bots and automated algorithms, the newly designed GPT-4 was able to identify ambiguous images and decipher puzzles based on its trained image algorithm, critically counteracting the online procedure. Media such as YouTube and TikTok started to commercialize this method, and due to the spread of knowledge, many ReCAPTCHA formats became compromised in front of the cutting-edge LLM, concerning the effectiveness of the security norms society had. Though this achievement done by AI didn't harm people, it meant that humans just became far more vulnerable to exposure to AI and potentially misused versions of it.

Fast-forwarding to January of 2025, a different AI language model shocked the public with its content distortion and censorship, which was the DeepSeek AI developed in China. That particular AI model, before the censorship controversy, was acclaimed for its astounding performance compared to expensive AI model plans given by monopolized leaders such as OpenAI or Google, becoming a fast-growing model in the world. However, a nuance analysis of content accuracy revealed that DeepSeek failed to correctly answer questions that were somehow related to the Chinese government and the Communist Party in China. This limitation in an LLM concerned many users about the potential distortion or refusal that the AI would make to convey certain propaganda and assertions. The censorship applied to the model triggered users to be reluctant to use that particular model, leading to a constricted state of freedom of expression.

It was not solely surveillance that was withered by the advent of AI. In 2025, several content-driven by AI were condemned and removed from multiple platforms. In May of South Korea, multiple teenagers were arrested for creating and distributing more than 850 deepfake pornography images of celebrities and ordinary people using conventional AI applications. Even though the Korean police quickly engaged with traditional investigations and calls for judicial measures, aside from charges of sexual crime and generating fake content, no evident actions were taken on the usage of AI as a method of crime. Alongside, it was suggested that modernized laws are required for AI-specific crimes based on the conventional access to tools.

Away from sexual AI crimes, in early June, multiple YouTube channels were caught using AI-generated images and videos that presented fake trial rulings of the rapper “Diddi” and several other celebrities, including Will Smith, Justin Bieber, etc. The videos attempted to integrate newscast formats and thumbnails of judges alongside celebrities to make it as if the videos were qualified as truth or to be believed by users. In response to these truth-concealing videos, YouTube made radical decisions to pause revenue from the channels and even delete those accounts that spread fake content under its policy of scam, fraud, and misuse of information. It was much more apparent that AI could be an accessible method for information to be distorted and deceitful, potentially being a threat to users.

Conclusively, AI technologies are in a state where they are capable of being used as a medium in cracking privacy, enhancing surveillance, distorting truth, and many other malicious purposes, but are not critically restrained by legal means in many countries or organizations. Since AI is a rapidly changing technology, unpredictable even by human statistics, it is hard to standardize a regulation that could effectively restrain these actions from pausing. Also, its convenience and accessibility contribute to an easier occurrence of crime and violations, which obscures how people could be protected against a complex and flexible software.

Current State of Affairs

In 2025, a number of recent and ongoing events are driving the discussion over human rights protection in the AI era. The need to create annual responses has increased due to the exponential spread of deep fakes, especially those that impersonate political figures. Concerns about political deepfakes and vote manipulation have increased as a result of the major elections that will be held around the globe this year. As a result, several states, including the US, have passed or enhanced legislation requiring the revelation or outlawing certain uses of deepfakes during campaigns. Ratification of the AI and Human Rights Convention of the Council of Europe is a new legal standard-setting benchmark that can be applied to national and regional approaches to AI regulation on a global scale.

Protecting human rights concerning artificial intelligence, particularly as it relates to surveillance and deepfakes, is a developing area that is constantly evolving. There are some areas, such as Europe, that have full frameworks such as the EU Artificial Intelligence Act, as well as the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law set up by the Council of Europe, with a focus on encouraging the development and use of artificial intelligence while upholding fundamental rights and freedoms, including privacy, freedom of expression, and protection. Other areas, such as Australia and the United States, are changing existing laws to deal with related issues such as privacy and online safety, and risks associated with deep fakes, amongst others.

Nevertheless, the landscape for deepfakes is full of ongoing gaps and inefficiencies. Too many jurisdictions continue to build upon fragmented or increasingly restrictive sets of regulations - especially those outside the European Union. For example, while the US has laws addressing issues around deepfakes, it only responds to certain harms, is state-specific, and only deals with certain states of harm, such as unwarranted pornography deepfakes, interference in elections, etc., not the full potential for abuse. Law enforcement still remains an issue, particularly in relation to transnational crime and the ease of distribution of content online in real time. Even more concerning are the deeply entrenched 'technical and legal' grey areas, also very blurry with deepfakes and other AI for surveillance systems, which could force victims to rely only on less stable sources of law, like privacy, defamation, and fraud law, which have little to do with some of the unique challenges of AI-based content. In addition, holding AI systems accountable and developing transparency for automated decision-making remains an issue, particularly regarding algorithmic bias, lack of explainability, and few means for redress or redress options.

The implementation of the EU AI Act and the AI and human rights treaty, which are significant turning points for the advancement of legal protection and unified regulation, are noteworthy regulatory improvements on the development front. However, deepfake technology and AI-based surveillance are becoming more powerful, sophisticated, and widespread, usually outpacing technological countermeasures and regulations. As a result, even though the law is being strengthened in some areas, the whole picture is still not complete, and new problems—like the collection of large amounts of biometric data and even more convincing deep fakes—continue to emerge, sometimes more quickly than the law does.

However, on the other hand, AI is seen as a good and safe technology, when responsibly developed and used, because it creates access to justice, creates better healthcare outcomes and emergency response, fosters economic growth through increasing productivity, democratizes access to advanced technologies, and provides efficient, fair, and transparent public services. In education, it offers personalized learning and improved access to education, which, as we saw when access to education during the pandemic improved inclusion and participation. Legal and ethical frameworks like the EU AI Act provide a critical baseline safeguard for accountable and transparent AI systems that respect human rights. While there remain challenges and risks, there is a growing global consensus of engagement through the active involvement of various stakeholders, that responsible AI offers social benefits that increase human rights while enabling innovation - a way to summarize the benefits of responsible AI that we can leverage by embracing risks, challenges, and regulation.

Stances of Parties

Australia

Australia takes a conservative and constraining approach to allowing AI within its nation. As early as 2019, the Australian government set AI Ethics Principles nationally as a proactive standard for universal AI applications before the AI boom in the 2020s. Even before AI-affiliated measures, Australia has established firm grounds in deepfake content, sanctioning the eSafety Commissioner policy since 2015 to issue takedown orders on any deepfake content if reported by viewers.

Canada

Canada takes a neutralistic and more liberalized stance on the utilization of AI. Canada has progressed to pass the C-63 Act- or the Online Harms Act- to prevent any forms of malicious online content with efficacy and adaptability. Canada, in line with the booming AI trends, has introduced the Artificial Intelligence and Data Act (AIDA), making it one of the first countries to respond to the commercialization of AI products and the rise of generalized AI software.

China

China, extending its Great Firewall of China (Online "Greatwall" from third-party content) infrastructure, takes advantage of surveillance and control, applying AI as a medium for a flexible and delicate system for precautionary measures to counteract crime and potential factors to unrest. Recently, China's proposal of affordable yet comparable AI Models (e.g., Deepseek) has sparked global interest in the technological sector, driving the ever-expanding market for LLM models. China's subtle technological advancements and strict structure are epitomizing the application of AI as security measures, prior to some privacy concerns.

Democratic People's Republic of Korea (DPRK)

DPRK, despite obscuring its technical strategies of surveillance and content spread, attempts to put restrictive measures of surveillance and install AI, which are comparatively innovative for North Korea in its deprived state of advancing beyond international norms. North Korea, adhering to traditional stances to the international community, has recently developed to exploit and hack cryptocurrencies using AI algorithm crackers and generate fake profiles to automate more mining processes of the same cryptocurrency through the use of AI.

Egypt

Egypt, despite having issues with digital violence against women, takes a rather stagnant position on the particular agenda due to its focus on solving more fundamental problems in the nation, such as debts and urbanization flaws. However, Egypt has taken some preliminary steps to apply AI safely, such as holding an Egyptian AI Charter in April of 2023 and incorporating an AI-powered deep and dark web monitoring system in the Cyber and Information Security Exhibition and Conference (CAISEC).

France

Associated with the European Union (EU) AI Act regulating and standardizing the use of AI in European nations, France has imposed a stance where it would take risk-based procedures on any potential forms of AI exploitation and malicious uses. Simultaneously, France has applied AI surveillance technology during its hosting period for the Olympic Games to enhance security against human errors that might occur. Overall, France is taking a more positive approach to acquiring AI infrastructure.

Germany

Germany takes AI in a comparably objective manner to other neighboring countries, restricting AI surveillance programs starting from government cabinets, police, and some organizations, but also constraining the accessibilities and permissions surrounding the AI implementation. The Bundesrat (German Federal Council) has already established legislation to punish cyber-violence, including deepfakes. Nevertheless, there are some holes for the German nation to fill in on par with AI-generated fake media.

India

India states that it is extending their Central Monitoring System (CMS), enabling interceptions of telecommunication across the web, which recently implemented AI algorithms to enhance network traffic analysis as a method of preventing crime. That said, India still lacks a mature description of bills or legislation that would prevent AI from being used as a medium for crime, requiring the nation to adopt a robust resolution that would allow it to address deepfakes regarding the tech.

Japan

Japan has primarily supported the freedom in utilizing AI across public and private sectors of the nation, but at the same time, it has enacted multiple regulations to safeguard the users as well. As recently as May of 2025, Japan's parliament enacted an "AI Bill," the first ever in the country to regulate AI according to the protection of citizens and further inspect the proliferation of AI technologies throughout society. Prior to that, Japan's cybersecurity laws also have reinforced penalties and binding powers upon the misuse of deepfake and AI generative content if provoked by the victim.

Pakistan

Pakistan pertains to the affirmative sides in engaging AI as a tool to moderate and track anomalies in society. Exemplified by the Face Trace System (FTS), under Punjab Province's Police Sector, AI motion and facial recognition systems have been in existence in modern-day Pakistan, perhaps showing signs of a lack of transparency or personal privacy overlooked by public observations. On deepfakes, Pakistan, experiencing fake AI content obstructing the 2024 election, has made Pakistan show deep remarks in advancing its laws to deal with deepfakes and punish with more severe decisions.

Phillippines

The Philippines has taken an almost stubborn position in using AI as a security measure since 2012. Even in the Data Privacy Act and Cybercrime Prevention Act prior to the widespread use of AI, CCTVs, and facial recognition systems that included any tracking algorithms, they were critically limited to the national level, not easily accessible to individuals or the private sector. With its House Bill 10567, which gives accountability to deepfake creations, the Philippines is taking a new approach to prevent sexual abuses or misinformation occurring through the flexible media

Poland

Poland, adhering to the EU AI Act, is looking at a new paradigm of technology as the nation seeks adamant military and integrity. Poland has integrated real-time AI surveillance and made it legal inside Poland, compared to the old, complex judicial process that AI infrastructures had to go through. Also, Poland is seeking to draft a new AI Act, independent of international organizations, expecting it to be implemented by 2027 for a standard of AI applications. However, their process to amend the technological laws is on hold due to conflicting aspects between EU requirements.

Russia

Russia, with its ongoing war with Ukraine, is taking the position as an absolute advocate for operating national surveillance systems. As of 2024, over 1 million CCTVs were engaged with facial recognition systems and connected to the System for Operative Investigative Activities(SORM) framework to enable quicker and efficient investigations on cases online that could potentially prevent crime. Urban surveillance systems with AI facial recognition technology were rationalized by privacy rights, more adhesively allowing an organization to control the public. Also, regarding deepfake punishments, it is estimated that Russia would update its Criminal Code based on newly added amendments on content made with AI that may be distorting and confusing for citizens.

Republic of Korea(RoK)

The Republic of Korea, a nation focused on telecommunications and hardware, had a comparatively slow start to conform to the trend of allowing AI moderation & surveillance systems, while simultaneously banning and tightening regulations on AI misuse. Due to political polarization and recent martial law, the National Assembly is in a stagnant position to enact regulations on sexual and fraud-associated crimes derived from AI. Also, internal reports have shown concerns and statistical growth of AI deepfake crimes in universal ways, despite its increased efforts to enhance cybersecurity.

Saudi Arabia

Saudi Arabia, along the lines of its Vision 2030- an innovative project and NEOM Initiative, aims to rework its online infrastructure and laws through the Saudi Data & AI Authority(SDAIA). In this path, Saudi stands for applying AI surveillance in CCTVs and online networks as a method of enhancing public safety, and simultaneously, a detection method for deepfakes. As Saudi Arabia is approaching its advancement as an AI powerhouse, they are experimenting with multiple AI measures and stopping illicit generative content.

South Africa

South Africa, already implementing multiple AI-driven surveillance technologies, implies their position of conforming to the general flow of adapting AI into societies. Facial recognition, license plate readers, and registration analyses are all predominant AI elements in the nation, showing the nation's strong association with AI surveillance. On the side of deepfakes, enforcing the Protection of Personal Information Act since 2013 has reflected South Africa's adamant opinion on criminalizing deepfake creations, securing its protective stance.

Turkey

Turkey uses facial recognition, which is one of the increasing applications of AI systems, in locations that sell items illegal to minors under the age of 18, such as cigarettes and liquor, in order to protect both the minors and appropriate shops from punishment. Also, Turkey has increasingly adopted AI-analyzed crowd control CCTVs after recent unrest and protest outbursts, taking its natural stand for AI surveillance acceptance. Additionally, any harmful acts of AI content creation that could lead to crime, such as fraud through deepfakes or generative content, could be harshly punished under articles 243-245A.

Ukraine

Ukraine takes the place of a strong advocate in applying AI across multiple sectors, even throughout battlefields, wielding AI as an unprecedented strategy to turn the tide in its current war. In its times of unrest, Ukraine has identified numerous Russian soldiers using facial recognition technology provided by national AI firms and some from third-party attributes. As deepfake content is thwarting the public to understanding of the tides of war, the nation is strictly banning false AI-generated deepfakes or any kind of misinformation, taking very extreme steps in AI utilization and punishments associated with exploitation.

United Kingdom(UK)

The UK amends its informational and cybersecurity laws to follow the modern trend of AI implementation, as it plans to add regulatory statements in its Data Protection and Digital Information Bill, and already has criminalized the creation of sexually explicit deepfakes done inconsistently. Due to the negative reputation that AI-driven deepfakes have generated, the UK takes a rather conservative stance in accepting AI content. Though the UK has taken reluctant views on AI content, it has started its implementation of AI CCTVs across public and government property since early 2025.

United States of America(USA)

Even though the United States owns most AI behemoths, no national-level AI policy on surveillance has been made. That said, several states established their own legislation upon utilizing AI surveillance systems. In addition to this absence, the United States needs a nationwide reformation of cybersecurity law that would illustrate correct usages of AI surveillance that still complements the existing AI conventions. On the contrary, the *Take It Down* Act on deepfakes was enacted in May 2025, which addresses the prohibition of nonconsensual online publications that carry visual depictions of individuals, critically counteracting the widespread creation of misinforming AI-generated media.

Vietnam

Vietnam is currently drafting a reformation of its technological and digital law, which mandates the labeling of AI-generated content and classifications for more accurate disclosure to users on the setback that AI content could misintentionally elicit. Currently, surveillance practices are done occasionally upon biometric authentication services using faces for genuineness, but due to risks of also of deepfakes, caution has been evoked amongst citizens since 2022.

Possible Solutions

Global Framework on Ethical AI Surveillance
The elaboration of such a framework may be undertaken with the UN, member states, and civil society organisations. It would attempt to establish legal and ethical guidelines for the use of AI, especially in government. This framework will cover transparency standards for independent oversight, public reporting on surveillance programs, and a record of data collection, storage, and use. Standard human rights impact assessments before a system is adopted may be required; the report advises on allowing an AI surveillance system to be adopted. Member states could also agree to international peer reviews and to have the findings published in an annual UN report as another means to foster accountability. Consider proposing that until moral and legal safeguards are established, technologies such as facial recognition should have limited access or be limited or paused for a period of time. This would protect privacy while minimizing the risk that AI would be misused to exploit the vulnerable, silence dissent, or violate free speech.

International Guidelines
The establishment of international norms and mechanisms to detect and contain the destructive proliferation of deepfakes. Member states could support the establishment of a Universal Deepfake Tagging Protocol (UDTP) to force AI-generated material to carry visible or invisible digital markers indicating it's synthetic. Alongside detection technology, citizen education campaigns might be rolled out to help people spot and make sense of manipulated media. Delegates could also propose the creation of a shared global fund to help developing countries access deepfake detection tools and training programs. This approach not only helps protect individuals from fraud, harassment, or political misinformation but also encourages cooperation between nations, tech companies, and researchers to uphold the integrity of information in the digital space.

Establishment of Independent AI Accountability and Oversight Bodies
It is vital to create autonomous accountability and oversight bodies for artificial intelligence (AI) at both national and international levels for holistic algorithm audits, impact assessments, and enforcement of human rights in AI surveillance and deepfake technologies. These bodies would review AI systems for bias and discrimination and transparency both pre and post-deployment, and require comprehensive documentation of AI decisions and assessments from the developers (while providing protections for proprietary information), require jurisdictions to provide documented means of complaints/remedies for individuals affected by adverse decisions (outlining responsibility), require that there are documented regular publications by developers/organizations to be publicly accessible (providing transparency), and would also notify the public of events where documented harms have impacted people (ensuring accountability). The bodies would also be able to liaise with any international bodies to harmonize frameworks and assist with cross-border jurisdictional dispute resolutions. These oversight institutions would address critical enforcement gaps, instill trust and integrity in AI deployments, and build and maintain accountability for deploying AI technologies consistent with human rights and ethical approaches, complementing broader multilateral frameworks and multi-stakeholder engagement forums.

Questions to Consider

1. How should societies balance the need for security and public safety provided by AI with the fundamental right to privacy and freedom?
    a. What precisely are the legal and technical measures that will potentially protect Privacy with AI as it pertains to surveillance and not jeopardize Security?
    b. In what ways will transparency of AI use build public trust and safeguard operational needs?

2. What ethical responsibilities do the government and tech companies have in an AI system perpetuating bias or discrimination?

3. How might your country's economic priorities, technological capabilities, or political system shape its approach to regulating AI surveillance and deep fakes?

4. What mechanisms do you need to ensure transparency and accountability for individuals harmed by AI or deep fake misuse?

5. How can international cooperation help solve cross-border challenges posed by AI-driven surveillance while still respecting national sovereignty?

6. How will educating people and public awareness patrols help facilitate informed and responsible use of AI?
    a. What strategies will increase literacy around AI and deepfakes, particularly in diverse populations?

7. What role should independent oversight and civil society organizations play in overseeing AI surveillance and deepfake technologies?
    a. What mechanisms will strengthen their ability to shape policy and hold actors accountable?

Bibliography

ADHRB. “AI Regulation in Saudi Arabia: Innovation over Human Rights | Americans for Democracy & Human Rights in Bahrain.” Americans for Democracy & Human Rights in Bahrain, 23 July 2024, www.adhrb.org/2024/07/ai-regulation-in-saudi-arabia-innovation-over-human-rights/? Accessed 13 July 2025.

AIPRM. “AI Laws Around the World.” Aiprm.com, 28 Apr. 2024, www.aiprm.com/ai-laws-around-the-world/. Accessed 13 July 2025.

Guardian staff reporter. “Fake, AI-Generated Videos about the Diddy Trial Are Raking in Millions of Views on YouTube.” The Guardian, The Guardian, 29 June 2025, www.theguardian.com/technology/2025/jun/29/fake-diddy-ai-videos-youtube. Accessed 25 July 2025.

Jabkhiro, Juliette, and Julien Pretot. “How France Plans to Use AI Surveillance to Keep Paris 2024 Olympics Safe.” Reuters, 8 Mar. 2024, www.reuters.com/sports/olympics-how-france-plans-use-ai-keep-paris-2024-safe-2024-03-08/?. Accessed 13 July 2025.

Jha, Raj Shekhar. “Eye Spy! Delhi Police Sharpens Surveillance in Digital Age AI Image; to Use Next-Gen Netra to Identify Potential Threats.” The Times of India, Times of India, 18 June 2025, timesofindia.indiatimes.com/city/delhi/eye-spy-delhi-police-sharpens-surveillance-in-digital-age-ai-image-to-use-next-gen-netra-to-identify-potential-threats/articleshow/121937425.cms?. Accessed 13 July 2025.

Justice, Ministry of. “Government Crackdown on Explicit Deepfakes.” GOV.UK, 7 Jan. 2025, www.gov.uk/government/news/government-crackdown-on-explicit-deepfakes?. Accessed 13 July 2025.

Law of Ukraine on Regulating Artificial Intelligence Technologies and Addressing the Issue of Deepfakes to Protect Democratic Processes in the Context of Russian Military Aggression Section I: General Provisions Section II: Regulation of Deepfake Creation and Distribution Article 3: Prohibition of Malicious Deepfakes. cotler.tau.ac.il/wp-content/uploads/2024/12/Law-of-Ukraine-on-Regulating-Artificial-Intelligence-Technologies-.pdf? Accessed 13 July 2025.

Law, SD. “Digital Evidence.” SD Law Cape Town Attorneys, 12 May 2025, www.sdlaw.co.za/articles/digital-evidence/? Accessed 13 July 2025.

Scassa, Teresa. “AI, Human Rights, and Canada’s Proposed AI and Data Act.” Teresascassa.ca, 2024, www.teresascassa.ca/index.php?Itemid=80&id=380%3Aai-human-rights-and-canadas-proposed-ai-and-data-act&option=com_k2&view=item. Accessed 13 July 2025.

Sharma, Shweta. “North Korea’s BlueNoroff Uses AI Deepfakes to Push Mac Malware in Fake Zoom Calls.” CSO Online, 19 June 2025, www.csoonline.com/article/4009603/north-koreas-bluenoroff-uses-ai-deepfakes-to-push-mac-malware-in-fake-zoom-calls.html?. Accessed 13 July 2025.

Tech, ID. “Poland Expands AI Surveillance Powers with New Public Security Law - ID Tech.” ID Tech, 27 May 2025, idtechwire.com/poland-expands-ai-surveillance-powers-with-new-public-security-law/. Accessed 13 July 2025.

Ted, Cruz,. “S.146 - 119th Congress (2025-2026): TAKE IT down Act.” Congress.gov, 2025, www.congress.gov/bill/119th-congress/senate-bill/146. Accessed 13 July 2025.

"AI Watch: Global Regulatory Tracker - Japan | White & Case LLP." Whitecase.com, 13 June 2025, www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-japan. Accessed 13 July 2025.

"DataGuidance." DataGuidance, 2025, www.dataguidance.com/jurisdictions/philippines?. Accessed 13 July 2025.

"EU Artificial Intelligence Act | Up-To-Date Developments and Analyses of the EU AI Act." Artificialintelligenceact.eu, 2024, artificialintelligenceact.eu/. Accessed 13 July 2025.

"Legal Framework for Artificial Intelligence Development under Vietnam's Draft Law on Digital Technology Industry." Apolat Legal, 4 July 2025, apolatlegal.com/legal-framework-for-artificial-intelligence-development-under-vietnams-draft-law-on-digital-technology-industry/? Accessed 13 July 2025.

"OECD AI Policy Observatory Portal." Oecd.ai, 2025, oecd.ai/en/incidents/2025-02-10-8bc8. Accessed 13 July 2025.

"Pakistan Implements AI-Powered Criminal Identification System | Digital Watch Observatory." Digital Watch Observatory, 29 Feb. 2024, dig.watch/updates/pakistan-implements-ai-powered-criminal-identification-system?. Accessed 13 July 2025.

"Regulating Privacy and Digital Identities in the Age of AI: The View from Egypt." The Cairo Review of Global Affairs, 21 June 2025, www.thecairoreview.com/essays/regulating-privacy-and-digital-identities-in-the-age-of-ai-the-view-from-egypt/?. Accessed 13 July 2025.

"The Basic Act on Cybersecurity - English - Japanese Law Translation." Japaneselawtranslation.go.jp, 2019, www.japaneselawtranslation.go.jp/en/laws/view/3677/en. Accessed 13 July 2025.

"Turkiye: Adopted Data Protection Authority Guidelines on Deepfakes." Digitalpolicyalert.org, 2024, digitalpolicyalert.org/event/17085-adoption-of-deepfake-information-note-by-turkish-data-protection-authority. Accessed 13 July 2025.

조선일보. "AI 딥페이크 합성 음란물 500개 제작, 유포한 10대 구속." 조선일보, 22 May 2025, www.chosun.com/national/regional/2025/05/22/EDCB3EKNPBEPTCHH2PGLJYOP6M/. Accessed 25 July 2025.

신상민. "딥페이크 성범죄 느는데 검거율 하락…AI 기본법 '한계.'" Newstomato.com, 27 Mar. 2025, www.newstomato.com/readnews.aspx?no=1257321. Accessed 25 July 2025.