Protecting Human Rights while Using Artificial Intelligence to Counter Terrorism

Position Paper

United Nations Special Rapporteur
on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism

December 2025

SPECIAL PROCEDURES
UNITED NATIONS
HUMAN RIGHTS COUNCIL

Contents

Acknowledgements

This Position Paper was primarily drafted by Jonathan Andrew, a consultant to the mandate of the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism, Research Fellow at The Geneva Academy of International Humanitarian Law and Human Rights, and co-editor of *Human Rights Responsibilities in the Digital Age: States, Companies and Individuals* (Bloomsbury, 2023). It was supervised and revised by the Special Rapporteur and reflects the mandate's position.

In preparing this Position Paper, the mandate adopted sought the views of stakeholders through a call for inputs and other outreach, including States, international and regional organisations, civil society, the private sector, academics, multi-stakeholder initiatives and other actors. Forty-eight submissions were received, including from 16 States, three international organisations, and 21 civil society organisations.[1] Submissions are published on the mandate's website,[2] except where confidentiality was requested.

The expert advice is also acknowledged of Professor Kubo Mačák (University of Exeter) and of the thematic sections on Counter-terrorism and human rights and on Technology and human rights of the Office of the High Commissioner for Human Rights (OHCHR) in Geneva.

The work on this topic was stimulated in part and informed by the Expert Meeting on the Rule of Law and Human Rights Aspects of Using Artificial Intelligence for Counter-Terrorism Purposes in Geneva on 8 May 2025, hosted by the Geneva Centre for Security Policy in cooperation with Switzerland and the Counter-Terrorism Executive Directorate.

This research was funded by the Government of Switzerland and The University of Sydney.

This project was assisted by personnel of the mandate, including Karen Reyes Tolosa (Human Rights Officer, OHCHR Geneva), Loana Benjamin (Associate Human Rights Officer, OHCHR Geneva), Alexandra Lily Kather (consultant Legal Adviser, New York), Alexandra Meagher (consultant Legal Adviser, Sydney), and Netra Hankins (Intern, OHCHR Geneva).

This Paper is part of the mandate's long-standing work on new technologies and counter-terrorism, including on mass surveillance, biometrics and facial recognition, spyware, travel and border management technologies, new financial technologies, drones, lethal autonomous weapons systems and cybercrime.[3] A position paper on facial recognition will be published in early 2026, and further work on online terrorist content moderation is planned for 2026.

This Paper is available online at: https://www.ohchr.org/en/documents/position-papers/protecting-human-rights-while-using-artificial-intelligence-counter

[1] Submissions are available at https://www.ohchr.org/en/calls-for-input/2025/call-input-position-paper-human-rights-impacts-using-artificial-intelligence. State submissions were received from Argentina, Azerbaijan, Bosnia and Herzegovina, Chile, Cuba, Guyana, North Macedonia, Norway, Spain, Switzerland, Venezuela and five States who requested confidentiality. International organisation submissions came from the Council of Europe, European Union and, in confidence, another regional organisation.

[2] https://www.ohchr.org/en/calls-for-input/2025/call-input-position-paper-human-rights-impacts-using-artificial-intelligence.

[3] See https://www.ohchr.org/en/special-procedures/sr-terrorism/new-technologies and https://www.ohchr.org/en/special-procedures/sr-terrorism/index-thematic-issues-2005-present.

Introduction: The promise and risks of artificial intelligence in counter-terrorism

Artificial intelligence (AI) originated more than half a century ago and so has been an "emerging" technology for many years.[1] Its application to enhance the autonomy of systems or to support decision-making is, however, comparatively recent, following advances in computing that have enabled significant new technologies to counter terrorism.[2]

Today, AI can potentially be deployed in a broad array of activities in counter-terrorism, with the potential in some settings to increase the precision, efficiency and effectiveness of law enforcement, minimise unnecessary intrusions, and mitigate human biases.[3] Its uses include in threat detection, risk assessment, physical and digital surveillance, predictive policing and force deployment, enhanced forensics, border management, judicial proceedings, detention and military decision-making, among others. Its use can also enhance other technologies and capabilities used in law enforcement, from scenario planning, simulations and resource allocation to modelling the logistics of operations and surveillance activities.[4]

AI systems can operate with varying degrees of autonomy: the more autonomous an AI system, the less direct human control there is over its actions. AI can assist decision-making in allowing for the more rapid collection and processing of large volumes of data and can potentially even displace human judgment in specified situations. These capacities raise fundamental questions as to the role and purpose of human agency, reasoning, rationality and judgment in often complex counter-terrorism contexts.[5]

Critically, certain uses of AI systems by State authorities in countering terrorism can profoundly infringe human rights. AI algorithms can aggregate and analyse highly personal or sensitive data such as arrest and criminal records, associations with family members, friends and colleagues, patterns of crime and policing, social media networks and posts, communications data, travel information, employment records and personal information held in government databases for social, health, education and other services.[6] Such information may be used to profile the alleged terrorist risks of individuals and groups and prompt invasive counter-terrorism measures, including surveillance, arrest and detention, administrative restrictions, kinetic operations, and other measures to prevent terrorism and violent extremism.

Yet, the AI systems are far from infallible. Their datasets and algorithms may be biased, too limited or unrepresentative, illegally sourced or inaccurate, potentially leading to discriminatory profiling, violations of the right to privacy, and counter-productive false positives. AI systems may lack adequate human control and analysis, resulting in over-reliance on automated decision-making lacking in nuance, context and responsiveness to human rights considerations. AI assessments are merely probabilistic, not perfect predictions.[7]

More fundamentally, many AI systems are still in developmental, experimental or testing phases, and their capabilities, including in real-world conditions, remain to be proved. There is a real risk that proponents of AI, including corporate actors fuelling the current global AI investment boom, may overstate the benefits and minimise the costs and risks. Since the number of actual terrorists in most populations is very small, serious dataset limitations can prevent meaningful extrapolation for the purpose of future threat detection, particularly when predicting the behaviour of individuals.[8] The proprietary and national security secrecy surrounding AI development and deployment can also shield the technology from independent scrutiny to validate its effectiveness and risks, including both false positive rates and failures to detect risks, as well as from effective operational oversight. Indeed, some AI security technologies may be entirely withheld from and unknown to the public. There is limited empirical data available about how widely AI systems are being used by counter-terrorism and law enforcement authorities globally, what kinds of systems are in use, objectively how effective they are, what rights violations and other harms they are causing, and where the balance between security and liberty is being set.

The diverse uses of AI in counter-terrorism contexts have the potential to violate numerous fundamental rights, including privacy and data protection; equality and non-discrimination; freedoms of expression, association, peaceful assembly and religion; political participation; liberty; security of person; life; and the rights of vulnerable, marginalised and disadvantaged groups.[9] The work and safety of human rights defenders is also at risk.[10]

In the experience of United Nations human rights mechanisms, the use of new technologies in counter-terrorism has unleashed new waves of human rights violations targeting civil society, human rights defenders, journalists and political opponents, and there is good reason to expect the same from AI unless old habits change. The Security Council has urged States to use new technologies to counter terrorism without paying sufficient regard to the human rights risks, including in countries lacking human rights protections, independent judiciaries, a rule of law culture, or democratic oversight. Democracies too have abused technological tools. The Council has also neglected to adequately address adverse or unintended consequences as its resolutions have been implemented. There has sometimes been insufficient human rights due diligence and safeguards in United Nations technical assistance and capacity building involving new technologies.[11]

Over-confidence in technological solutions, as part of the predominantly repressive rather than preventive approach to counter-terrorism by many States, can further detract from addressing the conditions conducive to terrorism, including State violations of human rights. Since counter-terrorism often serves as a laboratory for new technologies, there is the additional risk that AI developed to confront exceptional security threats will be normalised in other contexts, widening the threat to human rights.

The General Assembly and Human Rights Council have affirmed that AI systems must be developed and used in ways that respect, protect, and promote human rights and that Member States must refrain from deploying systems that cannot comply with international human rights law or that pose undue risks to rights.[12] Global calls for inclusive, transparent, and accountable AI governance[13] must address the particular risks in counter-terrorism and security contexts, including by rejecting the troubling exemptions in some recent legal frameworks for AI systems used for national, public and border security, military and law enforcement purposes.

This Position Paper focuses on the key human rights risks of developing and using relevant AI systems to counter terrorism. It addresses how the design, development and use of AI engage

international human rights law, how human rights must inform decisions relating to the operation of AI systems, and the responsibilities of public authorities and private actors. Specifically, the Paper outlines the characteristics and uses of AI systems that are most relevant in countering terrorism, as well as their limitations and main human rights risks, including to equality and non-discrimination, privacy and data protection, freedom of expression and access to information, fair trial, liberty and rights in detention, and effective remedies for rights violations. The Recommendations build on existing AI governance frameworks by identifying opportunities to strengthen regulation and accountability.

The Recommendations set out some essential elements of adequate legal frameworks to regulate AI systems to protect human rights in counter-terrorism:

- There must be clear and specific legal authorisation of AI, augmenting wider technology-neutral regulation. AI governance should be left to ordinary security, law enforcement or criminal justice rules, or to a laissez-faire free-market or industry self-regulation approach.
- It must be prohibited to develop, use and transfer and export AI systems that cannot meet human rights standards or pose an unacceptable risk to rights, and there must be heightened regulation of other AI systems that are permitted given the sensitivity of counter-terrorism contexts.
- Developers and deployers of AI systems, including public authorities and private actors, must conduct human rights due diligence assessments throughout their lifecycle,[14] with due attention to vulnerable groups (including minorities, children, women, and persons with disabilities), and avoid or mitigate risks through "safety by design".
- Effective human control must be exercised over AI systems, and high-risk technology that is incapable of meaningful human supervision must be prohibited.
- There must be stringent safeguards on data quality, testing and validation and assurance of the transparency and explainability of AI systems.
- Personal data protection and data security must be guaranteed.
- Applications of AI in specific areas, such as surveillance, moderation of online terrorist content, border management, judicial proceedings, detention and military operations, must be subject to specific measures to comply with international human rights law and other relevant frameworks, such as refugee law and international humanitarian law.
- AI systems must be subject to independent oversight bodies, and accountability mechanisms to provide effective and accessible remedies to victims of rights violations.

The speed of development of AI systems, uncertainty about their present and future capabilities, and a competitive sovereign, security and commercial AI "arms race" all present significant challenges to regulation and tend to promote "regulation lag". Nonetheless, States, the international community, the private sector and civil society should not fatalistically believe that humanity is irrepressibly hostage to the march of technology, commercial interests or absolutist security imperatives. The power to effectively regulate – and the moral and legal imperative to do so to protect human rights – are squarely within our hands.

1. AI in intelligence collection and analysis and threat detection

The integration of AI into intelligence gathering represents a significant shift for contemporary counter-terrorism capabilities, fundamentally augmenting the capacity to collect, process, and analyse vast quantities of data to detect threats.[15] Exponential growth in digital communications, online activities, and ever more interconnected information systems have generated unprecedented volumes of data that exceed traditional analytical capabilities.[16] AI technologies, particularly machine learning algorithms and natural language processing systems, can help to manage this inundation of data, potentially enabling security agencies to more rapidly, efficiently and effectively identify patterns, anomalies, and potential threats.[17]

Machine learning algorithms are particularly useful in processing heterogeneous data sources, including communications metadata, financial transactions, travel records, and social media activity. These systems can analyse millions of data points simultaneously, identifying correlations and behavioural patterns that may elude human analysts,[18] who are also limited in number. Through supervised learning techniques, AI systems can be trained on known terrorist profiles and attack methodologies and can apply this knowledge to detect similar characteristics within broader populations. Unsupervised learning approaches could further identify novel threat patterns that deviate from established profiles or models, potentially revealing emerging terrorist tactics or previously unidentified networks, online or in physical locations.

Advances in natural language processing are also increasingly driving the automated analysis of textual and verbal communications across multiple languages and dialects. These can be targeted to enhance the monitoring of online radicalisation, recruitment, and operational planning for terrorist attacks. These capabilities are also being used for so-called sentiment analysis, leveraging algorithms to assess the emotional tenor of communications, potentially flagging what may be shifts towards terrorist intent or violent extremist ideologies. Complementary developments in network analysis powered by AI enable the mapping of complex organisational structures, which can assist in identifying key nodes within terrorist networks and revealing previously obscure connections between individuals and groups.[19]

AI-enabled data processing enables assessments at greater speed, including real-time or near-real-time threat assessment. Traditional human intelligence analysis, often coupled with human resource, technology and language constraints, can result in substantial delays between collecting data and generating actionable intelligence from it.[20] AI systems could reduce these intervals significantly, potentially enabling the prevention of terrorist attacks.

The acceleration of data collection, processing and analysis, and the new insights it offers, brings serious risks of both intensifying existing patterns of human rights violations stemming

from surveillance and information collection generally, particularly concerning the rights to privacy, data protection and non-discrimination. Since the right to privacy is a "gateway" right protecting and enabling many other human rights,[21] AI's intensification of privacy violations can magnify other human rights harms, including damaging effects on civic space, political participation and democracy – which themselves build resilience against terrorist violence.

2. Advances in behavioural biometrics

Public authorities, including law enforcement, increasingly use biometrics in a wide range of applications to identify individuals. Uses range from verifying individuals' identities in public spaces to regulating access to large-scale events, to the identification of individuals upon arrest, subsequent criminal justice procedures and during detention.[22] In addition, biometrics are increasingly seen as integral to surveillance capabilities to identify individuals when monitoring activities and associations within groups and during social gatherings.

Such uses of biometrics reduce anonymity, limit autonomy and potentially infringe the right to privacy and freedoms of peaceful assembly, association and expression and the right to political participation.[23] In Myanmar, for example, the creation of a nationwide integrated biometric system covering the entire population, in order to tackle crime and protect national security, has raised serious concerns.[24]

Biometric systems are not infallible.[25] To date, little research has focused on how the greater use of in AI in biometrics facilitates human rights violations. The implications for vulnerable groups are of particular concern. Biometrics systems frame human identity in terms of a person's quantifiable physical and behavioural manifestations,[26] presuming that bodies are composed of stable, predictable, and controllable characteristics that record an innate or intrinsic fixed personhood. However, such premises are empirically problematic given human diversity and raise considerable human rights and international humanitarian law concerns.[27]

More widespread use of AI in developing biometrics capabilities may lead to the discriminatory treatment of persons from certain vulnerable groups whose physiological or behavioural characteristics are deemed atypical. For example, persons with limited mobility, older persons, children, neurodivergent individuals, the visually impaired and others with disabilities are more likely to experience challenges when confronted with biometrics in security contexts.[28] Those with cognitive impairments stemming from diseases such as dementia or Parkinson's disease, for example, could find that the notionally irregular features of their gait or body movements render their actions suspect, subjecting them to unjustified greater scrutiny and interference.

The belief that AI unequivocally enhances the accuracy of biometrics in identifying individuals is, as yet (or if ever), unjustified.[29] Historically, evident biases have stemmed from data quality concerns and a lack of equal representation in the datasets used to develop AI systems.[30] Research further indicates that other vulnerable groups such as minorities, migrants, Indigenous Peoples and women and girls are still far more likely to be subject to discriminatory and disproportionately poor treatment by AI-enabled systems that exploit biometrics.[31]

3. AI-driven facial recognition technology (FRT) and networked surveillance

The convergence of AI with facial recognition technology (FRT) is transforming surveillance and shaping security infrastructure, with profound implications for human rights. Deep learning architectures, particularly convolutional neural networks, which make use of three-dimensional image data to classify and recognise objects,[32] have enhanced FRT accuracy to levels often exceeding human performance in controlled environments, enabling automated identification across expansive datasets within milliseconds.[33] These advances have precipitated widespread deployment of AI-driven surveillance systems in public spaces, public transportation networks, and border control facilities, to continuously monitor crowds and flag suspect individuals.[34] The scalability of these systems permits simultaneous tracking of multiple subjects in different locations, creating new opportunities for law enforcement and intelligence agencies[35] and cross-border intelligence sharing.

However, the increased use of AI in FRT raises familiar concerns regarding algorithmic bias, as empirical studies have demonstrated disproportionate error rates across demographic groups, particularly affecting individuals with darker skin tones, women and marginalised communities.[36] Despite many years of research repeatedly identifying these persistent problems, such disparities stem from the continued reliance by developers on unrepresentative training datasets and the continued reinforcement of the underlying algorithmic architectures that perpetuate systemic inequities.[37] Furthermore, the opacity of AI decision-making processes in surveillance contexts challenges the accessibility and effectiveness of accountability frameworks and due process protections.[38] The persistent collection and retention of biometric data also creates risks of function creep, where systems deployed for exceptional security purposes expand into broader monitoring of society,[39] without protections for sensitive personal data and the right to privacy.

4. Enhanced mass surveillance, predictive policing and force deployment

While mass surveillance and bulk data retention are not new, advances in information and communication technologies continue to enhance monitoring capacities that risk interfering in fundamental rights.[40] AI-powered public surveillance systems increasingly monitor community activities. Law enforcement and border security agencies are expanding their use of AI-powered predictive analytics to identify leads and allocate resources, even as human rights groups and civil society organizations raise serious and often unaddressed concerns.[41] AI-driven data fusion capabilities exploit arrays of sensing and monitoring capabilities to develop ever more complex and invasive analysis of mobility patterns, interactions and associations between individuals and groups.[42]

Policies on the operation of vital public infrastructure are already being shaped by AI to develop more nuanced mapping and analysis of movement patterns, develop preventative measures in counter-terrorism operations[43] and secure major sporting and other public events, public spaces and critical infrastructure. Governments, public authorities and the European Union have funded research to develop AI-based technologies to enhance data processing from transport and communication networks for such purposes.[44] These studies have included monitoring the application of AI to better understand urban dynamics, public transportation networks and more nuanced personal and contextual services to enhance pre-emptive modelling, and operational responses, to potential terrorist threats.[45]

AI, in conjunction with increased data retention capacities, allows for more granular and nuanced information relating to individuals to be more quickly and efficiently derived from both personal and contextual and environmental data. Systems that exploit AI for surveillance may therefore accelerate and deepen the invasiveness of detection and screening activities. Behavioural details that relate to personal interests, activities and relationships can be inferred through increasingly complex processing of data. In many instances the data collected may concern completely innocuous activities and be unrelated to any, even minor, unlawful activity, let alone relate to a serious crime such as terrorism. In such cases, the resulting interferences in the right to privacy, and potentially other fundamental right such as liberty, would not meet the requirements of necessity and proportionality under international human rights law. As discussed in section 6 below, AI's intensification of surveillance profoundly threatens the rights to privacy and data protection unless stringent and effective safeguards are imposed.

5. AI as an enabler of systems integration: Vision-language models (VLMs)

Vision-language models are significantly advancing the complexity of AI applications. These multimodal systems can process and integrate visual and textual information to generate sophisticated interpretations, descriptions, and predictions.[46] They employ deep learning techniques to establish semantic connections between images and natural language, enabling applications ranging from automated image captioning to visual question answering and cross-modal retrieval.[47] An example is CLIP, an advanced AI model developed by OpenAI and the University of California Berkeley. The convergence of visual and linguistic data can facilitate a more nuanced understanding of complex scenes, objects, and contextual relationships.

VLMs can be used to detect terrorist threats by analysing multimedia content. They can automatically scan images, videos, and text across social media and online platforms to identify content such as depictions of terrorist violence, terrorist and violent extremist propaganda and recruitment materials. They can potentially assist security agencies in flagging suspicious communications and behaviour, detect patterns of radicalisation, and monitor terrorist networks at scale. They may also enhance the interpretation of visual symbols, recognise hate speech in multiple languages, and identify weapons or possible training facilities from satellite imagery.[48] Content moderation systems have already begun to deploy VLMs to identify prohibited material across social media platforms.

These capabilities can also threaten human rights and public interests. These systems typically require extensive training datasets comprising millions of image-text pairs, often scraped from publicly accessible internet sources without explicit consent from the individuals whose personal data is processed,[49] thus violating the right to privacy. The incorporation of photos, social media content, and other visual data into training corpora raises fundamental questions regarding data ownership, consent, and the right to control one's digital representation. Once trained, VLMs retain latent knowledge of patterns, faces, and contexts encountered during training, potentially enabling reconstruction or inference of sensitive information about individuals who contributed unwittingly to training datasets.[50]

The inferential capabilities of VLMs extend privacy risks beyond direct personal identification. These systems can deduce sensitive attributes including health status, socio-economic background, political affiliations, and personal relationships from visual cues combined with contextual information.[51] When deployed in surveillance contexts, VLMs enable granular and comprehensive behavioural analysis that transcends traditional monitoring and bare

identification, instead generating detailed narratives about individuals' movements, associations, and activities[52] and offering interpretative assessments of intentions and characteristics. Automated profiling may generate discriminatory classifications based on visual attributes combined with contextual information, perpetuating biases in training data.

Further, these models can be exploited for non-consensual image manipulation, creating fabricated or compromising content that could threaten human dignity, including sexually explicit representations or deepfakes designed to humiliate, delegitimise, threaten and deter political opponents, journalists, human rights defenders or civil society actors. Evidence could also be falsified in criminal prosecutions or to impose administrative security measures.

The potentially exploitative use of VLMs thus presents considerable human rights challenges. The opacity of these models' decision-making processes complicates accountability, as the basis for inferences or classifications may remain inscrutable even to system developers. Their use consequently necessitates robust governance frameworks addressing consent, data minimization, algorithmic transparency, and meaningful human oversight to mitigate privacy infringements and other rights violations while preserving legitimate technological benefits.[53]

6. Impacts of AI on online platforms, freedom of expression and civic space

AI systems can potentially both enable and restrict how online communication can facilitate human rights and civic and democratic participation. The digital sphere is nowadays vital to facilitating the rights to freedom of association, peaceful assembly and opinion and expression, and the right to participate in public affairs, both online and in the physical world.[54] The right to freedom of association protects a wide variety of groups, organisations and movements which form in civil society to bring together people with shared interests or convictions, including to defend human rights, protest, and criticise governments and corporate actors. Many associations' activities are mediated by online channels such as social media and messaging apps,[55] which transcend geographic barriers and facilitate local, national and transnational organisation.[56] The architecture of social media disseminates existing causes and generates new focal points for collective mobilization.[57] The immediacy and asynchronous nature of online communications are particularly effective for connecting people who were once disassociated or faced barriers to association, allowing people with shared interests to organise and often eliminating hierarchies or a formal mobilisation strategies.[58] This network effect and its dynamism also enable crowdsourcing and fund raising.[59] Encryption can provide protection from unjustified State interference in legitimate online civil and political freedoms.

Digital platforms can also potentially facilitate and strengthen the agency of vulnerable, disadvantaged and marginalized groups. For example, they can enhance communication among individuals whose disabilities or impairments may otherwise constrain their ability to interact face-to-face, while reducing exposure to offline harassment, intimidation and discriminatory treatment. Those with marginalized identities may leverage the anonymity and physical distancing characteristic of digital discourse as protective mechanisms to facilitate their participation and association online. Engagement in virtual associations is particularly significant for individuals with concealed or stigmatized identities, including those who lack the security to openly disclose their sexual orientation or politically radical viewpoints. These groups value online participation even more highly than their mainstream counterparts.[60]

AI systems can, however, potentially impede these rights-enabling capacities of online platforms. Most problematically, AI can be misused by State authorities to disrupt or distort online freedoms of association, peaceful assembly, expression and the right to political participation. Firstly, online channels can be readily monitored by AI systems to amplify or intensify unjustified mass surveillance. AI enables the authorities to identify in more granular detail information concerning individual and group communications and interactions online. Monitoring of digital communications, social media content and online behaviours allows for the comprehensive profiling of citizens' political opinions, associations, and ideological leanings. Automated sentiment analysis and keyword detection systems enable real-time identification of dissenting voices, the organisers of protests and opposition movements, and critical commentary of government practices, policies and leaders.

AI-driven mass surveillance – including under counter-terrorism powers – could be targeted by the authorities to identify individuals, groups, movements or patterns of expression and activity that is critical of the authorities or considered politically subversive, thus enabling the suppression of human rights defenders, activists, dissidents, minorities and journalists. This includes tracking their financial flows, areas of operation, and transnational associations, in order to restrict and isolate their activities. Predictive algorithms may also pre-emptively flag individuals as security threats based on their communications and interactions, leading to targeted harassment or prosecution. AI systems can thus enable unjustified interferences in not only the right to privacy but also freedoms of expression, association, peaceful assembly and religion, the right to participate in public affairs, and minority and cultural rights.[61]

Such pervasive monitoring capabilities can create self-censorship effects, as individuals modify their patterns of expression while anticipating government scrutiny, including journalists, human rights defenders and civil society actors, among others. These technologies fundamentally undermine the anonymity and spontaneity essential for robust democratic debate, transforming the public sphere into a surveilled environment where authentic political expression becomes increasingly constrained and citizens retreat into silence. The negative impacts on civil and political rights can have cascading effects on other human rights that their exercise protects and promotes. Examples include where human rights defenders and other civil society actors advocate, organise and demonstrate for labour and trade union rights, economic equality and poverty relief, health and education, environmental protection and climate justice, and Indigenous and minority rights, among others.

Secondly, as discussed in the next section, AI-powered tools could be used by the authorities, often in concert with technology companies, to censor, block or "take down" legitimate online expression, as well to identify individual accounts for restriction, suspension or disabling.[62] Thirdly, the authorities could deploy AI-generated misinformation, disinformation, "deepfakes" and deceptive counter-narratives to undermine and neutralise legitimate expression, association, peaceful assembly and civic and political participation, and to confuse and manipulate civic debate.

In addition to State misuse of AI in online environments, the design and operation of algorithms by technology companies can also interfere in civil and political freedoms online. Many of the digital communication channels upon which civil society and advocates depend lack transparency. Algorithmic determinations governing the reach and audience of messages or communications can impede effective engagement, networking and dissemination of information by individuals and civil society actors. The construction of user profiles predicated upon subjective interpretations of associations between individuals or groups involves significant risks.[63] Reductive models that attempt to infer relational dynamics among

association members and between different organisations may further fail to authentically represent the actual characteristics and complexities of those connections, potentially mischaracterizing associational networks and their social dynamics.[64] Critically, in engineering social interactions between individuals and groups, they may also further replicate or engender entirely new biases, discriminatory effects or forms of exclusion.

AI-generated terrorist content online

The capacity of digital forums to generate civic engagement has not been only constructive. Certain online platforms have been appropriated by terrorist and violent extremist actors to disseminate incendiary rhetoric and incitements to violence against perceived adversaries, including political opponents, migrants, and minorities,[65] thus enabling abuses of the rights of others.[66] The Secretary-General has warned of the online mobilization of "incels" with converging extremist ideologies including racism, misogyny, anti-feminism and homophobia, normalising violence against women,[67] which has potential to escalate into terrorist violence.

The anonymity and decentralized architecture of online forums can create environments conducive to amplifying extremist messaging, while weakening mechanisms for transparency, accountability and human rights-consistent content moderation.[68] So-called "filter bubbles" have always existed, reflecting confirmation bias and people's predisposition to often ignore competing viewpoints. Online platforms amplify the effects of disassociation and intensify disaffection, estrangement and hostility toward others. Closed discussion fora may insulate members of a group from diverse opinions and fact-checking, reinforcing the impact of disinformation in polarising opinion.[69]

The use of generative AI (GenAI) presents multifaceted challenges for counter-terrorism, since it transforms how terrorist and violent extremist ideologies can propagate and how individuals may radicalise.[70] GenAI can enable malicious actors to automate content production, creating sophisticated propaganda materials that adapt messaging strategies to target specific human vulnerabilities and cultural sensitivities across populations.[71] These capabilities can undermine the foundations necessary for informed public discourse, democratic resilience[72] and respect for human rights. GenAI can also facilitate hyper-personalized radicalisation pathways through algorithmic analysis of individual behavioural patterns, social media interactions, and demographic characteristics discernible from online interactions.[73]

In many countries online gaming now plays as great a role in connecting people, especially young people, as traditional social media. Initially designed solely for gameplay, gaming platforms now also serve as hubs for broader social interactions and creating communities with shared interests or beliefs. Platforms with large associative and interactive components include content communities such as YouTube, certain collaboration projects and the longer established gaming platforms such as Twitch, Fortnite and Minecraft. By 2025, the most popular online games had a combined following of more than one billion users worldwide.[74] Discussions on these fora are also frequently mirrored across other gaming-based platforms, such as Discord.[75] Some online players reportedly use gaming platforms for illicit activities linked to organised crime including people trafficking, the sale of materials potentially connected to nuclear proliferation, arms dealing and terrorism.[76] Other online services are also used for unlawful activities.[77] There are emerging concerns about the risks of the dissemination of terrorist content through such platforms, including through GenAI.

Responding to AI-generated content online

The nature, scale and pace of traditional counter-terrorism strategies may prove less effective against dynamically-generated online content that evolves in real-time to circumvent detection algorithms and exploit social tensions.[78] Responses to the proliferation of harmful material are increasingly drawing on AI to anticipate emerging narrative threats, moderate online content, and construct and deploy customised counter-narratives.

While automated content moderation systems have significant capabilities to detect explicit terrorist propaganda online, they are ill-suited to detecting and moderating implicit terrorist or violent extremist content online, namely content that "conveys harmful messages through coded language, irony, humour, or cultural references".[79] As the Special Rapporteur on Freedom of Expression observes, such systems tends to over-moderate political expression in sensitive contexts, including conflict situations, or fail to remove harmful content;[80] in some instances, algorithms have even amplified harmful content to maximise user engagement.[81] The problems of false positives and false negatives are aggravated by moderation in multilingual, under-resourced contexts where training data is sparse and moderation systems are underdeveloped.[82] There is also a tendency to automatically remove content, instead of taking a proportionate approach, such as by "shadow-banning" instead of removing. Even where online platforms are regulated by State authorities, vague definitions and criteria may not give platforms sufficient guidance and can allow private actors too much discretion.[83] While human moderation is essential for interpreting context, intent, and nuance, staff and resources devoted to moderation have been scaled down by major online platforms since 2022, including in pursuit of maximalist "free speech", and investment in AI moderation has expanded, along with reliance on unreliable "community" or crowd-sourced approaches.[84]

States must regulate online content moderation in accordance with the rights to freedom of expression, association, peaceful assembly, participation in public affairs, and non-discrimination. They must ensure that expression is restricted, whether by States authorities or private actors, only in accordance with the requirements of legality (thus requiring precision in the legal criteria for moderation), necessity, proportionality and legitimate aim, consistent with articles 19 and 20(1) of the International Covenant on Civil and Political Rights (ICCPR),[85] the Rabat Plan of Action against hate speech, and the recommendations on incitement to terrorism of the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism. The latter requires that there must be an intention to incite terrorism, as properly defined according to best practice international standards, as well as a reasonable probability or real risk that a terrorist act will be committed.[86]

Further, States must ensure that technology companies undertake and publicly disclose human rights due diligence and impact assessments, with heightened obligations in conflict and high-risk situations. Companies should also provide adequate, well-trained, unbiased human content moderation and human oversight of AI content moderation; ensure algorithmic transparency; monitor the effectiveness of mitigating actions; publish their moderation policies and detailed moderation reports; maintain adequate and well-trained in-house human rights staff and multilingual and culturally diverse expertise; and provide effective internal appeal mechanisms.[87] Investors should conduct due diligence on their investments and use their leverage to protect human rights.

Excessive content moderation risks weakening social resilience, fuelling social division and alienation from democratic political processes, and rendering populations more vulnerable to terrorist narratives.[88] Interventions must focus not merely on suppressing harmful content but on promoting pluralism of participation and view-points and ensuring public interest information remains available and accessible. As the Global Digital Compact reaffirms, States must "refrain from imposing restrictions on the free flow of information and ideas that are inconsistent with obligations under international law",[89] including the right to freedom of expression and opinion. AI should be harnessed to bolster democracy by improving access to public interest information, amplifying marginalized voices, and enabling civic participation.[90]

7. AI in border management: Preserving rights at the border

A number of border authorities are now integrating AI systems with the aim of enhancing the accuracy, efficiency and timeliness of decision-making. Digital technologies can assist individuals to conveniently and securely cross land, air and sea borders whilst assisting border, customs and law enforcement officials with the prevention of crimes, including identity fraud, terrorist travel, and trafficking of humans and illicit goods.[91] The use of new technologies for secure identity verification and notionally unobtrusive vehicle, luggage and cargo checks may help to eliminate delays to travellers and cargo while assisting to secure borders.

However, AI-powered technologies, including biometrics, machine learning algorithms, traveller information databases, entry and cargo screening equipment, and drones and other surveillance capabilities, can jeopardize the rights to privacy, non-refoulement and the right to asylum, the freedom of non-nationals from arbitrary expulsion or denial of entry, the right to family unity, freedom of movement, non-discrimination, and freedoms of peaceful assembly, association and expression. The aggregation of personal data across multiple government systems also creates data security vulnerabilities with respect to cyberattacks and unauthorized access. Effective governance must safeguard human rights by ensuring transparency in algorithmic decision-making, establishing robust oversight mechanisms, and implementing protections against discrimination, non-refoulement and other rights affected at borders.

Biometrics

Biometric authentication systems are a foundational application of AI in border management, employing facial recognition, iris scanning, and fingerprint analysis to verify traveller identities against watchlists, criminal information and intelligence databases,[92] as well as facilitating cross-border sharing of traveller information. However, the use of biometrics identification systems in border management can risk perpetuating inequality through biases and discriminatory profiling resulting from prejudices transcribed by historical data collection, processing and retention practices. This can lead to violations of other human rights, including freedom of movement (including the right of leave and re-enter one's State of nationality), the freedom of non-nationals from arbitrary expulsion, non-refoulement and the right to asylum, and non-discrimination by enabling discriminatory profiling based on ethnicity, religion, or other protected characteristics. Without adequate safeguards, these AI-powered systems can unjustifiably restrict mobility and disproportionately target or affect minority communities.[93] There are further concerns about the potentially unnecessary and disproportionate blanket over-collection of data about all travellers; inaccurate data; excessive data retention periods; unauthorized access to data; the legal "black box" of cross-border data sharing without adequate transparency and human rights safeguards; and the collection and sharing of data in armed conflicts involving counter-terrorism.[94] The Security Council's requirement on States to

collect and share certain traveller information, and the associated United Nations Countering Terrorist Travel Programme, have attracted particular human rights criticisms.[95]

Detecting fraud and deception

Machine learning algorithms are able to refine their accuracy through continuous exposure to diverse biometric datasets. Advanced systems incorporate behavioural biometrics, analysing gait patterns and micro-expressions to detect suspected deceptive behaviour or psychological stress indicators associated with malicious intent.[96] The accuracy of behavioural biometrics is significantly affected, however, by the inherent variability of human behaviour. Moreover, temporal inconsistencies, environmental factors, physiological states, and individuals' own intrinsic patterns of learning can all destabilise the measurements upon which biometrics are founded.[97] The resulting false positives can not only impinge on human rights but distract authorities from detecting and investigating genuine crimes and security threats.

Further, detecting narrative inconsistencies presents distinct challenges. AI systems must contextualize expressive statements within cultural frameworks, account for linguistic variations, and distinguish genuine memory lapses from deliberate deception.[98] The absence of standardized ground truth data for deceptive behaviours hampers model training. Ground truth data provides the accurately labelled, verified information needed to train supervised machine learning models, validate their performance and test their ability to generalize or make accurate predictions based on new data.[99]

AI has also played a wider transformative role in fraud detection, which is relevant to countering identity and document fraud at borders and in other migration processes. Conventional fraud detection has predominantly utilised rule-based frameworks combined with human oversight, which has often proved inadequate against the escalating complexity and scale of deceptive practices.[100] AI technologies, especially machine learning algorithms, have established innovative pathways to effectively identify and prevent fraudulent transactions in real-time. Machine learning's exceptional efficacy in fraud detection stems in large part from its ability to process vast datasets and recognize novel patterns.[101] In the customs and border management contexts, natural language processing applications may also be used to facilitate multilingual document verification. AI systems are currently under development to automatically detect identity and document fraud and inconsistencies in traveller narratives during interviews, but the reliability of such technologies remain to be proven, including the risks of false positives. The accuracy of the above systems is further challenged by sophisticated counterfeit documents that increasingly incorporate genuine security features, thus requiring AI models to distinguish subtle anomalies. Fraud techniques are also dynamic and evolving, meaning that continuous system adaptation is necessary, potentially compromising their reliability and effectiveness.

Predictive analytics and natural language processing

Predictive analytics powered by AI enable risk-based screening protocols to more efficiently allocate scarce border inspection resources. These systems aggregate data from multiple sources including, for example, travel histories, financial transactions, social media activity, and intelligence reports, to generate threat probability scores for individual travellers.[102] This approach is intended to allow border agencies to expedite low-risk crossings while intensifying scrutiny of perceived high-risk individuals, theoretically optimizing travellers' transit experiences, security outcomes and operational throughput.[103] Again, the reliability and human rights-consistency of these systems is only as good as its underlying data, and the AI-powered

aggregation and processing of large datasets from multiple sources has the potential to compound the underlying data inaccuracies and intensify discrimination and violations of privacy and other affected rights.

Automated entry systems

Automated entry systems are expected to play a significant role in improving border crossing experiences for travellers and border officials. Automated Border Control (ABC) gates represent the convergence of multiple AI technologies, enabling self-service processing that can reduce wait times and operational costs while maintaining security.[104] These systems integrate document readers, biometric sensors, and decision-support algorithms to authenticate travellers without direct human intervention, potentially freeing officers to focus on complex entry cases requiring more nuanced judgment. However, the deployment of AI in border management raises significant concerns regarding privacy erosion, algorithmic discrimination, and the potential for mission creep in surveillance capabilities. Research has documented disparities in facial recognition accuracy across demographic groups, potentially subjecting certain populations to unnecessary and disproportionate scrutiny.[105]

8. AI in judicial proceedings: Fair trial and judicial independence

Judicial proceedings increasingly use technology-based solutions,[106] including AI. These technologies can potentially enhance access to justice, remedies and accountability, increase equality before the courts and the fairness of trials, and strengthen judicial independence and impartiality.[107] For example, AI could assist victims of terrorism and counter-terrorism by providing them with comprehensible legal information, assisting them to file complaints and documents and represent themselves in proceedings, and aiding in translation and interpretation into their own language. It can potentially increase efficiencies in the administration of justice and judicial workloads by expediting case management and reducing delays, identifying and applying precedents, semi-automating basic decision-making, aiding legal drafting, enhancing judges' capacity to process evidence, correcting human biases concerning gender, race, religion or nationality, enhancing the capacity of legal aid lawyers, and improving the systemic monitoring and analysis of legal trends.[108] Some technologies that use AI include case management software, apps to facilitate engagement in justice processes, information retrieval and analysis technologies, data collection and processing, automated transcription, and the use of online communication and messaging for court hearings and disclosure of evidence.[109]

AI and automated systems are already being utilised in the criminal investigation and prosecution of terrorist offences.[110] More broadly, as discussed in this paper, these capabilities extend to risk assessment, biometric identification, the detection of fraud, automated analysis of subject matter of communications, and tools that analyse and evaluate content to determine its authenticity (including the detection of "deepfakes" in video and synthetic audio, for example). Some of these capabilities may lead to the production of evidence that is admitted in court to prove an accused person's guilt.

Despite the potential human rights benefits, the use of AI technologies also poses risks to the right to a fair trial and judicial independence, protected under article 14 of the ICCPR and customary international law, as well as rights that depend on these, including liberty and even the right to life where the death penalty is applied. The 2025 report of the Special Rapporteur on the Independence of Judges and Lawyers on AI in judicial systems warns against "techno-

solutionism" and exaggerated hype about what AI can accomplish to propel the adoption of AI systems carrying significant human rights risks.[111] Such tools may be affected by the usual problems of data errors, incompleteness or unrepresentativeness, and bias (resulting in discrimination including against minorities that are over-represented in the criminal justice system). The Secretary-General has cautioned that biased AI systems could result in discrimination and that inequality of arms may result from defendants being unaware of how AI systems affect their treatment,[112] and expert technical challenges to the reliability and use of AI may be expensive and beyond the means of many criminal defendants.

Further, automation could render judicial decision-making non-individualised, opaque and unchallengeable, and prompt "automation bias" or deference to technological solutions, thus violating fair trial,[113] eroding the individual nature of criminal responsibility, and undermining judicial discretion, capacities and independence. Even the judge may be unable to explain automated decision-making processes,[114] including due to the lack of transparency and "black box" or proprietary algorithms preventing judicial verification and effective challenges. The sheer volume of source data, such as thousands of prior judicial decisions, could also overwhelm any meaningful oversight by a "judge in the loop". AI-driven penalties could affect proportionality and judicial discretion in sentencing, while automation could reduce the contextual human empathy and judgment necessary to ensure penalties are proportionate in the light of the individual circumstances of the offender. AI tools can generate false positives and misleading correlations, and lack context and nuance, when applied to complex human behaviours. Judicial independence could be further compromised by political and corporate pressures on courts to use AI solutions.[115] The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law accordingly requires States parties to adopt measures to ensure AI systems are not used to undermine respect for judicial independence and access to justice (article 5). The principles of transparency, preserving contextual judicial judgment, due diligence, and accountability are especially important to upholding fair trial rights as justice processes are increasingly digitised.

Where AI is used in generating or evaluating evidence, the opacity of algorithms and data sources, and the risk of courts perceiving such "scientific" evidence as reliable, could undermine an accused person's right to prepare a defence, to effectively challenge the evidence against them (which presupposes adequate disclosure of the basis of the evidence), equality of arms, and the ability of their lawyers to assist and represent them. It could also limit the ability of juries and judges to effectively discharge their adjudicative functions in the trial process and to guarantee the overall fairness of a trial. Consequently, care is needed to ensure chain-of-custody of AI-enabled evidence processes, empirical validation, and expert explanation. AI evidence must be rigorously tested, non-discriminatory, subject to ordinary admissibility requirements (including to exclude evidence illegally obtained through AI systems), capable of being effectively challenged, and subject to judicial control. [116]

There are equally human rights concerns relating to counter-terrorism law enforcement adjacent to criminal justice processes. Predictive policing may undermine the presumption of innocence,[117] since it predicts the value of an outcome variable from existing data sources rather than testing causal explanations or describing specific data points[118] in relation to a person's actual behaviour. The Special Rapporteur on the Independence of Judges and Lawyers recommends that predictive analytics should never be used as a basis for prosecution or detention.[119] AI may also be used in algorithmic risk assessment tools, including to predict recidivism (whether at the stages of bail, sentencing, parole or post-release), or to identify and evaluate individuals who are believed to be vulnerable to terrorism, violent extremism or

radicalisation. Yet, all risk assessment and "behavioural insights" tools must be empirically validated and not given automatic or undue weight as probative of a person's risk or the need to adopt a particular counter-terrorism measure in response, and all relevant factors, including mitigating circumstances and mental health, must be holistically considered.[120]

9. AI in detention: liberty and humane and dignified detention conditions

Any deprivation of liberty renders individuals particularly vulnerable, as they depend on their captors for basic needs and are readily exposed to human rights violations. To date, the principal drivers to integrate digital technologies in places of detention include efforts to address the needs and welfare of detainees and staff. In some jurisdictions, detention facilities have introduced security systems which use AI and facial recognition, in response to increased and lethal violence,[121] including between detainees and involving the authorities. Less invasive monitoring includes mobile phone blockers, body scanning and biometrics. Monitoring technologies, including AI, can assist States to fulfil their duty to protect the rights to life (ICCPR, article 6), physical and mental security of person (ICCPR, article 9), freedom from torture and other ill-treatment (ICCPR, article 7 and Convention against Torture), humane and dignified conditions of detention (ICCPR, article 10), and health (International Covenant on Economic, Social and Cultural Rights, article 12).

Such technologies also bring risks for human rights. Firstly, if the authorities deploy technology as a substitute for direct direct human contact with detainees, they will also give up critical insights required for taking well-informed and timely decisions about detainee welfare,[122] whether in armed conflict, administrative detention in public emergences, or in the criminal justice system. States must therefore ensure that any AI-enabled decision support systems in detention do not adversely affect the treatment of detainees and the conditions of detention. The authorities must retain direct contact with detainees, which is essential to build trust, foster situational awareness, maintain order without force, identify problems early, and ensure that detention conditions comply with international law.[123]

Secondly, if AI-systems are used to detect and interpret subjective characteristics such as gait or body language, "sentiment", expression, habits or patterns of association with other detainees, there is a risk of misidentifying innocent conduct, including disabilities, as threatening or aberrant behaviour. Some research suggests that there is weak association of emotions with facial expressions, which vary across cultures and contexts.[124] These distortions could in turn lead to unjustified disciplinary or punitive measures, in violation of individuals' rights, as well as counter-productively generating grievances amongst the detainee population.

Thirdly, examples of the harms of digital technology used in detention continue to be reported. For example, the Center for Prisoners Rights, a Japanese NGO, reported in 2022 that people on death row may be held in solitary confinement and monitored constantly by surveillance cameras with no measures in place to ensure the right to privacy, including as regards using toilet facilities or changing clothes.[125] AI offers the possibility of perfecting the panopticon, particularly where data is retained for protracted periods or used for other purposes.

Another risk to human rights comes from any use of AI to support decisions on who should be detained in the first place.[126] The risks are heightened in situations of potential mass detention, such as public emergencies and armed conflict. As the ICRC observes, while technology deployed responsibly and with robust human oversight can contribute to compliance with international law, it also "carries risks including bias, lack of transparency, and faulty programming and analysis, all of which can undermine humane treatment and compliance".[127]

There may be considerable nuance and subjectivity in judgments about whether a person is, for example, participating in hostilities in armed conflict or otherwise presenting a security risk, whether a person's associations with a terrorist group in an emergency are suspicious. The right to liberty, including freedom from arbitrary detention (ICCPR, article 9), must be fully safeguarded against any AI technologies that cannot meet its requirements, including legality, individual decision-making, necessity, proportionality and non-discrimination.

10. Military uses of AI: Retaining human control

The application of AI in military contexts has been extensively discussed in recent years,[128] including in lethal autonomous weapons systems (LAWS) and AI-enabled decision support systems (AI DSS), including target identification and "deep sensing" to gather and analyse various data streams simultaneously to create real-time, evolving threat assessments.[129] AI has reportedly been used in recent conflicts in military decision making and targeting processes, as in Gaza, Ukraine, Iraq, Syria and Yemen, although the specifics of particular technologies and the circumstances and parameters of their use often remain opaque.

The design, development, deployment, and operation of military technologies that incorporate AI pose extensive risks for international humanitarian law (IHL) compliance and human rights protection. Compliance with international law is of critical importance as militaries develop and use AI and related new technologies in counter-terrorism operations during armed conflicts. The ICRC has already issued influential guidance to promote adherence to IHL.[130] States must consider how these legal frameworks specifically apply to their AI capabilities and how existing laws can be effectively implemented to ensure compliance, oversight, governance and accountability and remedies.

The deployment of AI-enabled systems has raised particular challenges for the attribution of responsibility and accountability in relation to violations of IHL and human rights law. Consequently, inadequate transparency resulting from using AI in military targeting risks creating accountability gaps, by making it harder to establish both individual criminal responsibility for war crimes or other international crimes and State responsibility for breaches of IHL and the right to life. These concerns are compounded by the necessity of examining human agency, responsibility, and accountability as discrete yet interrelated concepts, each requiring independent analytical consideration due to their distinct conceptual foundations and contributory roles in the broader accountability framework.[131] While acknowledging the expertise of militaries, civilian oversight of military AI systems is also imperative.[132]

Lethal autonomous weapons systems (LAWS)

The implications for human agency and accountability vary between LAWS and AI DSS. Thus far, much of the discussion on the military use of AI has been on LAWS. The principal concerns have been the issue of maintaining human control over weapons, the critical functions of weapons, attacks, the targeting process, and the final decision whether to authorise the use of force.[133] The operational advantages potentially conferred by LAWS could incentivise the proliferation of AI systems within such weapons across military forces, although not all LAWS involve AI (such as landmines or the Phalanx ship defence system).

The unconstrained development and use of LAWS pose significant risks to civilians and other persons in armed conflict and increase the prospect of conflict escalation and retaliatory non-

compliance with IHL. Additionally, there remain definitional ambiguities regarding what constitutes an autonomous weapon system, particularly given the variable degrees of human intervention and oversight involved in their deployment. In theory it might be possible for LAWS to achieve substantial operational independence in military targeting. However, in order to maintain transparent chains of responsibility to respect IHL, AI systems must not replace human decision-makers and must be limited to supporting them.[134] Establishing accountability mechanisms also necessitates a clear delineation of responsibilities between the different actors engaged: technology developers, operational personnel, and military commanders.

AI-enabled decision support systems (AI DSS)

AI DSS are computerised tools that use AI to aggregate various data sources to produce analyses, recommendations or predictions to support human decision-making in armed conflict. The opacity inherent in contemporary AI systems with respect to decision-making processes, wherein systems function essentially as algorithmic "black boxes", presents significant challenges to accountability frameworks given the lack of explainability,[135] when processes remain either opaque or inaccessible to human comprehension.[136] In order to maintain transparent chains of responsibility and thus compliance with IHL and human rights law, AI DSS must remain understandable to humans and preserve sufficient time and space for human deliberation, assessment and verification.[137]

The implementation of novel AI-systems potentially intensifies the responsibility upon human operators regarding the application of appropriate precautionary measures in military targeting. Satisfying such obligations may necessitate specialised training protocols and comprehensive technical understanding of the mechanisms by which the AI-system functions operationally. Nevertheless, some observers contend that the requisite precautionary scope remains consistent with existing conventional weapons systems and established intelligence sources, suggesting that no fundamental departure from existing operational standards is required.

Review and regulation of weapons

IHL requires States to review the IHL-consistency of any AI-enabled system that forms part of a weapon system or means of warfare, or that influences how such systems or means are used or expected to be used.[138] This includes not only States who manufacture such systems but also who purchase them.[139] Reviews should consider, for instance, whether the weapons can be used in a manner consistent with IHL's principles of distinction, proportionality and precautions, and the prohibition on indiscriminate attacks. Critical considerations concerning AI-based systems might include the periodicity of review (a contentious issue[140]) and the threshold of technological modification necessitating renewed legal evaluation. Consensus regarding standardized review methodologies regrettably remains absent, while State practices lack transparency and common disclosure protocols among international actors.[141]

Beyond IHL, the arms sector is regulated by many domestic, regional, and international legal instruments, including weapons control and elimination conventions and export control regimes that position States as the final gatekeepers of the arms trade. AI-powered LAWS that inherently cannot satisfy the principles of distinction and proportionality and the prohibition on indiscriminate attack may not satisfy the 2013 Arms Trade Treaty's prohibitions on transferring arms or munitions that either "would be used" in the commission of international crimes or grave breaches of the Geneva Conventions (article 6), or there is an overriding risk that they "could be used" in serious violations of IHL or human rights law (article 7).

11. AI research, development and commercialisation: The role of business

States are the primary users of AI technologies in counter-terrorism and public security[142] but they also promote the development of AI by financing private research and development and establishing public-private partnerships. Private entities, including defence contractors, tech companies and "start-ups", act as innovators and developers of AI technologies and deliver services such as training, deployment, and maintenance of devices and systems.[143] Investors are also driving AI development, with US$4.7 trillion of capital expenditure projected from 2026 to 2030, including on data centres, computer chips, power grids, and networking equipment. There is, however, volatility around the profitability of AI and prospects of an AI "bubble", and powerful open-source models continue to compete with proprietary systems.

States must regulate existing and emerging AI technologies, including private actors, by establishing legal frameworks, standards and accountability and remedy procedures. The 2011 United Nations Guiding Principles on Business and Human Rights address the responsibilities of businesses, including investors, to respect human rights wherever they operate and whatever their size or industry. This responsibility means companies must assess their actual or potential impacts, prevent and mitigate abuses, and address adverse impacts.[144] Institutional and other investors also bear responsibilities to ensure that their investments in companies developing or using AI technologies and supporting infrastructures are not enabling human rights violations.

The Guiding Principles can complement other instruments regulating arms to help States and business to prevent, mitigate, and remedy negative human rights impacts in armed conflict.[145] They can help to fill gaps in the formal regulation of arms, such as the Arms Trade Treaty being limited to eight categories of conventional weapons and not addressing dual-use goods[146] and weapons-enabling technologies such as AI. Some arms companies have failed to conduct adequate human rights due diligence, for instance by focusing on the risks of forced labour and other human rights concerns in their supply chains and workplaces but failing to identify the profound risks of the use of their products and services, including in conflicts.[147]

Given the specialised technical knowledge of technology companies involved in developing or deploying AI, including in military and security applications, they bear a heightened responsibility to identify the human rights risks of their products and services. This includes by conducting thorough human rights due diligence assessments of AI systems throughout their lifecycle, including design, sale and transfer, implementation, training, deployment, and post-use; mitigating risks where feasible (including through embedding safeguards – "technology by design") and terminating projects where risks cannot be adequately mitigated. Companies must further ensure the transparency and explainability of AI systems, including through appropriate disclosure to users, regulators, investors, affected persons and the public. They should also meaningfully engage with affected stakeholders, including civil society and national human rights institutions, on human rights. Finally, companies should establish, maintain and adequately resource accessible and effective complaints mechanisms for receiving, assessing and remedying violations of human rights resulting from their AI systems or investments in such systems.

A valuable proactive measure to reduce bias in training datasets is to ensure diversity of personnel within the engineering teams involved in its design, and to provide diversity training for management overseeing research and development.[148] These steps would reduce the likelihood of biased datasets and programming that would otherwise exacerbate discrimination.[149] More broadly, those most affected by potential oversights in AI systems' design, for example, minorities, migrants, women, LGBTIQ+ persons and persons with

disabilities, should be actively included in corporate research and development as well as in national and multilateral governance discussions. Inclusive stakeholder consultations must be regularly held, and mechanisms should be put in place to report potential biases.[150]

12. Future trends and risks in AI

The wide use of AI is still relatively new. While having been conceived of over half a century ago, it is only in the past few years that advances in computing have allowed for its effective operational use. AI has the capacity to expedite its own advancement and thus further transform security capabilities. At the same time, it is difficult to accurately predict how scientific and technological progress will evolve, and thus to forecast the future of AI in countering terrorism. The outcome of this potentially contested process, and to whom benefits are accrued and risks imposed, remains to be seen. While many States and the United Nations have been alert to the risks of terrorist misuse of new technologies,[151] including AI, there has been a greater reluctance to confront States misuse of AI while countering terrorism.[152]

In parallel, complementary developments in a multitude of aligned scientific spheres, such as quantum technology and neuroscience, will enable new possibilities, presenting benefits as well as risks for human rights. Advances in quantum technologies are particularly likely to influence the future of AI in counter-terrorism.[153] Quantum technology translates the principles of quantum physics into technological applications. Advances in quantum computing may enable a greater number of actors to decrypt classified or sensitive communications, which could benefit intelligence activities and facilitate many data-driven operational objectives.[154] The use of AI in conjunction with quantum computing might result in significantly more advanced machine learning methods, which could, for example, improve image recognition and enhance the development of virtual environments for counter-terrorism simulation and training.[155] To date, no major international organisations have undertaken to develop any formal initiatives devoted to monitoring or regulating the military, law enforcement or other applications of quantum technology.[156]

A further obstacle is the opacity of determining how dual-use items, including goods, computer hardware (such as advanced AI microchips) and software, may be exploited in the context of counter-terrorism. Dual-use technologies derive from civilian or defence industries and have both military and commercial end uses.[157] They are not covered by frameworks such as the Arms Trade Treaty or any other comprehensive regulatory framework. The human rights risks are magnified where States and business prioritise commercial advantages and obtaining a capability "edge" over competitors and wish to avoid regulation.

1. Rights to privacy and personal data protection

As the above discussion illustrates, a key privacy concern with the AI-enabled proliferation of data collection and analysis is the unprecedented ability to deliver novel insights into aspects of previously imperceptible human behaviour. This analysis which may be used to determine whether individual conduct conforms to certain standards, norms or patterns of personal behaviour, may exert a deep chilling effect: monitoring of this nature may coerce or influence broader changes in habits and practices that would constitute social engineering.[158] The right to privacy can be violated irrespective of whether the information derived from the processing of data by an AI system is considered sensitive, or the person concerned is aware of the monitoring or has been caused any inconvenience. AI regulation, privacy and data protection require a complementary framework that embeds the principles of legality, necessity, proportionality, non-discrimination, transparency and explainability, and human oversight.[159]

The right to privacy prohibits the State from indiscriminately monitoring the population en masse or groups within it, and requires the State to prohibit discrimination by private actors. As such, in relation to individual or bulk surveillance, the law must precisely define the basis on which individuals or groups are identified for monitoring by a public authority authorised by law to do so; and any interference must be necessary and proportionate in pursuit of a legitimate security aim, non-discriminatory, and subject to due process, independent oversight, judicial safeguards, and effective remedies.

Where AI facilitates bulk electronic surveillance, domestic law must set out with sufficient clarity, precision and detail the grounds upon which bulk interception might be authorised.[160] The authorization (such as a warrant) must identify the types or categories of "selectors" to be used, each selector must be justified as necessary and proportionate, and detailed records must be kept at every stage. The importance of supervision and review of bulk surveillance is amplified because of the inherent risk of abuse and the imperative of secrecy. The need for safeguards is at its highest at the end of the process, where information about a particular person is analysed and the content of communications is examined. The process must be subject to "end-to-end safeguards", meaning that an assessment must be made at each stage of the process of the necessity and proportionality of the measures being taken: bulk interception must be subject to independent authorisation at the outset; and its operation must be subject to continuous supervision and independent ex post facto review.

Further, legal frameworks must govern and ensure that data collection, access, processing, sharing, transfer, and storage practices are safe, secure and necessary and proportionate for legitimate counter-terrorism purposes, in accordance with international law.[161] Public authorities and intelligence agencies must implement data protection principles, including data minimization, purpose limitation requirements, and secure erasure procedures, while ensuring compliance with data protection, privacy legislation and adherence to human rights standards that safeguard other interdependent rights. Individuals should be notified when their personal data will become part of a dataset used by an AI system.[162] As a general rule individuals and groups should also be empowered to consider, give and withdraw their informed consent to the use of their data and to choose how their data is used.[163]

The conduct of thorough and effective reviews of the legality of the use of AI systems with respect to data protection and privacy rights becomes even more important and challenging where the future impacts of new capabilities may be difficult to project. As data processing, storage and analytical capabilities evolve, potentially protracted or indefinite data retention practices, and breaches of purpose specification principles, could lead to potentially long-term privacy and data protection violations, that may affect individuals long after initial data collection is undertaken.

2. Equality and non-discrimination

AI necessarily relies upon the collection, processing and retention of personal and other data and has the potential to perpetuate and intensify biases in datasets or to deepen discrimination where AI tools are deliberately tasked to profile groups on prohibited identify grounds. States must therefore integrate safeguards to ensure its use is consistent with the rights to equality and non-discrimination. International treaty and customary law prohibit discrimination on prohibited grounds, such as on the basis of race, colour, sex, ethnicity, age, language, religion, political or other opinion, national or social origin, disability, property, birth or other status. Non-discrimination and equality should be addressed within each phase of development of an AI system, including in its design, development, training and use,[164] including through ex ante and continuing due diligence assessments by States and business.[165] States must refrain from discriminatory uses of AI but also regulate private actors to protect against discrimination.

Guidance from the UN Committee on the Elimination of Racial Discrimination is especially relevant in the AI research and development lifecycle. The Committee recommends that "before procuring or deploying such systems States should adopt appropriate legislative, administrative and other measures to determine the purpose of their use and to regulate as accurately as possible the parameters and guarantees that prevent breaches of human rights".[166] Further, "when the results of an assessment of a technology indicate a high risk of discrimination or other human rights violations, States should take measures to avoid the use of such a technology".[167] Moreover, States "should take all appropriate measures to ensure transparency in the use of algorithmic profiling systems. This includes public disclosure of the use of such systems and meaningful explanations of the ways in which the systems work, the data sets that are being used, and the measures in place to prevent or mitigate human rights harms".[168]

3. Transparency, explainability and effective remedies for rights violations

The deployment of AI in countering terrorism presents significant challenges to effective access to remedies for rights violations. The right to an accessible, effective remedy is a fundamental component of international human rights law, yet the technical opacity of many AI-driven tools creates significant barriers.[169] Frequently, AI systems, particularly those employing deep machine learning architectures or proprietary algorithms, have been presented as "black boxes" that are complex to interpret and even more difficult to explain. This renders the decision-making process fundamentally opaque.[170] Such opacity prevents individuals from mounting effective challenges to erroneous or discriminatory classifications, thereby undermining their procedural rights, including to identify and attribute responsibility for a violation and to pursue remedies.

Transparency deficits compound these difficulties. When States deploy AI systems without adequate disclosure regarding their operational parameters, data sources, or accuracy rates, affected individuals often lack the information necessary to identify potential rights violations or otherwise to demonstrate that certain automated determinations warrant judicial or other forms of scrutiny. As previously mentioned, AI systems may mirror the existing customs, mores, biases and forms of discrimination that permeate our societies.[171] A significant, persistent challenge is therefore how the legacy of inadequate scrutiny of gaps in representation reflected in an array of data collection, processing and retention practices can be addressed, to ensure such distortions do not further perpetuate harm and infringe human rights.

As such, due diligence frameworks must be embedded throughout the AI lifecycle—from the preliminary development phase to final system deployment, including "privacy-by-design" and ethical guardrails, as well as in continuous review of the performance of systems once operational. Further, there must be sufficient disclosure to affected persons, as well to relevant oversight and regulatory bodies and the public, of all relevant elements of the AI-system, including model specifications and algorithmic configurations, data sources, known or foreseeable limitations, as well as what may be unknown. Where algorithms are proprietary or classified, enforceable exceptions enabling adequate disclosure must be made. The absence of transparency insulates algorithmic decision-making from effective oversight, creating accountability vacuums wherein obligations to remedy violations may be weakened, especially where responsibilities are diffused across multiple institutions or public authorities.[172]

Paradoxically, predictability and understandability are widely held to be vital qualities of AI systems: such systems should do what they are expected to do, and they must do so for intelligible reasons. The principle of explainability is particularly critical in enabling remedies and accountability for violations involving AI-systems. When algorithmic systems generate risk assessments, threat predictions, or targeting recommendations that may result in adverse actions against individuals, those affected must be able to understand the basis for such determinations to effectively exercise their right to a remedy.

As a result, better understanding of these black box models has become of paramount importance as AI systems continue to flourish in diverse real-world applications, not just in the military domain but across almost every facet of modern life.[173] Advancements in AI have made it increasingly pervasive in optimising performance for more nuanced tasks and in complex operational environments. This trend is observable in both the law enforcement and military domains, where automation has been identified as a potentially powerful force multiplier, extending human capabilities and improving both speed and accuracy.[174]

Automation is increasingly utilized in intelligence gathering, surveillance and reconnaissance (ISR) and military decision-making.[175]

In situations where emerging technologies may pose a high risk to human rights, existing oversight frameworks must be reviewed and new mechanisms to enhance transparency and accountability must be established. Efforts to elucidate the inner workings of technologies are becoming more innovative and yielding workable methods that can advance transparency and accountability. Lately, research to evolve and advance explainable AI (XAI) has grown considerably and proposed solutions to render AI more transparent, thereby potentially assisting its advance and adoption in critical high-risk domains[176] while improving the prospects for respecting human rights. Conversely, the intrinsic value of developing explainable AI (XAI) as a concept to address concerns over insufficient transparency and accountability has been questioned. XAI may in fact exacerbate the misuse of AI in critical scenarios where human-centred decision-making is vital to minimising the risk of harm,[177] and there is a risk that overconfidence in XAI will increase pressure for greater automation of decision-making, thus undermining human rights compliance.

Accountability is further complicated because established legal frameworks may struggle to attribute responsibility where AI systems are allowed to operate with varying degrees of autonomy from adequate human intervention.[178] Questions then inevitably arise as to whether accountability should be conferred upon those designing and creating algorithms or AI systems architecture, implementing agencies or authorities, or officials charged with supervision. Without clear accountability structures defined by law and procedure, and robust explainability requirements, the right to an effective remedy risks becoming illusory for those subjected to AI-mediated decisions or measures, potentially creating a new category of technologically-facilitated rights violations that effectively evade State obligations to provide legal redress.

4. Evolving governance frameworks

The deployment of AI in counter-terrorism presents extensive challenges for oversight and governance that intersect with the emergence in the past decade of national, regional and international frameworks addressing AI generally as well as new technologies at large. The OECD AI Principles (2019) emphasise human-centred values, transparency, and accountability.[179] Such principles may be contested in counter-terrorism contexts where confidentiality and national security imperatives may be perceived to conflict with transparency requirements. Further, where law enforcement and intelligence agencies deploy AI systems in areas such as predictive analytics and automated threat assessment systems, they may face particular difficulties in reconciling these principles with the clandestine nature of counter-terrorism operations.[180]

UNESCO's Recommendation on the Ethics of AI (2021) further elucidate the scope of responsibilities engaged in developing AI systems. It emphasises that AI systems must respect fundamental freedoms and proportionality, including when deployed in security and policing contexts.[181] Again, the challenge is to reconcile potential tensions between preventive security measures which could potentially conflict with human rights, particularly regarding AI capabilities connected to mass surveillance or the use of advanced behavioural biometrics for identification and profiling purposes.[182]

The landmark EU Artificial Intelligence Act (2024) specifies how AI can be used in the EU and aims to balance safeguarding core EU values, including fundamental rights, with allowing law enforcement and intelligence agencies to leverage the opportunities AI offers.[183] It mandates a rigorous compliance framework for conformity, incorporating mechanisms for human oversight.[184] It does, however, contain highly problematic exemptions for AI systems used exclusively for national security, defence or military purposes, and certain other exemptions or concessions in relation to law enforcement, border management and public security. Looking ahead, the EU AI Act's regulatory framework will likely substantially shape subsequent technological developments in AI in the region. It will necessitate enhanced collaboration between Europe's public sector entities, particularly law enforcement, and multidisciplinary stakeholders including AI technologists, system developers, ethicists, and data protection specialists, to guarantee regulatory compliance in emerging systems.[185]

Similarly, the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (2024) establishes legally binding obligations ensuring that all activities across the lifecycle of AI systems comply with human rights, democratic principles and the rule of law.[186] It also exempts AI in national defence (article 3(4)) and national security, in the latter case "with the understanding that such activities are conducted in a manner consistent with applicable international law, including international human rights law obligations, and with respect for … democratic institutions and processes" (article 3(2)).

The Global Digital Compact (2024) underscores international concern that AI systems must be fully human rights compliant, "safe, secure and trustworthy", and internationally governed through a balanced, inclusive and risk-based approach.[187] It further encourages transparency, accountability and robust human oversight. In resolution 79/325 of August 2025, the General Assembly followed up by establishing two key mechanisms for global AI governance: the Independent International Scientific Panel on AI and the Global Dialogue on AI Governance, with a strong emphasis on human rights, equality, and non-discrimination. The resolution calls for AI systems to be developed and deployed in ways that are safe, secure, trustworthy, and aligned with international human rights law. It highlights the need to prevent discriminatory impacts, ensure inclusive participation, and bridge the global digital divide, particularly by supporting the meaningful involvement of developing countries and marginalized communities. It also stresses that AI governance must be human-centric and rights-based.

Collectively, these international frameworks offer the promise of a pragmatic, sustained and consistent attempt to develop guardrails around AI-enabled systems, as long as excessive exceptions for national security, law enforcement and border management are avoided. Further, their eventual efficacy in safeguarding human rights will largely depend upon robust implementation and enforcement mechanisms. Given the reality that an ever-widening gap is likely to develop between existing legal oversight and the deployment of new technologies exploiting AI systems, it is imperative to address the emergent human rights concerns already identified and anticipate future trajectories and risks.

International cooperation and coordination

Advancing the use of AI systems in counter-terrorism operations necessitates coordinated multilateral action addressing the legal, technical, and operational dimensions of cross-border security cooperation. Fragmentation in regulatory and operational approaches currently undermines coherent responses to transnational threats while also potentially compromising human rights. Certain applications of AI in law enforcement activities may require both ex ante

and continuous lifecycle risk management so as to ensure potential human rights risks are identified, addressed, and mitigated,[188] and that remedies are accessible and effective.

Institutional frameworks should be further strengthened by establishing specialised bodies within existing structures dedicated to AI governance in security contexts. These institutions should work to harmonize technical standards, operational procedures, and human rights safeguards across jurisdictions, facilitating mutual support and cooperation in countering terrorism. Whilst initiatives aimed toward legal harmonization across jurisdictions and the global regions would appear beneficial, challenges will likely persist in efforts to achieve alignment.[189] Data localization laws are increasingly being passed to mandate that personal data collected from or about a country's citizens or residents must be stored and processed within the country's borders, limiting cross-border data transfers ostensibly to protect national security and human rights and assert data sovereignty.

Multilateral treaties addressing cross-border data transfers from AI-enabled surveillance and automated decision-making in counter-terrorism contexts could nonetheless be considered. Such instruments could assist in reconciling complex and divergent data protection and privacy regimes, constitutional provisions, and cultural perspectives on the acceptable limits of surveillance while maintaining operational effectiveness against increasingly sophisticated threat actors operating across jurisdictions. Any new arrangements should not merely maintain existing levels of human rights protection but strengthen safeguards, including transparency, explainability and accountability.

Certification frameworks

The development of human rights-compliant AI systems necessitates large volumes of representative, unbiased, and high-quality data. Within the domain of law enforcement, and particularly in counter-terrorism operations, this requirement presents considerable challenges owing to the highly sensitive and frequently classified nature of such data, thereby constraining its availability and use for research and innovation objectives.[190] Standards and procedures must therefore be developed to verify and authenticate classified data used in AI systems, while ensuring that adequate disclosure is afforded where necessary to protect human rights, such as in the contexts of the fair hearing right to know and challenge the evidence, whether in criminal trials or when seeking remedies for violations such as of the right to privacy.

International technical standards organizations can assist by prioritising the development of certification frameworks for AI systems employed in law enforcement operations. These standards should address data protection requirements such as data quality assurance and purpose specification, in addition to other requirements such as algorithmic transparency, bias mitigation and data security, while establishing mutual recognition of standards that enable secure information exchange between certified systems.

Capacity building, diversity and accessibility

Furthermore, investing in systematic capacity building initiatives can begin to address disparities in technical expertise. Early-stage stakeholder engagement across diverse demographics constitutes a particularly effective approach to mitigate risk within the AI systems development lifecycle. Creating more diverse and representative development teams, encompassing gender, age, ethnicity, disability status, and other demographic attributes, promotes multifaceted perspectives that are essential for detecting stereotypes and biases. It also better enables the formulation of targeted safeguards to mitigate potential risks and

enhance benefits for vulnerable populations. Critically, developers must ensure that systems are designed in accordance with universal accessibility principles, accommodating users irrespective of age, gender, ability, or other distinguishing characteristics.[191] Programs to develop competencies should prioritise AI literacy, strengthening institutional capacity, and technology transfer mechanisms to ensure equitable participation in global counter-terrorism efforts while maintaining rigorous human rights standards across all jurisdictions.[192]

Consultation, collaboration and information exchange

The international community including States, international organisations and civil society should work together to further strengthen human rights-based regulation of the future development and use of AI. Greater collaboration through engaging all relevant stakeholders can build on the existing normative frameworks to clarify and enhance the oversight of AI, especially to provide the necessary guidance in the areas of transparency, accountability, human rights due diligence and the more precise and nuanced diffusion of human rights norms throughout all phases of the AI development and deployment lifecycle. Frequent dialogue and discussions between States can also foster mutual understanding and develop greater trust. International organizations can facilitate these interactions and develop measures to build confidence and encourage exchanges that develop awareness and cultivate opportunities to collaborate in promoting and protecting human rights.

Sharing practices and knowledge developed from conducting legal and technical reviews of AI systems in counter-terrorism contexts, in addition to encouraging greater interaction between the scientific and technical communities and human rights advocates and practitioners, can further assist in more widely promoting the responsible use of emerging technologies in other domains beyond counter-terrorism. While counter-terrorism is often a laboratory for new security technologies, innovations in this domain can provide vital lessons to strengthen human rights protection against the risks of AI systems more generally.

A. Overarching principles

1. Respect for legality

The development and use of AI systems in counter-terrorism must be authorised by law. Legal authority to use AI systems in law enforcement, intelligence and military contexts must conform to the principle of legality, such that permitted uses must be precise and specific, reasonably certain and foreseeable, and avoid vague and overbroad powers that are prone to abuse. Blanket military, national security, law enforcement or border management exemptions must not be permitted;[193] any tailored exceptions must be strictly necessary and proportionate in pursuit of a legitimate aim and consistent with human rights law and must not deprive an individual of their right to an effective remedy. Given the risks to human rights, AI governance cannot be left to ordinary security, law enforcement or criminal justice rules, or to a laissez-faire free-market or industry self-regulation approach: effective State regulation is essential.

2. Respect for human rights

States must only deploy AI-systems that interfere in human rights where they are strictly necessary and proportionate in pursuit of a legitimate security aim. Their use must respect all relevant rights, including equality and non-discrimination, privacy and personal data protection, freedoms of association, peaceful assembly, expression and religion, information and media freedoms, participation in public affairs, liberty, and the right to effective remedies and reparation. Public and private sector developers and deployers of AI systems must conduct rigorous human rights due diligence assessments throughout the lifecycle of such systems.

3. Prohibitions on AI and moratoria

AI systems must be prohibited where they cannot be used in compliance with human rights law.[194] Examples include where:[195] the risk of an individual committing a terrorist offence is assessed solely based on profiling or personality traits, including based on biometrics or sentiment analysis, absent human assessments based on objective, verified facts linked to criminal activity; facial recognition databases are sourced from untargeted, non-consensual scraping of images from the internet or surveillance footage; “real-time” remote biometric identification is used to indiscriminately surveil in public places; biometric systems infer sensitive attributes; emotions are inferred in certain sensitive settings; and decisions about military targeting are fully automated. Where there is uncertainty about the human rights compliance of a particular AI system, there must be a moratorium on its use until it is demonstrated that it can be used in a manner compliant with human rights.

4. Risk-based regulation

AI systems in counter-terrorism, including those involving automated processing of personal data to evaluate individuals, should be subject to heightened regulation, which should include[196] enhanced transparency requirements; adequate data governance; provision of technical documentation and record-keeping to enable compliance; risk and quality management systems throughout the technology lifecycle; and assurance of human oversight. Particularly invasive AI systems, such as remote biometric identification, should be subject to prior judicial authorisation, with prompt post-facto authorisation permissible in emergencies.

5. Transparency and explainability

Developers should be required to design AI systems to maximise transparency and explainability in how they function, thus enabling their operators to adequately understand and interrogate the functioning of algorithmic processes and their outputs, as well as the sources, currency and weighting of underlying datasets.

6. Transfer of AI systems

AI systems which cannot be used in compliance with human rights must be prohibited from domestic or international sale or transfer. Where there is uncertainty about the human rights compliance of an AI system, it must not be transferred until it is established that it can be verifiably used in conformity with human rights. The transfer of other AI systems must be subject to rigorous human rights diligence, including a prohibition on transfer where there is a credible risk that they could be used to violate human rights in the place of deployment, including by targeting human rights defenders, journalists and other civil society actors.[197]

B. Research, development and human rights due diligence

7. Human rights due diligence by developers

States must require businesses or others developing AI systems for use in counter-terrorism to conduct rigorous, comprehensive human rights due diligence impact assessments, including to identify biases and impacts on vulnerable, disadvantaged and marginalised groups (including detainees, minorities, persons with disabilities, women, children, migrants, and LGBTIQ+ persons). Due diligence processes should engage all relevant stakeholders, including diverse civil society organisations.[198] Initial assessments require periodic review and evaluation, given that an AI system's functioning may necessitate adjustments in its performance. Assessments should be subject to external review and validation and be made publicly available. Investors in AI, whether in research, capital infrastructure or commercial activities, must also undertake relevant due diligence. Businesses must devote sufficient personnel, expertise, capacity and resources to facilitate due diligence. There must be effective and dissuasive financial and administrative penalties for failures to conduct adequate human rights due diligence.[199]

8. Data quality, testing and validation

AI system developers and deployers should ensure that datasets are representative, ethically sourced, accurate, current, and relevant to the intended purpose. AI systems should be rigorously tested and validated in environments that simulate their operational context,

including to assess risks such as automation bias (namely, a deployer's over-reliance on AI-enabled outputs); and they should be re-tested whenever they are modified or their intended purpose changes. Quality assurance or management systems should be established.

9. Ensure diversity of personnel and engage stakeholders

Developers should ensure diversity of personnel within AI design and engineering teams and provide diversity training for management. Inclusive stakeholder consultations should be regularly held, including with those potentially most affected by AI systems, and mechanisms should be put in place to effectively report any potential biases identified.[200]

10. Encourage research on human rights implications

All stakeholders should invest in research into the human rights implications of the use of AI, including in counter-terrorism, in particular to develop a more nuanced understanding of how the interdependency and indivisibility of human rights are engaged. Research should also consider the laws, policies, procedures and technical innovations necessary to ensure that human rights are fully respected and protected in AI systems. Collaboration between government agencies, academic institutions, and private sector entities should advance responsible AI innovation, including by prioritising explainable AI, bias mitigation, privacy-preserving technologies, and data security, while adhering to ethical research standards and protecting subjects in accordance with data protection and privacy laws.

11. Undertake long term planning

Technology developments are very likely to outpace efforts to regulate AI, which creates risks for the effectiveness of international norms to safeguard human rights. To address this concern, broad collaboration amongst stakeholders on new and emerging technologies should be forward-looking, including by conducting regular horizon scanning and scenario planning exercises, and encouraging, where feasible, technology-neutral language to legal standards and frameworks that may be applied consistently over time.

C. Controlling the use of AI

12. Human rights due diligence by deployers

Those who deploy AI systems must conduct human rights impact assessments before, during and after use, to establish and maintain clear boundaries for AI-assisted decision-making processes, mitigate potential adverse effects or refrain from using or terminate use of the system, and ensure alignment with international human rights law and national law. Due diligence must holistically consider the legal and technological environment in which they would be embedded,[201] including their integration with other new technologies. Due diligence processes should engage all relevant stakeholders, including diverse civil society organisations. Business must devote sufficient personnel, expertise, capacity and resources to facilitate due diligence and human rights compliance, proportionate to the nature and scale of the use of AI, including in areas such as online content moderation.

13. Mandate human-in-the-loop protocols

AI should support, not replace, human decision-making in counter-terrorism contexts. Qualified personnel must retain effective oversight of operational uses of AI, with heightened scrutiny where impact assessments have identified elevated risks to human rights. Human oversight should aim to prevent and minimise risks to human rights from the operation of AI systems, including from unintended or unanticipated consequences.

14. Ensure transparency and explainability

AI systems must function in a manner that allows for comprehensible explanations of their operation, inputs and outputs and recommendations. Algorithms functioning in an opaque manner should be replaced with interpretable models that enable understanding of reasoning processes, facilitating transparency and accountability. Governments should invest in public programs and community workshops that demystify algorithmic decision-making processes and fostering informed civic participation in AI governance discussions.

15. Invest in continuous training and education

Personnel operating AI systems require comprehensive training in both their technological capabilities and implications for human rights protection. Regular professional development programs, subject to ongoing evaluation and review, should address evolving AI technologies, legal requirements, cultural sensitivity, and decision-making responsibilities in complex operational environments. Training should also be available for those involved in oversight and accountability processes and mechanisms, including legal professionals and judges.

16. Protect human rights in AI-enabled surveillance

Indiscriminate AI-powered mass surveillance of the population must be prohibited. In relation to targeted individual or bulk surveillance, the law must precisely define the grounds for monitoring; and any interference must be necessary and proportionate in pursuit of a legitimate security aim, non-discriminatory, and subject to due process, independent oversight, judicial safeguards, and effective remedies. AI-enabled bulk electronic interception must be subject to prospective independent authorisation; the authorisation must identify the categories of "selectors" to be used; each selector must be justified as necessary and proportionate; and detailed records must be kept. There must be independent supervision and review of bulk surveillance with "end-to-end safeguards" to assess necessity and proportionality at each stage.

17. Rights-compliant AI-supported moderation of online terrorist content

States must regulate online content moderation in accordance with the rights to freedom of expression, association, peaceful assembly, participation in public affairs, and non-discrimination. Expression may only be restricted, whether by States or private actors, in accordance with the requirements of legality, necessity, proportionality and legitimate aim, consistent with articles 19 and 20(1) of the ICCPR,[202] the Rabat Plan of Action against hate speech, and the recommendations on incitement to terrorism of the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism.[203] States must ensure that companies undertake and publicly disclose human rights due diligence and impact assessments. Companies should also provide adequate, well-trained, unbiased human content moderation and oversight of AI content moderation; ensure algorithmic transparency; monitor the effectiveness of mitigating actions; publish moderation policies and detailed moderation reports; maintain

adequate and well-trained in-house human rights staff; and provide effective internal appeal mechanisms.[204] Investors should conduct due diligence on investments and use their leverage to protect human rights.

18. Protect human rights in AI-supported border management

Any use of AI in border management must provide meaningful human oversight and safeguards for the rights to privacy and data protection, non-refoulement and asylum, freedom of movement of nationals (to enter and leave one's own country), freedom of non-nationals from arbitrary expulsion or arbitrary denial of entry, the right to family unity, non-discrimination, liberty, and freedoms of expression, association and peaceful assembly. Border authorities must ensure transparency, explainability, dataset accuracy and non-discrimination in algorithms and implement robust data security. Unnecessary and disproportionate over-collection of data must be avoided, and there must be rigorous safeguards on cross-border data sharing. New technologies such as behavioural biometrics and linguistic analysis of narrative inconsistencies should not be used unless their reliability is validated. Authorities must carefully safeguard against inaccuracies and bias in facial recognition systems across demographic groups.

19. Ensure fair trial and judicial independence in AI-supported legal proceedings

The use of AI in judicial systems must be governed by judges, context-specific, and respectful of local needs, languages, culture and legal traditions.[205] Judiciaries should adopt guidelines on the use of AI (see also UNESCO Guidelines for the Use of AI Systems in Courts and Tribunals 2025) to ensure respect for the right to fair trial (including equality of arms, legal representation and the presumption of innocence), judicial independence, non-discrimination, liberty and other affected rights. Key information about judicial AI systems should be published, to enable legal challenges and civil society oversight. AI evidence, including digital and forensic evidence, must be subject to the ordinary rules of admissibility, disclosure and challenge. Predictive analytics should never be used as a basis for prosecution or detention and should be stringently controlled when used in civil or administrative proceedings to restrict rights.[206]

20. Safeguard human rights where AI is used in detention

States must ensure that any AI systems in detention do not adversely affect the conditions of detention and the treatment and discipline of detainees, and that detention authorities retain sufficient direct contact with detainees. The use of AI decision-support systems in determining eligibility for detention must remain subordinate to human decision-making, be consistent with the requirements of the right to liberty, including the individual necessity and proportionality of detention and judicial review, and ensure transparency, explainability and non-discrimination in relation to algorithms and datasets.

Data safeguards and data security

21. Maintain data governance standards

Legal frameworks must govern and ensure that data collection, access, processing, sharing, transfer, and storage practices are safe, secure and necessary and proportionate for legitimate counter-terrorism purposes, in accordance with international law.[207] Public authorities and intelligence agencies should implement data protection principles such as data minimization,

purpose limitation requirements, and secure erasure procedures while ensuring compliance with data protection, privacy legislation and adherence to human rights standards that safeguard other interdependent rights. Individuals should be notified when their personal data will become part of a dataset used by an AI system.[208] As a general rule individuals and groups should also be empowered to consider, give and withdraw their informed consent to the use of their data and to choose how their data is used.[209]

22. Implement robust cybersecurity measures

AI systems in counter-terrorism contexts present attractive targets for adversarial attacks, data poisoning activities such as the corruption of training data, and system manipulation. Multi-layered security architectures must protect against both external threats and internal misuse while maintaining system integrity and operational continuity.

D. Monitoring, oversight and accountability

23. Implement rigorous algorithmic testing and independent auditing mechanisms

Regular algorithmic audits must examine AI systems for bias, accuracy, adverse impacts and interferences with human rights. Independent oversight bodies should evaluate training data quality, machine learning model performance across different groups, and potential discriminatory outcomes, particularly regarding racial, religious, and ethnic profiling in surveillance and threat assessment applications.

24. Ensure effective record keeping and audit trails

The design of AI systems should incorporate adequate record-keeping capabilities, including automatic recording, in order to identify relevant events and risks and enable an audit trail. Public registers should also be maintained of AI systems used in counter-terrorism.

25. Create independent public oversight mechanisms

Accountability requires effective independent oversight of AI in law enforcement and intelligence, in accordance with best practice international standards on oversight.[210] Independent review bodies, legislative committees, and judicial oversight mechanisms should monitor the use of AI systems, investigate complaints, and ensure compliance with the law and human rights while respecting confidentiality safeguards. Oversight mechanisms must have guaranteed independence, sufficient human, financial and technical resources (including relevant AI expertise), and a broad mandate with adequate powers. In addition to a national human rights institution and privacy and data protection authorities, States should consider establishing a dedicated AI oversight body or regulator with competencies covering counter-terrorism, law enforcement, intelligence and security, border management, and the military. Oversight bodies must be given access to all relevant datasets, algorithms, stored data and audit trails, and have the capacity to review data retention and management practices and real-time digital operations, as well as inter-agency and cross-border data sharing.[211]

26. Ensure accessible and effective remedies and accountability for violations

State and private actors who deploy AI should transparently disclose when and how AI is used and publicise information about complaints and remedial processes. Companies should establish, maintain and adequately resource accessible and effective complaints mechanisms for remedying violations of human rights resulting from their AI systems. Any individual who believes that their rights have been infringed by the use of an AI system must be able to bring a complaint to an independent oversight institution, such as an ombudsman, national human rights institution or a court. Victims of unlawful actions, including rights violations, must have recourse to an institution that can provide a binding, effective remedy, including full reparation for the harm suffered. Procedures must ensure due process, including sufficient disclosure concerning algorithms, data sources and potential biases and access to legal representation. Justice must be accessible, including through legal aid in accordance with international law.

E. International cooperation and collaboration

27. Protect human rights in cross-border information sharing involving AI systems

International cooperation and coordination mechanisms should address jurisdictional challenges in AI-enabled counter-terrorism operations. Standardized protocols for information sharing, joint operations, mutual legal assistance and extradition should incorporate safeguards for individuals' rights across different legal systems while maintaining operational effectiveness. The privacy and other rights of non-citizens abroad should not be arbitrarily excluded from legal protection. Certification frameworks should be developed to enable secure information exchange between certified systems while ensuring compliance with human rights, including privacy and data protection. Data should not be shared where it is likely that it will be used for the purpose of violation of human rights.

28. United Nations human rights due diligence

All relevant United Nations entities must address the human rights implications of providing capacity-building and technical assistance capacity in relation to AI systems (including in relation to biometrics and border management tools), particularly regarding States with demonstrated records of human rights violations in counter-terrorism.[212] Both moratorium and suspension protocols on capacity-building and technical assistance should be established, not limited to situations covered by the Human Rights Due Diligence Policy on United Nations Support to Non-United Nations Security Forces.

29. Strengthen human rights-compliance global governance of AI

Member States and international organisations should ensure that processes to strengthen global governance of AI, including through the Global Dialogue on AI Governance and the Independent International Scientific Panel on AI, preserve and strengthen compliance with international human rights law, international humanitarian law and international refugee law, including through specific initiatives to particularise the effective application of these norms where AI systems are used in counter-terrorism or other security contexts.

F. Military applications

30. Strengthen the international regulation of autonomous weapons systems

States should urgently negotiate by 2026 a treaty regulating autonomous weapons systems, as advocated by the Secretary-General and the ICRC,[213] including to limit where, when and for how long they are used, the types of targets they strike and the scale of force used, and to ensure effective human supervision and timely intervention and deactivation.

31. Endorse the ICRC's recommendations on AI-enabled decision support systems[214]

In particular, States and parties to armed conflicts must ensure that human control and judgement are preserved in decisions that pose risks to the life, liberty, and human dignity (ICRC Recommendation 1) and ensure lawful and responsible use in compliance with IHL (R11); systems should be designed to maximise transparency and explainability (R2); developers and users should mitigate bias to prevent discriminatory outcomes (R4); legal reviews should be conducted wherever AI DSS forms part of a weapons system (R7); users should assess information from all sources reasonably available when drawing on AI DSS outputs (R13); and the use of AI DSS in detention operations must not adversely affect the treatment of detainees and the conditions of detention (R14).

Endnotes

[1] See further: Lawrence Livermore National Laboratory, "The Birth of Artificial Intelligence (AI) Research", https://st.llnl.gov/news/look-back/birth-artificial-intelligence-ai-research.
[2] Ganor, B. (2021). Artificial or human: A new era of counterterrorism intelligence? *Studies in Conflict and Terrorism*, *44(7)*, 605. See also Schmid, S., Riebe, T. and Reuter, C. (2022). Dual-use and trustworthy? A mixed methods analysis of AI diffusion between civilian and defense R&D. *Science and Engineering Ethics*, *28(2)*, 12.
[3] Kathleen McKendrick, "Artificial Intelligence: Prediction and Counterterrorism", Chatham House Research Paper, 2019, 2.
[4] Saheb, T. (2023). Ethically contentious aspects of artificial intelligence surveillance: A social science perspective. *AI and Ethics*, *3(2)*, 369; Khan, F.A., Li, G., Khan, A. N., Khan, Q.W., Hadjouni, M. and Elmannai, H. (2023). AI-driven counter-terrorism: Enhancing global security through advanced predictive analytics. *IEEE Access*, *11*, 135864, https://ieeexplore.ieee.org/document/10328769; Campbell, L. et al (1997). The use of artificial intelligence in military simulations. *IEEE* International Conference on Systems, Man and Cybernetics: Computational Cybernetics and Simulation. *3*. For a general overview of applications see Morgan, F.E., et al "Military Applications of Artificial Intelligence" (RAND Corporation, 2020); and UNICRI and INTERPOL, "Toolkit for Responsible Artificial Intelligence Innovation in Law Enforcement" (2023), https://www.ai-lawenforcement.org.
[5] Submissions to this Position Paper by the EU, 10 and Bangladesh NGOs Network for Radio and Communication, 2.
[6] A/HRC/52/39, para. 23.
[7] A/HRC/52/39, para. 24.
[8] McKendrick, above note 3, 17.
[9] See e.g. Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism, "Human Rights Implications of the Development, Use and Transfer of New Technologies in the Context of Counter-terrorism and Countering and Preventing Violent Extremism", A/HRC/52/39; United Nations High Commissioner for Human Rights, "The Right to Privacy in the Digital Age", A/HRC/48/31. See also Submission to this Position Paper by the EU, 5.
[10] A/HRC/RES/58/23.
[11] E.g. the United Nations Countering Terrorist Travel Programme and goTravel Software Solution: https://www.ohchr.org/sites/default/files/documents/issues/terrorism/sr/statements/2023-10-30-a-ct-travel-gotravel-position-paper.pdf.
[12] A/RES/78/265, paras 1-2 and 13; A/HRC/RES/58/23, para. 6. See also Global Digital Compact, A/79/L.2, para. 8(i).
[13] A/RES/78/265 (2024), paras. 6 and 8.
[14] A/HRC/RES/58/23, para. 9.
[15] Khan et al, above note 4; Kalsooma, U., Arshad, S., Albarah, A., Siddiqi, I., Ullah, S., Mateen, A. and Amin, F. (2025). A big data driven multilevel deep learning framework for predicting terrorist attacks. *Scientific Reports*, *15(1)*, 23060.
[16] See also Confidential Submission 1 to this Position Paper by a regional organisation, 1-2; and Submissions by the Fundación Regional de Asesoría en Derechos Humanos (INREDH), 2-3; Gavin Sullivan and Nadia Jude (Edinburgh Law School), 2-4.
[17] OSCE, "Artificial Intelligence in the Context of Preventing and Countering Violent Extremism and Terrorism: Challenges, Risks and Opportunities", Event Summary Document, 2024, https://www.osce.org/files/f/documents/4/f/575877.pdf.
[18] Bridgelall, R. (2022). Applying unsupervised machine learning to counterterrorism. *Journal of Computational Social Science*, *5(2)*, 1099.
[19] Saini, J.K. and Bansal, D. (2024). Computational techniques to counter terrorism: a systematic survey. *Multimedia Tools and Applications*, 83(1), 1189; Cross, M.K.D. (2023). Counter-terrorism and the intelligence network in Europe. *International Journal of Law, Crime and Justice*, *72*, 100368.
[20] Montasari, R., "Machine learning and deep learning techniques in countering cyberterrorism", in Montasari, R., *Cyberspace, Cyberterrorism and the International Security in the Fourth Industrial Revolution: Threats, Assessment and Responses* (Springer International Publishing, 2024), 135; Pemmasani, P.K. (2023). AI in national security: Leveraging machine learning for threat intelligence and response. *The Computertech*, 1.
[21] A/HRC/52/39, para. 45.
[22] Abdulrahman, S.A. and Alhayani, B. (2023). A comprehensive survey on the biometric systems based on physiological and behavioural characteristics. *Materials Today: Proceedings*, *80*, 2642.
[23] Kieslich, K. and Lünich, M. (June 2024). Regulating AI-based remote biometric identification: Investigating the public demand for bans, audits and public database registrations. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability and Transparency*, 173.
[24] Submission to this Position Paper by Human Rights Myanmar, 2.

[25] EDPS, "Misunderstanding with regard to Biometric Data", June 2020, https://www.edps.europa.eu/sites/default/files/publication/joint_paper_14_misunderstandings_with_regard_to_identification_and_auuthentication_en.pdf.
[26] Submission to this Position Paper by ECNL, 1-2.
[27] Jain, A.K., Deb, D. and Engelsma, J.J. (2021). Biometrics: Trust, but verify. *IEEE Transactions on Biometrics, Behaviour and Identity Science*, *4(3)*, 303. See also Thomas, R. (2025). Biometrics, international migrants and human rights. *European Journal of Migration and Law*, *7(4)*, 377; Huszti-Orbán, K. and Ní Aoláin, F., "Use of Biometric Data to Identify Terrorists: Best Practice or Risky Business?" (University of Minnesota Human Rights Center, 2020).
[28] Al-Khateeb, M.T. (2021). Toward a rhetorical account of refugee encounters: Biometric screening technologies and failed promises of mobility. *Rhetoric Society Quarterly*, *51(1)*, 15.
[29] Ross, A., Banerjee, S. and Chowdhury, A. (2022). Deducing health cues from biometric data. *Computer Vision and Image Understanding*, *221*, 103438.
[30] Byeon, H., Raina, V., Sandhu, M., Shabaz, M., Keshta, I., Soni, M. and Lakshmi, T. V. (2024). Artificial intelligence-enabled deep learning model for multimodal biometric fusion. *Multimedia Tools and Applications*, *83(33)*, 80105.
[31] Michael, K., Abbas, R., Jayashree, P., Bandara, R. J. and Aloudat, A. (2022). Biometrics and AI bias. *IEEE Transactions on Technology and Society*, *3(1)*, 2.
[32] https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks.
[33] Almabdy, S. and Elrefaei, L. (2019). Deep convolutional neural network-based approaches for face recognition. *Applied Sciences*, *9(20)*, 4397.
[34] Remote biometric identification is often identified as a high-risk to human rights. See relevantly the review conducted under the EU's AI Act: Submission to this Position Paper by the EU, 4. However, civil society have criticised the EU AI Act: Submission to this Position Paper by Privacy International 4; see also EDRi, "EU's AI Act fails to set gold standard for human rights", https://edri.org/our-work/eu-ai-act-fails-to-set-gold-standard-for-human-rights/.
[35] Europol, "AI and Policing: The Benefits and Challenges of Artificial Intelligence for Law Enforcement", September 2024, https://www.europol.europa.eu/publication-events/main-reports/ai-and-policing; NYU School of Law, "How Policing Agencies Use AI", October 2024, https://www.policingproject.org/ai-explained-articles/2024/9/6/how-policing-agencies-use-ai. See also: Confidential Submission 2 to this Position Paper by an INGO, 5-6.
[36] Fountain, J.E. (2022). The moon, the ghetto and artificial intelligence: Reducing systemic racism in computational algorithms. *Government Information Quarterly*, *39(2)*, 101645.
[37] Abiade, S.F. (2025). Artificial Intelligence surveillance in counterterrorism: Assessing democratic accountability and civil liberties trade-offs. *International Journal of Science and Research Archive*, *16(1)*, 89.
[38] Submission to this Position Paper by Gavin Sullivan and Nadia Jude (Edinburgh Law School), 4-6.
[39] Limantė, A. (2024). Bias in facial recognition technologies used by law enforcement: Understanding the causes and searching for a way out. *Nordic Journal of Human Rights*, *42(2)*, 115; Wang, X., Wu, Y.C., Zhou, M. and Fu, H. (2024). Beyond surveillance: privacy, ethics and regulations in face recognition technology. *Frontiers in Big Data*, *7*, 1337465.
[40] Confidential Submission 2 to this Position Paper by an INGO, 1.
[41] Confidential Submission 1 to this Position Paper by a regional organisation, 2; UNICRI and UNOCT, "Countering Terrorism Online with Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia", 2021, https://unicri.org/Publications/Countering-Terrorism-Online-with-Artificial-Intelligence-%20SouthAsia-South-EastAsia, 23-33; Khan et al, above note 4. See also Confidential Submission 3 to this Position Paper by an INGO, 2; EDRi, "Use Cases: Impermissible AI and Fundamental Rights Breaches", August 2020, https://edri.org/wp-content/uploads/2021/06/Case-studies-Impermissible-AI-biometrics-September-2020.pdf.
[42] Submission to this Position Paper by Privacy International, 2; OHCHR, "The Right to Privacy in the Digital Age", A/HRC/48/31, 15 September 2021, https://www.ohchr.org/en/documents/thematic-reports/ahrc4831-right-privacy-digital-age-report-united-nations-high, para. 14. See further: Allam, Z. and Dhunny, Z.A. (2019). On big data, artificial intelligence and smart cities. *Cities*, *89*, 80; Fan, W.Q., Ismail, A.S., Mohammed, F. and Mukred, M., "AI-driven Smart City Security and Surveillance System: A Bibliometric Analysis" in Al Sharafi, M, et al (eds.), *Current and Future Trends on AI Applications: Volume 1* (Springer Nature Switzerland, 2025), 305.
[43] Thakur, N., Nagrath, P., Jain, R., Saini, D., Sharma, N. and Hemanth, D.J. (2021). Artificial intelligence techniques in smart cities surveillance using UAVs: A survey. *Machine Intelligence and Data Analytics for Sustainable Future Smart Cities*, 329.
[44] Burton, J. (2023). Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization and political violence. *Technology in Society*, *75*, 102262.
[45] Aminiyeganeh, K., Coutinho, R.W. and Boukerche, A. (2024). IoT video analytics for surveillance-based systems in smart cities. *Computer Communications*, *224*, 95; Cheng, X., et al (2017). Exploiting mobile big data: Sources, features and applications. *IEEE Network*, *31*, 72.
[46] Zhang, J., Huang, J., Jin, S. and Lu, S. (2024). Vision-language models for vision tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46(8)*, 5625.
[47] https://openai.com/index/clip/.
[48] Akilli, E. (2024). Artificial intelligence in counterterrorism: Navigating the intersection of security, ethics and privacy. *SETA Perspective, 73*, 4.
[49] Caldarella, S., Mancini, M., Ricci, E. and Aljundi, R. (September 2024). The phantom menace: Unmasking privacy leakages in vision-language models. *European Conference on Computer Vision,* 435.
[50] Ghosh, A., Acharya, A., Saha, S., Jain, V. and Chadha, A. (2024). Exploring the frontier of vision-language models: A survey of current methodologies and future directions. *arXiv preprint arXiv:2404.07214*.
[51] Karamolegkou, A., Rust, P., Cao, Y., Cui, R., Søgaard, A. and Hershcovich, D. (2024). Vision-language models under cultural and inclusive considerations. *arXiv preprint arXiv:2407.06177*.

[52] ACLU, "Machine Surveillance is Being Super-Charged by Large AI Models", March 2024, https://www.aclu.org/news/privacy-technology/machine-surveillance-is-being-super-charged-by-large-ai-models; IPO Forum, "Advancing Smart Surveillance: Vision Language Models (VLMs) that Comprehend Surveillance Camera Footage", 22 May 2025, https://www.ipoforum.org.tw/en/news/44; Liu, L. et al (2025). MINGLE: VLMs for semantically complex region detection in urban scenes, *arXiv preprint arXiv:2509.13484*.
[53] Confidential Submission 1 to this Position Paper by a regional organisation, 12. See also: Meta, "I-JEPA: The First AI Model Based on Yann LeCun's Vision for More Human-like AI", 13 June 2023, https://ai.meta.com/blog/yann-lecun-ai-model-i-jepa/.
[54] Gordon, E., Baldwin-Philippi, J. and Balestra, M., "Why We Engage: How Theories of Human Behavior Contribute to Our Understanding of Civic Engagement in a Digital Era", Berkman Center Research Publication No. 21, 2013; Jung, N., Kim Y. and de Zuniga, H. (2011). The mediating role of knowledge and efficacy in the effects of communication on political participation. *Mass Communication and Society*, 407.
[55] See e.g. Human Rights Council resolution 24/5 (2013); Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association, A/HRC/23/39 (2013); European Commission for Democracy through Law (Venice Commission) and OSCE Office for Democratic Institutions and Human Rights, "Joint Guidelines on Freedom of Association" (2014), 10-16; European Court of Human Rights Press Unit, "Factsheet: Political Parties and Associations" (2016); International Labour Organisation, "Freedom of Association: Compilation of Decisions of the Committee on Freedom of Association" (ILO, 2018); A/HRC/41/41 (2019).
[56] Marwick, A.E. and Boyd, D. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media and Society*, *16(7)*, 1051.
[57] Tahat, K., Habes, M., Mansoori, A., Naqbi, N., Al Ketbi, N., Maysari, I. and Altawil, A. (2024). Social media algorithms in countering cyber extremism: A systematic review. *Journal of Infrastructure, Policy and Development*, *8(8)*, 6632.
[58] Strandburg, K., "Surveillance of Emergent Associations" in Acquisti, A. et al, *Digital Privacy: Theory, Technologies and Practices* (Auerbach Publications, 2008), 435, 437; Runtzen, D. and Zenn, J. (2011). Association and assembly in the digital age, *The International Journal of Not-for-Profit Law*, *13(4)*, 53.
[59] Shapovalova, N., "Assessing Ukrainian Grassroots Activism Five Years After Euro-maidan", *Carnegie Europe*, 6 February 2019, https://carnegieeurope.eu/2019/02/06/assessing-ukrainian-grassroots-activism-five-years-after-euromaidan-pub-78248.
[60] Noman, H., "Arab Religious Skeptics Online: Anonymity, Autonomy and Discourse in a Hostile Environment", Berkman Center Research Publication No. 2015-2, 2015; Steinberg, S. (2017). Sharenting: Children's privacy in the age of social media, *Emory Law Journal*, *66*, 839; McKenna, K. and Bargh, J. (1998). Coming out in the age of the internet: Identity demarginalization" through virtual group participation. *Journal of Personality and Social Psychology*, *75*, 681; Watt, S., Lea, M. and Spears, R., "How Social is Internet Communication? A Reappraisal of Bandwidth and Anonymity Effects" in Woolgar, S. (ed.), *Virtual Society? Technology, Cyberbole, Reality* (Oxford University Press, 2002), 61.
[61] For deployments of AI systems for counter-terrorism purposes engaging freedom of expression concerns, see further: Submission to this Position Paper by SMEX, 3-4; see also Submission to this Position Paper by Starling Lab, 2-5.
[62] A/80/341, para. 65.
[63] Profiling of this nature applied to personality traits and characteristics for predictive purposes presents an especially grave risk. See further Submission to this Position Paper by the EU, 4.
[64] Barabasi, A.L., *Linked: The New Science of Networks* (Perseus, 2002); Carrington, P.J., Scott, J. and Wasserman, S., *Models and Methods in Social Network Analysis* (Cambridge University Press, 2005).
[65] Koehler, D., Fiebig, V. and Jugl, I. (2023). From gaming to hating: Extreme-right ideological indoctrination and mobilization for violence of children on online gaming platforms. *Political Psychology*, *44(2)*, 419; Holbrook, D. (2021). The terrorism information environment: Analysing terrorists' selection of ideological and facilitative media. *Terrorism and Political Violence*, *33(4)*, 697.
[66] Winder, S., Aquilino, M. C., Warnier, S., Paillé, P., Zürcher, E. and Toro, D., "Combatting New Forms of Extremism: RAND Europe Evidence Submission to the UK Parliament's Home Affairs Committee", 2025,, https://www.rand.org/content/dam/rand/pubs/testimonies/CTA4100/CTA4128-1/RAND_CTA4128-1.pdf.
[67] Secretary-General, "Intensification of efforts to eliminate all forms of violence against women and girls: technology-facilitated violence against women and girls", A/79/500 (2024), para. 29.
[68] Zhang, X. and Davis, M. (2024). E-extremism: A conceptual framework for studying the online far right. *New Media and Society*, *26(5)*, 2954. See also Submission to this Position Paper by ECNL, 2-3.
[69] Alsaad, A., Taamneh, A. and Al Jedaiah, M.N. (2018). Does social media increase racist behaviour? An examination of confirmation bias theory. *Technology in Society*, *55*, 41. See also Citron, D.K. (2020). Cyber mobs, disinformation and death videos: The internet as it is (and as it should be). *Michigan Law Review*, *118*, 1073, 1078. For changes made by Facebook following criticism of its 'Secret Groups' feature, see Gebhart, G., 'Understanding Public, Closed and Secret Facebook Groups", *Electronic Frontier Foundation*, 13 June 2017, www.eff.org/deeplinks/2017/06/understanding-public-closed-and-secret-facebook-groups.
[70] ICCT, "Under Pressure: Rethinking Comprehensive Approaches to CT and P/CVE in an Age of Austerity and Instability", 2025, https://icct.nl/publication/under-pressure-rethinking-comprehensive-approaches-ct-and-pcve-age-austerity-and.
[71] Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, *7(1)*, 549.
[72] Mylrea, M., "The Generative AI Weapon of Mass Destruction: Evolving Disinformation Threats, Vulnerabilities and Mitigation Frameworks" in Lawless, W. et al (eds.), *Interdependent Human-Machine Teams: The Path to Autonomy* (Academic Press, 2025) 315.
[73] DeCook, J.R. and Forestal, J. (2023). Of humans, machines and extremism: The role of platforms in facilitating undemocratic cognition. *American Behavioral Scientist*, *67(5)*, 629.
[74] Parent company and publisher Epic Games reported more than 650 million accounts worldwide for just one of its gaming entities, Fortnite.: https://www.statista.com/topics/5847/fortnite/.

[75] Gallagher, A., O'Connor, C., Vaux, P., Thomas, E. and Davey, J., "The Extreme Right on Discord", Institute for Strategic Dialogue, 2021, https://www.isdglobal.org/isd-publications/gaming-and-extremism-the-extreme-right-on-discord; Davey, J., "Extremism on Gaming (-adjacent) Platforms" in Schlegel, L. and Kowert, R. (eds.), *Gaming and Extremism: The Radicalization of Digital Playgrounds* (Routledge, 2024), 95.
[76] Freilich, J.D., Chermak, S.M., Arietti, R.A. and Turner, N.D. (2024). Terrorism, political extremism and crime and criminal justice. *Annual Review of Criminology*, *7(1)*, 187.
[77] Europol, "Internet Organised Crime Threat Assessment 2024", https://www.europol.europa.eu/publication-events/main-reports/internet-organised-crime-threat-assessment-iocta-2024; Andrews, S., Brewster, B. and Day, T. (2018). Organised crime and social media: A system for detecting, corroborating and visualising weak signals of organised crime online. *Security Informatics*, *7*, 3.
[78] Allchorn, W. (2024). Global far-right extremist exploitation of artificial intelligence and alt-tech. *Counter Terrorist Trends and Analyses*, *16(3)*, 13.
[79] Bibi van Ginkel, Yael Boerma and Tanya Mehra, "Reading Between the Lines: The Importance of Human Moderators for Online Implicit Extremist Content Moderation", ICCT, December 2025.
[80] Special Rapporteur on Freedom of Expression, "Threats to freedom of expression online in turbulent times", A/80/341, para. 61.
[81] van Ginkel, above note 79.
[82] A/80/341, para. 63.
[83] van Ginkel, above note 79.
[84] A/80/341, paras. 69-70; van Ginkel, ibid.
[85] A/80/341, paras. 102 and 109.
[86] A/HRC/16, 51, paras. 30-32; A/HRC/40/52, para. 37.
[87] A/80/341, paras. 73, 74, 109 and 110; van Ginkel, above note 79.
[88] Confidential Submission 1 to this Position Paper by a regional organisation, 6-8. See also: OSCE, "Artificial Intelligence in the Context of Preventing and Countering Violent Extremism and Terrorism: Challenges, Risks and Opportunities", Event Summary Document, 2024, https://www.osce.org/secretariat/575877.
[89] Global Digital Compact, A/79/L.2, para. 23(d).
[90] Submission to this Position Paper by the OSCCE, 14.
[91] Leese, M., Noori, S. and Scheel, S. (2022). Data matters: The politics and practices of digital border and migration management. *Geopolitics*, *27(1)*, 5.
[92] ECNL, "Technology and Counter-Terrorism: Mapping the Impact of Biometric Surveillance and Social Media Platforms on Civic Space", November 2022, https://ecnl.org/sites/default/files/2023-03/TECHNOLOGY%20AND%20COUNTER-TERRORISM_NOV%2022.pdf, 25.
[93] Confidential Submission 1 to this Position Paper by a regional organisation, 4 and Submission from the Mediterranean Centre for Peace and Security, 2-6.
[94] A/HRC/52/39, paras. 18-26.
[95] Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism, Position Paper of the on the United Nations Countering Terrorist Travel Programme and the goTravel Software Solution, 2023, https://www.ohchr.org/sites/default/files/documents/issues/terrorism/sr/statements/2023-10-30-a-ct-travel-gotravel-position-paper.pdf.
[96] Khan, N. and Efthymiou, M. (2021). The use of biometric technology at airports: The case of customs and border protection (CBP). *International Journal of Information Management Data Insights*, *1(2)*, 100049.
[97] Ryu, R., Yeom, S., Kim, S.H. and Herbert, D. (2021). Continuous multimodal biometric authentication schemes: a systematic review. *IEEE Access*, *9*, 34541.
[98] Korshunov, P. and Marcel, S. (September 2018). Speaker inconsistency detection in tampered video. In *IEEE 2018 26th European signal processing conference*, 2375.
[99] IBM, "What is ground truth?", 2025, https://www.ibm.com/think/topics/ground-truth#.
[100] Kou, Y., Lu, C.T., Sirwongwattana, S. and Huang, Y.P. (March 2004). Survey of fraud detection techniques. *IEEE International Conference on Networking, Sensing and Control*, *2*, 749.
[101] Odufisan, O.I., Abhulimen, O.V. and Ogunti, E.O. (2025). Harnessing artificial intelligence and machine learning for fraud detection and prevention in Nigeria. *Journal of Economic Criminology*, *7*, 100127.
[102] Ylönen, M. and Aven, T. (2023). A new perspective for the integration of intelligence and risk management in a customs and border control context. *Journal of Risk Research*, *26(4)*, 433.
[103] La Fors, K. and Meissner, F. (2022). Contesting border artificial intelligence: Applying the guidance-ethics approach as a responsible design lens. *Data and Policy*, *4*, e36.
[104] See e.g. European Commission, "Automated Border Control", https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/automated-border-control-abc_en.
[105] US Commission on Civil Rights, "The Civil Rights Implications of the Federal Use of Facial Recognition Technology", September 2024, https://www.usccr.gov/files/2024-09/civil-rights-implications-of-frt_0.pdf; Confidential Submission to this Position Paper by an INGO, 5-6.
[106] The use of new technologies extends to justice sector strategies; process re-engineering; automation, data collection, processing and retention; integration of legacy systems with modern functionalities; and online dispute resolution, e-filing, remote court process and technologies used to digitise and provide access to legal documents and evidence: UNDP, "Emerging Technologies and Judicial Integrity in ASEAN", 2021, https://www.undp.org/sites/g/files/zskgke326/files/2022-03/UNDP-RBAP-Emerging-Technologies-and-Judicial-Integrity-in-ASEAN-2021.pdf.
[107] See A/80/169 (2025).

[108] Simshaw, D. (2022). Access to AI justice: Avoiding an inequitable two-tiered system of legal services. *Yale Journal of Law and Technology*, *24*, 150.
[109] Ejjami, R. (2024). AI-driven justice: Evaluating the impact of artificial intelligence on legal systems. *International Journal of Multidisciplinary Research*, *6(3)*, 1.
[110] Završnik, A. (2020). Criminal justice, artificial intelligence systems and human rights. *ERA Forum*, *20(4)*, 567.
[111] A/80/169 (2025), para. 63.
[112] See A/79/296.
[113] See A/79/296.
[114] UNESCO Global Toolkit on AI and the Rule of Law for the Judiciary 2023.
[115] A/80/169 (2025), para. 12.
[116] A/80/169 (2025), para. 68(a), (c) and (d).
[117] Dubai's Oyoon program, for example, integrates over 300,000 cameras with advanced facial recognition and data analytics capabilities Submissions to this Position Paper by SMEX, 4; Durmaz, M., "AI Investments in the Gulf: Opportunities and Surveillance Risks", SMEX, 19 May 2025, www.smex.org/ai-investments-in-the-gulf-opportunities-and-surveillance-risks/. See further: Sachoulidou, A. (2023). Going beyond the 'common suspects': To be presumed innocent in the era of algorithms, big data and artificial intelligence. *Artificial Intelligence and Law*, 1, DOI10.1007/s10506-023-09347-w. See also: Ajil, A. and Staubli, S. (2023). Predictive policing and negotiations of (in) formality: Exploring the Swiss case. *International Journal of Law, Crime and Justice*, *74*: 100605; Martin, K. (2023). Predatory predictions and the ethics of predictive analytics. *Journal of the Association for Information Science and Technology*, *74(5)*, 531.
[118] Hälterlein, J. (2021). Epistemologies of predictive policing: Mathematical social science, social physics and machine learning. *Big Data and Society*, *8(1)*, 20539517211003120.
[119] A/80/169 (2025), para. 68(b).
[120] A/80/284, para. 6(f); A/HRC/52/39, para. 43.
[121] "Ecuador instala en las cárceles del país un sistema de seguridad con inteligencia artificial", *Notiamerica*, 22 November 2022, https://www.notimerica.com/politica/noticia-ecuador-ecuador-instala-carceles-pais-sistema-seguridad-inteligencia-artificial-20221122220042.html.
[122] ICRC, "Submission to the United Nations Secretary-General on Artificial Intelligence in the Military Domain", April 2025, 6, https://www.icrc.org/sites/default/files/2025-04/ICRC_Report_Submission_to_UNSG_on_AI_in_military_domain.pdf.
[123] Ibid, Recommendation 14.
[124] A/HRC/52/39, para. 29.
[125] International Federation for Human Rights and Center for Prisoners' Rights, "End Solitary Confinement and Video Surveillance of Death Row Prisoners", https://www.fidh.org/en/region/asia/japan/japan-end-solitary-confinement-and-video-surveillance-of-death-row.
[126] ICRC, "Submission to the United Nations Secretary-General", above note 122, 6.
[127] Ibid.
[128] See, e.g. Geiß, R. and Lahmann, H. (eds), Research Handbook on Warfare and Artificial Intelligence (Edward Elgar, 2024); Boothby, W.H. (2024). AI warfare and the law. *International Law Studies*, *104*.; Bruun, L. and Bo, M., "Bias in Military Artificial Intelligence and Compliance with International Humanitarian Law" (SIPRI, 2025); and Mačák, K. (2025). Artificial intelligence in armed conflict Cycon 2025 series: Introduction. *Articles of War (Lieber Institute)*, https://lieber.westpoint.edu/artificial-intelligence-armed-conflict-cycon-2025-series-introduction/.
[129] Submission to this Position Paper by Privacy International, 5.
[130] ICRC, "Submission to the United Nations Secretary-General", above note 122; ICRC, "International Humanitarian Law and the Challenges of Contemporary Armed Conflicts: Building a Culture of Compliance for IHL to Protect Humanity in Today's and Future Conflicts", 2024, sections V.2 and V.3, https://www.icrc.org/en/publication/international-humanitarian-law-and-challenges-contemporary-armed-conflicts-building; ICRC (2021)., Artificial intelligence and machine learning in armed conflict: A human-centred approach. *International Review of the Red Cross*, *102(913)*, 463; ICRC, "Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-Making and ICRC Observations on External Report", May 2024, https://www.icrc.org/en/publication/decisions-decisions-decisions-computation-and-artificial-intelligence-military-decision; ; ICRC and Geneva Academy, Artificial Intelligence and Related Technologies in Military Decision-making: Expert Consultation Report", May 2024; ICRC, "What you need to know about AI in armed conflict", October 2023, https://www.icrc.org/en/document/what-you-need-know-about-artificial-intelligence-armed-conflict.
[131] Chan, A., Salganik, R., Markelius, A., Pang, C., Rajkumar, N., Krasheninnikov, D. and Maharaj, T. (June 2023). Harms from increasingly agentic algorithmic systems. *Proceedings of the 2023 ACM Conference on Fairness, Accountability and Transparency*, 651.
[132] OHCHR, "Briefer on Human Rights and Artificial Intelligence in the Military Domain", 7.
[133] Report of the 2023 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, CCW/GGE.1/2023/2, 23 May 2023, https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_2_Advance_version.pdf.
[134] ICRC, "International Humanitarian Law and the Challenges of Contemporary Armed Conflicts", above note 130.
[135] Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K. and Hussain, A. (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation*, *16(1)*, 45; Wood, N.G. (2024). Explainable AI in the military domain. *Ethics and Information Technology*, *26(2)*, 29.
[136] Christie, E.H., Ertan, A., Adomaitis, L. and Klaus, M. (2024). Regulating lethal autonomous weapon systems: Exploring the challenges of explainability and traceability. *AI and Ethics*, *4(2)*, 229.
[137] ICRC, "International Humanitarian Law and the Challenges of Contemporary Armed Conflicts", above note 130, 64–67.

[138] Additional Protocol I of 1977 (API), article 36.
[139] ICRC Commentary of 1987 to article 36 of API, para. 1473.
[140] Copeland, D., Liivoja, R. and Sanders, L. (2023). The utility of weapons reviews in addressing concerns raised by autonomous weapon systems. *Journal of Conflict and Security Law*, *28(2)*, 285.
[141] Schwarz, E. (2021). Autonomous weapons systems, artificial intelligence and the problem of meaningful human control. *Philosophical Journal of Conflict and Violence*, *5*, DOI: 10.22618/TP.PJCV.20215.1.139004.
[142] Submission to this Position Paper by Centro de Estudios Legales y Sociales (CELS), 2.
[143] Refer to, e.g., the discussion of Microsoft's provision of data storage services to the Israeli military: "Microsoft Cuts Off Some Services to the Israeli Military", *Financial Times*, 25 September 2025, https://www.ft.com/content/32062a65-e458-4c04-b464-9af5a37cd16a; "Microsoft Disables Some Cloud Services Used by Israel's Defense Ministry", *Wall Street Journal*, 25 September 2025, https://www.wsj.com/tech/microsoft-cuts-back-work-with-israels-defense-ministry-bd4fae2a?mod=hp_lead_pos1. See also: Access Now, "Artificial Genocidal Intelligence: How Israel is Automating Human Rights Abuses and War Crimes", 9 May 2024, https://www.accessnow.org/publication/artificial-genocidal-intelligence-israel-gaza/; "'Lavender': The AI Machine Directing Israel's Bombing Spree in Gaza", *+972 Magazine*, 3 April 2024, https://www.972mag.com/lavender-ai-israeli-army-gaza/.
[144] UN Working Group on Business and Human Rights, "Guiding Principles on Business and Human Rights: An Introduction", 2024, https://www.ohchr.org/sites/default/files/Documents/Issues/Business/Intro_Guiding_PrinciplesBusinessHR.pdf.
[145] UN Working Group on Business and Human Rights, "Responsible Business Conduct in the Arms Sector: Ensuring Business Practice in Line with the UN Guiding Principles on Business and Human Rights", Information Note, 2021, https://www.ohchr.org/sites/default/files/2022-08/BHR-Arms-sector-info-note.pdf, 1.
[146] Ibid. See also the EU Commission's discussion of the sale and transfer of AI technologies in countering terrorism: Submission to this Position Paper by the EU, 9.
[147] UN Working Group on Business and Human Rights, ibid, 6.
[148] Submission to this Position Paper by INREDH, 6.
[149] Confidential Submission 2 to this Position Paper by an INGO, 4.
[150] See further Submission to this Position Paper by the EU, 9.
[151] E.g. in the Pact for the Future 2024, A/RES/79/1; Eighth Review of the Global Counter-Terrorism Strategy 2023, A/RES/77/298; Delhi Declaration on countering the use of new and emerging technologies for terrorist purposes 2022; Abu Dhabi Guiding Principles on countering the use of new and emerging technologies for terrorist purposes, S/2023/1035; and various Security Council resolutions.
[152] See A/HRC/52/39.
[153] Lele, A., *Quantum Technologies and Military Strategy* (Springer International Publishing, 2021).
[154] Krelina, M. (2021). Quantum technology for military applications. *EPJ Quantum Technology* 8(24), https://doi.org/10.1140/epjqt/s40507-021-00113-y; Der Derian, J. and Rollo, S., "Quantum Warfare" in Gruszczak and Kaempf, S. (eds.), *Routledge Handbook of the Future of Warfare* (Routledge, 2023), 319.
[155] Lovreglio, R., Ngassa, D.C., Rahouti, A., Paes, D., Feng, Z. and Shipman, A. (2022). Prototyping and testing a virtual reality counterterrorism serious game for active shooting. *International Journal of Disaster Risk Reduction*, *82*, 103283.
[156] McKinsey, Quantum Technology Monitor, 2024, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-year-of-quantum-from-concept-to-reality-in-2025. On 7 June 2024, the United Nations proclaimed 2025 as the International Year of Quantum Science and Technology: https://quantum2025.org.
[157] Ienca, M. and Vayena, E. (2018). Dual use in the 21st century: Emerging risks and global governance. *Swiss Medical Weekly*, *148(4748)*, w14688; Schulzke, M. (2019), Drone proliferation and the challenge of regulating dual-use technologies. *International Studies Review*, 21(3), 497.
[158] Khan, M.I., Arif, A. and Khan, A.R.A. (2024). AI's revolutionary role in cyber defense and social engineering. *International Journal of Multidisciplinary Sciences and Arts*, *3(4)*, 57.
[159] For example, the EU's data protection framework is guided by the complementarity of its AI Act with the General Data Protection Regulation (GDPR): Submission to this Position Paper by the EU, 5.
[160] See *Big Brother Watch and Others v. the United Kingdom, Application No.* 58170/13, 62322/14 and 24960/15, European Court of Human Rights (Grand Chamber), Judgment, 25 May 2021, para. 361.
[161] Global Digital Compact, A/79/L.2, para. 39(c).
[162] A/HRC/48/31, para. 55.
[163] Global Digital Compact, A/79/L.2, para. 39(d).
[164] Submission to this Position Paper by Bangladesh NGOs Network for Radio and Communication, 2 and Confidential Submission 2 by an INGO, 4.
[165] See e.g. A/HRC/56/68, para. 12.
[166] UN Committee against Racial Discrimination, General Recommendation No. 36 (2020), para. 58.
[167] Ibid, para. 62.
[168] Ibid, para. 61.
[169] See e.g. UNESCO, "Ethics of Artificial Intelligence", https://www.unesco.org/en/artificial-intelligence/recommendation-ethics; African Union, "The African Union Continental Artificial Intelligence Strategy", 9 August 2024, https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy, 34.
[170] Hassija, V. et al (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation*, *16(1)*, 45.
[171] Bertrand, A. et al (2022). How cognitive biases affect XAI-assisted decision-making: A systematic review. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics and Society*.
[172] Aizenberg, E. and Van Den Hoven, J. (2020). Designing for human rights in AI. *Big Data and Society*, *7(2)*, 2053951720949566.

[173] Islam, M.R. et al (2022). A systematic review of explainable artificial intelligence in terms of different application domains and tasks. *Applied Sciences*, 12(3), 1353.
[174] Kwik, J. and Van Engers, T. (2021). Algorithmic fog of war: When lack of transparency violates the law of armed conflict. *Journal of Future Robot Life*, *2(1-2)*, 43.
[175] Bryce, H. and Parakilas, J., "Conclusions and Recommendations" in Cummings, M.L., Roff, H.M., Cukier, K., Parakilas, J. and Bryce, H. (eds.), *Artificial Intelligence and International Affairs: Disruption Anticipated* (Chatham House, 2018), 43; ICRC, "Artificial Intelligence and Machine Learning", above note 130.
[176] Holland Michel, A., "The Black Box, Unlocked: Predictability and Understandability in Military AI", United Nations Institute for Disarmament Research, 2020, https://doi.org/10.37559/SecTec/20/AI1; Das, A. and Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371*.
[177] Wood, N.G. (2024). Explainable AI in the military domain. *Ethics and Information Technology*, *26*, https://doi.org/10.1007/s10676-024-09762-w.
[178] Almada, M. (June 2019,). Human intervention in automated decision-making: Toward the construction of contestable systems. *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 2.
[179] The OECD AI Principles 2019 are now adhered to by 48 States and the European Union: OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#adherents. They call on actors to respect the rule of law, human rights, democratic and human-centred values throughout the AI system lifecycle, and to safeguard and ensure non-discrimination and equality, freedom, dignity, autonomy, privacy and data protection, diversity, fairness, social justice and international labour rights.
[180] Montasari, R., "Addressing Ethical, Legal, Technical and Operational Challenges in Counterterrorism with Machine Learning: Recommendations and Strategies" in in Montasari, R., *Cyberspace, Cyberterrorism and the International Security in the Fourth Industrial Revolution: Threats, Assessment and Responses* (Springer International Publishing, 2024), 199.
[181] UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2021, https://unesdoc.unesco.org/ark:/48223/pf0000380455.
[182] Hu, X., Neupane, B., Echaiz, L.F., Sibal, P. and Rivera Lam, M., *Steering AI and Advanced ICTs for knowledge societies: A Rights, Openness, Access and Multi-stakeholder Perspective* (UNESCO Publishing, 2019), 33-67.
[183] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.
[184] Council of the European Union, "The EU's Response to Terrorism, 2025", https://www.consilium.europa.eu/en/policies/fight-against-terrorism/.
[185] Europol, above note 35, 7-8.
[186] Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, 2024, https://rm.coe.int/1680afae3c.
[187] Global Digital Compact, A/79/L.2, paras. 50-63.
[188] See, e.g., discussion of AI human rights impact assessments in the EU: Submission to this Position Paper by the EU, 6.
[189] See further ibid, 9.
[190] See further the discussion of data protection concerns highlighted by the EU Commission: ibid, 8.
[191] See, e.g., the dialogue of the EU Commission concerning risk mitigation in AI systems development: ibid, 8.
[192] Confidential Submission 1 to this Position Paper by a regional organisation, 4.
[193] A/HRC/52/39, para. 40.
[194] OHCHR, "The right to privacy in the digital age", A/HRC/48/31, para. 59(c); A/80/169, para. 6.
[195] EU AI Act, chapter II, article 5.
[196] EU AI Act, chapter III, articles 8-17.
[197] See also A/HRC/RES/58/23, para. 6; A/HRC/52/39, para. 57(c); A/HRC/41/35, para. 49; and A/HRC/44/24, para. 40.
[198] See also Global Digital Compact, A/79/L.2, para. 32(a).
[199] See also A/HRC/52/39, para. 59.
[200] See further Submission to this Position Paper by the EU, 9.
[201] A/HRC/51/17, para. 56(c).
[202] A/80/341, paras. 102 and 109.
[203] A/HRC/16, 51, paras. 30-32; A/HRC/40/52, para. 37.
[204] A/80/341, paras. 73, 74, 109 and 110; van Ginkel, above note 79.
[205] A/80/169 (2025), paras. 64-71.
[206] Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism, "Best practices to protect human rights while using administrative measures", A/80/284, para. 6(f).
[207] Global Digital Compact, A/79/L.2, para. 39(c).
[208] A/HRC/48/31, para. 55.
[209] Global Digital Compact, A/79/L.2, para. 39(d).
[210] Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism, "Compilation of good practices on legal and institutional frameworks and measures that ensure respect for human rights by intelligence agencies", A/HRC/14/46, 2010; Global Counter-Terrorism Forum, "Brussels Memorandum on Good Practices for Oversight and Accountability in Mechanisms in Counterterrorism" 2024.
[211] See also A/HRC/52/39, para. 50.
[212] See also A/HRC/52/39, para. 58.
[213] "UN and Red Cross call for restrictions on autonomous weapon systems to protect humanity", UN News, 5 October 2024, https://news.un.org/en/story/2023/10/1141922.
[214] ICRC, "Submission to the United Nations Secretary-General", above note 122.