TAMSAAL
ISSN Print: 3007-5939
ISSN Online: 3007-5947
Volume 4, Number 7, 2026, Pages 505 – 517
Journal Home Page
https://tamsaal.pk/index.php/tamsaal/index

Responsibility to Protect 2.0: AI Surveillance, Human Rights Law, and State

Abdullah Shah[1], Shafiya Qadeer Memon[2], Dr. Sajid Sultan[3], Major Muhammad Irfan(R)[4], Dr. Surraya Jamal[5]

[1]*Department of Public Administration, Shah Abdul Latif University Khairpur, Sindh, Pakistan, Email: abdullah.shah@salu.edu.pk*
[2]*Assistant Professor, Department of Software Engineering, Mehran University of Enigneering and Technology Jamshoro Sindh Pakistan, Email: shafiya.memon@faculty.muet.edu.pk , ORCID: 0009-0007-3900-3259*
[3]*Lecturer in Law, BZU Sub Campus Vehari, Email: sajidsultan@bzu.edu.pk*
[4]*National Logistics Corporation, Email: muhammaderfan25@gmail.com*
[5]*Assistant Professor, Department of Management Sciences, Qurtuba University of Science and Information Technology Peshawar, Pakistan, Email: surayyajml@gmail.com*

ARTICLE INFO

Article History:

Received:	*June*	*19, 2026*
Revised:	*July*	*15, 2026*
Accepted:	*July*	*21, 2026*
Available Online:	*July*	*30, 2026*

Keywords:

Responsibility to Protect, artificial intelligence, mass atrocity prevention, surveillance, international human rights law, sovereignty, state intervention, algorithmic accountability

Corresponding Author:
Abdullah Shah
Email:
abdullah.shah@salu.edu.pk

DOI: *https://doi.org/10.59075/tamsaal.v4i7.160*

ABSTRACT

This study critically explored the connection between the Responsibility to Protect (R2P) and the use of artificial intelligence (AI) for surveillance, international human rights law, and intervention by state authorities by adopting a qualitative doctrinal legal research method, which involved analysis of international treaties, U.N. resolutions and reports, judicial rulings and scholarly works. The study tracked the origin and development of R2P from the World Summit Outcome document of 2005 to the three-pillar operationalisation of the doctrine, and it explored the use of AI surveillance technologies by states and international actors for early warning, atrocity prevention, intelligence collection and population surveillance, as well as the frequent violation of international human rights law, such as privacy, freedom of expression, non-discrimination and due process. Based on the documented use of social media surveillance in the atrocities being committed against the Rohingya population in Myanmar, the study demonstrated how AI-based technologies can be at the same time used as instruments to prevent atrocities and as conduits to the same atrocities. The study relied on doctrinal, comparative and critical legal analysis to highlight the four main gaps in current international legal rules and standards: the lack of explicit doctrinal guidance on AI surveillance in R2P; the non-binding nature of existing human rights guidance; weak accountability mechanisms for diffuse and privatized surveillance tools; and the absence of a framework that balances atrocity prevention with human rights protection. The study, in turn, suggested 'Responsibility to Protect 2.0,' a legally-based approach to atrocity prevention in which AI surveillance can be included, as long as it abides by the principles of legality, necessity, proportionality, accountability, transparency, human oversight, and remedy. The study introduces a doctrinally informed, rights-based approach that can inform the future of integrating AI surveillance technologies into international atrocity prevention efforts.

Introduction

The doctrine of the Responsibility to Protect (R2P) was born of the conviction that state sovereignty is not absolute protection against international concern regarding the well-being of populations subjected to genocide, war crimes, ethnic cleansing and crimes against humanity (International Commission on Intervention and State Sovereignty [ICISS], 2001). R2P was adopted by unanimous vote by world leaders at the World Summit in 2005 and stated that the primary responsibility for preventing these four atrocity crimes lies with every state in the world, while the international community has a residual responsibility to help and intervene where necessary when a state is manifestly failing in its duty to protect its people (UN General Assembly, 2005).

Since this adoption, R2P has been implemented via a three-pillar approach: the state's responsibility to prevent, to assist, to build capacity, and timely and timely collective response (UN Secretary-General, 2009). The ability to identify early warning signs of impending atrocities is central to the effective functioning of all three pillars, and has traditionally been reliant on diplomatic reporting, human rights monitoring and on-the-ground fact-finding (UN Secretary-General, 2025).

Artificial intelligence (AI)-based surveillance technologies have revolutionized this landscape by providing states and intergovernmental actors with unprecedented capabilities to identify trends of mass violence, hate speech and population displacement in near real-time (Feldstein, 2019). Meanwhile, these same technologies are being used by governments to monitor vast populations, to predict how to stop crime, and to squelch dissent, and thus there is a risk that AI surveillance could be a tool to prevent atrocities and a trigger for the atrocities R2P seeks to prevent (Zwitter, 2014).

This dual-use aspect was well demonstrated in Myanmar, where the UN Independent Independent Fact-Finding Mission reported that social media, and Facebook in particular, also helped spread a hate speech campaign that contributed to violence against the Rohingya population; and that these platforms, and related data, in principle, helped to inform earlier warnings of the growing danger (UN Human Rights Council, 2018). The Gambia v. Myanmar case before the International Court of Justice (ICJ) highlighted the importance of the obligations of a state under the Genocide Convention in just this context (International Court of Justice, 2020).

AI surveillance for atrocity prevention and intervention, therefore, raises a host of questions under the international human rights law, including the rights of privacy, freedom of expression, non-discrimination and due process, which are guaranteed by binding international instruments (International Covenant on Civil and Political Rights [ICCPR], 1966; Universal Declaration of Human Rights [UDHR], 1948). The right to privacy extends to the digital age: The UN Human Rights Council has repeatedly declared that (UN Human Rights Council, 2017, 2019).

Despite this expanding body of ‘soft law' guidance, the existing R2P framework, which was developed before the widespread use of AI-enabled surveillance, is silent on how AI-enabled surveillance tools can be incorporated into R2P efforts, and on the legal protections that must be

included to prevent their misuse by states (McGregor, Murray & Ng 2019). In this way, it is a huge gap in the international legal architecture as AI surveillance is being used by states and international actors in its use and abuse that it is meant to address.

The following gap informed the present study, which critically analysed the relationship between R2P, AI-enabled surveillance, international human rights law and state intervention, in order to propose a legally-based framework, called 'Responsibility to Protect 2.0', for supporting the responsible use of AI surveillance in atrocity prevention while remaining firmly grounded in existing international human rights law. The study sought to explore the legal underpinnings of R2P, their links to relevant international human rights standards, and assess the legal and ethical implications of state and international uses of AI surveillance techniques for early warning and population monitoring; and to identify a set of legal principles and safeguards that balance the protective value of AI surveillance with the rights it poses risks to, building on the doctrinal literature on algorithmic accountability under international human rights law (McGregor, Murray, & Ng, 2019) and the evolving normative robustness of R2P itself (Welsh, 2019).

Literature Review

The Responsibility to Protect has its origins in the work of the International Commission on Intervention and State Sovereignty, and the failure of the international community to act in Rwanda when the genocide was unfolding and for a long time after the Srebrenica massacre (ICISS, 2001). Also during this timeframe, Feinstein and Slaughter (2004) made similar claims, that the international community had a 'duty to prevent' in their view, an illegitimate regime to acquire weapons of mass destruction and to commit mass atrocities, suggesting yet another shift in thinking: away from the concept of sovereignty and toward populations as being 'protected'. The international community, in the form of heads of state and government at the 2005 World Summit, confirmed the formal endorsement of these conceptual foundations, by declaring that each individual state has a responsibility to protect its population from genocide, war crimes, ethnic cleansing, and crimes against humanity (UN General Assembly, 2005). This consensus was documented by Bellamy (2009) to be the result of political manoeuvres, in which the Summit's final formulation was significantly less broad than the ICISS's original proposal, so as to avoid the implications of the doctrine for state sovereignty.

R2P has been operationalized in a three-pillar approach that is outlined in the report of the Secretary-General in 2009: first, the responsibility of the state to protect (Pillar I); second, international assistance and capacity-building (Pillar II); and third, timely and decisive collective action through the United Nations (Pillar III) (UN Secretary-General, 2009). UN Security Council resolution relating to the protection of civilians in armed conflict confirmed the doctrine by adding strength to it, formally endorsing the principles stated in paragraphs 138 and 139 of the World Summit Outcome (UN Security Council, 2006). Following this, the General Assembly passed a resolution to proceed with its deliberations on R2P, demonstrating its ongoing but limited institutionalisation (UN General Assembly, 2009). On that point, Hakimi (2014) has warned that R2P is not a legally binding obligation as such, but rather is a political commitment and a framework for structured deliberation, because the three response mechanisms outlined in the R2P are not legally binding as such, but can be constrained by the existing mechanisms of the UN Charter, such as the Security Council's discretion under Chapter VII.

International Human Rights law pertaining to surveillance.

International human rights law provides a range of rights that are directly affected by state surveillance. Article 12 of the Universal Declaration of Human Rights (1948) and Article 17 of the International Covenant on Civil and Political Rights (1966) protect the right to privacy, guaranteeing freedom from arbitrary or unlawful interference in an individual's privacy, family, home or correspondence. The General Comment No. 16 of the UN Human Rights Committee (1988) has also been directly relevant to the use of AI-enabled surveillance systems, which are gathering and processing enormous amounts of personal data, by requiring states to take effective measures to prevent the unlawful collection, retention, and use of personal data by public authorities. The ICCPR also ensures freedom of expression in Article 19 and bans discrimination in Article 26, all of which are relevant to the monitoring, profiling, and limitations on the exercise of these rights by AI systems.The ICCPR also protects freedom of expression under Article 19, and prevents discrimination under Article 26, which is applicable to the monitoring, profiling and limitations on the exercise of these rights by AI systems (ICCPR, 1966). The UN Human Rights Council has established that these protections apply, just as strongly, online as well as offline, and has urged States to ensure that surveillance is carried out in compliance with the principles of legality, necessity and proportionality (UN Human Rights Council, 2017).

Artificial Intelligence (AI) in Atrocity Prevention

States and international organizations have increasingly used AI surveillance tools for border control, public safety and security, and more recently for atrocity risk monitoring and early warning, such as through the use of facial recognition, satellite imagery analysis, and natural language processing of social media posts (Feldstein, 2019). However, as noted by Zwiter (2014), the analytical power of big data and AI technologies has a real potential to enhance EWS, a central theme of R2P's preventive pillar, and facilitate the identification of trends, such as the spread of hate speech or population displacement or the strengthening of military mobilisation, that might otherwise remain undetected by the international community until it is too late. The capacity to promote data collection for hate speech analysis and action can be both powerful and dangerous, as in the case of Myanmar, where platforms used to spread hate towards the Rohingya also provided data that the UN Fact-Finding Mission believes would have helped lead to a more timely international response, had it been analysed and acted upon at an earlier stage (UN Human Rights Council, 2018). More generally, Amoore (2020) has warned that the algorithmic systems being deployed for security and surveillance often conceal their own algorithmic decision-making processes under the guise of "technical neutrality," which makes it difficult to prosecute their state and corporate users.

The Legal and Ethical Implications of AI Surveillance

International human rights law presents significant legal and ethical challenges associated with the use of AI surveillance technologies by states. In 2019, McGregor, Murray and Ng highlighted algorithmic bias and opacity and the lack of effective mechanisms for redress as key issues in reconciling this human right approach with machine-driven decision-making processes, noting that such processes are often opaque and operate in ways that are hard to challenge with the tools of traditional law designed to challenge human decision-makers. The UN Special Rapporteur on freedom of opinion and expression also noted that the largely unregulated private

surveillance industry has enabled the export of sophisticated surveillance technologies to states where human rights are poorly respected, with poorly regulated or non-existent transparency and accountability over the use of such technologies and their impact on the rights of individuals (UN Human Rights Council, 2019). In a separate report, the Special Rapporteur on freedom of opinion and expression highlighted the broader implications of AI technologies for the information environment, and warned that without adequate human rights protection AI-based technologies for content moderation and monitoring could be used as tools to discriminate against individuals and groups and to censor their content (UN General Assembly, 2018). The Special Rapporteur on the right to privacy also noted that the potential of AI to infringe on privacy rights, particularly in the areas of automated profiling, biometric identification and predictive analytics, and called for binding legal protections before their implementation by states (UN Human Rights Council, 2021).

AI surveillance in the R2P framework also poses more basic issues regarding the new dynamics between sovereignty, intervention and technology's capacity. R2P was designed, however, not to replace the concept of sovereignty but to complement international responsibility, as AI surveillance technologies might threaten that balance by allowing for some types of international monitoring, and potentially intervention-adjacent action, without the need for traditional "peacekeeping" or military intervention. In a study that is directly relevant to whether R2P can productively include AI-powered instruments without compromising its heart and soul of human rights protections, Welsh (2019) investigated the issue of norm robustness in relation to R2P. The case of the Bosnian Genocide case of the International Court of Justice, which held that States have a due diligence obligation to take action based on information showing a serious risk of genocide, is relevant here too, and, arguably, extends to information that is produced or made available by monitoring systems that utilise artificial intelligence (International Court of Justice, 2007).

To Algorithmic Accountability: Existing Framework Gaps

Several researchers have pointed to certain deficiencies in the current global legal landscape relating to AI surveillance and state interference. As discussed in an edited volume on new technologies and human rights practice (Land and Aronson 2018), international human rights law was created in many instances before the advent of AI and big-data technologies, resulting in significant gaps in interpretation as to how existing human rights frameworks should be applied to algorithmic decision-making, automated surveillance, and population monitoring. McGregor, Murray, and Ng (2019) also pointed out that current human rights accountability mechanisms, which focus on specific decisions and discernable state actors, are ill-equipped to handle AI surveillance systems that operate in a diffuse and opaque manner, with responsibility spread across state agencies, private technology firms, and international institutions. However, Hakimi (2014) pointed out that R2P does not have a well-established framework of enforcement or accountability even in its classical version, and this may be further complicated, not resolved, by the adoption of technologically advanced surveillance technologies without the parallel development of doctrine-enforcing legal and institutional safeguards.

Theoretical Framework

This study draws on the theoretical concept of sovereignty as responsibility which forms the basis of R2P (ICISS, 2001) in conjunction with the due diligence framework laid out by the

International Court of Justice when considering the prevention of genocide (International Court of Justice, 2007). In this integrated approach, the states and international community are said to have a dynamic due diligence responsibility to act in a reasonable manner and take ‘every possible measure' to prevent atrocity crimes, including through the use of AI-powered early warning systems, and are subject to the body of international human rights law on matters such as surveillance, privacy, expression and non-discrimination. This theoretical justification helps to support the central hypothesis of the study: the R2P's prevention pillar can and should be interpreted as including responsible capacity in the surveillance of AI, which should be subject to the same legality, necessity, proportionality, and accountability requirements that are applied to other uses of state and international power in international human rights law.

Research Questions

The above doctrinal and theoretical review guided the study, which was guided by the following research questions:

RQ1: What are the legal bases of R2P and how do they relate to international human rights law pertaining to surveillance?

RQ2: What are the legal and ethical challenges of using AI surveillance systems in atrocity prevention, early warning and population monitoring?

RQ3: Will the current international legal framework sufficiently cover opportunities and risks of AI surveillance in the R2P architecture, and if not, what legal principles and protection are required to be included to fill the gaps?

Methodology

Research Approach

The methodology used in this research is doctrinal legal research with a qualitative approach. This study used a qualitative doctrinal legal research methodology to critically analyze the relationship between the Responsibility to Protect (R2P), Artificial Intelligence (AI) surveillance, international human rights law and the intervention of the state. Since the study involved interpretation and critical analysis of current legal rules, principles and instruments, without any empirical testing of hypotheses, doctrinal legal research was deemed the proper methodology.

Sources of Data

The research mainly relied on secondary legal sources such as international treaties and conventions, United Nations General Assembly and Security Council resolutions and reports, resolutions of the Human Rights Council, reports of the Special Rapporteurs, International Court of Justice judicial decisions, principles of customary international law, books, peer-reviewed journal articles and various policy documents. Special focus on the legal basis of R2P as enshrined in the 2005 World Summit Outcome, international human rights norms regarding privacy, freedom of expression, non-discrimination, due process and documented cases where AI and social media surveillance technologies were linked to the risk of atrocities, including in Myanmar.

Doctrinal Legal Analysis

A doctrinal analysis was undertaken to aid in the interpretation of the existing legal rules and principles about R2P and international human rights protection regarding surveillance, such as locating the particular treaty provisions, resolutions, and judicial decisions that address states' obligations with respect to atrocity prevention, privacy, and non-discrimination. The analysis covered the potential uses, or uses, of AI surveillance technologies by states and international organisations for early warning, atrocity prevention, intelligence gathering, and population monitoring, as well as the legal and ethical dangers of algorithmic bias, mass surveillance, privacy infringement and abuse by states' surveillance bodies.

Comparative Analysis

International legal, institutional and selected state practices and approaches were examined to assess how the use of technology was challenging the traditional concept of responsibility, sovereignty and intervention. A comparative approach was used to show the opportunities for prevention and the potential for human rights violations in the context of AI surveillance from documented state and international experiences and reports by the UN Special Rapporteur.

Using a critical legal analysis, gaps in the current international legal framework in relation to AI-based surveillance and state intervention were identified, such as with respect to accountability, transparency and remedy for violations of human rights related to AI. AI surveillance generated opportunities and risks that were examined through a systematic review of relevant legal developments and policy debates to assess whether the current R2R framework was sufficient to meet these needs. The findings were then combined to formulate a concept of ‘Responsibility to Protect 2.0' that introduces a legally-based concept in which AI technologies may be used to assist atrocity prevention and early warning, and which can be placed under the umbrella of international human rights law, legality, necessity, proportionality, accountability, transparency, and meaningful safeguards against misuse.

Scope and Limitations

This study only examined publicly accessible legal and scholarly resources in English, and was not designed to conduct an original empirical evaluation of specific AI surveillance systems that are being used by a specific state, as much of this information is classified and/or proprietary. While the analysis was not designed to be comprehensive about the domestic legal frameworks of all jurisdictions, it did encompass the international legal framework applicable to AI surveillance and R2P generally, to which illustrative reference is made where relevant.

Ethical Considerations

The study was only conducted on publicly accessible legal and scholarly literature, and did not include any human subjects, therefore, there was no need for empirical data collection or human-subject ethical approval. Academic referencing for all secondary sources used was done appropriately.

Legal Analysis and Findings

R2P's Three Pillars and the Role of AI Surveillance

The mapping of AI surveillance abilities to the three pillars of R2P sheds new light on the potential of AI in the prevention of mass atrocities and its susceptibility to mis-use. As seen in Table 1, different AI surveillance uses are associated with the different human rights risks involved with each of the three pillars of R2P. While the state is responsible to protect its citizens, it may use AI surveillance domestically to identify when communities are in danger of communal violence, displacement or hate speech, and simultaneously use the same tools to surveil and suppress minority populations or political dissidents by the State that is supposed to be protecting them. International assistance and capacity building, under Pillar II, can include the support of international actors to monitor at risk populations and to build the evidence base for early warning using AI application of open-source and satellite data, but this can raise questions regarding human rights compliance of the data collection methods and the responsibilities of international actors under human rights law. In Pillar III, the reliability, provenance and potential bias of this evidence and analysis generated by AI can raise questions about its reliability, and it can be used to inform the deliberation of the Security Council and the international community, but only after careful legal scrutiny for its legitimacy to support collective action of a coercive nature.

Table 1. R2P Pillars, Illustrative AI Surveillance Applications, and Associated Human Rights Risks

R2P Pillar	Illustrative AI Surveillance Application	Associated Human Rights Risk
Pillar I: State responsibility	Domestic early-warning monitoring of hate speech, displacement, and mobilization	Risk of repurposing for surveillance or repression of minorities and dissidents
Pillar II: International assistance	Satellite imagery and open-source/social media analysis to support at-risk population monitoring	Privacy and data protection concerns; questions of consent and data provenance
Pillar III: Timely and decisive response	AI-generated analysis informing Security Council and international body deliberation	Algorithmic bias, opacity, and reliability of evidence underlying coercive action

Human Rights Law Engaged by AI Surveillance

AI surveillance in each of these pillars involves a different range of international human rights protections as outlined in Table 2. The right to privacy is invoked in all cases in which AISs gather, store or analyse personal information, such as communications and biometric information (ICCPR, 1966; UDHR, 1948). AI surveillance may be used to monitor, flag or limit online speech, which is action that engages the right to freedom of expression, especially where there is a “chilling effect”, that is, on legitimate speech (UN General Assembly, 2018). The right to non-discrimination is implicated when algorithmic bias in AI surveillance systems unfairly targets or misidentifies people based on their race, religion or other protected attributes (ICCPR 1966). Wherever AI-based evaluations come into play to inform and drive decisions that impact

negatively on individuals and groups, meaningful processes for explanation, challenge and remedy are involved (McGregor, Murray, & Ng, 2019).

Table 2. International Human Rights Instruments Engaged by AI Surveillance

Human Right	Governing Instrument(s)	Relevance to AI Surveillance
Privacy	UDHR Art. 12; ICCPR Art. 17; HRC Res. 34/7 (2017); HRC Res. 42/15 (2019)	Collection, retention, and processing of personal, biometric, and communications data
Freedom of expression	UDHR Art. 19; ICCPR Art. 19	Monitoring, flagging, or restricting online speech; chilling effects on expression
Non-discrimination	ICCPR Art. 26	Algorithmic bias disproportionately affecting protected groups
Due process	ICCPR Art. 14 (by analogy); general principles of accountability	Absence of explanation, challenge, or remedy for AI-informed adverse decisions

Case Illustration: Myanmar and the Limits of Existing Practice

Myanmar is an example of the doctrinal conflicts that lie at the core of this study. In the context of the climate of violence against the Rohingya population, the UN Independent International Fact-Finding Mission observed that Facebook was "a powerful tool for spreading hate speech" and "slow and ineffective" in combating it and that Facebook failed to provide the Mission with sufficient data to measure the extent of the problem (UN Human Rights Council, 2018). The international law gravity of the situation under the Genocide Convention (International Court of Justice, 2020) was further demonstrated by the provisional measures order issued by the International Court of Justice in The Gambia v. Myanmar. This case exposes both aspects of the dilemma of AI surveillance as central to this study: on the one hand, in principle, social-media data could be utilized to help detect the escalation of atrocity risk, but, on the other hand, in its observable manifestations, it can be used to hasten the escalation of atrocity risk without sufficient human rights protection or transparency and accountability mechanisms, both on the platform and at the state level.

Existing Legal Framework Identified Gaps

In light of the above analysis, the current study called out four main gaps in the current legal framework of R2P and the application of AI surveillance. Second, R2P's key documents, created before the proliferation of AI technologies, fail to explicitly mention AI surveillance and provide no guidance on what is allowed in the doctrine's preventive and/or responsive pillars. Second, existing international human rights law is in principle applicable to algorithmic and automated surveillance, but has not been carefully examined or formally adopted into binding international law in the context of AI and algorithmic and automated surveillance. Second, existing international human rights law is in principle relevant to AI surveillance, but has not been studied and formally incorporated into binding international law in the specific context of algorithmic and automated surveillance. Third, current accountability systems are largely unsuitable for the fragmented and often private nature of AI surveillance, which rests with a variety of state bodies, private tech firms, and even international bodies in some instances (McGregor, Murray, & Ng, 2019). Most importantly for this study, no conceptual framework

explicitly accounts for the preventive potential of AI surveillance as part of an early-warning function of R2P while simultaneously examining the human rights obligations that limit its use, thus creating a gap that this study aimed to fill.

Toward Responsibility to Protect 2.0: A Proposed Framework

This study suggests that AI surveillance technologies can contribute to atrocity prevention and early warning, in line with R2P, while respecting a number of core legal principles and safeguards that are outlined in Table 3 below. These principles are based on existing international human rights legal standards on permissible restrictions on rights including privacy and expression, applied to the particular technical and institutional aspects of AI surveillance systems.

Table 3. Proposed Responsibility to Protect 2.0 Principles and Safeguards

Principle	Core Requirement	Legal Basis
Legality	AI surveillance for atrocity prevention must be authorized by clear, publicly accessible legal rules	ICCPR Arts. 17, 19; HRC Res. 34/7 (2017)
Necessity	Use must be demonstrably necessary to a legitimate atrocity-prevention purpose, not merely convenient	ICCPR Art. 17; HRC Res. A/HRC/41/35 (2019)
Proportionality	The scale and intrusiveness of surveillance must be proportionate tothe atrocity risk being addressed	Human Rights Committee General Comment No. 16 (1988); ICCPR Art. 17
Accountability	Clear attribution of responsibility across state, international, and private actors deploying AI surveillance	McGregor, Murray, & Ng (2019)
Transparency	Public disclosure of the existence, purpose, and safeguards of AI surveillance systems used for atrocity prevention	A/HRC/46/37 (2021); A/73/348 (2018)
Human oversight	Meaningful human review of AI-generated assessments before they inform prevention or response decisions	Due diligence standard, ICJ Bosnian Genocide case (2007)
Remedy	Accessible mechanisms for individuals and groups to challenge and seek redress for AI surveillance-related rights violations	ICCPR Art. 2(3); McGregor, Murray, & Ng (2019)

Discussion

The results of this doctrine analysis solve both of the research questions in this study. As for RQ1, this study has revealed significant overlaps between the legal basis of R2P and the field of international human rights law in respect of the 'due diligence' obligations of States in preventing atrocity crimes, and the rights based restrictions on the means available to fulfil these obligations (UN General Assembly, 2005; International Court of Justice, 2007). Regarding RQ2, this study identified significant legal and ethical challenges to the use of AI surveillance technologies in international human rights law, namely issues of privacy, freedom of expression, non-discrimination, and due process, which are exacerbated where such technologies are

implemented in states with poor human rights records or by opaque public-private surveillance mechanisms (UN Human Rights Council, 2019; McGregor, Murray, & Ng, 2019). Regarding RQ3, this study revealed that the current R2P framework fails to adequately respond to the opportunities and risks brought about by AI surveillance and determined four specific gaps that the proposed Responsibility to Protect 2.0 framework aims to address.

The Myanmar experience is illustrative of the importance of this gap in practice: the lack of clear legal requirements for the platforms and the state on surveillance and content moderation settings created the context where technology which could have helped with early warning instead became a contributor to atrocity risk (UN Human Rights Council 2018). The R2P 2.0 framework proposed in this study builds on the established human rights principles of legality, necessity and proportionality, and also incorporates the newer requirements of accountability, transparency, human oversight, and remedy in the context of algorithmic systems, which must be elaborated and elaborated further through state practice, judicial interpretation and international institutional development (Welsh 2019; Land & Aronson 2018).

Conclusion

This study aimed at critically analysing the relationship between the Responsibility to Protect, use of AI technology in surveillance, international human rights law and state intervention through the qualitative doctrinal legal research methodology based on the analysis of international treaties, resolutions, judicial decisions and academic literature. The study revealed that while R2P does not establish a new, standalone legal duty to intervene, there is a strong reinforcement from states' due diligence obligations under international human rights and genocide-prevention law and that AI surveillance technologies also have substantive potential, albeit currently underdeveloped, to enhance the early-warning function at the core of the R2P's preventive pillar. In parallel, the study revealed significant threats to privacy, freedom of expression, non-discrimination and due process and concrete examples of the role that social media surveillance and dissemination of content had played in the Myanmar atrocities. This study, after identifying specific gaps in the current legal framework, proposed ‘Responsibility to Protect 2.0’, a doctrinally based framework where AI surveillance could be legitimately part of atrocity prevention, subject to the principles of legality, necessity, proportionality, accountability, transparency, human oversight and remedy. It is a contribution to the R2P and international human rights law literature because it directly tackles a gap left open by the doctrine's original architects, and provides a basis for further legal and institutional development in an era of developing AI surveillance technology.

Recommendations

The findings of this study should further encourage the United Nations and relevant international organizations to seriously consider, through new guidance from the Office of the Special Adviser on the Prevention of Genocide and the Special Adviser on the Responsibility to Protect, the potential for integrating AI surveillance technologies into R2P's early-warning architecture, in accordance with the principles of legality, necessity, proportionality, accountability, transparency, human oversight and remedy outlined in the proposed Responsibility to Protect 2.0 framework. States using AI surveillance technologies, for domestic or international security or terrorism prevention, should implement domestic laws that mandate human rights impact assessments, transparency about the use and purpose of AI surveillance systems, and redress to

individuals and groups affected by such systems, in line with the recommendations of the Special Rapporteur (UN Human Rights Council, 2019, 2021). The platforms and tools of technology companies that could be used to commit atrocities should be subject to clear and firmly established human rights due process obligations, such as clear and transparent data-sharing with relevant international Fact Finding and Early Warning mechanisms where appropriate measures can be taken to ensure those platforms and tools are not used in a way that is likely to cause harm or ease the commission of atrocities, as in the Myanmar case. This study was necessarily based on publicly available legal texts and scholarly works and did not conduct empirical legal analyses of particular AI surveillance systems, and future research should aim to further test and refine the doctrinal framework proposed here with regard to the current practice of operational AI surveillance systems in states and international bodies in contexts where they are prone to be used to commit atrocities.

References

1. Amoore, L. (2020). Cloud ethics: Algorithms and the attributes of ourselves and others. Duke University Press.
2. Bellamy, A. J. (2009). Responsibility to protect: The global effort to end mass atrocities. Polity Press.
3. Convention on the Prevention and Punishment of the Crime of Genocide, Dec. 9, 1948, 78 U.N.T.S. 277.
4. Feinstein, L., & Slaughter, A.-M. (2004). A duty to prevent. Foreign Affairs, 83(1), 136–150.
5. Feldstein, S. (2019). The global expansion of AI surveillance. Carnegie Endowment for International Peace.
6. Hakimi, M. (2014). Toward a legal theory on the responsibility to protect. Yale Journal of International Law, 39(2), 247–282.
7. International Commission on Intervention and State Sovereignty. (2001). The responsibility to protect: Report of the International Commission on Intervention and State Sovereignty. International Development Research Centre.
8. International Court of Justice. (2007). Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro), Judgment, I.C.J. Reports 2007, p. 43.
9. International Court of Justice. (2020). Application of the Convention on the Prevention and Punishment of the Crime of Genocide (The Gambia v. Myanmar), Provisional Measures, Order of 23 January 2020, I.C.J. Reports 2020, p. 3.
10. International Covenant on Civil and Political Rights, Dec. 16, 1966, 999 U.N.T.S. 171.
11. Kaye, D. (2019). Speech police: The global struggle to govern the internet. Columbia Global Reports.
12. Land, M. K., & Aronson, J. D. (Eds.). (2018). New technologies for human rights law and practice. Cambridge University Press.
13. McGregor, L., Murray, D., & Ng, V. (2019). International human rights law as a framework for algorithmic accountability. International and Comparative Law Quarterly, 68(2), 309–343.
14. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), 2016 O.J. (L 119) 1.
15. Rome Statute of the International Criminal Court, Jul. 17, 1998, 2187 U.N.T.S. 3.

16. UN General Assembly. (2005). 2005 World Summit Outcome (A/RES/60/1).
17. UN General Assembly. (2009). The responsibility to protect (A/RES/63/308).
18. UN General Assembly. (2018). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression: Artificial intelligence technologies and implications for freedom of expression and the information environment (A/73/348).
19. UN Human Rights Committee. (1988). General comment No. 16: Article 17 (Right to privacy).
20. UN Human Rights Council. (2017). The right to privacy in the digital age (A/HRC/RES/34/7).
21. UN Human Rights Council. (2018). Report of the Independent International Fact-Finding Mission on Myanmar (A/HRC/39/64).
22. UN Human Rights Council. (2019). The right to privacy in the digital age (A/HRC/RES/42/15).
23. UN Human Rights Council. (2019). Surveillance and human rights: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/HRC/41/35).
24. UN Human Rights Council. (2021). Report of the Special Rapporteur on the right to privacy: Artificial intelligence and privacy, and children's privacy (A/HRC/46/37).
25. UN Secretary-General. (2009). Implementing the responsibility to protect: Report of the Secretary-General (A/63/677).
26. UN Secretary-General. (2025). The responsibility to protect: 20 years of commitment to principled and collective action: Report of the Secretary-General (A/79/875-S/2025/248).
27. UN Security Council. (2006). Resolution 1674 (2006) on the protection of civilians in armed conflict (S/RES/1674).
28. Universal Declaration of Human Rights, G.A. Res. 217 (III) A, U.N. Doc. A/810 (Dec. 10, 1948).
29. Welsh, J. M. (2019). Norm robustness and the responsibility to protect. Journal of Global Security Studies, 4(1), 53–72.
30. Zwitter, A. (2014). Big data and international relations. Ethics & International Affairs, 28(4), 377–389.