SECURITY AND HUMAN RIGHTS MONITOR

Preventive Counterterrorism in the Age of AI: Strategic Effectiveness and Human Rights Risks

Jeonghye Cho

Police researcher, Public Safey Research Office, Public Safety Research Department, Police Science Institute, Korean National Police University

cokcok@police.go.kr

DOI: 10.58866/YARN2489

Abstract

This paper examines how AI reshapes the balance between national security and human rights in counterterrorism. AI-driven strategies rely on mass data collection and surveillance, creating tensions with privacy, due process, and non-discrimination. Using OHCHR principles - legality, necessity, and proportionality - it analyzes the EU AI Act and OSCE frameworks. It identifies three key risks: privacy violations through surveillance, discriminatory targeting from biased algorithms, and reduced due process due to limited explainability. The EU framework also contains gaps via national security exceptions. The paper argues that human rights protection is essential, requiring clear legal bases, transparency, accountability, and independent oversight.

Keywords

Counterterrorism - Human Rights - Artificial Intelligence

I. Introduction: Changes in the Environment of National Counterterrorism Strategies in the Age of AI

Rapid advances in digital technologies have profoundly transformed contemporary terrorism. Terrorist organizations are increasingly exploiting artificial intelligence (AI), encrypted communications, social media platforms, and other emerging technologies to facilitate recruitment, propaganda, financing, operational planning, and attacks. In response, governments have adopted AI-driven surveillance, biometric identification, and large-scale data analytics to strengthen their counterterrorism efforts. While these technologies improve operational effectiveness, they also raise complex legal and ethical questions concerning privacy, accountability, and the protection of fundamental human rights.[1]

AI-enabled systems use algorithmic classification and predictive analysis to identify potential threats before attacks occur. This makes preventive intervention a central feature of modern counterterrorism. Rather than independently determining what constitutes terrorism, AI systems operationalize legal and policy definitions through risk assessment and algorithmic classification. Therefore, definitional choices directly influence how AI systems identify potential threats, prioritize surveillance, and support preventive counterterrorism measures. These definitional choices have significant implications for strategic effectiveness and the protection of fundamental rights.

Although terrorism is generally understood to be violence intended to intimidate or coerce that is politically or ideologically motivated, the international community has never reached a universally accepted definition.[2] Various attempts to define terrorism have one thing in common: terrorism is defined as an act of violence that is planned and carried out to achieve political objectives.[3] The primary strategy is to instill fear in order to convey a message to the public.[4]

In the 1994 Declaration on Measures to Eliminate International Terrorism, the General Assembly distinguished terrorism from other forms of violence due to its motivation 'for political purposes'.[5] The United Nations General Assembly Resolution A/RES/49/60 (1994) and Security Council Resolutions 1373 (2001) and 1566 (2004) identified three core

1 United Nations General Assembly, 48/31, The right to privacy in the digital age, 2021, https://digitallibrary.un.org/record/3946475?v=pdf.

2 A.P. Schmid, Defining Terrorism, ICCT Report, International Center for Counter-Terrorism, 2023, p. 2.

3 P. Waldmann, Terrorismus: Provokation der Macht, 1998, pp. 10 and 13.

4 P. Waldmann, Terrorismus: Provokation der Macht, 1998, p. 12.

5 B. Saul, Defining 'Terrorism' to Protect Human Rights, Sydney Law School Legal Studies Research Paper No.09/125, 2008, p. 4.

elements of terrorism: organized violence targeting civilians, political or ideological motives, and the intent to instill fear.[6] While these provisions establish a minimum common standard, the concept's breadth and ambiguity leave room for interpretation. This ambiguity carries the potential for political abuse.

In the context of counterterrorism, striking a balance between national security and the protection of human rights is a central issue in security policy research. States have an active duty to protect their citizens from terrorist threats while also having legal and normative obligations to respect and guarantee human rights. Although countries have expanded their legal authorities and surveillance capabilities to bolster counterterrorism capabilities, this has sparked ongoing debates about potential human rights violations.

A broadly defined concept of terrorism creates structural conditions that allow states to expand restrictions on human rights while justifying counterterrorism measures. In response, the Office of the United Nations High Commissioner for Human Rights (OHCHR) views effective counterterrorism policies as complementary to human rights protection rather than conflicting with it.[7] Furthermore, both the Council of Europe and the Organization for Security and Co-operation in Europe (OSCE) emphasize that counterterrorism policies must be based on the rule of law and the protection of fundamental rights.[8]

The current terrorist landscape is undergoing structural changes alongside the digital transformation. In particular, the online space plays a crucial role in radicalization and mobilization. For example, in Malaysia, approximately 17 percent of ISIL's recruitment was conducted through social media, while in Singapore, the majority of terrorism-related detainees are reportedly radicalized by online propaganda.[9]These trends demonstrate that individuals can embrace extremist ideologies and act on them without physical contact. Radicalized users can leverage social media to disseminate their ideologies, forge alliances, and orchestrate and publicize heinous acts.[10]

At the same time, the nature of terrorist acts is changing. In recent years, Western countries have experienced an increase in attacks carried out by lone terrorists operating independently of an organized command structure.[11] The significant increase in attacks by lone terrorists over a short period suggests that the terrorist threat is becoming increasingly decentralized and unpredictable.[12] This demonstrates that traditional, organization-centric surveillance and response strategies alone are insufficient for effectively responding to the evolving terrorist threat or securing

6 United Nations General Assembly, 49/60, Measures to Eliminate International Terrorism, 9 December 1994. Retrieved 1 February 2026, https://digitallibrary.un.org/record/189017; United Nations Security Council, Resolution 1373, 2001. Retrieved 1 February 2026, digitallibrary.un.org/record/449020?v=pdf; United Nations Security Council, Resolution 1566, 2004. Retrieved 1 February 2026, digitallibrary.un.org/record/532676?v=pdf.

7 United Nations General Assembly, Report of the United Nations High Commissioner for Human Rights on the protection of human rights and fundamental freedoms while countering terrorism, A/HRC/4/88, 2007. Retrieved 6 July 2026, https://digitallibrary.un.org/record/595012?v=pdf.

8 Council of Europe, Fight against terrorism. Retrieved 28 January 2026, coe.int/en/web/commissioner/thematic-work/fight-against-terrorism; ODIHR, Counter Terrorism, Protecting Human Rights, The Warsaw, 2008. Retrieved 28 January 2026, odihr.osce.org/sites/default/files/f/documents/d/6/29103.pdf.

9 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia , New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

10 A.Shaw, Social Media, extremis, and radicalization, 2023. Retrieved 9 July 2026, https://www.science.org/doi/10.1126/sciadv.adk2031.

11 Institute for Economics & Peace, Global Terrorism Index 2025: Measuring the Impact of Terrorim, Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf.

12 Institute for Economics & Peace, Global Terrorism Index 2025: Measuring the Impact of Terrorim, Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf.

sufficient deterrence.

Misinformation and conspiracy theories spread rapidly through online platforms, blurring the line between mainstream and extremist discourse. Despite efforts to reduce exposure to extremist content, social media platforms continue to provide an environment conducive to the promotion of violence, hatred, and conspiracy theories.[13] This trend is amplified during times of social crisis. In some cases, terrorists and violent extremists have exploited disinformation and manipulated content to expand their support base and intensify recruitment efforts.[14]

These changes require a complete restructuring of counterterrorism strategies. While counterterrorism has long aimed to prevent and combat terrorism,[15] contemporary strategies are shifting toward an increasingly preventive approach.[16] However, this strategic shift highlights the limitations of existing human-centered analytical systems. The volume and velocity of data generated in the digital environment exceeds the capabilities of traditional analysis, creating structural constraints that prevent intelligence and law enforcement agencies from identifying threats in a timely manner.[17]

In this context, AI is emerging as a core technology in national counterterrorism strategies. AI can compensate for the cognitive and temporal limitations of human-centered analysis by processing large-scale data in real time and detecting potential risk signals. In fact, the Global Terrorism Index (GTI) suggests that terrorist organizations and counterterrorism agencies will likely utilize AI in the future.[18]

However, AI-based counterterrorism strategies require extensive information gathering and continuous surveillance. This creates structural tensions with fundamental human rights, such as the rights to privacy and due process and the prohibition of discrimination. Methods that predict and intervene in an individuals' risk before a crime occurs, in particular, carry the potential to conflict with the presumption of innocence and the principles of the rule of law. Thus, while AI enhances the efficiency of counterterrorism strategies, it also necessitates reexamining their legitimacy. These concerns are important in the context of AI-enabled counterterrorism because expanded surveillance capabilities can increase both operational effectiveness and the risk of violating fundamental rights.

These challenges have significant implications for countries developing AI governance and counterterrorism frameworks. South Korea, for example, is expanding its digital governance capabilities and exploring the application of AI technologies in public security. However, this effort is accompanied by the broader institutional challenge of ensur-

13 A. Shaw, Social Media, extremis, and radicalization, 2023. Retrieved 9 July 2026, https://www.science.org/doi/10.1126/sciadv.adk2031.

14 Institute for Economics & Peace, Global Terrorism Index 2022: Measuring the Impact of Terrorism, Sydney, 2022. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2022/03/GTI-2022-web-04112022.pdf.

15 United Nations General Assembly, 60/288. The United Nations Global Counter-Terrorism Strategy, 2006, Retrieved 6 July 2026, https://docs.un.org/en/A/RES/60/288.

16 United Nations General Assembly, Letter dated 22 December 2015 from the Secretary-General to the President of the General Assembly, 2015, Retrieved 6 July 2026, https://docs.un.org/en/A/70/675; United Nations Development Programme, Preventing Violent Extremism through promoting inclusive development, tolerance and respect for diversity, 2016. https://www.undp.org/sites/g/files/zskgke326/files/publications/Discussion%20Paper%20-%20Preventing%20Violent%20Extremism%20by%20Promoting%20Inclusive%20%20Development.pdf.

17 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia, New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

18 Institute for Economics & Peace, Global Terrorism Index 2025: Measuring the Impact of Terrorim, Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf.

ing that enhanced security capabilities remain consistent with human rights protections and democratic oversight. Therefore, ensuring that AI-based counterterrorism strategies are accompanied by clear legal standards and institutional safeguards is essential for maintaining security effectiveness and public trust.

There is a relative lack of research that comprehensively analyzes how the adoption of AI changes the operational logic of counterterrorism strategies, as well as how human rights protection relates to strategic effectiveness in this process. However, Martini (2024) systematically addresses the issue of the exception logic caused by the application of AI in the counterterrorism domain, as well as the need for standardization at the level of international organizations.[19] While technology-focused research tends to pay relatively little attention to the human rights implications of counterterrorism technologies, law- and human rights–centered research has primarily focused on the legality of counterterrorism measures and their compatibility with human rights standards, while paying relatively less attention to how respect for human rights may affect the effectiveness of counterterrorism strategies.[20]

What sets this article apart is its integration of legal definitions, technological design, and institutional safeguards into a single analytical framework. It shows that governing AI-enabled preventive counterterrorism requires balancing strategic effectiveness with protecting fundamental rights. To this end, the article views the protection of human rights as a key element for ensuring the long-term effectiveness and social legitimacy of counterterrorism strategies, not merely as an external normative constraint. From this perspective, this article considers security and human rights to be mutually complementary elements rather than conflicting interests and poses the following research questions:

First, how does the introduction of AI alter the operational logic of national counterterrorism strategies? Second, what human rights risks are associated with AI-based preventive counterterrorism strategies? Third, what strategic constraints do these human rights risks impose on the effectiveness and legitimacy of counterterrorism strategies?

The article examines shifts in the operational logic of AI-based counterterrorism strategies and their human rights implications using a methodology combining normative and comparative policy analyses. Rather than relying on empirical quantitative analysis, the analysis is based on an interpretive and structural analysis of international norms, policy documents, and existing academic literature. The research framework consists of two complementary analytical layers.

The first stage is to analyze how the adoption of AI is transforming the operational logic of counterterrorism strategies and to establish normative standards. Counterterrorism strategies are shifting to an increasingly preventive model. Based on this, we aim to examine AI-based counterterrorism strategies by applying the principles of legality, necessity, and proportionality. These principles serve as normative standards for AI's surveillance-prediction-intervention mechanisms.

The second stage involves comparative policy analysis. This article examines the European Union Artificial Intelligence Act (EU AI Act, 2024) and the OSCE's counterterrorism and human rights framework as case studies to assess how the previously established normative standards are implemented or constrained in actual institutional design. It also examines whether national security exceptions conflict with human rights protection principles.

19 A. Martini, 'AI, counter-terrorism and global governance: state of the art', in Revista de Paz y Conflictos, 2024, vol. 17, pp. 205-221.

20 A. Martini, 'AI, counter-terrorism and global governance: state of the art', in Revista de Paz y Conflictos, 2024, vol. 17, pp. 205-221.

The purpose of this two-stage analysis goes beyond a simple normative evaluation to explain how human rights risks affect the effectiveness and legitimacy of counterterrorism strategies. These gaps lead to strategic inefficiencies, including reduced policy credibility, excessive surveillance, and increased false positives.

2. Structural Changes in Contemporary Terrorism

Modern terrorism is undergoing fundamental restructuring in terms of its forms and organizational structures, as well as in response to changes in the technological environment. This chapter analyzes these changes and provides the structural context for shifts in the operational logic of AI-based counterterrorism strategies, which are discussed in subsequent sections.

2.1. Decentralized and Individualized Terrorism

Modern terrorism is shifting away from traditional, organization-centric structures toward a decentralized, digital model. While terrorist organizations used to operate through physical training camps and hierarchical command structures, today, individuals can acquire ideological attack tactics through online spaces and carry out acts of violence independently without belonging to any formal organization.[21]

Voluntary radicalization and 'lone wolf terrorism' are particularly prevalent among young people, who are often motivated by ideological sympathy and an internalized sense of identity rather than direct control from an organization.[22] Rather than emerging from a single ideology, radicalization occurs through the blending and reconfiguration of extremist narratives and conspiracy theories. Individuals combine these narratives to justify their violent acts.[23]

Consequently, modern terrorists have evolved into autonomous actors within a networked information environment rather than operating under the direct control of a single organization. This suggests that existing counterterrorism strategies, which rely on identifying organizations, tracing command structures, and dismantling organizational networks, are limited in their ability to respond effectively to the evolving threat landscape. The pace and impact of technological change, particularly that of AI, will likely have a profound effect on the strategic environment and the intelligence cycle needed to understand it.[24]

2.2. Algorithmic Radicalization and Digital Environments

The evolution of modern terrorism goes beyond weakening organizational structures. The proliferation of generative AI further amplifies this trend. Generative AI automates the creation and translation of propaganda content. It also enables the production of more sophisticated disinformation, automated propaganda campaigns, and more advanced strategies for circumventing censorship, thereby significantly increasing the speed and adaptability of

21 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia, New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

22 Institute for Economics & Peace, Lone Wolf and Youth Terrorism: Evolving Patterns of Terrorism & Radicalisation in Western Democracies, Sydney, 2025. Retrieved 6 February, 2026, visionofhumanity.org/wp-content/uploads/2025/03/Lone-Wolf-and-Youth-Terrorism.pdf.

23 C. Polizzi, Youth Radicalization: A New Frontier in Terrorism and Security, in Global Terrorism Index 2025, Institute for Economics & Peace, Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf.

24 P. Bury and M. Chertoff, 'New Intelligence Strategies for a New Decade', The Rusi Journal, 2020, p. 8.

radicalization processes.[25]

Some terrorist organizations use multilingual digital content to target different regions and cultural spheres simultaneously.[26] This demonstrates a mode of expansion that differs from traditional, region-based organizational models.[27] Consequently, radicalization is continuously reproduced within algorithm-based information environments rather than through physical contact or local networks.

Furthermore, the proliferation of encrypted messaging services and private networks has intensified this trend. These closed communication environments enable the sharing of ideologies and the coordination of operations while evading external surveillance. This makes the radicalization process even more difficult to detect.[28] This undermines the effectiveness of traditional intelligence gathering methods and limits the timing and scope of state intervention.

2.3. Concealment and Strategic Uncertainty of Terrorist Preparations

Terrorist organizations are reorganizing their operational methods around digital platforms. As individuals adopt extremist ideologies and acquire operational knowledge online, they may leave fewer observable physical indicators. This makes it more difficult to detect potential threats before an attack using conventional suspect identification methods.

For example, some groups use online propaganda and digital financial tools to raise funds and expand their influence without physical infrastructure,[29] thereby weakening geographical constraints on terrorist activities. Consequently, small-scale attacks with limited resources can amplify their strategic impact through political and psychological repercussions, such as diplomatic tensions, economic instability, and social distrust.[30] The decline in visibility and the rise in uncertainty are prompting nations to increasingly adopt preventive approaches that identify and manage potential risks at an earlier stage.

The next section examines how these pressures are reshaping the operational logic of counterterrorism strategies through the adoption of AI and how international normative frameworks attempt to constrain this transformation.

3. AI-based Counterterrorism: Operational Logic and Normative Framework

This section analyzes how the introduction of AI reconfigures the operational logic of national counterterrorism strategies. This analysis corresponds to the first layer of the article's framework, which focuses on operational logic. Focusing on the transition to data-driven preventive risk management models, the section examines changes in

25 J. Wrigley, From Vulnerability to Strength: The Growing Role of Strategic Communication in Preventing Terrorism and Violent Extremism, in Global Terrorism Index 2025, Institute for Economics & Peace, The Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf.

26 Institute for Economics & Peace, Global Terrorism Index 2025: Measuring the Impact of Terrorim, The Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf.

27 L. Webber and L. Meloy, Perspectives, ISKP Intensifying Online Propaganda Targeting Russia and Central Asia, Eurasianet, 2024. Retrieved 6 February, 2026, eurasianet.org/perspectives-iskp-intensifying-online-propaganda-targeting-russia-and-central-asia.

28 Institute for Economics & Peace, Global Terrorism Index 2025: Measuring the Impact of Terrorim, Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf.

29 L. Webber and L. Meloy, Perspectives, ISKP Intensifying Online Propaganda Targeting Russia and Central Asia, Eurasianet, 2024. Retrieved 6 February, 2026, eurasianet.org/perspectives-iskp-intensifying-online-propaganda-targeting-russia-and-central-asia.

30 BBC News, Kabul Airport Attack: What Do We Know?, 26 August 2021. Retrieved 1 February 2026, bbc.com/news/world-asia-58349010.

information production methods, decision-making structures, and intervention timing.

3.1. Preventive Logic and Data-Driven Intelligence

As discussed in Section 2, modern terrorist threats are becoming increasingly decentralized and individualized, making early detection more challenging for intelligence and law enforcement agencies. These operational challenges are prompting counterterrorism strategies to place greater emphasis on preventive and data-driven approaches.

In the face of these changes, countries are bolstering preventive strategies to identify and address potential risks before terrorist attacks occur. This approach focuses on managing the likelihood of a threat occurring rather than specific incidents or acts themselves. It involves techniques such as early warning systems, anomaly detection, and behavioral pattern analysis.

The adoption of AI raises concerns about privacy, data protection, transparency, accountability, and algorithmic bias, underscoring the importance of ensuring that AI systems are developed and deployed in accordance with ethical principles and human rights.[31] This shows that AI-based counterterrorism strategies are not just a technical issues, but are also closely linked to democratic oversight and protection of human rights.

The foundation of AI-based counterterrorism strategies is the development of a data-centric intelligence production system.[32] AI can process vast amounts of data in real time and identify patterns, enabling the detection of risk signals that were previously difficult to capture. In this process, the concept of 'risk' is shifting from human intuition to a data-driven approach. Policy decisions are increasingly based on the results of data analyses.

Law enforcement and intelligence agencies use AI to analyze different types of data to calculate risk scores and identify potential threats early on. In fact, the Chicago Police Department uses an algorithm-based risk assessment system to calculate individual 'threat scores' and implement preventive intervention and surveillance strategies based on these scores.[33] This demonstrates that AI-based risk prediction is being applied to public safety and security operations beyond simple data analysis.[34] Thanks to new technology, countries can now monitor and analyze the digital environment at a scale and speed that would have been difficult to achieve using traditional, human-centered analytical frameworks.

Along with this shift, decision-making structures are also changing. Risk assessments generated by AI systems are incorporated directly into the policymaking process and, in some cases, are linked to automated warning and intervention systems. Although AI-assisted decision-making can improve processes, it raises concerns about transparency, accountability, and the need for human oversight.[35] Data bias or flawed correlation analysis, for example, can result in the overclassification of individuals or groups as high-risk targets.[36] Thus, although data-driven deci-

31 UNESCO, Recommendation on the ethics of artificial intelligence, Retrieved 7 July 2026, https://unesdoc.unesco.org/ark:/48223/pf0000380455.

32 Boaz Ganor, 'Artificial or Human: A New Era of Counterterrorism Intelligence?', Studies in Conflict & Terrorism, 2019, Vol. 44. Issue 7. p. 605.

33 Time, The Police are using Computer Algorithms to tell if you're a Threat, 2017. Retrieved 19 May 2026, https://time.com/4966125/police-departments-algorithms-chicago.

34 Time, The Police are using Computer Algorithms to tell if you're a Threat, 2017. Retrieved 19 May 2026, https://time.com/4966125/police-departments-algorithms-chicago.

35 European Commission, Ethics Guidelines for Trustworthy AI, 2019, Retrieved 7 July, 2026, pp. 14-20.

36 K. Lum and W. Issac, To predict and serve?, in Significance, 2016, Vol. 13, Issue 5. academic.oup.com/jrssig/article/13/5/14/7029190.

sion-making is efficient, it also carries the potential for new forms of error.

3.2. Normative Tensions: The OSCE and the EU AI Act

The operational logic of AI-based counterterrorism strategies is shaped by an international normative environment that extends beyond the policy choices of individual nations. The OSCE has consistently emphasized the concept of 'comprehensive and co-operative security', defining counterterrorism and human rights protection as complementary obligations rather than mutually exclusive ones.[37] Similarly, OSCE's ODIHR has always emphasized the importance of implementing counterterrorism and countering violent extremism measures in accordance with human rights and the rule of law.[38]

This normative stance is evolving into more concrete policy discussions, particularly in light of recent advancements in AI technology. In March 2024, the OSCE's Transnational Threats Department and the Representative on Freedom of the Media (RFoM) hosted an expert meeting on the use of AI for preventing violent extremism and radicalization (P/CVERLT) in collaboration with the International Centre for Counter-Terrorism (ICCT).[39] The meeting emphasized that AI can serve as an effective policy tool for law enforcement agencies and civil society when used in a manner centered on human rights. The OSCE identified the promotion of media and information literacy as a key policy area and presented respect for human rights as an essential prerequisite for the use of AI.[40] Furthermore, in July 2024, the OSCE Parliamentary Assembly issued a resolution recommending the use of AI for counterterrorism.[41]

The resolution emphasized the importance of sharing regulatory principles and best practices that are based on respect for human rights and the rule of law. Thus, the OSCE framework can be understood as an attempt to integrate AI-based counterterrorism strategies with human rights norms. This demonstrates how the normative standards explained in this article are adopted in policy discourse.

Another example of the implementation of these standards in institutional design can be seen in the EU AI Act. The Act is designed to regulate high-risk AI systems and attempts to embed human rights protections, such as a general prohibition on the use of real-time remote biometric recognition in public spaces, into institutions.[42]

However, the Act permits significant exceptions in the context of counterterrorism and national security. For example, real-time biometric recognition technology may be used to prevent 'specific, serious, and imminent threats', such as terrorist attacks.[43] Furthermore, the use of AI for national security purposes is largely excluded from the

37 OSCE, The OSCE Concept of Comprehensive and Co-operative Security, 2009. Retrieved 1 March 2026, cdn.osce.org/sites/default/files/f/documents/c/0/37592.pdf.

38 ODIHR, Counter Terrorism, Protecting Human Rights, The Warsaw, 2008. Retrieved 28 January 2026, odihr.osce.org/sites/default/files/f/documents/d/6/29103.pdf.

39 OSCE, OSCE and ICCT hold expert seminar on artificial intelligence, preventing and countering violent extremism to discuss challenges and opportunitie', 2024. Retrieved 1 February 2026, osce.org/secretariat/564863.

40 OSCE, OSCE and ICCT hold expert seminar on artificial intelligence, preventing and countering violent extremism to discuss challenges and opportunitie', 2024. Retrieved 1 February 2026, osce.org/secretariat/564863.

41 OSCE PA, Resolution on Artificial Intelligence and the fight against Terrorism, 2024. Retrieved 3 March 2026, oscepa.org/en/documents/ad-hoc-committees-and-working-groups/ad-hoc-committee-on-countering-terrorism/resolutions-and-publications/5040-resolution-on-artificial-intelligence-and-the-fight-against-terrorism-adopted-at-the-31st-annual-session-bucharest-29-june-to-3-july-2024/file.

42 Future of Life Institute, EU Artificial Intelligence Act. Retrieved 4 April 2026, artificialintelligenceact.eu/article/2/.

43 Future of Life Institute, EU Artificial Intelligence Act. Retrieved 4 April 2026, artificialintelligenceact.eu/article/5/.

scope of regulation.[44] Additionally, AI systems that assist human judgment based on objective facts directly related to crime are exempt from certain regulations.

Article 2(3) of the EU AI Act notably excludes 'AI systems used exclusively for military, defense, or national security purposes' from its regulatory scope,[45] creating a significant normative gap. The problem is that today's AI based surveillance infrastructure largely relies on private technology companies and digital platforms. In this context, broad exemption clauses could be used to circumvent AI regulations by interpreting 'national security' broadly.

This framework of exceptions creates structural tension with the core principles of international human rights law - legality, necessity, and proportionality - that serve as normative standards for this article. First, broad exceptions for national security purposes can obscure the scope of the law, thereby undermining clarity and predictability. This occurs, in part, by weakening the legal standards required by the principle of legality regarding the circumstances under which AI-based surveillance is justified. Second, the criterion of 'imminent threat' is highly open to interpretation and carries the risk of weakening rigorous scrutiny of the principle of necessity during actual policy implementation. Third, the use of highly intrusive technologies, such as biometric recognition, may result in the excessive restriction of fundamental rights relative to their effectiveness, posing a risk of conflict with the principle of proportionality.

Thus, the EU AI Act exhibits a dual structure: it attempts to institutionally incorporate human rights protections, while simultaneously allowing these norms to be restricted through counter-terrorism and national security exceptions. This is a prime example of how the gap between normative principles and institutional implementation manifests in actual policy design. As one of the first comprehensive regulatory frameworks governing AI, the EU AI Act is likely to serve as an important reference for countries that are still developing AI governance frameworks.

This is particularly relevant in South Korea, where the introduction of the AI Basic Act has made AI governance a key policy priority.[46] South Korea is an analytically significant case because it combines a historical legacy of security-driven surveillance with ongoing debates about balancing counterterrorism efforts with fundamental rights. This tension has been evident since the passage of the Anti-Terrorism Act in 2016. At the same time, South Korea has swiftly adopted AI for public security purposes. Recent institutional reforms, such as transferring the National Intelligence Service's counter-espionage investigative authority to the police in 2024, demonstrate the country's efforts to enhance democratic oversight of intelligence and national security. Against this backdrop, the EU's experience foreshadows the normative and institutional challenges that South Korea may encounter when developing AI-based counterterrorism strategies with regard to reconciling the institutionalization of fundamental rights protections with security-related exceptions that could limit them.

3.3. Operational Applications and Multi-Level Governance

AI-based counterterrorism strategies are expanding beyond the independent efforts of individual nations and are becoming part of a multilayered governance structure involving international organizations, governments, and private companies. The United Nations Global Counterterrorism Strategy (A/RES/60/288) emphasizes the importance

44 Future of Life Institute, EU Artificial Intelligence Act. Retrieved 4 April 2026, artificialintelligenceact.eu/article/2/.

45 Future of Life Institute, EU Artificial Intelligence Act. Retrieved 4 April 2026, artificialintelligenceact.eu/article/2/.

46 Korea Legislation Research Institute, Framework Act on the Development of Artificial Intellegence and the Creation of a Foundation for Trust (Act No. 20676), Retrieved 5 August 2026, https://elaw.klri.re.kr/kor_service/lawView.do?hseq=71019&lang=ENG.

of international cooperation and strengthening national capacities in the fight against terrorism, and identifies human rights and the rule of law as core principles of counterterrorism policy.[47]

In the digital environment, platform companies play a key role in managing the flow of information. Although these companies have the technical and informational resources necessary to combat terrorism, such as online content, user data, and algorithmic systems, public agencies often have difficulty accessing these resources directly. Consequently, countries are finding it difficult to effectively counter the spread of extremism and online radicalization in the digital space without cooperating with private-sector platform companies. Consequently, counterterrorism strategies are being restructured around public-private partnership models.

At the same time, this collaborative framework gives rise to new normative issues. AI-based surveillance and content control depend heavily on the algorithms and internal policies of private platforms. However, it is difficult to hold private companies to the same standards of accountability and democratic oversight as government agencies. In particular, legal and ethical tensions intensify when freedom of expression and privacy protection intersect with national security.

In the field of counterterrorism, AI serves as a core analytical infrastructure that supports national strategic decision-making. It goes beyond being merely a technological tool. It is used to integrate, refine, and analyze data from various sources to identify potential risk signals, visualize networks of relationships among actors, and establish intervention priorities.[48] These capabilities lay the groundwork for a fundamental transformation in the operational approach of counterterrorism strategies by enabling the more efficient allocation of limited investigative resources and personnel.

Meanwhile, at the operational level, AI is used in three main functional areas.

The first function is controlling the information environment. AI detects false and distorted information, helping to prevent terrorist and violent extremist narratives from entering mainstream discourse.[49] The fact that extremist groups leveraged conspiracy theories and disinformation to expand their online influence during the pandemic demonstrates that countering information manipulation has become a core component of counterterrorism strategy.[50]

For instance, the Government Communications Headquarters (GCHQ), the UK's intelligence agency, has introduced AI-based tools to detect fake accounts spreading disinformation.[51] This shows that, as hostile actors use AI to disrupt the information environment, states strengthen their response capabilities using the same technology.

---

47 United Nations General Assembly. 60/288. The United Nationsl Global Counter-Terrorism Strategy. 2006. Retrieved 6 February 2026, docs.un.org/en/A/RES/60/288.

48 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia , New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

49 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia, New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

50 Institute for Economics & Peace, Global Terrorism Index 2022: Measuring the Impact of Terrorism, The Sydney, 2022. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2022/03/GTI-2022-web-04112022.pdf.

51 Reuters, Britain's GCHQ cyber spies embrace the AI revolution, 2021. Retrieved 6 February 2026, reuters.com/article/technology/britains-gchq-cyber-spies-embrace-the-ai-revolution-idUSKBN2AO2W7/.

Second is the function of platform-based intervention and control. AI is used to automate content review and removal through methods such as 'deplatforming' or 'shadow-banning'.[52] While these measures are somewhat effective at curbing the spread of terrorism-related content, concerns have been raised that, in the absence of a clear international definition of what constitutes terrorism or extremism, platforms may apply varying and potentially biased criteria.[53] INTERPOL has also emphasized the dangers of data bias and algorithmic design in this context.[54]

Third is the function of behavioral intervention and narrative countering. AI is used not only to block content, but also to generate and disseminate alternative messages that counter extremist narratives. The Institute for Strategic Dialogue (ISD) in London, for example, used AI-assisted advertising tools and counter-narrative initiatives to target audiences vulnerable to ISIL propaganda.[55]

These diverse applications have significant normative implications. As automated decision-making and intervention become more prevalent, it becomes increasingly difficult to verify how the principles of legality, necessity, and proportionality are applied in actual operations. This issue is addressed again in the analysis of human rights risks in Section 4 below.

4. Human Rights Risks and Strategic Limitations

This section analyzes the human rights risks inherent in the operational logic of AI-based counterterrorism strategies. These risks are derived from the principles of legality, necessity, and proportionality proposed by the Office of the United Nations High Commissioner for Human Rights, as discussed in Chapter 3. Furthermore, it examines how these risks impose structural constraints on the effectiveness and legitimacy of the strategies, affecting more than just violations of norms.

4.1. Legality: Surveillance and Legal Boundary Ambiguity

State power should be exercised based on clear and predictable legal grounds. However, AI-based preventive counterterrorism strategies rely on extensive data collection and continuous surveillance, which has the potential to blur legal boundaries.[56] However, these experiences demonstrate the continuing importance of ensuring accountability in the use of state authority.

South Korea's counterterrorism legislation illustrates the challenges of regulating broad surveillance powers. For example, South Korea's counterterrorism legislation grants intelligence agencies broad authority through vague

52 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia, New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

53 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia, New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

54 INTERPOL and UNICRI, Towards Responsible AI Innovation, Torino, 2020. Retrieved 8 February 2026, unicri.org/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf.

55 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia, New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

56 J. Hardy, 'The Rise of the modern Intelligence State', in A. Henschke, A. Reed, S. Robbins and S. Miller (eds.), Counter-Terrorism, Ethics and Technology, 2021, link.springer.com/chapter/10.1007/978-3-030-90221-6_7.

concepts such as 'tracking' or 'information collection'.[57] These provisions could encompass not only physical surveillance but also extensive digital surveillance. Consequently, individuals' online activities may be subject to surveillance, even when reasonable suspicion has not been established. Recent institutional reforms, such as transferring the National Intelligence Service's counter-espionage investigative authority to the police in 2024, reflect efforts to strengthen oversight of national security powers. In this context, the introduction of AI-based surveillance and risk assessment technologies requires an examination of whether existing legal and institutional safeguards are sufficient to prevent potential abuses of power.

By the adoption of AI technology, many governments have increased their use of automated intelligence to collect and evaluate data in an effort to ensure national security.[58] Government agencies are incentivized to collect and store more data under the pretext of improving predictive accuracy. In this process, data accumulates regardless of an individual's consent. There is a risk that surveillance will shift from an exceptional measure to a routine administrative activity. This could have a chilling effect on not only the right to privacy, but also on fundamental rights such as freedom of expression.[59]

As facial recognition technology becomes more widely used, significant ethical issues requiring careful consideration are emerging. These issues relate primarily to the tension between security needs and individual privacy, autonomy, and democratic accountability.[60] These technologies offer the efficiency of automatically identifying individuals in public spaces while enabling widespread surveillance of the general public.

Although some pilot implementations have confirmed technical efficiency,[61] concerns have been raised regarding algorithmic bias and the potential for misidentification. This shows that technical accuracy does not guarantee legal legitimacy. Nevertheless, INTERPOL operates a global facial recognition system that utilizes facial images provided by over 179 countries.[62] The Norwegian Police Directorate aims to enhance transparency by anonymizing images of individuals in surveillance footage using AI.[63] This allows the anonymized data to be used for pattern recognition.[64]

4.2. Necessity and Proportionality: Prediction, Bias and Rights Restrictions

According to the principle of necessity, restrictions on fundamental rights imposed by the state are only permitted

57 Korean Law information Center, Act on Counter-Terrorism for the Protection of citizens and Public Security (Act No. 15608) Artikel 9 (Collection of Information on Terrorism suspects). Retrieved 14 April, 2026. https://www.law.go.kr/lsInfoP.do?lsiSeq=203233&urlMode=engLsInfoR&viewCls=engLsInfoR#0000.

58 H.M. Verhelst, A.W. Stannat, G. Mecacci, Machine Learning against Terrorism: How Big Data Collection and Analysis Influences the Privacy-Security Dilemmna, in Science and Engineering Ethics, 2020, Vol. 26, p. 2976.

59 J.W. Penney, Understanding Chilling Effects, Minnesota Law Review, 2022, Vol.106, Issue 3, scholarship.law.umn.edu/minnlrev/vol106/iss3/6/.

60 M. Smith and S. Miller, The ethical application of biometric facial recognition technology, in AI & Society, 2022, Vol. 37, link.springer.com/article/10.1007/s00146-021-01199-9.

61 A. Engel, Facial Recognition Technologies in the Public Sector, in The Cambridge Handbook of Facial Recognition in the Modern State, Cambridge University Press, 2024. Retrieved 6 February, 2026, cambridge.org/core/books/cambridge-handbook-of-facial-recognition-in-the-modern-state/facial-recognition-technologies-in-the-public-sector/F62BC3D7147EBED04239D020AF245412.

62 INTERPOL, Facial Recognition, Retrieved 4 February, 2026, https://www.interpol.int/en/How-we-work/Forensics/Facial-Recognition.

63 INTERPOL and UNICRI, Towards Responsible AI Innovation, Torino, 2020. Retrieved 8 February 2026, unicri.org/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf.

64 INTERPOL and UNICRI, Towards Responsible AI Innovation, Torino, 2020. Retrieved 8 February 2026, unicri.org/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf.

when they are indispensable for achieving a legitimate purpose. However, since AI-based counterterrorism strategies rely on risk prediction, the criteria used to determine the necessity of intervention change.

AI predictive models inherently operate based on historical data. If certain groups or behavioral patterns are overrepresented in the data, there is a risk that those characteristics will be classified disproportionately as risk factors. This can lead to innocent individuals being treated as potential threats, which structurally undermines the presumption of innocence and due process.

Furthermore, the massive data collection required to improve predictive accuracy conflicts with the principle of necessity. If large-scale data accumulation expands without verifying its essentiality for threat prevention, surveillance may justify itself beyond its original purpose. With the explosive growth of data on individuals' online behavior on social media, there has been increased interest in using this data to predict terrorist activity. However, human rights experts and civil society organizations have pointed out that this process creates structural entry points that can lead to discriminatory judgments.[65]

To mitigate these issues, analysis based on anonymized or aggregated data has been proposed as an alternative.[66] This approach has the potential to better align with the principle of necessity, which requires minimizing the scope of intervention, as it allows for the identification of risk patterns without directly targeting specific individuals. In other words, rather than preemptively intervening against specific individuals, this approach enables more limited and indirect intervention through trend analysis and the detection of risk signals at the group level.

This approach is gradually being adopted in actual policy and technical initiatives. For instance, the STONE (Shaping Terrorist Online Network Efficacy) model simulates how a terrorist network reorganizes when certain individuals are eliminated.[67] This helps focus limited resources on the most effective intervention points. This approach limits the scope of intervention to the minimum necessary rather than expanding indiscriminate surveillance.

Similarly, the Central Office for Information Technology in Security (ZITiS) in Germany has developed a system that uses recommendation algorithms to filter relevant information from vast amounts of data for investigations.[68] Since the system improves the efficiency of searching for and analyzing information rather than making direct risk assessments of individuals, it functions as a support tool in the pre-intervention stage.

This trend is also evident at the policy program level. The RED-Alert Project funded by the EU aims to detect early signs of radicalization while maintaining privacy standards.[69] AI company Moonshot takes a non-coercive approach to intervention that focuses on strengthening individuals' resilience to extremist and manipulative online content

65 INTERPOL and UNICRI, Towards Responsible AI Innovation, Torino, 2020. Retrieved 8 February 2026, unicri.org/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf.

66 INTERPOL and UNICRI, Towards Responsible AI Innovation, Torino, 2020. Retrieved 8 February 2026, unicri.org/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf.

67 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia, New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

68 INTERPOL and UNICRI, Towards Responsible AI Innovation, Torino, 2020. Retrieved 8 February 2026, unicri.org/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf.

69 European Commission, Real-time Early Detection and Alert System for Online Terrorist Content based on Natural Language Processing, Social Network Analysis, Artificial Intelligence and Complex Event Processing, Retrieved 5 February 2026, cordis.europa.eu/project/id/740688.

instead of imposing sanctions or restrictive measures.[70] Instead of identifying individuals for punitive intervention, this approach aims to reduce vulnerability to extremist messaging by improving media literacy and raising awareness of online manipulation techniques.[71]

Additionally, since 2019, the Monitoring System and Transfer Platform Radicalizaion (MOTRA) has been studying radicalization processes in Germany.[72] The project focuses on preventing and combating violent radicalization. The project comprehensively and systematically observes various forms of radicalization in German society, with the ultimate goal of analyzing the processes that lead to terrorist acts.

These examples demonstrate that AI-based analysis does not necessarily result in direct surveillance and intervention against individuals. Rather, it can be designed to better satisfy the principle of necessity. However, these approaches have fundamental limitations. Even when based on anonymized or aggregated data, when risk scores or classification results generated by algorithms lead to policy interventions, they can indirectly regulate individuals' thoughts and expressions.

Consequently, prediction-based intervention has a dual nature. First, it contributes to the principle of necessity by enabling authorities to allocate resources toward individuals or situations assessed as posing a higher risk rather than applying broad, indiscriminate measures. On the other hand, prediction-based intervention can reinforce the necessity assessment by enabling authorities to justify interventions through risk-based evaluations rather than generalized suspicion.

The principle of proportionality requires that restrictions on fundamental rights not be excessive in relation to the objective they seek to achieve. However, while AI-based counterterrorism strategies offer high efficiency, they also carry the potential to result in excessive restrictions on rights.[73]

The use of AI in the online environment, in particular, clearly highlights this tension. Although automated content detection and removal systems are effective at curbing the spread of content, they risk restricting legitimate political expression or social criticism if they do not sufficiently consider the context of the content.[74] Moreover, since these systems operate automatically via algorithms, errors are difficult to correct. In the absence of transparent appeal procedures, rights violations may persist.

Another issue with algorithm-based decision-making is that it amplifies the biases inherent in the data. Social and institutional biases present in the training data can be reinforced during automated analysis, which can result in specific groups being assigned higher risk scores. This conflicts directly with the principle of non-discrimination and

70 UNOCT and UNICRI, Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia, New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

71 Moonshot, Bulding Resilience to Misinformation in Europe, 2025, Retrieved 7 July 2026, https://moonshotteam.com/wp-content/uploads/Building-Resilience-to-Misinformation-in-Europe.pdf.

72 Bundeskriminalamt, Forchungsprojekte Terrorismus/Extremismus, Retrieved 1 April 2026, bka.de/DE/UnsereAufgaben/Forschung/ForschungsprojekteUndErgebnisse/TerrorismusExtremismus/Forschungsprojekte/forschungsprojekte.html; MOTRA, Projektbeschreibung, Retrieved 1 April 2026, motra.info/motra-im-profil/projektbeschreibung/.

73 INTERPOL and UNICRI, Towards Responsible AI Innovation, Torino, 2020. Retrieved 8 February 2026, unicri.org/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf.

74 E.J. Llansó, No amout of AI in content moderation will solve filtering›s prior-restraint problem, in Big Data & Society, 2020, Vol. 7, Issue 1. journals.sagepub.com/doi/epub/10.1177/2053951720920686.

suggests that technological efficiency can paradoxically exacerbate social inequality.

These biases are especially likely to be reinforced in the context of predictive policing.[75] Groups that have been subject to more intensive surveillance generate more data. This data is then used to train models that assign higher risk scores to those groups. This structure progressively amplifies initial biases, systematically targeting specific ethnic and religious groups for excessive surveillance.[76] Thus, false positives are not merely technical errors, but rather, they can function as mechanisms that perpetuate social inequality within technological systems.

The UN General Assembly and the Human Rights Council have emphasized that the obligation to protect human rights applies without exception in counterterrorism contexts.[77] Without appropriate safeguards, AI-enabled surveillance and automated decision-making processes could threaten the right to privacy and other fundamental rights, including freedom of expression, the right to equality and non-discrimination.[78] Furthermore, the Counter-Terrorism Committee Executive Directorate's 'Tech Against Terrorism' initiative recommends establishing clear definitions and processing standards for illegal content.[79] These definitions and standards serve as a minimum institutional safeguard, preventing automated content regulation from being applied excessively and infringing upon freedom of expression.

4.3. Definitional ambiguity and Strategic Constraints

The human rights risks associated with AI-based counterterrorism strategies are amplified when combined with the ambiguity of the term 'terrorism'. Without a unified international definition of terrorism, individual nations apply different legal standards and interpretations, undermining the consistency and predictability of counterterrorism law enforcement.

This issue is clearly evident in the case of the United States. Although the U.S. provides a basic definition of domestic terrorism at the federal level, it permits individual states to establish their own counterterrorism laws and criminal elements.[80] Consequently, the same act may be interpreted differently depending on the jurisdiction, and the criteria for what constitutes terrorism lack uniformity.[81]

The problem is that this legal diversity extends beyond mere differences, resulting in legislation that encompasses

75 K. Lum and W. Issac, To predict and serve?, in Significance, 2016, Vol. 13, Issue 5. academic.oup.com/jrssig/article/13/5/14/7029190.

76 P. Ugwudike, AI audits for assessing design logics and building ethical systems: the case of predictive policing algorithms, in AI and Ethics, 2021, Vol. 2, link.springer.com/article/10.1007/s43681-021-00117-5.

77 United Nations, Human Rights, Retrieved May 19 2026, un.org/counterterrorism/en/cct/human-rights.

78 United Nations General Assembly, The right to privacy in the digital age, A/HRC/48/31, 2021, https://digitallibrary.un.org/record/3946475?v=pdf.

79 Tech Against Terrorism, Content Personalisation and the Online Dissemination of Terrorist and Violent Extremist Content, London, 2021. Retrieved February 6 2026, techagainstterrorism.org/hubfs/210120-TAT-Position-Paper-content-personalisation-and-online-dissemination-of-terrorist-content-JB-vFINAL-DA.pdf.

80 International Center for Not-for-Profit Law (ICNL), 'State Domestic Terrorism Laws in the United States: A Growing Threat to First Amendment Rights', Washington, D.C, 2024. Retrieved 3 March 2026, icnl.org/wp-content/uploads/US-state-terrorism-laws-report-29-Aug-2024-update.pdf; W. S. McCann and N. Pimley, Mixed Mandates: Issues Concering Organizational and Statutory Definitions of Terrorism in the United States, in Terrorism and Political Violence, 2020. Vol. 32 no. 4, p. 807.

81 International Center for Not-for-Profit Law (ICNL), 'State Domestic Terrorism Laws in the United States: A Growing Threat to First Amendment Rights', The ashington, D.C, 2024. Retrieved 3 March 2026, icnl.org/wp-content/uploads/US-state-terrorism-laws-report-29-Aug-2024-update.pdf.

broad and vague concepts.[82] For example, many state anti-terrorism laws include abstract criteria such as 'threats', 'causing chaos', and 'interfering with government functions'.[83] These laws are designed to allow for severe criminal penalties even when no actual acts of violence have occurred. This structure has the potential to be applied to individuals who criticize government policies, such as protesters, social activists, and civil society organizations, contrary to the original legislative intent.

Such conceptual uncertainty is further exacerbated by the introduction of AI-based counterterrorism systems. Since these systems learn from and apply the criteria defined by existing laws and policies, the more ambiguous the legal concepts, the more this uncertainty is reproduced and amplified. In the online environment, where distinguishing between political discourse, satire, legitimate social movements, and expressions of extremism is difficult, it is hard to rule out the possibility that normal civic participation could be classified as a terrorist threat.

Consequently, counterterrorism strategies face constraints on two practical levels. First, at the operational level, the distinction between terrorist content and legitimate expression becomes blurred, forcing law enforcement agencies and platform operators to repeatedly make uncertain judgments between over-regulation and regulatory evasion. Second, at the normative level, there is an increased risk of a chilling effect on fundamental rights, such as freedom of expression and political participation.

To address these issues, Tech Against Terrorism emphasizes the importance of clear policies and a sophisticated designation system.[84] However, expanding technical measures without first establishing clear conceptual criteria will likely yield limited results.

The ambiguity of the terrorism definition extends beyond legal debate to become a key structural variable that defines the scope and legitimacy of AI-based counterterrorism strategies. Without clear conceptual boundaries, expanded technical surveillance and analytical capabilities risk turning counterterrorism strategies into tools for political control and social management, causing them to deviate from their original purpose of responding to actual violent threats. This is not an inherent problem of AI technology, but rather a structural limitation arising from the combination of unclear norms and power.

For a long time, achieving a universally accepted definition of terrorism has been difficult. Nevertheless, there is still a need for precise legal definitions and standards that determine what constitutes terrorism and when counterterrorism measures can be implemented. Broad or ambiguous definitions may expand the scope of state intervention, creating the risk of arbitrary or disproportionate restrictions on fundamental rights, especially when combined with AI-based surveillance and predictive risk assessment technologies.

82 Ibid.; J.J. Crowley and T.E. Posada, HB452-Domestic Terrorism, in Georgia State University Law Review, 2017. Vol. 34 Issue 1. Retrieved 4 April 2026, readingroom.law.gsu.edu/cgi/viewcontent.cgi?article=2902&context=gsulr.

83 International Center for Not-for-Profit Law (ICNL), 'State Domestic Terrorism Laws in the United States: A Growing Threat to First Amendment Rights', Washington, D.C, 2024. Retrieved 3 March 2026, icnl.org/wp-content/uploads/US-state-terrorism-laws-report-29-Aug-2024-update.pdf.

84 Tech Against Terrorism, Content Personalisation and the Online Dissemination of Terrorist and Violent Extremist Content, London, 2021. Retrieved February 6 2026, techagainstterrorism.org/hubfs/210120-TAT-Position-Paper-content-personalisation-and-online-dissemination-of-terrorist-content-JB-vFINAL-DA.pdf.

5. Conclusion: Balancing Counterterrorism Strategies and Human Rights Protection in the AI Era

Premised on the notion that structural changes in terrorism in the AI era demand a fundamental shift in national counterterrorism strategies, this article has comprehensively analyzed the relationship between the strategic effectiveness and human rights risks of AI-based preventive counterterrorism strategies. As terrorism becomes more personalized and operates invisibly in the digital environment, states are shifting to increasingly preventive strategies that proactively identify and intervene in risks. This shift is substantially supported by AI technologies that enable real-time processing and pattern analysis of large volumes of data.[85]

The key finding of this article is that this technological shift increases efficiency and fundamentally reshapes the operational logic of counterterrorism strategies. Through the principles of legality, necessity, and proportionality, this article shows that AI-based preventive counterterrorism strategies present new challenges to ensuring compliance with these standards. Regarding legality, AI-enabled surveillance may extend state authority beyond clearly defined legal boundaries. Regarding necessity, predictive interventions may encourage broader data collection and earlier forms of intervention based on probabilistic assessments. Regarding proportionality, restrictions on fundamental rights may become excessive when the intrusiveness of AI-based measures outweighs their demonstrated effectiveness.

When synthesizing these analyses, the gap between normative standards and institutional implementation emerges as a structural factor that limits the long-term effectiveness of counterterrorism strategies, affecting more than just human rights issues. This gap is not merely a matter of legal consistency. Rather, it limits the effectiveness of counterterrorism policies in three ways at the strategic level. First, widespread surveillance and unclear criteria undermine social trust in the policy, thereby weakening the foundation for cooperation. Second, prediction errors and false positives can lead to the inefficient allocation of limited resources, thereby reducing the capacity to respond to actual threats. Third, excessive restrictions on rights can lead to the alienation and backlash of specific groups, which poses a risk of exacerbating the environment for radicalization in the long term.

Therefore, human rights issues in AI-based counterterrorism strategies should be understood as key variables that determine the strategy's effectiveness and sustainability, not as externally imposed normative constraints. This suggests that security and human rights are not mutually exclusive goals but can form a mutually reinforcing relationship under certain conditions.

From this perspective, the article emphasizes the importance of the institutional design of AI technology utilization rather than questioning the adoption of AI technology itself. AI has the potential to substantially enhance counterterrorism capabilities in various areas, such as detecting signs of radicalization, analyzing online extremist content, and identifying risk networks. Therefore, the policy challenge is not to deny the legitimacy of AI use but to clearly define the conditions under which it can be justified.

The following institutional conditions are essential to achieving this balance. First, clear and specific legal grounds for AI-based counterterrorism measures must be established to satisfy the principle of legality. Second, the scope of data collection and analysis should be limited to what is necessary for legitimate counterterrorism purposes, and there should be institutionalized mechanisms for assessing necessity. Third, any restrictions on fundamental

85 V.R. Depa, Real-time lot data processing using cloud-based AI and streamline analytics, in International Journal of Engineering Technology Research & Management, 2023, Vol. 7 Issue 9, ijetrm.com/issues/files/Jan-2023-12-1736664084-30September%202023.pdf.

rights must be proportionate to the objectives pursued and require continuous evaluation of the effectiveness and intrusiveness of AI-based measures.

The Anti-Terrorism Act, enacted in South Korea in 2016, for example, provides the National Intelligence Service with the authority to collect information on individuals suspected of posing a terrorist threat. However, the act has raised concerns about the potential for excessive information collection and whether existing oversight mechanisms are sufficient to ensure that intelligence-gathering activities remain consistent. Therefore, discussions on AI-based counterterrorism strategies in South Korea should address three institutional challenges alongside technological adoption: establishing clear legal standards and limits for AI-based surveillance and information collection; ensuring that data collection, risk assessment, and preventive interventions are necessary and proportionate to legitimate security objectives; and strengthening independent oversight and accountability mechanisms to ensure compliance with these principles.

In conclusion, preventive counterterrorism strategies in the AI era transcend technical issues, constituting a complex policy domain where norms and strategy converge. In this domain, protecting human rights is not an optional consideration, but rather a core component that defines the legitimacy and effectiveness of the strategy. Only when this shift in perspective occurs can a sustainable counterterrorism strategy be realized.

Finally, this article has limitations because it is based on normative and theoretical analysis. Future research should empirically verify the relationship between the effectiveness of AI-based counterterrorism strategies and their impact on human rights by focusing on actual policy implementation cases. This will make it possible to identify under what conditions a balance between strategic efficiency and human rights protection can be achieved.

Bibliography

P. Bury and M. Chertoff, New Intelligence Strategies for a New Decade, ***The Rusi Journal***, 2020.

BBC News, '***Kabul Airport Attack: What Do We Know?***, August 26 2021. Retrieved 1 February 2026, bbc.com/news/world-asia-58349010.

Bundeskriminalamt, Forchungsprojekte Terrorismus/Extremismus, Retrieved 1 January 2026, https://www.bka.de/DE/UnsereAufgaben/Forschung/ForschungsprojekteUndErgebnisse/TerrorismusExtremismus/Forschungsprojekte/forschungsprojekte.html; MOTRA, Projektbeschreibung, Retrieved 1 April 2026, https://www.motra.info/motra-im-profil/projektbeschreibung/.

Council of Europe, ***Fight against terrorism***. Retrieved 28 January 2026, coe.int/en/web/commissioner/thematic-work/fight-against-terrorism.

J.J. Crowley and T.E. Posada, HB452-Domestic Terrorism, in ***Georgia State University Law Review***, 2017. Vol. 34 Issue 1. Retrieved 4 April 2026, readingroom.law.gsu.edu/cgi/viewcontent.cgi?article=2902&context=gsulr.

V.R. Depa, Real-time lot data processing using cloud-based AI and streamline analytics, in ***International Journal of Engineering Technology Research & Management***, 2023, Vol. 7 Issue 9, ijetrm.com/issues/files/Jan-2023-12-1736664084-30September%202023.pdf.

European Commission, 'Real-time Early Detection and Alert System for Online Terrorist Content based on Natural Language Processing, Social Network Analysis, Artificial Intelligence and Complex Event Processing, Retrieved 5 February 2026, https://cordis.europa.eu/project/id/740688.

European Commission, ***Ethics Guidelines for Trustworthy AI***, 2019, Retrieved 7 July, 2026, pp. 14-20.

A. Engel, ***Facial Recognition Technologies in the Public Sector***, in The Cambridge Handbook of Facial Recognition in the Modern State, Cambridge University Press, 2024. Retrieved 6 February, 2026, cambridge.org/core/books/cambridge-handbook-of-facial-recognition-in-the-modern-state/facial-recognition-technologies-in-the-public-sector/F62BC3D7147EBED04239D020AF245412.

Future of Life Institute, EU Artificial Intelligence Act. Retrieved 4 April 2026, artificialintelligenceact.eu/article/2/

Future of Life Institute, EU Artificial Intelligence Act. Retrieved 4 April 2026, artificialintelligenceact.eu/article/5/

B. Ganor, 'Artificial or Human: A New Era of Counterterrorism Intelligence?', Retrieved 10. February 2026. ***Studies in Conflict & Terrorism***, 2019, Vol. 44. Issue 7. pp. 605-624.

Institute for Economics & Peace, ***Global Terrorism Index 2025: Measuring the Impact of Terrorism***, Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf

Institute for Economics & Peace, ***Global Terrorism Index 2022: Measuring the Impact of Terrorism***, Sydney, 2022. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2022/03/GTI-2022-web-04112022.pdf

Institute for Economics & Peace, ***Lone Wolf and Youth Terrorism: Evolving Patterns of Terrorism & Radicalisation in Western Democracies***, Sydney, 2025. Retrieved 6 February, 2026, visionofhumanity.org/wp-content/uploads/2025/03/Lone-Wolf-and-Youth-Terrorism.pdf

International Center for Not-for-Profit Law (ICNL), 'State Domestic Terrorism Laws in the United States: A Growing Threat to First Amendment Rights', Washington, D.C, 2024, Retrieved 3 March 2026, https://www.icnl.org/wp-content/uploads/US-state-terrorism-laws-report-29-Aug-2024-update.pdf

INTERPOL and United Nations Interregional Crime and Justice Research Institute (UNICRI), 'Towards Responsible AI Innovation', Torino, 2020, Retrieved 8 February 2026, https://unicri.org/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf.

INTERPOL, ***Facial Recognition***, Retrieved 4 February, 2026, https://www.interpol.int/en/How-we-work/Forensics/Facial-Recognition.

J. Hardy, 'The Rise of the modern Intelligence State', in A. Henschke, A. Reed, S. Robbins and S. Miller (eds.), ***Counter-Terrorism, Ethics and Technology***, 2021, link.springer.com/chapter/10.1007/978-3-030-90221-6_7.

Korean Law information Center, Act on Counter-Terrorism for the Protection of citizens and Public Security (Act No. 15608) Artikel 9 (Collection of Information on Terrorism suspects). Retrieved 14 April, 2026, https://www.law.go.kr/lsInfoP.do?lsiSeq=203233&urlMode=engLsInfoR&viewCls=engLsInfoR#0000.

Korea Legislation Research Institute, Framework Act on the Development of Artificial Intellegence and the Creation of a Foundation for Trust (Act No. 20676), Retrieved 5 August 2026, https://elaw.klri.re.kr/kor_service/lawView.do?hseq=71019&lang=ENG.

K. Lum and W. Issac, To predict and serve?, in ***Significance***, 2016, Vol. 13, Issue 5. https://academic.oup.com/jrssig/article/13/5/14/7029190.

E.J. Llansó, No amout of AI in content moderation will solve filtering›s prior-restraint problem, in ***Big Data & Society***, 2020, Vol.7 Issue 1. journals.sagepub.com/doi/epub/10.1177/2053951720920686.

A. Martini, 'AI, counter-terrorism and global governance: state of the art', in ***Revista de Paz y Conflictos***, 2024, Vol. 17, pp. 205-221.

W.S. McCann and N. Pimley, Mixed Mandates: Issues Concering Organizational and Statutory Definitions of Terrorism in the United States, in ***Terrorism and Political Violence***, 2020. Vol. 32 no 4, pp. 807-830.

Moonshot, ***Bulding Resilience to Misinformation in Europe***, 2025, Retrieved 7 July 2026, https://moonshotteam.com/wp-content/uploads/Building-Resilience-to-Misinformation-in-Europe.pdf.

ODIHR, ***Counter Terrorism, Protecting Human Rights***, Warsaw, 2008. Retrieved 28 January 2026, odihr.osce.org/sites/default/files/f/documents/d/6/29103.pdf.

OSCE, ***OSCE and ICCT hold expert seminar on artificial intelligence, preventing and countering violent extremism to discuss challenges and opportunities,*** 2024. Retrieved 1 February 2026, osce.org/secretariat/564863.

OSCE, ***The OSCE Concept of Comprehensive and Co-operative Security***, 2009. Retrieved 1 March 2026, cdn.osce.org/sites/default/files/f/documents/c/0/37592.pdf.

OSCE PA, ***Resolution on Artificial Intelligence and the fight against Terrorism, 2024***. Retrieved 3 March 2026, oscepa.org/en/documents/ad-hoc-committees-and-working-groups/ad-hoc-committee-on-countering-terrorism/resolutions-and-publications/5040-resolution-on-artificial-intelligence-and-the-fight-against-terrorism-adopt-

ed-at-the-31st-annual-session-bucharest-29-june-to-3-july-2024/file.

C. Polizzi, ***Youth Radicalization: A New Frontier in Terrorism and Security***, in Global Terrorism Index 2025, Institute for Economics & Peace, Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf.

J.W. Penney, Understanding Chilling Effects, ***Minnesota Law Review***, 2022, Vol. 106, Issue 3, scholarship.law.umn.edu/minnlrev/vol106/iss3/6/.

Reuters, ***Britain's GCHQ cyber spies embrace the AI revolution***, 2021. Retrieved 6 February 2026, reuters.com/article/technology/britains-gchq-cyber-spies-embrace-the-ai-revolution-idUSKBN2AO2W7/.

B. Saul, ***Defining 'Terrorism' to Protect Human Rights,*** Sydney Law School Legal Studies Research Paper No.09/125, 2008.

A.P. Schmid, ***Defining Terrorism***, ICCT Report, International Centre for Counter-Terrorism, 2023.

A. Shaw, Social Media, extremis, and radicalization, ***Science Advances***, 2023, https://www.science.org/doi/10.1126/sciadv.adk2031.

M. Smith and S. Miller, The ethical application of biometric facial recognition technology, in ***AI & Society***, 2022, Vol. 37, link.springer.com/article/10.1007/s00146-021-01199-9.

Tech Against Terrorism, ***Content Personalisation and the Online Dissemination of Terrorist and Violent Extremist Content***, London, 2021. Retrieved February 6 2026, techagainstterrorism.org/hubfs/210120-TAT-Position-Paper-content-personalisation-and-online-dissemination-of-terrorist-content-JB-vFINAL-DA.pdf.

Time, ***The Police are using Computer Algorithms to tell if you're a Threat***, 2017. Retrieved 19 May 2026, time.com/4966125/police-departments-algorithms-chicago.

P. Ugwudike, AI audits for assessing design logics and building ethical systems: the case of predictive policing algorithms, in ***AI and Ethics***, 2021, Vol. 2, link.springer.com/article/10.1007/s43681-021-00117-5.

UNESCO, ***Recommendation on the ethics of artificial intelligence***, Retrieved 7 July 2026, https://unesdoc.unesco.org/ark:/48223/pf0000380455.

United Nations, Human Rights, Retrieved 19 May 2026, un.org/counterterrorism/en/cct/human-rights.

United Nations General Assembly, 49/60, ***Measures to Eliminate International Terrorism***, 1994, Retrieved 1 February, 2026, https://digitallibrary.un.org/record/189017.

United Nations General Assembly, 60/288, ***The United Nations Global Counter-Terrorism Strategy,*** 2006, Retrieved 6 February 2026, docs.un.org/en/A/RES/60/288.

United Nations General Assembly, 4/88, ***Report of the United Nations High Commissioner for Human Rights on the protection of human rights and fundamental freedoms while countering terrorism***, 2007, Retrieved 6 July 2026, https://digitallibrary.un.org/record/595012?v=pdf.

United Nations General Assembly, ***Letter dated 22 December 2015 from the Secretary-General to the President of the General Assembly***, 2015, Retrieved 6 July 2026, https://docs.un.org/en/A/70/675; United Nations Development

Programme, Preventing Violent Extremism through promoting inclusive development, tolerance and respect for diversity, 2016. https://www.undp.org/sites/g/files/zskgke326/files/publications/Discussion%20Paper%20-%20Preventing%20Violent%20Extremism%20by%20Promoting%20Inclusive%20%20Development.pdf.

United Nations General Assembly, 48/31. ***The right to privacy in the digital age***. 2021. Retrieved 7 July 2026, https://digitallibrary.un.org/record/3946475?v=pdf.

United Nations Security Council, ***'Resolution 1373'***, 2001. Retrieved 1 February 2026, digitallibrary.un.org/record/449020?v=pdf; United Nations Security Council, ***'Resolution 1566'***, 2004. Retrieved 1 February 2026, digitallibrary.un.org/record/532676?v=pdf.

UNOCT and UNICRI, ***Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia***, New York, 2021. Retrieved 6 February 2026, un.org/counterterrorism/sites/default/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

H.M. Verhelst, A.W. Stannat, G. Mecacci, Machine Learning against Terrorism: How Big Data Collection and Analysis Influences the Privacy-Security Dilemmna, in ***Science and Engineering Ethics***, 2020, Vol. 26, pp. 2975-2984.

P. Waldmann, ***Terrorismus: Provokation der Macht***, 1998, Gerling Akademie Verlag.

L. Webber and L. Meloy, ***Perspectives: ISKP Intensifying Online Propaganda Targeting Russia and Central Asia***, Eurasianet, 2024. Retrieved 6 February, 2026, https://eurasianet.org/perspectives-iskp-intensifying-online-propaganda-targeting-russia-and-central-asia.

J. Wrigley, ***From Vulnerability to Strength: The Growing Role of Strategic Communication in Preventing Terrorism and Violent Extremism***, in Global Terrorism Index 2025, Institute for Economics & Peace, Sydney, 2025. Retrieved 3 March 2026, visionofhumanity.org/wp-content/uploads/2025/03/Global-Terrorism-Index-2025.pdf.

This article was published by the Security and Human Rights Monitor (SHRM).

Security and Human Rights (formerly Helsinki Monitor) is a journal devoted to issues inspired by the work and principles of the Organization for Security and Cooperation in Europe (OSCE). It looks at the challenge of building security through cooperation across the northern hemisphere, from Vancouver to Vladivostok, as well as how this experience can be applied to other parts of the world. It aims to stimulate thinking on the question of protecting and promoting human rights in a world faced with serious threats to security.

Netherlands Helsinki Committee
Het Nutshuis
Riviervismarkt 4
2513 AM The Hague
The Netherlands

© Netherlands Helsinki Committee. All rights reserved.

www.nhc.nl