ARTIFICIAL INTELLIGENCE: EMERGING CHALLENGES IN IMPLEMENTING UNITED NATIONS SECURITY COUNCIL RESOLUTION 1540

© Chris Yang - Certain characteristics of AI would make end-user a more prominent risk factor

ABSTRACT

This article explores challenges related to the transfer of emerging technologies, specifically artificial intelligence (AI), and the potential impact it has on the framework for strategic trade controls (STC). The widespread impact of AI poses significant challenges for national authorities tasked with implementing United Nations Security Council resolution 1540 (2004) (UNSCR 1540). These challenges include defining controlled AI items, verifying the declared end use, and establishing AI control schemes that extend beyond mere destination. These factors make end-user risk assessment crucial, but they also complicate international harmonization of controls, potentially shifting the focus of UNSCR 1540 from non-State actors to State actors. The article argues that the international community must develop a shared understanding of how to address these emerging risks without compromising global security.

THE AUTHOR:

Hyuk Kim

Hyuk Kim is a Consultant and Non-resident Fellow at 38 North of the Stimson Center. Prior to joining 38 North, Mr. Kim served as a Research Fellow at the James Martin Center for Nonproliferation Studies (CNS) at the Middlebury Institute of International Studies. He has also worked as a Nonproliferation and Nuclear Security Fellow at the Pacific Forum and as a guest researcher at the Stockholm International Peace Research Institute (SIPRI).

Over the past two decades, the international community has witnessed global progress in efforts to curb the illicit procurement of weapons of mass destruction (WMD) and related materials by non-State actors. Following the adoption of UNSCR 1540 in 2004, there was a tremendous increase in the awareness of the potential threat associated with misappropriation of WMD-related materials.[1] Meanwhile, as States placed increased emphasis on controlling these goods, the evolving proliferation risks associated with emerging technologies also became an important consideration. Evidence that these issues warranted careful attention in the implementation of UNSCR 1540 is seen in UNSC resolution 2325 (2016) and its related debates.[2] This article explores challenges related to the transfer of emerging technologies, specifically AI, and the potential impact it has on the framework for STC.

1 "Letter dated 1 December 2016 from the Permanent Representative of Spain to the United Nations addressed to the Secretary-General," S/2016/1013, *United Nations Digital Library*, 1 December 2016. Available at: https://documents.un.org/doc/undoc/gen/n16/410/58/pdf/n1641058.pdf?token=ceUN7G9ifOlqK859Nf&fe=true

2 Resolution 2325 (2016) / adopted by the Security Council at its 7837th meeting, on 15 December 2016," *United Nations Digital Library*, 2016. Available at: https://digitallibrary.un.org/record/852037?ln=en

In general, a national regulatory authority for STC implementation takes at least four factors into account in risk assessment for an intended export: an item's inherent WMD or military potential, its destination, and its stated end-use and end user. In most cases, each element alone cannot constitute a definitive criterion for authorizing a transaction, and the authority should have a holistic view by considering all factors to ascertain the permissibility of the proposed export. However, certain characteristics of AI would make end user a more prominent risk factor than the others as stated below.

First and foremost, akin to a field of study, AI is a horizontal technology. It encompasses a broad range of techniques that can be integrated at any stage of the technology lifecycle across various industrial sectors. In general, AI alone does not exist as a specific controlled item, but becomes subject to controls when it converges with other controlled items in their lifecycle. For instance, the multinational STC arrangements, such as the Nuclear Suppliers Group (NSG), define controlled technology as "information required for the development, production, or use of any item" on the control lists.[3] In this instance, AI becomes a technology transfer concern if an engineer well-versed in techniques for anomaly detection develops software that identifies the malfunctioning status of production equipment for controlled materials, even though those anomaly detection techniques can be used for legitimate civilian applications.

The ubiquitous nature of AI could pose challenges for a national authority seeking to proactively reduce potential AI-driven proliferation risks by setting a list of controlled AI items and related parameters. For instance, in 2020, the US government announced an interim rule to control software for geospatial analysis.[4] The specific criteria for the controlled software included the use of a "Deep Convolutional Neural Network (CNN) to detect the object of interest from positive and negative samples." However, the US private sector, such as ride-handling services, was hesitant to view these criteria as control parameters. They perceived it as merely a description of CNN, a technique widely used, primarily for image classification purposes.[5]

Second, many AI models, such as artificial neural networks (ANNs) can be fine-tuned, rendering end-use less critical for risk assessment. Specifically, a pre-trained ANN model can be further trained through a process called transfer learning to address specific needs or better perform in a particular context.[6] This means that an end user could fine-tune a pre-trained model with a general

3 "Communication received from the Permanent Mission of the Argentine Republic to the International Atomic Energy Agency regarding Certain Member States' Guidelines for Transfers of Nuclear-related Dual-use Equipment, Materials, Software and Related Technology," INFCIRC/254/Rev.12/Part2, *International Atomic Energy Agency*, 29 July 2022. Available at: https://www.iaea.org/sites/default/files/publications/documents/infcircs/1978/infcirc254r12p2.pdf

4 "Addition of Software Specially Designed To Automate the Analysis of Geospatial Imagery to the Export Control Classification Number 0Y521 Series," 85 FR 489, *Federal Register*, 6 January 20020. Available at: https://www.federalregister.gov/documents/2020/01/06/2019-27649/addition-of-software-specially-designed-to-automate-the-analysis-of-geospatial-imagery-to-the-export

5 "AH89 Public Comment 21," *Regulations.gov*, 10 March 2020. Available at: https://www.regulations.gov/document/BIS-2019-0031-0022

6 Hyuk Kim, "North Korea's Artificial Intelligence Research: Trends and Potential Civilian and Military Applications," *38 North*, 23 January 2024. Available at: https://www.38north.org/2024/01/north-koreas-artificial-intelligence-research-trends-and-potential-civilian-and-military-applications/

© Bolivia Inteligente - AI is a horizontal technology that encompasses a broad range of techniques

object detection function to meet specific military requirements. Similarly, a military-oriented pre-trained model could be enhanced to handle more complex environments.[7] In addition, transfer learning does not require the whole dataset used for a pre-trained model. Instead, only the relevant data of interest to the end user is necessary, reducing hardware demands, such as storage and computational power.

A recent study indicates that the object detection capabilities of a pre-trained surveillance model can be improved using transfer learning.[8] The study found that a daytime surveillance system initially trained on RGB images from public datasets can be further trained for night-time surveillance using infrared images.[9] Remarkably, this transferred learning process relies solely on a set of infrared images, avoiding the need for the entire original dataset. Likewise, it is feasible that an end user could repurpose civilian drone object detection software for military target detection via transfer learning. Therefore, the stated end-use in an export license application may not always serve as a reliable risk indicator.

Third, AI is predominantly a software-based technology that can be transferred to the end user via intangible means, called intangible transfer of

7 Toubman, J. J. Roessingh, P. Spronck, A. Plaat and J. Van Den Herik, "Transfer Learning of Air Combat Behavior," *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, USA, 2015, pp. 226-231, doi: 10.1109/ICMLA.2015.61.

8 E. S. Kim, W. Kim, J. Park and K. Yeo, "Human Detection in Infrared Image Using Daytime Model-Based Transfer Learning for Military Surveillance System," *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, Korea, Republic of, 2023, pp. 1306-1308, doi: 10.1109/ICTC58733.2023.10393353.

9 Ibid.

technology (ITT).[10] In contrast to the trading of physical goods, AI software can be transferred via various digital platforms, including email and cloud computing environments. Moreover, many cloud computing services provide AI development environments supported by the latest AI-dedicated chipsets, with which a proliferator could exploit such platforms for malicious AI development without relying on importing AI-related hardware.[11] This means that restrictions focused only on the physical destination of AI hardware are insufficient to fully address proliferation risks associated with AI. In this respect, for national authorities, the characteristic of AI as a software-based technology necessitates a shift in focus beyond the cross-border movement of goods. Rather than concentrating solely on the physical destination of exports, cross-nationality transactions, such as international scientific collaboration, also become a crucial consideration in any jurisdiction.

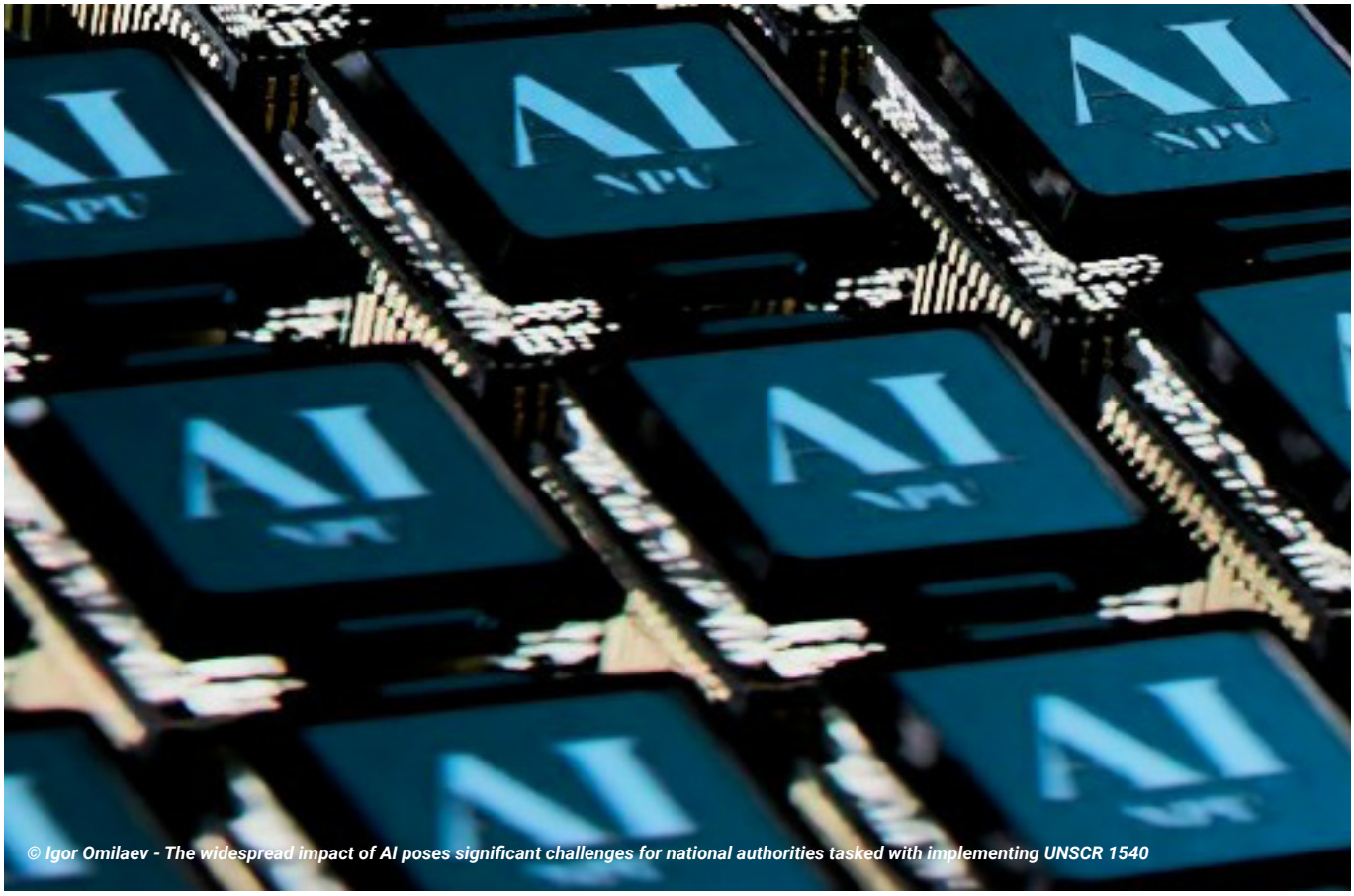
Consequently, the end user, especially one intending to divert AI toward WMD applications, is likely to become a more prominent risk factor compared to other elements. However, end-user-focused STC could be seen as conflicting with the principle of focusing on the threat posed by non-State actors, which was a key factor in the adoption of UNSCR 1540. During the debates on the draft resolution, some States expressed concerns that the obligations under UNSCR 1540 could have significant implications for national security and the right to self-defence. More importantly, they pointed out that coercive measures could indeed be imposed not only on non-State actors, but also on States themselves.[12]

10 Hyuk Kim, "Intangible Transfer of Technology (ITT): Open-source Information Analysis for the Implementation of Sanctions on North Korea," *38 North*, 10 March 2023. Available at: https://www.38north.org/2023/03/intangible-transfer-of-technology-itt-open-source-information-analysis-for-the-implementation-of-sanctions-on-north-korea/

11 Hyuk Kim, "North Korea's Artificial Intelligence Research: Trends and Potential Civilian and Military Applications."

12 UN Security Council 4950th Meeting, S/PV.4950, *Security Council Report*, 22 April 2004. Available at https://www.securitycouncilreport.org/un-documents/document/1540-spv-4950.php

For national authorities, the characteristic of AI as a software-based technology necessitates a shift in focus beyond the cross-border movement of goods.



© Igor Omilaev - The widespread impact of AI poses significant challenges for national authorities tasked with implementing UNSCR 1540

Likewise, it would be challenging for national authorities to achieve global harmonization in their risk assessments of certain end users, except for those related to prominent terrorist organizations. This is because threat perceptions may vary among States, depending on their individual national security and foreign policy interests. If a governmental agency responsible for foreign affairs has more influence in interagency coordination for the export licensing process than a non-proliferation-focused strategic trade control authority, the decision to grant a license could be heavily influenced by the State's geopolitical interests. Consequently, the intended recipient State might interpret such a decision as a reflection of the exporting State's perception of them as a competing State, rather than of the intended importer as a non-State actor.

The widespread impact of AI poses significant challenges for national authorities tasked with implementing UNSCR 1540. These challenges include defining controlled AI items, verifying the declared end-use, and establishing AI control schemes that extend beyond mere destination. Concurrently, a potential increase in reliance on the end user as a risk indicator could alter the spirit of UNSCR 1540 by gradually shifting the primary focus of the resolution from non-State actors to States. These end-user-focused controls could raise political concerns for many States about becoming implicated in regional rivalries or strategic competition. In this regard, to address AI-driven proliferation risks, the international community needs to build a common understanding on how to distinguish non-State actors from State actors.