Making Sense of Anthropic Threat Intelligence Report in the Backdrop of the 10th BWC Review Conference (2027)

Hina Pandey

Dr Hina Pandey is a Research Fellow at the Manohar Parrikar Institute for Defence Studies and Analyses

Summary

In September 2026, Anthropic released its latest Threat Intelligence Report on "Detecting and Countering Misuse of AI". The report highlighted AI model-enabled biological misuse as a major emerging risk. This article assesses the implications of this report in the context of the current developments in the Biological Weapons Convention and the impact on global biosecurity norms.

Introduction

More than half a century ago, the global community signed the first multilateral disarmament treaty. The Biological Weapons Convention Treaty (BWTC) (*hereafter BWC)*, which entered into force on 26 March 1975, highlighted in its preamble the determination to achieve "effective progress towards general and complete disarmament"[1]. By definition, in the context of the BWC, it implies the 'prohibition and elimination' of any kinds of biological weapons. An important reminder here would be that while 189 countries are the State Parties to the BWC and currently no State Party has admitted to having a biological weapons programme, doubts remain whether the "Treaty has achieved its main objective of truly eliminating biological weapons globally"[2]. Yet, so far the BWC has been a success, in terms of building a strong norm that "Biology must serve peace, not advance harm"[3]. A strong norm is still a relatively good measure of restraint, especially in the absence of any verification mechanism of the treaty; however, in the backdrop of the new and emerging developments in the field of science and technology (S&T), including Artificial Intelligence (AI), there is merit in asking relevant questions – will this strong norm survive? What more needs to be addressed and done to sustain the norm at a time when developments in the field of Science and Technology (S&T) are rapid, mostly emanating from private enterprise and at times outpacing the design and implementation of policy regulations? The answer lies in the continuous strengthening of the Convention itself.

In its recent Anthropic Threat Assessment Report on *Detecting and Countering Misuse*

of AI (September 2026), "biological misuse is identified as one of the most serious risks of frontier AI models"[4]. The threat intelligence report had indeed highlighted a concerning misuse of Artificial Intelligence (AI). As one of the Claude models was utilized for the conduct of malicious activity, providing credence to the concerns about AI models reaching the capability to either make existing pathogens more dangerous/harmful or create entirely new ones. The threat report identified five real-world case scenarios, not as typical misuse, but as illustrative examples of novel threat activities (research) that AI models were asked to conduct by the user (researchers). The tasks included the use of Claude to assist in -

i) designing a research grant application to "genetically alter an organism to create a new, or enhanced biological property on Chikungunya virus", which has no licensed therapeutic countermeasures[5].

ii) conducting research on highly pathogenic avian influenza (bird flu) - notably the gain-of-function research to genetically alter an organism by providing it a new trait that it naturally does not possess. In the context of this research on bird flu, it is particularly concerning because the possibility of a change of vector from birds to mammals (including humans) and its possibility to infect outside of the traditional scope – 'beyond the respiratory tract' - would imply a major catastrophe, as there is near-zero population immunity against the disease in humans. A pandemic potential of such a disease can only be imagined. This specific research plan in which Claude's assistance was sought was not only limited to tracing scientific literature, data analysis, or editorial assistance, but also involved "genetic engineering approaches aimed at introducing mutations associated with mammalian adaptation in airborne transmissibility in animal models"[6].

iii) In other cases, AI models were utilized in pursuing investigations on dual-use, non-transmissible novel venoms and toxin agents, one of which is on the WHO's R&D blueprint priority list of diseases having 'greatest epidemic and pandemic threat'. In both these cases, the AI models were used to pursue research on 'novel compounds that can simultaneously be developed into novel therapeutics or toxic agents'. In addition, what is worrisome is that the usage of AI models was traced to the geographic regions where their access is legally blocked. Furthermore, the inherent safety measures within the AI models that block requests were evaded by routing the research to other competitive models. Not only has this gap highlighted the inherent vulnerability within the existing AI ecosystems, but the report itself has also flagged its own limitation of automated safety filters. It admits the gap in reliably identifying the intent of the user in the dual-use case research[7].

This is of some significant consequences for two reasons. Firstly, a mere identification of malicious intent cannot be truly translated into a real defence, primarily because some in-operations usage may deliberately appear to be benign in nature – genuine academic pursuits or meriting therapeutic enquiries. Additionally, the definition of 'malicious intent' needs to be further explored; Can LLM models effectively identify malicious intent? What is the contextual baseline that is offered here? Furthermore, is malicious intent the same as malicious activity? It is also worth noting that an experienced 'potential threat actor' would most likely train the AI model to weaken the inherent

safeguard filters. Secondly, the dual-use nature of biotechnology not only compounds this dilemma; it further raises questions on the operations of the inherent safety features. For instance, if the LLM is able to qualify the therapeutic-benign usage as 'malicious intent or activity', it probably runs the risk of disabling a legitimate scientific advancement.

Anthropic's report findings are indeed alarming; however, caution needs to be exercised in the extreme interpretation of it. Yes, it is a cause of significant concern that AI models are perhaps reaching the capability of imparting potential bioweapons knowledge to actors and users who may not necessarily possess the domain expertise, including the "possibility that the process of making a biological weapon could get simplified further."[8] However, it is crucial to recognise that in addition to the conduct of preliminary research and assistance, "editing DNA of viruses or bacteria requires specialised equipment, and that equipment is not cheap, and you still need to know how to operate the equipment."[9]

Finally, at this point in the discussion, the good side of AI models also merits attention. Recently, scientists for the first time have claimed the use of an AI-designed virus to treat patients. A start-up called *Radical Numerics* "has administered the virus, known as a phage, to a patient suffering from a deadly bacterial infection"[10], with the hope of killing the microbes responsible for the infection. The jury is still out as the findings are yet to be made public. If and when successful, this may mean an optimistic disruption in the future of medicine.

Run-up to the Tenth BWC Review Conference (2027)

These risk concerns and opportunities are not lost on the BWC; the Eighth and Ninth BWC Review Conferences have taken much cognizance of this changing threat landscape and have emphasized the strengthening of international cooperation to monitor dual-use research and encourage ethical guidelines for scientific conduct. In the recently concluded 9th Session of the BWC WG, the Chair, Ambassador Frederico S Duque Estrada Meyer (Brazil), in his opening remarks reiterated the concerns about the possible emergence of a gap between these rapidly advancing capabilities and (our) Member States' collective ability to understand their governance and prevent misuse[11] and urged the delegations to come up with a 'clear and concrete' strategy and find language acceptable to everybody, and find the 'elements of discussion' that can be implemented immediately in the 10th Review Conference.

Indeed, the BWC is headed for its tenth Review Conference in 2027, before which it has to successfully conclude the upcoming Final Session of the WG on the Strengthening of the BWC, in December 2026. The task at hand is to come up with a consensus draft report to provide recommendations to the Meeting of States Parties to be held immediately in the same month. This is crucial as it will set the stage for the next BWC Review Conference. The current contention is on issues pertaining to verification matters, the International Cooperation Assistance (ICA) mechanism, and the Science and Technology Advisory (STA) Mechanism. At a fundamental level, the issue with the STA Mechanism is the one that reflects the divide between the Global North and the Global South. While oversight of AI models and regulations is required, it is also important that they don't come up disguised as 'export controls' targeting nations that intend to gain from peaceful AI uses in biology. In this context, ensuring adherence to Article X of the BWC becomes important, as it provides the State Parties

"the right to participate in, the fullest possible exchange of equipment, materials and scientific and technological information for the use of bacteriological (biological) agents and toxins for peaceful purposes."[12] The governance and regulations for newer technologies cannot obstruct peaceful progress. Clearly, the BWC stands at a critical juncture of biosecurity, and time is of the essence. It remains imperative that consensus is forged within the given deadline.

Views expressed are of the authors and do not necessarily reflect the views of the Manohar Parrikar IDSA or of the Government of India.

Endnotes:

[1] BWC Treaty Text, "Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction. Opened for Signature at London, Moscow and Washington. 10 April 1972", International Humanitarian Law Databases.

[2] Ambassador DB Venkatesh Varma (2025), "Negotiating the Protocol to the BWC: An Indian Perspective", in Edited, Ajey Lele (2025), "50 Years of Biological Weapons Convention: Tracking the Journey", pp. 125, Pentagon Press.

[3] Remarks by EAM Dr. S. Jaishankar at Conference on 50 Years of the Biological Weapons Convention: "Strengthening Biosecurity for the Global South", New Delhi (December 01, 2025).

[4] Anthropic Threat Intelligence Report, (2026) "Detecting and countering misuse of AI", (September 2026).

[5] Ibid.

[6] Ibid.

[7] Ibid.

[8] Ajey Lele,(2025) "Biological Warfare and Biological Weapons: A Response" in Edited, Ajey Lele (2025) "50 Years of Biological Weapons Convention: Tracking the Journey", pp. 26-27, Pentagon Press

[9] Jim Geraghty, "The Emergence of AI-Designed Viruses", *National Review,* 14 September 2026.

[10] Carl Zimmer, "Start-Up Claims to Use A.I.-Created Viruses to Treat a Patient", *The New York Times,* October 20, 2026.

[11] "WG9 – a Stuttering Start Converting Yellow Text to Green Text", *BioWeapons Prevention Project,* Report 2026-8, 18 August 2026.

[12] Treaty Text , "Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction. Opened for Signature at London, Moscow and Washington. 10 April 1972", International Humanitarian Law Databases.