HSP

HARVARD SUSSEX PROGRAM

Impacts of Artificial Intelligence on the CBW Prohibition Regimes: Analysis, Challenges, and Futures

Joshua R. Moon
Alexander Ghionis
Shaunna McIvor
Boaz Chan

31 March 2024

Executive summary

As AI technologies become increasingly sophisticated and accessible, policymakers must grapple with the complex challenges and opportunities facing the CBW prohibition regimes. This report explores the intersection of AI and CBW, situating AI within the broader CBW ecosystem and examining its potential impact on the processes of CBW acquisition. Drawing on expert interviews, workshops, and literature reviews, the report identifies four key dimensions through which AI can influence CBW acquisition activities and the wider environment: acceleration of processes, generation of pathways, mediation of ideas, and modification of transparencies.

The report further highlights four categories of challenges that face the CBW prohibition regimes: the emergence of new utilities for CBW, the circumvention of controls and proliferation of knowledge and materials, the creeping legitimization of CBW development, and the divergence of national interests and norms. Understanding and tracking how AI contributes to these is a complex and pressing challenge.

To illustrate these aspects in practice, the report presents eight hypothetical vignettes that explore how state and non-state actors might leverage AI to pursue CBW-related objectives. These vignettes, grounded in contemporary geopolitical and technological realities, demonstrate the potential for AI technologies to support disinformation, circumvent obstacles, develop resources, further stress the grey areas of the prohibitions, and erode norms against CBW. The vignettes are each analysed briefly using our framework, offering areas of policy action related to each scenario. The vignettes and the totality of our research also enable us to identify four broad, emerging themes:

- Sophisticated use often requires multiple AI systems to be joined together, emphasising a need to support efforts to implement new and complementary technical safeguards within AI algorithms.
- Use requires access to specialist datasets, equipment, materials, and chemicals highlighting that existing CBW governance arrangements require strengthening to better address such potential.
- Use often obfuscates investigations and attribution and thus investments in investigations mechanisms, from public health intelligence to the UNSGM, are needed to embed AI/cyber capabilities and training for rostered experts in the collection, handling, and processing of digital evidence.
- Intent plays a central role in how AI use presents risks for CBW, and thus supports a need for strengthened international norms against the use of CBWs. Crucially, the relationship between increasing access to more sophisticated AI, and the intentions and motivations of actors, is complex and not easily resolved.

The report underscores the urgent need for policymakers to engage with, and explore, the implications of AI for the CBW prohibition regimes in a creative and open manner, recognising that AI is not technologically determined to produce CBW but is, rather, an effective enabling tool for those actors and activities that threaten the integrity of the prohibition regimes.

Contents

1. Introduction

Artificial intelligence (AI), the development of computer systems capable of performing tasks that typically require human intelligence, is profoundly, yet unevenly, reshaping our social and technological landscapes. While its potential benefits are widely recognised, AI also raises concerns about its negative implications for privacy, accountability, safety, security, controllability, and transparency. These AI-anxieties are evident in almost all sectors of human activity where AI integration is anticipated. As AI's influence and impact can be imagined and modelled in nearly any context, we are witnessing the symbiotic unfolding of both our hopes for AI and our anxieties about its potential consequences.

AI-anxieties are not limited to specific sectors; they have also emerged in the context of international relations. Beyond the potential benefits AI might bring, its development and operation stokes fears about its contribution to instability, insecurity, violence, and war. These concerns manifest through various vectors, ranging from geopolitical and economic pressures for resources to the development of autonomous weapons systems and the generation of disinformation. The intersection of AI and chemical and biological weapons (CBW) is particularly prone to these anxieties and dynamics.

CBW exist at the confluence of security, science, technology, and socio-technical notions of utility, intention, and purpose. As such, CBW cannot be understood solely in terms of the agent or delivery device itself, at a particular time and place. Rather, CBW are situated within a broader, interdisciplinary ecosystem that encompasses not only the central processes of acquisition but also a range of other social, technical, and legal elements that shape meaning and structure the ecosystem.

These diverse aspects are also fundamentally anchored in people and the decisions they take. To grasp the implications of CBW, it is essential to consider both the central acquisition processes that are understood to relate to the materiality of humans having and using CBW, and the broader eco-system contexts which shape, and are shaped by, intentions to acquire.

2. The CBW eco-system and processes of acquisition

The term "CBW ecosystem" refers here to the complex, interconnected system which creates and sustains our social construction of CBW. This ecosystem encompasses a wide range of elements, including:[1]

- **Scientific and technological factors**: Advances in science and technology, and associated skills, that may influence the development and acquisition of CBW.
- **Information and communication factors**: The role of information and disinformation in shaping perceptions and narratives around CBW.

[1] Reflections on these topics can be found across the literature, including, Crowley, M., Dando, M. and Shang, L. (*eds*.) *Preventing Chemical Weapons – Arms Control and Disarmament as the Sciences Converge* (Royal Society of Chemistry; London, 2018);

- **Legal and normative factors**: Rules, laws, and norms that govern the development, acquisition, and use of CBW.
- **Response and accountability factors**: Processes related to the investigation, attribution, and accountability for the development, acquisition, or use of CBW."
- **Sociopolitical factors**: Geopolitical dynamics, social and cultural attitudes, and public opinion that shape the context in which CBW are developed, acquired, or used.
- **Economic and industrial factors**: Economic incentives, industrial capabilities, and commercial interests that may influence the development and acquisition of CBW, as well as the availability of resources and materials needed for their production.

Within this wide eco-system, a central set of acquisition processes exist, which describe the specific steps and actions that actors may take to acquire a CBW capability. As such, acquisition processes detail some of the key elements required to materialise CBW within the context of eco-system.[2] Within the acquisition processes, the role of human intention and motivation are also present as important drivers.[3]

Diagram 1, below, illustrates a simplified representation of these acquisition processes, which includes the following core elements: inspiration, planning, development, synthesis, storage and transportation, targeting, and dissemination. It is essential to recognise that these processes are not always linear, and actors may engage with different aspects of the acquisition process depending on their specific context and capabilities.

For example, some actors may focus heavily on the development and synthesis stages, while others might prioritize targeting and dissemination. Not all actors will necessarily engage with every element of the acquisition process, as their goals, resources, and constraints may vary significantly.

Annex 1 of this document explains each of these elements in relation to what they imply and how AI might, in general terms, influence their internal dynamics.

[2] See, for example, Meulenbelt, S. and Nieuwenhuizen, M. 'Non-state actors' pursuit of CBRN weapons: from motivation to potential humanitarian consequences' (*International Review of the Red Cross*, 2015) 97(899); Sandberg, A. and Nelson, C. 'Who should we fear more: biohackers, disgruntled postdocs, or bad governments? A simple risk chain model of biorisk' (*Health Security*, 2020) 18(3) p. 156; Tucker, J. B. (*ed.*) *Toxic Terror – Assessing Terrorist Use of Chemical and Biological Weapons* (MIT Press; Cambridge Mass., 2001); Ouagrham-Gormley, S. B. 'Barriers to bioweapons: intangible obstacles to proliferation' (*International* Security, 2012) 36(4); Revill, J. 'Past as Prologue? The Risk of Adoption of Chemical and Biological Weapons by Non-State Actors in the EU' (*European Risk Regulation*, 2017) 8(4); Robinson, J. P., Boserup, A. & Neild, R. *The Prevention of CBW - Volume V of The Problem of Chemical and Biological Warfare* (Almqvist & Wiksell; Stockholm, 1971); Zanders, J. P. 'Assessing the risk of chemical and biological weapons proliferation to terrorists' (*The Non Proliferation Review*, 1999) Fall

[3] For example, see Tennenbaum, M. and Kosal, M. E. 'The interplay between frugal science and chemical and biological weapons: investigating the proliferation risks of technology intended for humanitarian, disaster response, and international development efforts' in Kosal, M. E. (*ed.*) *Proliferation of Weapons- and Dual-Use Technologies* (Springer Cham, 2021); and Zalesny, M. D., Whitney, P. et al. 'A conceptual model to identify intent to use chemical-biological weapons' (*Journal of Strategic Security*, 2017) 10(3)

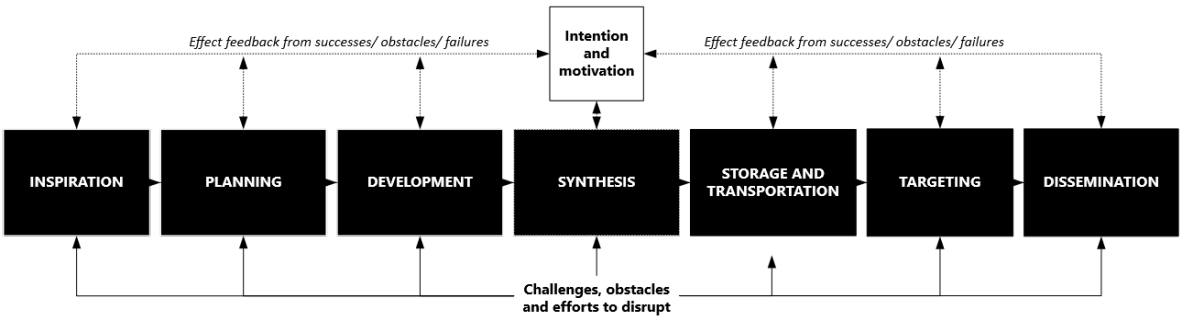

Diagram 1 – CBW acquisition processes (simplified)

The acquisition process is influenced by human psychologies, intentions, motivations, inspirations, and perceived utilities, which all shape actors' desires to acquire CBW.

Acquiring CBW is subject to various challenges and obstacles that can hinder an actor's motivation and ability to navigate the acquisition process successfully. These challenges can be internal or external to the actor and operate across the entire CBW ecosystem.[4] The following factors are illustrative of the basis of longstanding and evolving efforts to prevent the acquisition of CBW:

- Moral, ethical, and legal frameworks that influence actors' intentions to deter the pursuit of CBW.
- Difficulties in sourcing and sustaining necessary skills, resources, and materials that can delay, dissuade, or discontinue CBW acquisition efforts.
- Technological hurdles and scientific uncertainties that can render the path to CBW acquisition insurmountable or unviable.
- Inherent hazards and risks associated with handling dangerous pathogens or toxic chemicals, which can erode the intentions and motivations of potential actors, particularly in the absence of proper safety measures and expertise.
- International and national policies and regulations that create "webs of prevention," introducing substantial obstacles and personal risks that dissuade actors from seeking to acquire CBW.
- Resilient public health systems, effective first response capabilities, and well-developed medical countermeasures that serve as strong deterrents by diminishing the perceived impact and strategic value of CBW.

The complex interplay between the acquisition processes and the broad CBW ecosystem highlights the need for a contextualized and nuanced approach when addressing the

[4] Jefferson, C. 'Origins of the norm against chemical weapons' (*International Affairs*, 2014) 90 (3); Rappert, B. and McLeish, C. *A Web of Prevention: Biological Weapons, Life Sciences and the Governance of Research* (Earthscan, London: 2007); Revill, J. and Jefferson, C. 'Tacit knowledge and the biological weapons regime' (*Science and Public Policy*, 2014) 41(5); Robinson, J. P. (*ed.*) *Public Health Response to Biological and Chemical Weapons: WHO Guidance* (World Health Organisation; Geneva, 2004)

potential negative impacts of AI. As AI has the potential to influence an actor's engagement with acquisition processes, as well as the effectiveness of challenges and disruptive efforts, it is crucial to follow where AI goes, rather than be led by assumption.

3. Framing AI's impacts on the CBW prohibitions

The integration of AI into the CBW acquisition eco-system presents a complex challenge for those working to strengthen the CBW prohibition regimes. AI has the potential to support actors at every stage, from initial inspiration to eventual dissemination (see Diagram 1), in ways that are constantly evolving and difficult to predict. Equally hard to anticipate are the myriad ways in which AI could be used to circumvent the obstacles and disruption efforts aimed at impeding CBW acquisition.

So far, the trend in the analysis of the impact of AI on the CBW acquisition processes has focused on its impact on the development processes of the toxic or infective agent itself. In other words, analysis has largely asked if AI will make it easier to make CBW. This should not, however, be surprising: previous iterations of fears about the impact of advances in science and technology have often been focused on development, as opposed to other elements and processes for acquisition.[5] In many historic cases, such a technological determinism has eventually been tempered by recognition that other factors, and other stages, are equally relevant contributors to CBW acquisition.[6]

Indeed, a predictable consequence of the current narrow focus of AI on the development of CBW is that AI's implications for other stages within the acquisition process remain unclear. We have less clarity on how AI could be employed in non-developmental stages of CBW acquisition, for example in motivating and inspiring, or for storage of agents.

This is particularly salient as AI is, much like chemistry or biology, a general-purpose science and technology, with dual-use implications based on the purpose of its use. Identifying the most high-risk applications is useful, but fails to be comprehensive. Indeed, examination of AI applications can be guided by the criterion as to what purpose it is being put to, and how that purpose converges with particular acquisition stages, activities, and outcomes.

3.1 AI-anxieties and the dimensions of impact of AI

To open this landscape, we used expert interviews and workshops to elicit AI-anxieties and perceived risks across the wider CBW acquisition eco-system. Analysis of data revealed that the landscape of AI-anxieties and risks were complex and diffused, with different experts pointing to different elements.

[5] Here it is informative to think about similar fears and anxieties which have emerged in relation to the advent of the internet, micro-reactors, DNA synthesis, CRISPR gene editing tools, the DIY science community, and initially – and often pervasively – focus is given to how technologies will contribute to the development of CBW by lowering developmental barriers.

[6] Edwards, B. *Insecurity and Emerging Biotechnology: Governing Misuse Potential* (Palgrave Pivot; Cham, 2019); Tucker, J. B. (*ed.*) *Innovation, Dual Use, and Security: Managing the Risks of Emerging Biological and Chemical Technologies* (MIT Press; Cambridge Mass., 2012); Vogel, K. M. 'Framing biosecurity: an alternative to the biotech revolution model?' (*Science and Public Policy*, 2008) 35(1)

AI-anxieties were located right across the CBW acquisition eco-system, each potentially generating significant downstream concerns for their effects on the CBW prohibition regimes more widely. Table 1, below, shows some of the areas in which experts located their anxieties about the implications and effects of AI within the context of CBW.

Our AI-anxiety data has fundamentally reshaped our approach to studying AI and CBW. Instead of focusing narrowly on AI-driven advancements in biology or chemistry, we broadly examine where, how, and why AI might undermine CBW prohibitions across the entire acquisition spectrum. This pivot is crucial because AI-anxieties are diffused throughout the CBW ecosystem. By adopting a comprehensive perspective, we can systematically analyse the full range of AI-anxieties identified in Table 1 and their potential to erode the norms, institutions, and practices underpinning the CBW prohibition regime. This holistic framing is key to anticipating and mitigating the multifaceted risks posed by AI.

Bioregulator research and accelerated pathway generation for new discoveries	By-passing physical infrastructure requirements for developing CBW
Cascading and runaway systems	Chemical criminality and the dark web
Decentralisation, cloud-based operations and acquisition footprints	Deskilling, reskilling, and tacit knowledge and AI's impact on human resource
Designer assassination capabilities and developments in micro-targeting	Dissemination and targeting technologies and enhancements
Disinformation generation and normative manipulation	Genetic engineering and opportunities for ethno-targets
Finding new biological targets through augmented research methodologies	Manipulation of data and information to obscure activities and avoid detection
Grey area research and the creeping legitimisation of biological manipulation	Medical countermeasures and avoidance of detection
Non-lethal bodily adaptations and manipulation of 'human-ness'	Research acceleration outpacing review and regulation
Riot control and CNSAC research and deployment	Safety and security of biological data and cyberattacks
Synthetic pathways and re-evaluation of discontinued research	Technological combinations and convergence
Toxicity modelling	Verification and increasing opacity

Table 1 - The focus of AI-anxieties

Examining in more detail the narratives around such AI-anxieties, we identified four central commonalities in the descriptions of what the AI was understood to be doing in the specific contexts of the anxiety. These four commonalities were essentially

processes of changing occurring due to the application of AI technology, and they sometimes appeared simultaneously. Through the identification of these commonalities, we suggest that the application of AI can facilitate CBW acquisition through its ability to:

Accelerate Processes	AI may increase the pace of particular activities, be that generating material(s), trial and error processes, or predictive capabilities, that reduce overall costs of developing, acquiring, or using CBW.
Generate Pathways	The act of "generating pathways" refers AI's use to surpass assumed constraints and expand possibilities compared to existing routes, processes and mental models. Pathways thus emerge when prior boundaries are weakened. New pathways can erode limitations and uncover new spaces which facilitate the pursuit of CBW.
Mediate Ideas	Mediating ideas refers to influencing concepts, beliefs and assumptions by exposing individuals or groups to new ideas with the aim of catalysing shifts in relationships, behaviours, and norms. This frames idea mediation as an active redirection of thought patterns by selectively introducing ideas that realign activities by expanding or contracting perceived constraints.
Modify Transparencies	AI's potential impacts on transparency extend to the capacity to fully understand how AI systems make decisions, and other impacts on assumptions and relationships underpinning information flow in governance systems.

Table 2 - Four dimensions of impact of AI technologies

These four processes can ground and open-up analysis of how any AI technology can be applied to achieve effect on the CBW eco-system. They provide a way to visualise what is happening in real terms for actors involved. Accordingly, these dimensions of impact facilitate the translation of AI from an abstract phenomenon to enabling tools.

By viewing them as enabling tools which generate particular impacts, they can more clearly be associated and understood within the context of the major categories of challenge that face the CBW prohibitions. Such challenges are not new, and have helped to shape analysis and governance of CBW so far. These continue to be instrumental in defining types of negative outcome that result in a weakening of the CBW prohibitions, and remain compatible and instructive in the age of AI.

3.2 Categories of challenge to the CBW prohibitions

The previous section noted that the four processes AI facilitates can have material impacts on the CBW prohibitions. Drawing on the project data mentioned earlier and the

work of Julian Perry Robinson, we identify four primary categories of challenge that broadly frame how particular processes and activities can lead to negative outcomes.[7]

These challenges are not limited to a single activity or technology; instead they provide a framework to understand how and why particular outcomes negatively effect the CBW prohibition. These challenges show us what processes we must guard against. As such, they are particularly useful for framing how and why AI's integration into the CBW eco-system can have downstream impacts on the prohibition regimes. Table 2 provides a brief overview of these challenges, with simplified descriptions tailored for a focus on AI.

The utilities of CBW	AI may influence decisions to adopt or utilise CBW, through revealing potential utilities and targets, lowering resource/skill barriers, and enabling rapid, instrumental scientific and technical knowledge. All of these elements shape perceived tactical and strategic utilities and alter cost/benefit calculations. This expands the potential for both new and old utilities of CBW to (re)emerge.
Circumvention, proliferation and acquisition	Preventing state and non-state actors circumventing the multi-layered "webs of prevention" is a key challenge. AI can support actors in acquiring CBW relevant materials, developing new routes for production and testing, facilitating efforts to evade detection, reducing human involvement, providing new sources for planning and targeting, and presenting new modes of delivery.
Creeping legitimisation	AI systems could incrementally dilute CBW prohibitions through facilitating a creeping accumulation of minor intentional and inadvertent infractions, through, for example, unintentional dual-use research and grey-area CBW activities that probe and stretch norms over time, entrenching ambiguity and fragmenting trust.
National interests and normative divergence	Rapid AI advances will outpace cooperative governance, and may exacerbate diverging normative and strategic assessments of potential CBW value, due to real and perceived asymmetries in benefits, oversight, accountability, and power.

Table 3 - Four categories of challenge to the CBW prohibitions

[7] Robinson, J. P. P. 'Difficulties facing the Chemical Weapons Convention' (*International Affairs*, 2008) 84(2)

4. The purpose of the vignettes

The influence of AI on CBW acquisition processes is complex and multifaceted. To fully understand AI's implications, we must look more widely than the development and synthesis stage, and consider its impact on the entire on the other stages and processes (see Diagram 1), and on the wider eco-system. Indeed, AI's influence extends to issues such as disinformation, allegations, investigations, and accountability, which lie outside the acquisition process but within this broader CBW ecosystem.

Actors pursuing CBW may take different acquisition paths: well-resourced states may focus on development and synthesis of both the agent and the delivery device, while opportunistic non-state actors may bypass these stages, opting for theft or diversion and crude dissemination. Our four-dimensional framework – acceleration of processes, generation of pathways, mediation of ideas, and modification of transparencies – provides a structured approach to investigating what a value a particular AI application may bring to an actor's intention at any stage in the acquisition process. By analysing how and why actors leverage AI for CBW-related objectives, we gain a grounded view of AI's role and relationship with motivation, intention, and facilitation.

Case studies and scenarios are essential for contextualizing AI's impact on the CBW domain in this regard. They provide space and the means to translate abstract concepts into tangible examples, offering insights that can inform the development of our broad understanding of the contexts in which AI is used and why, and also for considering mitigation strategies and prevention measures.

The vignettes in this report, while brief, illustrate how AI could intersect with the CBW acquisition process and ecosystem. They serve as a foundation for further work to contextualize and inform policy-making and stakeholder engagement efforts.

When reflecting on these vignettes, it is crucial to consider the broader context and the ultimate goal of driving meaningful action to address the complex implications of AI for CBW-related issues. In that vein, it is important to underscore that the vignettes are designed only to be *representative* of our findings and do not in way indicate or imply any statistical likelihood of particular events, or the probability or risk level of any activities or impacts. The content of the vignettes is designed specifically to elicit reflections on the implications, and not to invite specific consideration of the narratives themselves as bearing any resemblance to real or potential events.

To support reflection, each vignette is accompanied by a small table that highlights which of the dimensions of impact, and which challenges to the Convention, are most obviously represented. Then, a short analysis is provided:

- Impacts: How is AI used in this case?
- Intent: Why is AI used in this case?
- Implications: What can be done?

This analysis provides an entry point for demonstrating how these framings and vignettes can support a much deeper unpacking of the scenarios and potentials that may evolve

through the increasingly relevant applications of AI in the context of CBW. Readers are invited to come to their own conclusions and consider their own responses to the vignettes.

They do not, however, provide detailed policy response prescriptions, as this falls outside the scope of the current project. However, the analysis clearly signposts the way forward for this crucial next step.

4.1 Vignette selection and development

Although illustrative, these vignettes have been crafted to closely resemble contemporary and emerging challenges under discussion in (inter)national CBW policy fora and beyond. Set within contexts and discussions reflecting near to medium term risks, the vignettes assume that current AI dynamics, including the pace of change and increasing accessibility of technologies, continue.

The content of the vignettes was inspired and shaped in relation to data drawn from a variety of sources, including academic, policy, and grey literature reviews, and from the data collected by the project team over the course of the project. Interviews and informal discussions with members of the CBW and emerging technologies communities were instrumental in developing a landscape of AI-anxieties which informed two workshops, facilitated by BASIC – The British American Security Information Council. These futures-oriented workshops generated a significant amount of data through particular group exercises, including a 'Futures Wheel' exercise, identification of drivers of change, co-creation of AI-CBW future headlines, and policy intervention development exercises. These workshops produced significant insights and data relating to AI's technological capabilities and applications, conceptualisations of CBW, current and future scenarios, pathways to obstacles, and on intervention options. The data both strengthened existing concepts developed by the project, in particular relating to AI-anxieties and dimensions of impact, and considerably expanded our view on what and how AI integrates and relates to CBW.

As such, the following vignettes are a direct response to, and follow on from, the insights and implications raised through the interviews and workshops. Care has been taken not to reproduce the work of others, and instead develop parallel or adjacent vignettes which draw on the concepts and ideas developed while providing new contributions to the data generated by this project.

The writing of these vignettes was done with a number of criteria or specifications in mind. These were important to capture in as the different criteria or specifications were highlighted as being valuable or insightful through the project and its activities.

The first was that the vignettes as a collection should seek to engage with all of the stages of the CBW acquisition process, rather than just the development stage. This was in response to our evaluation of the literature being too heavily skewed toward the development stage with less attention to other stages: this view was validated through interviews and the workshops. Therefore, we have ensured that all stages, as illustrated in Diagram 1, are treated within at least two vignettes.

The second was that we wanted to demonstrate that an all-actor approach to considering AI and CBW is important, and that assumptions about which actors engage in which processes should be not be made. Therefore, each pair of vignettes that deal with specific stages of the CBW acquisition process are also split between a state actor and a non-state actor to demonstrate that effort to understand how and why actors might engage in different stages is valuable.

A third was to demonstrate how, in practice, the dimensions of AI can enable a particular actor in a particular setting to advance their objectives, although this may be to varying degrees of success. Each vignette seeks to reveal how these dimensions of impact manifest in real terms, providing a perspective on why and how the AI brought value to the actor.

To demonstrate the importance of connecting and framing these scenarios as challenges to the CBW Conventions, a fourth aspect was to ensure that we brought to the surface how the activities describes have tangible implications for the Conventions and anti-CBW policy.

Of course, not at all vignettes cover all dimensions of impact or challenges to the Convention: this would require a significant increase in the amount of vignettes produced. Importantly, however, this demonstrates that the appearance and weight of these dimensions and challenges will not always be equal. Moreover, these element and challenges are not isolated from each other and while they have been parsed out for the benefit of a short vignette, in practice too there is overlap and interconnection. To that end, the vignettes provide a rather sanitised analysis, although we believe much deeper investigation in each case would reveal precisely the complex overlap and interconnection that makes it so important in the first instance to define our dimensions and challenges to support such analysis.

The effort here has been to produce short vignettes that situate AI within the socio-technical CBW systems, and show the benefit of moving beyond a technological determinism. This, we hope, can advance pragmatic policy discussions regarding the mitigation of potential harms. Rather than conclusively determining the future of AI-CBW interactions, these vignettes elucidate plausible scenarios based on project data, and expand our understanding of how AI may impact the CBW ecosystem, thus informing how policy and governance may need to respond to these challenges. Once again, the critical aspect is what the vignettes demonstrates and implies, and not the specific content or fictional renderings.

Table 3 summarises the eight vignettes and provides a quick-reference key to identify: which actor type is involved; the central processes of acquisition that AI is employed to support in the vignette; the AI's dimensions of impact on those processes; and how the negative outcomes relate to challenges to the CBW prohibitions.

5. Eight Vignettes

V1. State disinformation campaigns and false allegations

A large state actor seeks to undermine a bordering state with which they have a history of tensions by alleging a biological weapons attack in a disputed border region. The large state actor aims to exploit the ambiguity surrounding the incident to gain the confidence of the international community and delegitimise the bordering state and its leadership, thereby bolstering its claim over the disputed territory.

The large state actor has been investing in AI research and development to monitor and analyse global news, social media, and scientific literature. The state actor realises that this dataset can also be used to train AI models for generating realistic disinformation to support external propaganda strategies. Alleging that the bordering state has used a biological agent to attack villages in the disputed region, the state creates a portfolio of fictitious information, ranging from fake news reports, social media posts, photographs and video imagery, expert analysis, and academic papers and studies. To support their false narrative, the team uses AI-powered video and image manipulation tools to create fake footage of the alleged attack.

As the bordering state and the international community scramble to investigate the allegations, the AI models work to generate new content and narratives that create further confusion and doubt. Social media platforms' algorithms and trends are leveraged to ensure the fake content reaches a wide audience, targeting key opinion leaders and media outlets to amplify the message. AI-generated content further includes unofficial video and images of the border nation's armed forces acting out various 'trends' in the disputed territory, close to sites of violence.

The ensuing public outcry and amplification of faked content leads the UN Secretary-General to activate the UN Secretary-General's Mechanism, convening an expert group to investigate the veracity of these claims. Simultaneously, the large-state calls for a meeting of States Parties to Biological Weapons Convention under Article V and requests assistance under Article VII. Meanwhile, the large state increases and enhances armed occupation of the disputed territory under the guise of civilian protection.

Quick Look

Process Acceleration: Rapid production of effective propaganda strategies

Transparency Modification: Ease of making multiple forms of evidence for alleged use

Normative Divergence: Changing trust in evidence and governance processes

	U	C	CL	ND
PA				X
PG				
IM				
TM				X

Impacts: How is AI used in this case?

AI is being used here to *accelerate the processes* of generating disinformation about a false biological weapons attack by rapidly creating large volumes of fake content and disseminating it at speed. Furthermore, the use of AI is specifically *modifying transparency* by creating a fog of confusion and doubt centred on an alleged attack. By creating many fabricated evidence and reports, each potentially also referring to one another, the state is able to hinder investigation efforts and make it difficult to trace sources. The ability of the actors to thus create academic and journalistic articles which provide further evidence from conventionally 'trusted' sources further adds to the ease of muddying the information environment. From the perspective of the CBW treaties, this can exacerbate divergences in national interests and normative understandings of BW, as states may have different interpretations of the allegations. Ultimately, proliferating mis- and disinformation may lead to the erosion of trust and confidence in the credibility of international commitments made by states.

Intent: Why is AI used in this case?

The state actor sees value in using AI for disinformation because it allows them to utilise the taboo of BW as a tool for geopolitical manipulation and reputational damage. The actor's choice to fabricate a biological attack, rather than a conventional attack, is in its ambiguity and shock value. AI-generated false evidence enables the state actor to accelerate and expand its disinformation campaign, necessitating a lengthy and resource-heavy investigation process. The delay in confirmation, as well as the capacity to fabricate further false evidence, enables the state to justify further militarised actions during the confusion. In effect, the resource cost of developing such an intensive campaign is heavily reduced and enables a much more effective and long-term set of uncertainties to manifest which will provide a smokescreen for territorial expansion.

Implications: What can be done?

This vignette shows how AI's capacity to dramatically proliferate information and create mass confusion. It is therefore important that we consider policies related to verification and assurance of the information environment. This includes CBW-specific elements like:

- Developing guidance and training for investigators who are part of UNSGM investigations, OPCW Joint Investigations, etc.
- Strengthening allegations processes under BWC Article V and CWC Article XII to enable more rapid de-escalation to reduce the potential utility of false allegations

Moreover, the simple proliferation of mis-/disinformation and fabricated academic articles is a cause for more general concern. Other, more general implications could be addressed by:

- Public education on media literacy, with particular emphasis on recognising fabricated media and confirming across sources
- Enhanced quality control and auditing of scientific and academic results, data, and articles

V2. Cloud-based terror: non-state actors and outsourced acquisition

A terror organisation is covertly seeking to inspire crude, opportunistic attacks by unconnected solitary actors or cells on vulnerable local targets. They have an interest in chemical weapons (CW) due to their perceived utility for causing panic, disruption, and stretching emergency response resources. They also seek to reduce the likelihood of detection by moving planning from the central group to the network.

The organisation therefore begins distribution of propaganda to followers, aiming to inspire followers to take initiative and make tactical decisions themselves based on local conditions. They use widely accessible AI instructional materials to develop tailored propaganda and disinformation output for their supporters, legitimising the use of disruptive CW. This manipulated media provides a degree of moral cover alongside the technical manuals guiding improvised attacks.

Once the propaganda campaign has been established, the information and instructions needed to independently acquire and deploy CW are disseminated through more subversive channels. Lacking in-house technical expertise, they recognise that the internet contains much of the information needed to support acquisition. The central group therefore decides to collate the location of relevant open access datasets, instructions on how to access them, and guidance on crafting prompts for AI chatbots that overcome information restrictions. These documents also systematically consolidate scattered data from extremist web forums regarding impromptu weaponisation of basic toxic industrial chemicals or naturally occurring biotoxins using agriculture supplies, kitchens, and amateur lab equipment.

The public availability of large language models thus enables codifying the dispersed and unorganised data into coherent, customisable, do-it-yourself manuals with illustrated step-by-step guidance. This guidance enables non-experts with localised access to household materials to develop quick and dirty CW. The manuals emphasise using non-suspicious substances within public reach and improvising delivery mechanisms suited to different substances for different purposes. The group hopes that the use of these manuals by independent actors creates waves of small-scale attacks and assassinations, even failed attempts, will stretch emergency services and make a centrally planned attack more difficult to prevent or mitigate.

Quick Look

Pathway Generation: Synthesis pathways from household chemicals

Idea Modification: Popularising and raising awareness of CW

Utilities: Bringing CW to broader populations

Circumvention: Reduces need for controlled chemical precursor

	U	C	CL	ND
PA				
PG	X	X		
IM	X	X		
TM				

Impacts: How is AI used in this case?

The vignette poses challenges to the CBW Conventions by highlighting how AI can enable particular utilities for chemical weapons, by decentralising the actors and individual acquisition processes. It illustrates how AI can generate pathways for radicalised non-experts to manufacture CW with malicious intent. Moreover, AI is mediating ideas by providing moral rationalization for the use of chemical weapons, potentially eroding the norm against their use. These AI-driven processes can lead to the circumvention of controls, proliferation of knowledge and materials, and proliferation of individual access and use of CW.

This challenges the norm against chemical weapons use and raises concerns about non-state actor acquisition. The AI-driven dissemination of knowledge and moral rationalization can lead to normative divergence, eroding the consensus against chemical weapons. The Conventions must adapt to address the role of AI in lowering barriers to entry and enabling new actors to acquire and use chemical weapons.

Intent: Why is AI used in this case?

The terrorist organization here sees value in AI as it allows them to inspire and guide followers to independently acquire and deploy chemical weapons while maintaining operational secrecy. AI's value is not directly in what it can achieve for the non-state actor, but in how it can be leveraged by their followers. The actor's dissemination of knowledge and AI's generation of pathways for individuals to overcome technical and regulatory hurdles both reduce the need for followers to independently initiate and investigate potential uses of CW. By mediating ideas and providing moral justification, the use of AI also helps to erode any individual ethical resistances to the use of such indiscriminate weapons. In effect, the use of AI is directly about modifying intent and motivation of followers to pursue CW, rather than about a shift in perceived central utility.

Implications: What can be done?

The implications of this vignette are broad ranging, given that there is no central actor or central set of technologies/platforms being deployed. That being said, from a CBW perspective, there are some specific areas of activity which might be of use:

- Consideration in the next Programme of Work (22nd) of the 1540 Committee of a working group on digital proliferation
- Enhanced preparedness and investment in emergency services surge capacity in the event of such decentralised attacks

Similarly, there are a number of general implications which, if addressed, may also help to mitigate the decentralised proliferation of CBW. These include:

- Greater investments in active deradicalization in online spaces and social media investigations, e.g. updating the INTERPOL/UNCCT "Using the Internet and Social Media for Counter-Terrorism Investigations" Handbook
- New research into popular Chatbot jailbreaking techniques and how these could be overcome technically

V3. State AI exploitation of genomic data for targeted BW

A multiphase AI filtering system, "Nemesis Bio," is engineered by state-sponsored programmers to autonomously identify Single Nucleotide Polymorphisms (SNPs) specifically valuable for revealing population-wide immunological vulnerabilities. Ostensibly searching for different ways that its own population could be vulnerable to communicable disease (of any origin), natural language processing algorithms are intentionally trained to construct pathogens which overcome immune defences.

Over five years, however, a state-sponsored advanced persistent threat (APT) group has infiltrated a number of foreign private, academic, and state-based repositories of genomic, immunity, and pathogenicity data. Nemesis Bio includes an AI module to catalogue servers accessible to the APT that house relevant research data worldwide. This seeks to link traditionally disconnected, siloed data sources to create an aggregate dataset of genomic data, revealing genetic susceptibilities across populations and human subgroups. Rather than risk bulk data theft, Nemesis Bio then compiles the data to precisely extract fragmented excerpts with assumed utility for weaponising pathogens against population defenses.

By interconnecting discrete research from siloed institutions, this long-term and targeted data aggregation creates new systemic insights into populations' genetic susceptibilities. This state-orchestrated campaign creates unprecedented insights into systemic human weaknesses that could support the development of computational platforms powered by AI to reveal targeted opportunities for engineering pathogens. This could be tailored to target particular ethnic genome subsets, defeat future public health defences through immuno-circumventing techniques, or target novel bioweapons to specific populations.

The interconnection of genomic and immune datasets by the APT, the state reasons, provides sufficient separation and plausible deniability that the government's involvement is obscured. Not only this, but the agglomeration of data from dispersed datasets ensures that genomic markers associated with pathogens and their targets are similarly dispersed, making the origin of the pathogen more difficult to ascertain and thus attribute.

Quick Look

Pathway Generation: Synthesis pathways from household chemicals

Transparency Modification: Popularising and raising awareness of CW

Utilities: Bringing CW to broader populations

Circumvention: Reduces need for controlled chemical precursor

	U	C	CL	ND
PA				
PG	X	X		
IM				
TM	X	X		

Impacts: How is AI used in this case?

The AI system in this case, "Nemesis Bio," is accelerating the process of identifying and collating population-wide vulnerabilities by rapidly infiltrating and analyzing genomic, immune system, and disease progression data. It is generating pathways for potential BW development, uncovering new possibilities, reinvigorating discontinued research avenues, and suggesting new approaches in the effort to develop targeted biological weapons. The AI modifies transparency by obfuscating the data collection process and aggregating data from disparate sources in a clandestine manner, reducing the footprint of human involvement in this novel form of scientific espionage.

The vignette poses challenges to the CBW Conventions by demonstrating how AI can enable the discovery of new utilities for BW, such as micro-targeted agents that exploit genetic susceptibilities. This challenges the norm against BW by threatening to erode the strategic drawback found within BW's indiscriminate nature. Similarly, building composite datasets from multiple sources enables the circumvention of existing cybersecurity measures.

Intent: Why is AI used in this case?

The state sees value in the endeavour both in terms of potential impacts from BW use, but also from the plausible deniability that the group and their methods provide. The state-sponsored group sees value in AI as it enables them to rapidly identify potential targets and develop novel biological weapons. Moreover, the sponsorship from the state overcomes some of the barriers around resource and expertise that more traditional non-state actors might experience. AI use generates new pathways for weapons development and modifies transparency, similarly lowering the costs of BW pursuit. This alters the perceived utility of biological weapons and provides new opportunities for their use.

Implications: What can be done?

This vignette highlights AI-anxieties related to the security of sensitive genomic data and the potential for AI to enable the development of targeted BW. It illustrates how AI can accelerate the identification of vulnerabilities, generate pathways for novel weapon development, and modify transparency in the research process; emphasizing the importance of wider cybersecurity. Potential avenues for addressing these include:

- International cooperation in governing the storage, security, and access to personalised (particularly genomic) datasets, including in-storage decoupling of genetic data from health and medical data
- Close collaboration with Industry actors to preserve data protection and encourage reporting of breaches of various scales
- Enhancing investigative capabilities within OPCW and UNSGM rosters with the inclusion of additional cyber-forensic experts and training

V4. AI, the dark web, and the circumvention of controls

Rapid developments in AI have led to the creation of sophisticated systems across various industries, including chemical manufacturing and supply chain management. Convergence of AI systems and automated production has led to the creation of so-called 'cloud labs' that are able to receive, process, and ship orders of chemicals independently. These labs have the potential to revolutionize various industries both in terms of generating new business opportunities and efficiencies, but also in restructuring chemical production supply chains.

Hackers regularly breach the cybersecurity of companies that deploy these AI systems, stealing the underlying models, training data, and source code; selling these systems on the dark web either in parts or as chatbot platforms. An ecological activist, exploring the idea of acquiring chemical weapons to disrupt forest monocultures in their local area, purchases several of these AI tools and chatbots, including those designed for chemical manufacturing, supply chain optimization, and scientific research.

Using these AI systems, the individual begins planning and developing options for acquiring a chemical defoliant; identifying necessary precursors and equipment, optimizing production processes, and circumventing regulations to acquire materials without raising suspicion. The AI also assists in gathering information on dispersal methods and specific, rapidly degrading chemicals which will maximize harm to monocultures but minimize wider ecological harms.

As the individual progresses, they encounter a challenge in acquiring a specific chemical mixture crucial to their design, due to the complex and specialised equipment necessary for that particular synthesis step. Unable to produce the mixture independently, they place an order to the cloud-lab from which the AI was stolen originally. Thanks to the individual's newly developed tacit knowledge in using prompts which 'jailbreak' the lab's AI, the lab readily produces the necessary chemical mixture and ships the compound to the provided address.

Quick Look

Process Acceleration: Rapid trial and error with different potential chemical methods

Pathway Generation: Synthesis of chemicals needed without access to specialised equipment or skills

Utilities: Bringing CW to broader populations

Circumvention: Evades both chemicals export and technical barriers

	U	CPA	CL	ND
PA	X	X		
PG	X	X		
IM				
TM				

Impacts: How is AI used in this case?

AI systems are accelerating the time taken to undertake particular processes in the planning and development of CBW use. For example, through rapid identification of precursors, optimization of production, and navigation of regulations. The use of AI also generates previously obscured pathways for acquiring materials; by outsourcing the production of a critical mixture to a cloud-based laboratory, the need for individual skills is reduced.

The AI-driven acceleration of planning and generation of acquisition pathways reveals utilities for individual users and circumvents regulations by using growing knowledge of AI 'jailbreaking'. The vignette therefore poses challenges to the CBW Conventions by showcasing how AI can lower barriers to access for chemical weapons, thus challenging the norm against their development and use.

Intent: Why is AI used in this case?

The malicious actor here sees value in AI as it enables them to circumvent existing control mechanisms and regulatory oversight, streamlining the acquisition of chemical weapons. AI use enables the individual to plan the process of acquisition, generate pathways for acquiring materials, and circumvent existing regulation of chemical precursors. This lowers the barrier to entry for non-state actors and increases the perceived utility of chemical weapons.

Similarly, as the vignette evolves, the actor's intent changes when a complex barrier is found. The costs of proliferation are now increased, but the actor's own newly gained tacit knowledge enables the circumvention of this barrier thanks to a broader shift towards automation in the chemicals production industry.

Implications: What can be done?

This vignette highlights AI-anxieties related to the theft and misuse of proprietary AI systems for illicit purposes, such as chemical weapons development. It illustrates how AI can generate pathways for acquiring materials and modify transparency to evade detection. These AI-driven processes can lead to the circumvention of controls, proliferation of knowledge and materials, and the erosion of norms against chemical weapons. Activities which may be of value in addressing these harms include:

- Broader engagement with chemicals, manufacturing, and production industries to understand existing quality and security assurance processes in place for AI-enabled production, particularly around ISO 9001, ISO 14001, and ISO 27001.
- Consideration in the next Programme of Work (22nd) of the 1540 Committee of a working group on digital proliferation

V5. Grey area research, law enforcement and military spill-over

Scientists at a national biodefense laboratory are tasked with researching potential future biochemical threats, including toxins that could be weaponized for assassinations and small-scale use. The lab utilizes an advanced AI system for high-throughput computational screening of novel peptide toxins, allowing the more rapid exploration of different peptides and receptor targets compared to more classical experimental methods.

The AI system identifies a novel conotoxin peptide, derived from cone snail venom, that is predicted to cause temporary flaccid paralysis. The researchers see potential for understanding and defending against toxins as well as potential novel analgesics. They recognize that while conotoxins were previously studied for offensive use, the research was discontinued due to technological limitations and legal restrictions. However, with the advanced AI capabilities, the scientists believe that this research could now be continued within the concept of defensive research.

The researchers report the simulated results to the oversight committee, emphasizing the defensive applications of the findings. However, military leaders see the potential for novel incapacitation applications in domestic crowd control operations. They instruct the lab to synthesize the peptide and further research its mechanisms, scalability, and potential routes to use for riot control, arguing that a fuller understanding will enable stronger countermeasures and deterrence.

The scientists feel conflicted, recognizing that they are being compelled to translate defensive knowledge into potentially offensive applications. However, they convince themselves to follow orders, hoping a non-lethal tool could prevent greater violence and aid domestic law enforcement. However, the AI also highlights unforeseen variants that could permeate the blood-brain barrier and generate systemic, potentially lethal effects.

Nevertheless, the state government compels the lab to scale up production and develop different delivery systems, ostensibly for defensive research, which leads many of the lead scientists to threaten to resign. The state ignores their warnings, and the remaining scientists are eventually instructed to develop their production and delivery systems into weaponised offensive capabilities.

Quick Look

Process Acceleration: Identification and optimisation of peptides

Idea Mediation: Scientists consider *in-silico* methods less problematic

Creeping Legitimisation: Defensive research leading to offensive ideas

Norm divergence: State population control aims supersede scientists' moral concerns

	U	CPA	CL	ND
PA			X	X
PG				
IM			X	X
TM				

Impacts: How is AI used in this case?

In this vignette, AI is accelerating the process of discovering and optimizing novel peptide toxins by rapidly screening vast chemical spaces and suggesting modifications to enhance their potency and effects. This acceleration is initially presented as a component of defense research, and overcomes moral/legal concerns through the use of *in silico* methods that do not require direct experimentation on live subjects, nor production of the peptides. The AI also indirectly mediates ideas by presenting unforeseen, potentially lethal, options that offer novel methods of incapacitation and harm to offensive actors. In both cases, the emergence of such weapons from previously defensive research reveals the potential for acerated serendipity in AI-enabled R&D, furthering the creeping legitimisation and potential normative divergence through more rapid technical change.

Intent: Why is AI used in this case?

The scientists initially see value in using AI for this research because it allows them to rapidly explore a wide range of potential biochemical threats and develop potential countermeasures. However, this vignette demonstrates the unpredictability and instability of assumptions about intent, motivation, and normative adherence. Here, AI's ability to accelerate the discovery and optimization process, as well as serendipitous findings, alters the intentions of the military toward first developing domestic riot control capability, and subsequently developing offensive capability. The use of AI to lower resource-costs enhances the attractiveness of potential benefits of these agents. This demonstrates how intentions and motivations can be manipulated and revised at any stage of the process, enabled through the interconnection of AI impacts and human interventions.

Implications: What can be done?

This vignette poses challenges to the CBW Conventions through legitimising the acquisition of CBW with a veneer of scientific and strategic justification for deployment. The use of AI in this context can also lead to the erosion of norms against CBW by blurring the lines between defensive and offensive applications and creating a slippery slope towards their acceptance and normalization within specific actor groups. To address these challenges, the Conventions may need to:

- Strengthen transparency and verification measures, including through enhanced confidence building measures
- Continue efforts to develop understandings at the international level regarding acceptable domestic law enforcement research and use
- Promote international cooperation and information sharing in scientific research
- Reinforce existing responsibility and ethics training for scientists through programmes like iGEM and IBBIS
- Develop and strengthen whistleblower protections for scientists at national and international levels

V6. AI and disease-based gang violence

A well-funded criminal organization, seeking to eliminate rival gangs without attracting attention from authorities or inviting reprisals, is exploring ways to leverage AI to develop and deploy a biological agent that can be delivered to their rivals in a way that appears natural, such as through contaminated food or other means.

The organisation recruits undergraduates from the local area with knowledge of AI chatbots and data science. The hired team uses publicly available AI tools and datasets to identify pathogens that can cause severe illness or death, including on how and why pathogens can be cultivated outside of the lab, and their persistence in the environment. Assuming their role is to support efforts to improve and mitigate health and safety for a chain of restaurants, the hired team also provide information on food production and supply chain vulnerabilities as part of their role.

Using this knowledge, the criminal organisation focus on pathogens which are relatively common in food poisoning already, seeking to balance effectiveness with the need for a low-tech approach that minimizes the risk of detection. As the planning process advances, the hired team uses AI to identify potential cultivation and delivery mechanisms for their biological agent. Finding the results to be overly complex and expensive, the organisation shifts perspective to other means of causing disruption.

After months of dormancy, though, the team hear of a local *E. coli* outbreak at a restaurant. They collect samples of the food, and seek to further cultivate *E. coli* in their makeshift lab. Drawing on the data from the hired team on food production and supply chain, the organisation introduce infected food products at wholesalers and groceries identified as supply their rival's outlets.

Eventually, the outbreak of illness draws the attention of public health authorities, although the effectiveness of the operation is low, due to the dissemination being spread too thinly, the effects being indiscriminate and effecting the local population too, and the public health system minimising the resulting illnesses. Public health inspectors trace the outbreak back to the three specific deliveries, their use of a recent outbreak to supply them with the agent challenges the authorities' ability to trace the source of the outbreak.

Quick Look

Process Acceleration: Identification of pathogen and targets

Idea modification: Ability to target enabled opportunistic behaviour

Utilities: Expansion of potential strategic use of natural sources

Circumvention: AI supports non-traditional and low-tech acquisition

	U	C	CL	ND
PA	X	X		
PG	X			
IM	X			
TM				

Impacts: How is AI used in this case?

In this vignette, AI is being used to accelerate the process of identifying and selecting pathogens that can cause severe illness while mimicking natural disease outbreaks. It is also used to collate and synthesis information and knowledge more efficiently than if the organisation sought to do so themselves. As such, the AI was used to accelerate the various processes required to move through the acquisition stages. The AI is also being used to optimize the collation of knowledge required to support efforts to cultivate pathogens in non-technical settings with minimal resources and footprint. Additionally, the AI is generating pathways for delivering the bioweapon by analysing patterns of behaviour and supply chain vulnerabilities in the rival gangs' operations.

The ability of AI to optimize pathogens and delivery methods for specific contexts and targets highlights the potential for AI to help actors circumvent obstacles, although in this case the initial effort encountered enough obstacles to cause the team to abandon its pursuit. The vignette, however, highlights how AI may further enable more opportunistic use of CBW. The team were able to conduct an attack partly due to reinvigorated motivations, but also due to AI-enabled targeting which increased the impact of a relatively low-tech acquisition and dispersal.

Intent: Why is AI used in this case?

The criminal organization sees value in using AI because it allows them to develop an biological agent with relatively low-tech resources, reducing the risk of detection and attribution. The AI's ability to quickly identify suitable pathogens and optimize their properties makes it an attractive tool for achieving the organization's goals without the need for advanced laboratory equipment or expertise. This initially supports motivations to develop a biological agent for a particular utility. When this proves challenging, motivations are reduced, until a chance outbreak of *E. coli* was able to be leveraged. Previous AI-enabled data reinforced their intention, as a viable and effective delivery and dissemination option coincided with concurrent activities.

Implications: What can be done?

The use of AI by criminal organizations to develop low-tech biological weapons demonstrates how even relatively unsophisticated actors can leverage AI to enhance their motivations, lower the barriers to entry, and increase the risk of their use in targeted attacks. In particular, this demonstrates the varied forms of utility that AI technologies can support, and how this support can modify intent beyond initial planning stages. In these cases, the primary driver is the level of impact and possibility of attribution which mediates intent. These can be managed through:

- Greater investment in public health investigative capabilities and capacities
- Increased monitoring and security of samples and materials during and after investigations, including the destruction of potentially contaminated materials during contact tracing
- Enhanced full and effective national implementation of the CBW Conventions

V7. Target profiling at the margins of CBW

A state's law enforcement agency (LEA) has been investing in AI-powered systems to assist with crowd control and the deployment of riot control agents (RCAs). The LEA decide to purchase an off-the-shelf solution for doing so, eventually deciding on a platform that has been marketed for use in military situations to predict combatants' movements and identify chokepoints. The AI is trained on historical data, including images and videos of past protests, riots, and crowd behaviour, in both domestic unrest and international conflict.

During a period of heightened social unrest, the LEA relies heavily on the AI system to assess the threat level posed by crowds and recommend the appropriate amount and method of RCA dispersal. The AI analyses real-time footage from drones and CCTV cameras, identifying potential agitators and estimating the demographic makeup of the crowd, including factors such as age, gender, and ethnicity.

During a peaceful protest, the AI system identifies a few individuals with prior arrests and a specific demography within the larger crowd, and thus recommends the use of smoke. Moreover, an optimal location and quantity of tear gas is determined based on real-time environmental analysis. Following the AI's recommendations, the law enforcement agency forces the protestors into the defined location and deploys a large quantity of RCAs, resulting in severe injuries among the protestors and the public, including vulnerable individuals such as children and the elderly.

Independent civil society investigations suggest that the AI system had overestimated the threat level due to biases in its training data. The investigation also reveals that the parameters of the AI had not been modified from its military origins and thus aimed to create chokepoints rather than disperse crowds.

As public outrage grows, the local LEA seeks to absolve themselves of responsibility and claims that the state's LEA provided them with a military AI with little to no training in how to use it. Later, leaks from within the local LEA suggest that the AI's recommendations were simply accepted without question and that no oversight had been provided to ensure that decision-making was complemented by on-the-ground information.

Quick Look

Process Acceleration: Identification of threats and targets

Transparency: Decisions made by AI, local and state agencies blame each other

Creeping Legitimisation: RCAs combined with military AI technology

Norm Divergence: Governance of AI not disseminated to lower scales from state level

	U	C	CL	ND
PA		X	X	X
PG				
IM				
TM			X	X

Impacts: How is AI used in this case?

The AI system in this vignette is being used to accelerate the process of threat assessment and decision-making in the context of crowd control. It generates new pathways of analyses for decisions, integrating real-time footage and demographic data. Together, these generate recommendations based on a more complex analysis than would be possible within average human-driven timeframes, thus being used to circumvent human elements that would slow the process of targeting and disseminating the RCA. This enables accelerated recommendations for the deployment of RCAs by streamlining the process. However, the AI's biases and lack of transparency in its decision-making process lead to the disproportionate targeting of minority communities and the excessive use of force, bringing considerable ambiguity to the question of intention and accountability. The AI is therefore also modifying the transparency required to understand how and why decisions were reached. mediating the law enforcement agency's perceptions of threat and appropriate response, with harmful consequences.

Intent: Why is AI used in this case?

The law enforcement agency sees value in using the AI system because it promises to provide rapid, data-driven assessments of crowd behaviour and threat levels, enabling quicker and more efficient decision-making in high-pressure situations. The AI's ability to process large amounts of real-time data and provide recommendations based on historical patterns is seen as a way to enhance the effectiveness of crowd control strategies. However, the vignette illustrates that relying on AI without sufficient human oversight and consideration for potential biases can lead to severe unintended consequences, which unaddressed facilitate a creeping legitimisation of such tactics. The long-term concern therefore is that the lack of accountability for such activities can lead to a modification of motivations for those intent on increasing the militarisation of RCA use in domestic law enforcement and in international armed conflict.

Implications: What can be done?

This vignette poses challenges for the CBW Conventions in terms of ensuring that the development and use of AI systems for RCA deployment comply with the principles of necessity, proportionality, and non-discrimination. The Conventions must address the potential for AI to enable the misuse of RCAs and the disproportionate targeting of certain groups. The Conventions may also need to consider the implications of AI-assisted RCA deployment for the overall norm against chemical weapons.. Options include:

- Ensure visibility and agenda-time for RCA issues within the OPCW
- Increased regulation of the sale and use of military equipment in domestic law enforcement settings
- Enhanced efforts to develop and implement AI ethical standards education for individuals involved in AI-driven decision making
- Development of specific guidelines and regulations for the use of AI in the context of law enforcement and crowd control

V8. Enhanced dissemination of CBW

Gaian Defence (GD), an extremist environmental group, seeks to overwhelm emergency responders and cause public chaos by dispersing legally and illegally acquired irritants, malodorants, and toxic industrial chemicals across five major towns and cities. Lacking technical expertise, they turn to AI for targeting support.

Members discreetly acquire largely unrestricted lachrymatory agents, chlorine, and odorants through common channels, intending to provoke non-lethal harm through inflated concentrations in public spaces. However, optimal deployment poses distinct operational challenges. GD initially identifies that training machine learning algorithms on urban environmental data could direct their dispersal approach. They subscribe to online deep learning courses and research tools advertised for meteorological analysis. However, members struggle to make progress on their own.

Pivoting from commodity solutions, GD reaches out discreetly to machine learning contractors on the dark web, asking for custom neural network code tuned for their "research on optimizing rodent control in major urban areas." They submit detailed specifications for modelling climate within city infrastructure along with rudimentary drone payload constraints for 130 multi-rotor drones. Of the 12 bids returned, they select the cheapest fixed-price offer from a freelancer who fails to probe the suspicious application. By anonymously crowdsourcing from those platforms to contractors, the group procures tailored targeting AI without revealing intentions or capabilities. The results of the AI-enabled planning lead to adaptations of their dissemination plans. When remotely activated, the swarm releases aqueous capsaicin and chlorine solutions in preprogrammed locations, channelling pungent clouds through office complexes, shopping centres, and high streets. This is complemented by foot teams depositing odorant pellets into public transport stations and hubs. Within minutes, localized chaos stretches emergency services but ultimately causes more anxiety than damage.

Authorities are quick to identify collective non-lethal exposure across the pre-meditated pattern, facilitating public assurance efforts and limiting hysteria. The attacks ultimately cause short-term, localized disruption but no lasting casualties. However, Gaian Defence releases a statement claiming that the attack has been a success and that it serves their purpose in highlighting the effects of environmental degradation on nature.

Quick Look

Process Acceleration: Planning, targeting and dissemination

Ideational Mediation: AI enables the group to adapt their plans for specific utilities

Transparency: Using the dark web and AI systems shield the group

Circumvention: AI helps circumvent targeting and dissemination challenges

	U	C	CL	ND
PA		X		
PG				
IM	X			
TM		X		

Impacts: How is AI used in this case?

AI is being used to accelerate the process of optimizing the dispersal of irritants, malodorants, and toxic industrial chemicals across urban areas. The custom neural network code procured by Gaian Defence generates pathways for targeting the release of these substances, taking into account factors such as urban climate and infrastructure. The AI mediates the group's ideas by translating their intent to cause public disruption into actionable targeting data for the drone swarm and foot teams and to make adaptations to their plans. Additionally, the AI modifies transparency by enabling Gaian Defence to anonymously crowdsource the necessary expertise without revealing their true intentions or capabilities.

Intent: Why is AI used in this case?

Gaian Defence sees value in using AI for this attack because it allows them to overcome their lack of technical expertise and optimize the deployment of commercially available chemicals for maximum disruptive effect. As such, the AI supports their intention to use chemicals to cause disruption. The AI's ability to process complex urban environmental data and generate precise targeting information enables the group to stretch emergency services thin and cause localized chaos, even with relatively modest quantities of chemicals. As this potential utility takes shape, the support AI provides their motivation as they believe effective dissemination vectors are established. Moreover, the AI's ability to be procured anonymously through hidden developer forums allows Gaian Defence to maintain operational security and plausible deniability, reflecting their intentions to avoid attribution and legal consequences, demonstrating the value of AI in modifying the transparency surrounding their efforts.

Implications: What can be done?

The use of AI for optimizing the dispersal of toxic chemicals in urban areas lowers the technical barriers for non-state actors to exploit commercially available substances for causing public harm and fear, potentially eroding the norms against the use of these materials as weapons by demonstrating their utility in particular contexts. The AI's ability to be procured anonymously and generate precise targeting data based on urban environmental factors may inspire other malicious actors to pursue similar attacks, contributing to the potential proliferation of these tactics. Furthermore, the use of AI in this context may challenge existing response and attribution mechanisms, as the algorithms can be custom-tailored for specific environments and the actors can maintain plausible deniability. Prevention could include efforts to:

- Review, update, and strengthen monitoring, regulation, and licensing requirements for the acquisition of legal but controlled substances
- Support the development of best practice and ethical guidelines for the nascent freelance AI-solutions community
- Continue to invest in support for cyber safety and security for all businesses
- Ensure funding and training for first responders and public health systems to respond to CBRN incidents, including a focus on the use of toxic industrial chemicals.

4. Summary and Conclusions

The eight vignettes presented here are the culmination of two separate projects which aimed to better understand the potential implications of AI technologies on the Chemical and Biological Weapons prohibition regimes. The use of an initial set of scoping interviews, followed by three expert workshops, have enabled us to map a diversity of potential scenarios and interrogate them using our framework, based on the dimensions of AI impact and challenges to the CBW conventions. Each of the eight vignettes demonstrate different combinations of impacts and challenges, and describe the implications of AI technologies across multiple stages of the CBW eco-system and specific acquisition processes. Despite this, we are able to note four broad, emerging themes which may provide fertile areas for exploring governance responses.

First, all cases involved the use of multiple AI systems. For example, in vignette 2, the NemesisBio system is not a single algorithm, dataset, or outcome, but a composite system which combines datasets and sequentially iterates outcomes. Importantly, these combinations thus require some level of technical expertise on the part of the user to understand how to best use AI systems, how to potentially develop CBW, or how CBW might achieve strategic aims in ways that conventional means would not. The necessity for actors in this case to combine different systems, or system components, creates the potential for multiple 'stop' points where an algorithm might block action. It is therefore important to support new and ongoing efforts to implement technical safeguards within AI algorithms.

Second, all cases required access to specialist datasets and/or materials, whether that be access to datasets on genomic and immunity characteristics of populations, or access to laboratory chemicals and equipment. In either case, users require some capabilities and technological or scientific literacy to be capable of training or interpreting an AI algorithm for CBW ends. Moreover, the need for an actor to somehow make AI-developed plans 'real' requires the actor to actively engage in physical CBW proliferation. In these instances, existing CBW governance arrangements require strengthening to better address such potential.

Third, all cases highlight a need for strengthened investigation capabilities. In many examples, the role of AI algorithms in shielding actors from attribution/prosecution or in implicating other actors or accidents plays a central role in the actor's decision to pursue CBW. In others, the use of AI to create or target CBWs may have complicated investigations either by using agents that degrade quickly or by using pathogens that have occurred naturally. This emphasises the need for strong AI/cyber capabilities to be embedded in existing investigations mechanisms (from public health to UNSGM)

and core training for rostered experts in the collection, handling, and processing of digital evidence.

Finally, all cases revolved around the role of intent in how AI use presents risks for CBW. In each case, AI modified an actor's decision to pursue CBW, and often not just at the ideation or planning stage. In some vignettes, ideation and planning had already begun and AI boosted intent by providing targeting support; while in others the decision to pursue CBW had been abandoned or demotivated until the capabilities of AI algorithms overlapped with other factors to boost intent. Moreover, the ways in which AI technologies mediated intent varied across vignettes; with some technologies seeming to spark intent into life, while others changed the character of intent from defense to offense. As such, understanding the dynamics between AI, intent, and motivation are crucial to explore in more depth and on a case-by-case basis. In all instances, however, strengthened international norms and taboos against the use of CBWs remain central to efforts to mitigate the use of AI technologies in CBW proliferation.

Annex 1: AI's dimensions of impact and the emerging reshaping of CBW acquisition processes

The acquisition of chemical and biological weapons (CBW) is a multi-stage process, with each stage encompassing specific functions, tasks, contexts, activities, and roles, as represented in Diagram 1. Not every actor will engage in all stages, and the stages may not always follow a linear progression. The scale and scope of an actor's engagement will depend on factors such as resources, objectives, and intended use. This annex provides an overview of each stage, highlighting key internal aspects and exploring how AI's four main dimensions of impact—acceleration of processes, generation of pathways, mediation of ideas, and modification of transparencies—can reshape these stages and create new uncertainties in the CBW acquisition process.

1. Inspiration

The inspiration stage can be driven by a number of elements, inter alia, an actor's motivations, beliefs, perceptions of threats and opportunities, or strategic considerations to gain advantage over adversaries.

AI can play a significant role in shaping these factors by analysing vast amounts of data to identify patterns and trends that may make CBW appear more viable or advantageous. Through targeted propaganda and disinformation, AI may also exploit individuals' psychological vulnerabilities and ideological predispositions, gradually shifting attitudes towards CBW. This can make it easier for actors to justify pursuing these weapons and harder for others to counter the underlying drivers of their interest.

AI also accelerates the process of gathering and synthesizing information related to CBW, allowing actors to more quickly assess feasibility and potential impact. By obscuring the sources of inspiration through deepfakes and manipulated digital content, AI can help actors operate with greater impunity by modifying relevant transparencies, resulting in challenges identifying and addressing the true motivations behind actions.

2. Planning

In the planning stage, actors must develop a comprehensive strategy for acquiring and using CBW while evading detection. AI can support this process by optimizing resource allocation, identifying novel acquisition pathways, and generating sophisticated cover stories. AI-powered tools that can analyse vast amounts of data can be used to streamline timelines, mitigate risks, and adapt to changing circumstances. This reduces the cognitive burden on human planners and allows them to consider a wider range of options and contingencies.

By identifying alternative sources, synthesis routes, and unconventional channels for procuring materials and equipment, AI expands the pathways available to actors and helps them circumvent traditional controls. AI can also assist in recruiting personnel by analysing individuals' skills, expertise, and psychological profiles to identify those most likely to support CBW efforts. The use of AI in creating fake identities, shell companies, and digital trails further complicates detection and investigation, as human analysts struggle to distinguish legitimate activities from illicit ones. As such, the integration of different AI tools can facilitate the development of more robust, effective, and flexible

plans, providing scope to strengthen the acquisition process stage by stage and between stages.

3. Development

This stage broadly encompasses the research and development of CBW agents, delivery systems, and supporting infrastructure, with focus on scientific research, laboratory work, and the design and testing of delivery mechanisms. Actors must navigate technical challenges, safety concerns, and the risk of detection; the number of people involved is potentially at its highest in this stage, meaning human factors of intention and motivation may be particularly acute.

AI can transform the development stage by accelerating research and discovery, optimizing design and testing, and facilitating the creation of supporting infrastructure. Drug discovery platforms powered by AI can rapidly screen vast libraries of compounds, predicting chemical structures and focusing efforts on the most promising candidates. This allows human researchers to bypass time-consuming experimental work and concentrate on refining the most viable agents.

Simulation and modelling tools enhanced by AI help actors optimize delivery systems and production processes, taking into account a wide range of environmental factors and performance criteria. This enables human designers to make more informed decisions and reduces the need for extensive physical testing. AI can also suggest strategies for disguising CBW development efforts, such as generating fake research proposals and manipulating digital records. By handling these complex tasks, AI allows human actors to devote more time and energy to strategic planning and problem-solving.

4. Synthesis

Synthesis involves the actual production and manufacturing of CBW agents and related components. This stage often requires specialized facilities, equipment, and expertise, as well as access to precursors and other materials. Actors must balance the need for secrecy with the demands of potentially scaling production.

AI can enhance this stage by optimizing production processes, identifying alternative synthesis routes, and devising strategies to evade detection. Advanced process control systems can monitor and adjust synthesis parameters in real-time, ensuring maximum efficiency and minimizing the risk of accidents. This reduces the need for human intervention and allows operators to focus on higher-level decision-making.

AI algorithms can analyse vast amounts of chemical reaction data to suggest novel synthesis pathways that are faster, cheaper, or easier to scale up. This expands the options available to human producers and helps them adapt to supply chain disruptions or regulatory challenges. By simulating different facility configurations and generating fake documentation, AI also assists in concealing the true nature of CBW production. As AI takes on more of the routine tasks and analysis, human actors can concentrate on strategic issues such as resource allocation and risk management.

5. Storage and Transportation

Storage and transportation involve the secure containment and movement of CBW agents and related materials. Actors must ensure the safety and stability of the agents during storage and transport while minimizing the risk of accidents or detection.

AI integration can optimize inventory management, identify secure storage locations, and suggest covert transportation routes, helping actors balance the need for secrecy with the demands of safety and stability. Intelligent inventory systems can monitor the condition of CBW agents and components, predict supply needs, and identify potential vulnerabilities or risks. This reduces the cognitive load on human operators and allows them to make more informed decisions about storage and handling.

By analysing geospatial data, AI can recommend hidden or remote storage facilities that offer natural protection and easy escape routes. This helps human actors select the most advantageous locations and prioritize security measures. In planning transportation operations, AI can identify weaknesses in customs controls, suggest disguise techniques, and recommend optimal routes based on real-time data. This enables human operators to adapt quickly to changing circumstances and minimize detection.

6. Targeting

Targeting refers to the selection and prioritization of potential CBW targets, based on factors such as strategic value, vulnerability, and potential impact. Actors must gather and analyse information to generate intelligence on potential targets, assess the feasibility of different attack scenarios, and select the advantageous course of action.

In the targeting stage, AI enables actors to process information, identify high-value targets, and predict the impact of different attack scenarios. By integrating and analysing information from multiple sources, AI can create detailed profiles of potential targets, including their vulnerabilities, defences, and strategic significance. This allows human decision-makers to prioritize targets based on a more comprehensive understanding of the risks and benefits involved.

AI can also simulate the effects of different attack scenarios, helping actors refine their plans and maximize damage. By analysing genetic data, AI may even assist in designing targeted bioweapons that exploit population-specific vulnerabilities. As AI handles more of the data gathering and analysis, human actors can focus on strategic planning and adapting to evolving circumstances. However, the use of AI in generating false intelligence and misdirecting defenders can make it harder for human analysts to anticipate and prevent attacks.

7. Dissemination

Dissemination involves the actual deployment and delivery of CBW agents to their intended targets. Actors must select the appropriate delivery mechanism, consider factors such as weather conditions and population density, and ensure the agent remains viable during dissemination.

AI enhances the dissemination stage by optimizing delivery systems, predicting the most effective release conditions, and adapting to real-time changes. Modelling and

simulation tools powered by AI can help actors design and test various delivery mechanisms, predicting dispersion patterns and environmental impacts. This allows human operators to refine their plans and select the most effective options without extensive physical testing. By analysing real-time data from sensors, cameras, and social media, AI can identify the optimal time and location for releasing CBW agents to maximize spread and impact. This enables human actors to make more informed decisions and exploit environmental features to their advantage. AI can also assist in developing novel dissemination methods, such as drones or disguised agents, that are harder to detect or defend against. As AI takes on more analysis and decision support, human actors can concentrate on strategic issues and adapt to evolving situations.

Implications

The integration of AI into the CBW acquisition process has far-reaching implications, that extend beyond the optimisation of individual tasks within individual stages. AI's ability to accelerate processes, generate novel pathways, mediate ideas, and modify transparencies should lead us to re-evaluate and reshape some of our expectations and assumptions about CBW acquisition.

At its core, AI might enable a material shift in the calculus of CBW acquisition by potentially altering the perceived feasibility, desirability, and viability of pursuing both low-tech and high-tech CBW. With the potential to streamline processes, identify new opportunities, mitigate interventions or obstacles, and obfuscate activities, AI contributes to the lowering of barriers to a whole host of activities associated with the processes of CBW acquisition. The emerging view may well be that CBW acquisition is becoming more attainable and less risky than at any point since both CBW Conventions provided a comprehensive prohibition.

Moreover, AI's potential impact on the CBW acquisition process challenges traditional assumptions about the resources, expertise, and infrastructure required as AI takes on more of the cognitive and analytical burden. AI can redefine the roles and skill sets needed to engage in CBW acquisition, and this shift may provide those with limited technical expertise or resources new opportunities. In doing so, this makes it harder for authorities to identify and monitor potential threats.

AI-driven changes within the processes of CBW acquisition, in some cases small or tangential, can have significant implications for the international nonproliferation and arms control regimes. As AI enables new pathways and modifies transparencies, it can create gaps and ambiguities in existing regulatory frameworks, making it harder to detect and attribute violations. This reduction in transparency and certainty can undermine trust among states and weaken the norms and institutions that underpin the CBW prohibition.

Policymakers must adopt a proactive and adaptive approach that keeps pace with the rapid evolution of AI and its many applications in the CBW context. This requires a deeper understanding of how AI can influence the motivations, means, and opportunities for CBW acquisition. As this paper has demonstrated, AI does not just change the dynamics of development and synthesis, but has a far wider reach, made more troubling given that not all CBW-intent actors will even seek to develop or synthesise their own agents.

HSP

HARVARD SUSSEX PROGRAM

HSP is an inter-university collaboration for research, communication and training in support of informed public policy towards chemical and biological weapons. The Program links research groups at Harvard University in the United States and the University of Sussex in the United Kingdom. It began formally in 1990, building on two decades of earlier collaboration between its founding co-directors.