AI-anxieties, dimensions of impact, and challenges to the CBW Conventions

Alexander Ghionis
Research Fellow in Chemical and Biological Security

Anxieties about the potential impacts of artificial intelligence (AI) on the Chemical and Biological Weapons (CBW) Conventions are increasing. These novel 'AI-anxieties' often stem from the complex black-box of AI and the uncertainty as to outcomes and impacts.

However, these AI-anxieties tend to reveal more about our fears of AI than the tangible implications for the CBW Conventions. This is due, in part, to a focus on AI's role in the development of highly toxic or infective CBW, essentially concentrating our concerns on how AI will 'make it easier to develop a CBW'.

Yet, attention to the historical record suggests we need to remain vigilant to a much wider risk and challenge spectrum. Accordingly, there is a need to grasp a more general understanding of what AI enables, and to better conceive of those implications for anti-CBW policy.

This working paper proposes two framings to help guide this effort.

1. **Dimensions of impact**: Dimensions of impact guide a non-technical assessment of 'what' AI is doing in any particular case. Four dimensions of impact are identified that help explain why a particular application of AI raises concerns about its potential to facilitate CBW acquisition or undermine the prohibitions.

2. **Categories of challenge to the CBW Conventions**: Four are identified that frame what CBW policy has historically and contemporaneously guarded against. Broadly, these are socio-technical and political in character. This typology frames key areas where (un)intentional advances made by state or non-state actors should raise the alarm, providing a common language for assessing and locating AI's potential to generate negative outcomes.

For CBW policy experts, these frames abstract from the complex and highly technical discussions about AI which currently dominate discourse, aiding efforts to translate AI-anxieties into familiar policy languages. This approach provides opportunities to clarify implications and consider how mitigations may be effectively designed and employed from the perspective of the CBW Conventions.

NB: this working paper is subject to updates; version of 28 March 2024

1. Introductory observations

Growing concerns about the potential of artificial intelligence (AI) to undermine the Chemical and Biological Weapons (CBW) Conventions have given rise to what we may term 'AI-anxieties.' These AI-anxieties stem from a pervasive sense that current structures and practices designed to maintain and strengthen the CBW prohibition regimes may be insufficient in the face of rapid and uncertain technological change. Recent research by the Harvard Sussex Program found that AI-anxieties in the CBW community are widely diffused across the anti-CBW policy space.[1]

For example, while common concerns exist about AI-driven scientific tools supporting research and development into highly toxic or infective agents, many AI-anxieties were also clustered around concerns about AI's role in disinformation, indoctrination and radicalisation, knowledge acquisition and sharing, evading scrutiny and regulation, planning and dissemination support, reinspiring previously discontinued research, and reshaping strategic and political calculations. Moreover, significant anxieties exist that AI tools will create a greater *recognition* of the potential utilities of CBW beyond mass casualty, thus motivating actors to use CBW for a range of objectives, raising concerns about AI's role in actor intent. Additionally, low-tech applications of AI (such as 'chatbots') were as concerning as

[1] This research was supported by the UK Foreign, Commonwealth, and Development Office between 2022-2024

the use of high-tech applications of AI, such as in pharmaceutical drug development, by relatively small groups of well-resourced actors.

As such, we have located AI-anxieties not just in the high-tech 'development' or 'synthesis' stage of the CBW acquisition model, but also within processes across all other stages of the CBW acquisition model. Diagram 1, below, provides a simplified view of these different stages that an actor might need to engage with to acquire CBW. The diagram should be broadly familiar to those working on CBW issues.

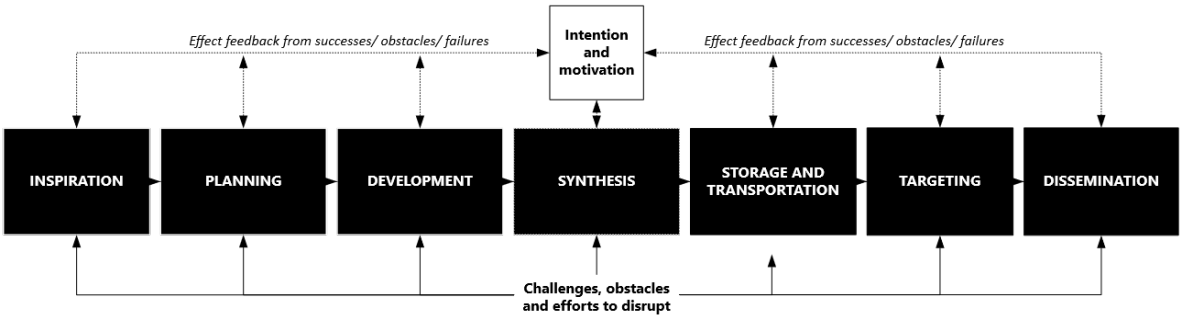

Diagram 1 – CBW acquisition stages and processes (simplified)

However, the dominant discourse on AI and CBW appears not to reflect this spectrum of concerns, and tends not to locate AI-anxieties to any great extent outside of the high-tech development and synthesis stages illustrated above. A general survey of the literature – and the nature of discussions at various academic and policy events – suggests we are witnessing a much narrower view receive the most attention. This could potentially obscure or limit our view of the risks, challenges and effective responses. What follows reflects on two elements that might rebalance this perceived weakness.

1.1 Identifying and generalising AI's impacts across contexts

First, efforts to categorise *what* any particular AI system is doing appears to be limited. An effort towards some generalisability of AI's dimensions of impact (as we may now refer to them) seems an important step in moving beyond the narrow treatment of specific high-tech applications of AI in the development of highly toxic or infective CBW. This could move us towards a more comprehensive picture that brings into focus AI applications in other settings and stages of the CBW acquisition process model. In other words, this would allow us to better interrogate a much broader range of AI-anxieties in a way which seeks to be consistent. This may help us to better articulate what the AI system is actually facilitating for its human operator.

We may wish to ask ourselves: What is it about the function or outcome of the applications of AI that inherently challenge the existing governance, norms, and activities underpinning the CBW Conventions? Is it the algorithm itself, or what the algorithm enables humans to do that raises concern? If the latter, what is the AI enabling the human to do? Are we worried about what the AI is doing in a technical

sense, or what it enables the actor to achieve? At this stage, we are moving away from a 'technological determinism' and bringing back questions of *intention* and *purpose* and human accountability.

The answers are not simple, and the answers are not singular. Yet the effort is important. By focusing only on very specific cases in which we put focus on the technological capability of the AI, and not on what it is actually facilitating for the actor, we run the risk of becoming overwhelmed in technical and arcane details that only advance us a certain distance in our understanding of the relationship between AI and CBW. We must balance our view of AI by grounding ourselves in our CBW perspective. This must be the case because we (those of us who identify as being in the CBW community) are not seeking to necessarily become AI experts, or to build AI governance writ large, but rather to strengthen our CBW regimes against technological surprise and uncertainty, of which AI is currently contributing. So, to achieve a balance, we must find ways to describe and understand what the AI is *doing* and *enabling* in various contexts so we can understand why this is relevant in relation to what our CBW governance efforts currently guard against. This helps us to consider where the blind spots, loopholes, or omissions currently exist. Likewise, we may also identify where governance is adaptable and able to respond. The picture will likely be an evolving patchwork.

To facilitate this grounding of AI within our CBW perspective, we may find that contextualising AI in terms of its *general effects* is a useful starting point. This effort to identify and categorise the 'dimensions of impact' may help us move beyond the narrow but dominant focus on specific AI use-case scenarios and towards a more nuanced understanding of how AI may influence the CBW prohibition regimes in different settings and times. The fundamental view created by the CBW Conventions is that the *intention* of actors, and their ability to fulfil that intention, should dictate how we design and implement anti-CBW policy. Therefore, we must treat AI no differently: how and why it interacts with actors' intentions should be a key approach in its analysis. The technical 'under the bonnet' approaches complement this, and connects different communities, but it should not dictate our view.

In terms of complementarity and communities, an effort to categorise AI by its dimensions of impact can bridge between technical and specific discussions about particular AI applications, and the CBW policy-oriented conversations. By translating the complex and often intimidating aspects of AI into more accessible and relatable terms, we may facilitate better communication and collaboration between the AI and CBW policy communities.[2] We may realise that while AI in general terms is novel and often intangible, what it enables and what it facilitates may in fact be rather more familiar to us as processes which existing governance already seeks to control or constrain. This may sound vague, but as will be explored below, the proposed dimensions of impact allow us to understand how intangible AI applications can enable actors (in various ways) to advance processes

[2] Recognising, of course, that these communities themselves are not homogenous or static, but that there is enough of a distinction in general terms to speak of different communities in the context of this paper.

that are familiar to those working on anti-CBW policy. In turn, this can lead to more informed decision-making and the development of more effective responses. We might be able to say with more confidence that we do not need to reinvent the wheel and that, in fact, the CBW Conventions can respond to the challenge of AI, as they have with other technological fears. We may also say that CBW policy experts do not need to become AI experts.

1.2 Characterising challenges to the CBW Conventions

The second element contends that a refocusing on exploration and articulation of the precise nature of the challenges AI applications, through their dimensions of impact, pose to the CBW Conventions would be valuable. A good portion of the current discourse, as mentioned, has focused on technical capabilities and potential implications within narrow examples, primarily AI's role in the development and synthesis of CBW agents.[3] Such assessments provide expert analysis of highly specific cases, most of which are framed as high risk and high impact, and which often place focus on the technical aspects of the AI system itself with a gradual lessening of detail as concrete categorisation of the challenges to the CBW Conventions are formulated.

While much of this dominant discourse increases our knowledge about a specific AI, the articulation of the emerging challenges for anti-CBW policy loses focus. This 'specific-AI, unspecific-challenge' discourse limits our view of where the threat lies from the perspective of the Conventions. It enables a deterministic view that AI inherently makes CBW creation easier, leading to the linear conclusion that AI will increase the likelihood of CBW use in a general sense. It may be suggested then, that the discourse can be strengthened through increased exploration or identification of the additional complex assumptions or factors that would be needed to respond to the generalised challenges. It may also be suggested that the current focus of the discourse frames particular types of risks, at the expense of presenting a wider perspective that takes into account challenges to emerge from low-tech or indirect AI impacts.

Consequently, current assessments generate general assumptions: AI will lower barriers to access, enabling highly sophisticated CBW development which could bring mass destruction. Of course, these assumptions may hold some truth. However, the history of CBW has been replete with similar analysis about new technologies, with little evidence to suggest highly sophisticated, novel, mass-casualty CBW development or utility has been sought by actors. It could very well be that AI may change the trajectory of history, and make these predictions true; so

[3] See, for example, Carter, S. R., Wheeler, N., et al. 'The convergence of artificial intelligence and the life sciences' Report (NTI, 2023); Dresp-Langley, B. 'The weaponization of artificial intelligence: what the public needs to be aware of' (*Frontiers in Artificial Intelligence*, 2023) 6; Ekins, S., Lentzos, F., Brackmann, M. & Invernizzi, C. 'There's a "ChatGPT" for biology: What could go wrong?' website of Bulletin of the Atomic Scientists, 2023; Lentzos, F. 'How to protect the world from ultra-targeted biological weapons' (*Bulletin of the Atomic Scientists*, 2020) 76(6); Rose, S. and Nelson, C. 'Understanding AI-facilitated biological weapon development' Report (The Centre for Long-Term Resilience, 2023); Sandbrink, J. B. 'Artificial intelligence and biological misuse: differentiating risks of language models and biological design tools' arXiv:2306.13952 12/08/2023; Urbina, F., Lentzos, F., Invernizzi, C. & Ekins, S. 'Dual Use of Artificial Intelligence-powered Drug Discovery' (*Nature Machine Intelligence*, 2022) 4(3)

far, we have seen recourse to the scruffy use of older CBW technologies, such as chlorine barrel bombs, or the more traditional nerve agents, used opportunistically and crudely in Syria, or in assassinations attempts by the Russian Federation. How these types and utilities make sense in relation to AI is under-examined.

So, it seems that current discourses do not bring us closer to understanding the conditions under which AI's impacts on the Conventions may be realised in their fullest extent. We find limited insight into AI's fundamental implications for CBW beyond the sophisticated development stage, and yet our AI-anxieties suggest there is a spectrum of challenges for the CBW prohibitions that could emerge from AI.

This spectrum of AI-anxieties makes it difficult to hold in one's mind a singular way in which AI can lead to negative outcomes. Are the AI-driven challenges likely to be the use of a sophisticated, high casualty CBW, or might they be less obvious, skirting in the grey-areas, eroding norms incrementally, and remaining difficult to predict as we look between the expanding capabilities of AI and the intentions of actors to use specific utilities of CBW for specific objectives? A more nuanced understanding of the challenges AI generates can help us anticipate and prepare for future challenges as AI continues to evolve and new applications emerge.

1.3 Remarks

Suffice to say this is not a critique of those authors who have written on this topic. Rather, it is a recognition that expanded approaches and framings could be developed to deepen the discussion. As mentioned above, the call is for a complementary set of interlocking approaches. A single approach that prioritises specific cases and technical renderings of AI forces a broad range of policy communities to try to coalesce around a single language and view. This can hardly be fruitful when different communities have different knowledges, languages, priorities, and objectives. The question then is how to create a wider, initial set of approaches and framings that can facilitate those who may identify as operating from the grounded perspective of the CBW Conventions to engage and understand what is relevant for them and why.

While AI's role in developmental stages is crucial, it represents only one segment of activities contributing to CBW acquisition (see Diagram 1). AI is not monolithic, and assessments of its role in specific developmental tasks do not provide a comprehensive template for addressing wider risks to the prohibitions. AI can integrate into almost any activity, and analysing each in detail is herculean. A useful complementary approach may be to understand what a particular AI application enables for a particular actor in a particular context. How might an actor's AI use lead to negative outcomes for the CBW Conventions, and why?

That AI is only one part (although an increasingly important part) of a broad portfolio of risks and challenges facing the CBW Conventions is important to underline. Understanding if, how, and why AI integrates into this wide portfolio of risks and challenges is becoming increasingly urgent. This can be direct, for example our fears about AI's ability to identify highly toxic compounds. It can also be indirect, for example how AI relates to disinformation, or how AI's ability to reopen previously infeasible, and thus discontinued, lines of grey-area research

might result in shifting strategic views of CBW over time. It strikes me that the vast indirect or non-obvious uses of AI are the ones we might need to guard against.

Indeed, as AI rapidly infiltrates most aspects of human activity, particularly in the Western world, there is a pressing need to understand how it integrates into all aspects of regime implementation so as to guard against both surreptitious and blatant challenges to the prohibitions. This requires moving beyond isolated, highly technical applications to grasp AI's implications in a broader context.

To conceive of the emerging challenges AI poses from the perspective of the CBW Conventions, the following sections outline two framings and explain how and why they are relevant for this task. In particular, what follows first looks at how we might conceive of what AI 'does' in different cases through categorising its dimensions of impact. Then, the paper considers how best to categorise and frame the challenges that AI might contribute to, helping to structure and embed often quite unfamiliar anxieties within a language that may be considerably more familiar to those working on anti-CBW policy and their attendant Conventions.

2. AI's dimensions of impact

Real-world examples of relevant, potentially problematic AI applications illustrate the range of capabilities that experts are concerned about, broadly understood as emerging 'AI-anxieties'. To abstract from specifics, such AI-anxieties can be analysed with a view to identifying commonalities and themes in their dimensions of impact. These dimensions denote identifiable impacts which can be used to better articulate *why* and *how* a particular AI system contribute to an actor's activities or outcomes that pose challenges for the prohibitions.

By examining the AI-anxieties generated through project data collections, it has been possible to broadly categorise based on an AI's dimensions of impact. Thus, a typology of four such dimensions of impact in the context of CBW are suggested, understood as an AI's ability to

- Accelerate processes;
- Generate pathways;
- Mediate ideas;
- Modify transparencies.

The purpose, in simple terms, is to suggest that the routes by which tangible challenges emerge, often starting from highly technical AI systems, tend to be recognisable and approachable when we abstract out of the technical details and into an impact-defined reference point. What seems almost intangible conveys its impacts through mechanisms that are, to some degree, familiar. These dimensions are not intangible because they have a real impact upon an actor's objective. The dimensions of impact help us to better grasp what the AI is doing and therefore why and how – and potentially what - it is *enabling* a CBW-intent actor to achieve.

Perhaps what is most useful is that these dimensions of impact are already, in simple terms, representative of other technologies and phenomena that have been recognised to raise concerns in relation to the CBW Conventions.

For example, consider the extensive and varied discussions found in the reports of the Scientific Advisory Board (SAB), which examine scientific and technological (S&T) developments and consider the potential challenges and opportunities these present. The SAB reports submitted to the five OPCW review conferences, and those published during the intersessional periods, provide rich detail on a significant number of areas of S&T, and for non-experts this technical discussion can be overwhelming due to the granularity of detail and expertise. It is possible to say that discussions about AI are similarly technical and potentially overwhelming.

Yet, by stepping back and abstracting areas of potentially problematic emerging S&T covered by the SAB, one finds that often the challenges that are thought to emerge tend to do so through one or more of the dimensions of impact proposed here. Often, the topics covered tend to result in anxieties not because of the materiality of a particular technological artefact itself – i.e. not the *thing* – but rather by what it potentially *enables* and how it may be *applied*.

There is somewhat of a parallel here to the concept underpinning the General Purpose Criterion: it is not necessarily the artefact (for example, a chemical weapon itself) but rather how chemistry or biology are used and for what purpose.

By stepping back from the artefact, we often see that the route to generating negative impacts is often facilitated through a particular artefact's ability to help an actor, either through the acceleration of research and associated processes; through the generation of new pathways and unforeseen potentials; through the facilitation, changing or proliferating of ideas, knowledge, or objects; or through altering levels of transparency usually (but not exclusively) with the intention of reducing clarity or oversight.

These dimensions of impact are not perfect typologies, and they overlap and interact. They are also rather broad, and almost any example can be viewed through this framing. Yet, they form a first step in abstracting from highly specific and technical AI-anxieties, helping us to understand in general terms what a particular AI system is doing in a particular context and *why* and *how* it brings value to a particular actor.

Therefore, this typology may be useful in contextualising AI within the much wider S&T review efforts that are undertaken in relation to the CBW Conventions, including efforts where we wish to understand why and how AI actually brings *benefits* to prohibition efforts. Additionally, the typology can help us link up these emerging AI-anxieties with technological anxieties that have come before, allowing us to frame technologies across time and space. However, these additional elements are outside the scope of the current paper.

What follows are very brief descriptions of these dimensions of impact, contextualised to AI. These will be expanded upon in later iterations of the paper.

2.1 Process acceleration

AI, in theory, can function as a catalyst to expedite and enhance the pace and efficiency of processes and activities associated with almost any aspect of the stages associated with the acquisition of CBW.

Particular attention has been drawn to processes and activities associated with the discovery, development, diffusion and deployment of scientific research. AI can accelerate process required to complete specific steps within research and development, and between steps.

Some examples that are relatively familiar include how AI can design new materials and pharmaceuticals orders of magnitude faster than human teams through rapid computational experiments, enabling accelerated development and deployment. Machine learning can crunch decades of chemistry publications, detecting new reactivity patterns in hours that scientists could spend months finding, diffusing insights at unprecedented speeds. Open-source simulation platforms can allow much faster replication of designs worldwide, increasing the speed at which participation in scientific endeavour takes place. Automated robotic labs directed by AI optimize experiments round the clock, achieving years of progress in days by tirelessly iterating. Natural language processing consumes vast text corpora, linking findings across disciplines in seconds versus the years it would take manual review.

Beyond those processes associated with research and development, this dimension of impact speaks more broadly to other processes and activities in which there is evidence that AI's primary impact is that of expediting a particular task or arrival at outcomes. This could include speeding up processes that design or generate disinformation, or that map digital or physical vulnerabilities in infrastructure. AI can accelerate processes that underpin non-scientific knowledge acquisition, or that identify individuals or threats from real-time video feeds. In a general-sense, AI can accelerate the time taken for an actor to reach a decision. Almost any activity has the potential to be accelerated by the integration of relevant AI systems. Thus, the concept of acceleration is wider than the focus on research and development in scientific contexts and can be identified in many examples of AI's application that one may examine.

Process acceleration is not necessarily limited to individual or unconnected tasks, but can also be used to describe in a more general sense how the integration of AI systems within different activities, and across different activities, can increase the speed which actors are able to advance along a particular path. The implications of increased speed in most contexts of activity are far reaching, but for those examining how AI relates to our understandings of risks and threats emerging in relation to CBW, we must recognise that efforts to govern, control, prevent and respond may concomitantly need to be adapted to be able to operate at a greater pace than was once assumed.

2.2 Pathway generation

AI acts as a generator of new potential technical directions by drawing novel connections across disciplines, transferring capabilities to new domains, and actively exploring high-risk, high-reward strategies. This expands possibilities for scientific inquiry and technological development, potentially re-opening previously infeasible lines of research or providing entirely new opportunities.

We might consider how AI could identify non-obvious connections between genetics and materials science, transferring insights to create self-replicating

organisms unlike any in nature. Deep learning can recursively test millions of hypothetical cyber incursion pathways against industrial controls systems, revealing novel attack vectors. AI exploration of novel biochemistry reactions can lead to the discovery of unique pathogenic mechanisms. AI modelling of supply chains could reveal clandestine procurement pathways for controlled materials avoiding existing safeguards. Autonomous experimentation could uncover dangerous gene editing techniques through iterative pathways which may be unfathomable without AI.

Certainly, in terms of scientific development, AI can enable the systematic realisation of technical paths thought too unpredictable, complex, and interdisciplinary for humans alone to envisage, and accordingly may open up new paths or encourage the revaluation of previously discontinued research.

Yet, pathway generation is clearly applicable outside of the scientific research setting when we understand the generation of new pathways to encompass the broader notion of 'providing new ways to overcome obstacles'. Obstacles could include regulations, trade controls, or physical security; they could include less tangible aspects such as normative or psychological obstacles to action or motivation, or obstacles in finding or synthesising information outside of the context of scientific development. Obstacles could relate to problems finding trustworthy people to support a CBW acquisition effort, or obstacles relating to limited resources.

In this broad view, AI applications can play particular roles in enabling actors to problem-solve their way around particular obstacles that stand in the way of achieving certain objectives. Again, the contexts or cases are diverse, but in many hypothetical or real cases we may encounter, understanding how and why a particular AI application supports an actor to circumvent challenges – big or small – can be illustrative in understanding how the AI relates to CBW.

2.3 Ideational mediation

AI has the potential to significantly influence the spread of ideas, shape narratives, and mold perspectives on a massive scale through its advanced capabilities in natural language processing, sentiment analysis, and content curation. The ability of AI to generate and disseminate persuasive content seamlessly across multiple platforms and languages can have profound implications for public opinion, policy debates, and societal norms.

For instance, AI-powered chatbots and text generators can produce and distribute vast amounts of manipulative or misleading content, engaging with humans in convincingly natural conversations. These interactions can subtly shape individuals' beliefs and attitudes, potentially eroding support for CBW prohibitions or sowing doubt about the effectiveness of existing governance mechanisms.

Similarly, AI algorithms designed to optimize user engagement and content virality on social media platforms can amplify the spread of disinformation and conspiracy theories related to CBW. By exploiting cognitive biases and leveraging micro-targeting techniques, these algorithms can create echo chambers and filter bubbles that reinforce false narratives and undermine efforts to promote fact-based discourse.

Moreover, AI-generated deepfakes and manipulated media can be used to create highly realistic and emotionally compelling content that distorts public perceptions of CBW-related events or policies. For example, a deepfake video purporting to show a secret CBW facility or a staged attack could be used to generate outrage, confusion, and mistrust, making it harder for authorities to maintain transparency and accountability. By manipulating search results, news feeds, and content recommendations, AI systems can effectively limit or increase the visibility and reach of content which, in particular contexts, may influence perceptions of the prohibitions.

From this perspective, anti-CBW policy must guard against the potential for AI systems to either intentionally or unintentionally shape ideas, beliefs, and perceptions. This can be as broad as polarising social views around an allegation or – on the micro-level – increasing the motivation or intent of a particular actor either through psychological manipulation or through convincing an actor that the obstacles they face may not be as insurmountable as they assumed. This latter point has been underexplored so far and has received little attention. Yet, this may be one of the biggest dimensions of impact that AI systems generate: the belief (mistaken or not) that by using AI for a specific task, an actor may increase their chances of success. As such, AI systems indirectly shape ideas about what is possible by being constantly framed as opening up a world of new potentials.

As such, the relationship between AI's four dimensions of impact, and human intention and motivation, is highly complex and hard to discern. However, we can be sure that the relationship between what AI is perceived to do by an actor, and an actor's own motivations, is formulated by AI's inherent ability to mediate and shape ideas about its own relationship with humans.

2.4 Transparency modification

AI systems' inherent opacity, complexity, and proprietary nature can pose significant challenges to transparency, accountability, and oversight. The 'black box' problem, where the internal workings of AI algorithms are inscrutable to external observers, can make it exceedingly difficult to assess AI-driven processes and decisions.

In the early stages of CBW acquisition, such as those processes underpinning inspiration and planning, the use of opaque AI can enable actors to explore and develop concepts and strategies while evading detection. For example, AI-driven extremist material or disinformation may motivate actors to pursue CBW with little external visibility as to who is driving them, or how they function.

As noted above, AI can accelerate scientific research and generate new pathways: it can also modify transparencies around it too. Advanced machine learning models used for tasks such as chemical synthesis optimization, delivery system design, or biological agent selection may operate in ways that are not fully understood or controllable by their human creators. This lack of transparency creates blind spots and uncertainties for those seeking to ensure that research and development activities remain ethical and legal. In the event of an actual or alleged CBW attack, this lack of transparency could greatly complicate efforts to investigate and attribute responsibility.

Within this domain, the 'reproducibility crisis' of AI-driven science and research has inspired calls to develop best practices to ensure reproducibility of findings.[4] Therefore we may consider that the challenge of transparency is in fact multidirectional: increased opacity has clear implications for judging intentions, however increased transparency also raises concerns about information hazards, and access to data, information, and knowledge.

Perhaps of most wide concern is the potential to circumvent oversight. AI-powered tools could be used by actors to either plan, design or generate false or misleading documentation, such as purchase orders, shipping manifests, or inventory records. These could well be sophisticated enough to pass scrutiny by regulators, customs officials, or other oversight bodies, allowing illicit materials and equipment to be procured or stored without detection. Similarly, AI could be employed to create convincing cover stories and alibis for actors reducing transparency, as noted above, mediating understandings and ideas. Of course, this is all possible without AI, but here we see two dimensions converging: AI can greatly accelerate the process of circumvention through its enabling of transparency modification.

Indeed, it is also clear that much illegal activity benefits from AI-driven encryption and obfuscation techniques, allowing actors to hide their communications and activities from external monitoring. In the realm of financial transactions, AI-powered money laundering and fraud detection evasion pose serious challenges. We must assume those pursuing a CBW capability will likely seek to make use of such tools given their increasing availability and access. In the era of digital open-source investigations, these sort of efforts to modify transparencies by malign actors may be particularly appealing.

2.5 Remarks

These dimensions of impact become useful for expanding on and analysing the ways in which, and the precise nature of, AI's impacts on the challenges to the Convention. These can help us understand how AI might contribute to fears around new utilities; circumvention and acquisition; creeping legitimisation; and normative divergence, by guiding our construction of scenarios to expose not just what is novel about AI but also about what is more recognisable and, perhaps, longstanding. In essence they provide a way for us to view what value an AI brings to an actor and thus how and why AI might contribute to motivations and intentions.

3. Challenges to the CBW Conventions

Many AI-anxieties imply that AI will make it easier for actors to develop CBW, in one way or another. It seems, however, that the potential challenges for the CBW Conventions are broader than this, interconnected with many non-AI factors, and go further and deeper than singular concerns about the ease with which CBW can be developed. Therefore, to go beyond the 'ease of development' anxiety we must examine what it is we mean by challenges to the CBW Conventions so as to open

[4] Artrith, N. et al. 'Best practices in machine learning for chemistry' (*Nature Chemistry*, 2021) 13; Heil, B. J. et al 'Reproducibility standards for machine learning in the life sciences' (*Nature methods*, 2021) 18(10);

up our perspective. In other words, there are more complex and varied dynamics and challenges that also demand attention.

For this, there is also an expert literature, most of which was written and formulated in a 'pre-AI' time, or in a 'non-AI' perspective. However, much of it remains entirely relevant. There is not space here to go into all of the reasons as to why this literature remains relevant. Suffice to say, for many who wrote of the major challenges facing the CBW Conventions, the understanding that developments in science and technology (S&T) would be a *continuous* influence on implementation of the Conventions resulted in categories and framings of challenges which could integrate and make sense of such changes. Indeed, evolving S&T is well embedded within the CBW Conventions, in particular the recognition that it can bring both challenges and opportunities – as indeed AI does.[5]

This is all to say that the existing work on framings and categories of challenge were designed to be useful for those seeking to strengthen the CBW Conventions, even as fast-paced S&T developments rendered it difficult to fully ascertain those specific challenges and opportunities emerging at the frontiers of science. In this view, AI should not make us re-examine wholesale our fundamental understandings and expectations in regard of the Conventions' potentials.

Drawing on the work of Robinson in particular, the specific categories of challenge to be used as a framing device are presented below.[6] His work is particularly helpful because many of the challenges and risks associated with the CBW Conventions, as reported widely in one way or another in the literature, can be understood within his typologies. Much can be incorporated into them, either by abstracting or by contextualising, as required. All of our fears and concerns as to ensuring the permanent elimination of CBW, be they AI-driven or otherwise, can be explored with reference to, and anchored in, this typology of categories of challenge. These help us situate and understand *why* something affronts these prohibitions, helping us to render myriad risks in the language and perspective of the Conventions themselves. In doing so, they may help us to understand where, how, and why particular interventions, mitigations, and responses could be useful.

While they may not be exhaustive, they do cover within each of them a very significant area in terms of themes and issues. Moreover, they are not temporally defined, meaning that these are not tied to a particular socio-techno-political time. This makes them additionally relevant, mirroring, as they do, the permanence

---

[5] Where the use of AI can bring benefit see, for example, Jeong, K. Lee, J-Y. et al. 'Vapor Pressure and Toxicity Prediction for Novichok Agent Candidates Using Machine Learning Model: Preparation for Unascertained Nerve Agents after Chemical Weapons Convention Schedule 1 Update' (*Chemical Research in Toxicology*, 2022) 35(3); Reinhold, T. and Schöring, N. *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm* (Spring; Cham, 2022)

[6] See, for example, Robinson, J. 'Near-Term Development of the Governance Regime for Biological and Chemical Weapons' (SPRU – Science Policy Research Unit, Item 456, dated 4 November 2006); Robinson, J. 'Categories of Challenge now facing the Chemical Weapons Convention' Discussion paper for Pugwash Meeting no. 324 at the 52nd Pugwash CBW Workshop *10 Years of the OPCW: Taking Stock and Looking Forward* Noordwijk, The Netherlands, 17-18 March 2007; Robinson, J. 'Difficulties facing the Chemical Weapons Convention' (*International Affairs*, 2008) 84(2); Robinson, J. 'Chemical and Biological Weapons' in Busch, N. and Joyner, D. H. (*eds*.) *Combating Weapons of Mass Destruction: The Future of International Non-Proliferation Policy* (University of Georgia Press; Athens, 2009)

envisioned within both Conventions. The following sub-sections briefly present these categories, and it is worth consider as these are presented how AI's dimensions of impacts discussed above could result in negative outcomes within these categories.

3.1 New and old utilities for chemical weapons

Despite the now long-standing prohibitions, the changing nature of conflict may create renewed incentives to pursue the acquisition of CBW. The underlying assumptions during the negotiation of the CBW Conventions, in particular the CWC, were that these weapons would be used on a large scale between the professional armies of the competing Cold War blocs.[7] Evidence of this can be found in the assumptions underpinning the verification regime of the CWC in which 'militarily significant quantities' define declaration thresholds: these are of such an extent that only a relatively well resourced state entity could really consider acquiring such volumes.[8]

Moreover, the chemicals listed in the Schedules of Chemicals are broadly representative of highly toxic chemical armament historically associated with states capabilities.[9] The General Purpose Criterion, found in Article II of the CWC, which ultimately speaks to a universal prohibition of specific *intended uses* of chemicals, rather than on chemical *things*, is the key to making the Convention relevant outside of those Cold War constraints.[10] Nonetheless, the persistent logic embodied within the OPCW's international disarmament and verifications practices is one that correlates chemical weapons as being comprised of

---

[7] Cordesman, A. 'One half cheer for the CWC: Military Perspectives' in Roberts, B. (*ed.*) *Ratifying the Chemical Weapons Convention* (Centre for Strategic and International Studies; Washington DC, 1994); Moodie, M. 'Confronting the Biological and Chemical Weapons Challenge: The Need for an Intellectual Infrastructure (*Fletcher Forum of World Affairs*, 2004) 28

[8] On the concept and history of militarily significant quantities see, for example: Ballard, J. 'Reassessing chemical weapon threats' in Su, F. and Anthony, I. *Reassessing CBRN Threats in a Changing Environment* (SIPRI; Stockholm, 2019) p. 14; Bartlett, J. & Hamilton, M. 'Proposals for Establishing limits and thresholds in the CWC, with special reference to Schedule 2B' AMD 19/7/91 PTN TG 1090/10/AMD/91; Robinson, J. P. 'Alleged Use of Chemical Weapons in Syria' Occasional Paper #4 (Harvard Sussex Program, 2013) p.34; Tucker, J. 'The role of the Chemical Weapons Convention in Countering Chemical Terrorism' (*Terrorism and Political Violence*, 2012) 24(1) p. 114; United Kingdom 'Proposals for Establishing Thresholds in the Chemical Weapons Convention: Schedule 2B' August 1991 CD/CW/WP.358 pg. 2

[9] Robinson, J. 'The chemical industry and chemical warfare disarmament: Categorizing chemicals for the purposes of the projected Chemical Weapons Convention' *SIPRI Chemical & Biological Warfare Studies* no 4 (SIPRI; Stockholm, 1986) pp 55-104

[10] "The GPC is part of the language the Convention uses to enunciate its scope. Thus, the Convention defines the chemical weapons that it prohibits, not in concrete terms (such as physical construction or chemical composition) that could become out of date as technology advances, but in terms of intent. So toxic chemicals and their precursors become banned weapons if they fail to meet the criterion of being --in the words of Article II.l (a) of the Convention -intended for purposes not prohibited under this Convention, as long as the types and quantities are consistent with such purposes. A definition of "purposes not prohibited under this Convention" appears in Art II.9, which identifies and details four broad categories of purpose to which dual-use chemicals may properly be applied." Robinson, J. 'The General Purpose Criterion of the Chemical Weapons Convention' (SPRU – Science Policy Research Unit, Item 398, dated 12 October, 2001); see also: United Kingdom 'The Comprehensive Nature of the CWC with Respect to Verification and National Implementation' RC-1/NAT.16 dated 29 April 2003; UK 'The Comprehensive Nature of the CWC and Scientific and Technological Change: the General Purpose Criterion' RC-2/NAT.24 dated 18 April 2008

> toxic chemicals that [are] so intensely aggressive in their effects that weapons disseminating them would be competitive, in quantitative casualty-producing terms or other such measures of tactical efficacy, with modern conventional weapons.[11]

This Cold War view on what it means to develop or use chemical weapons, in some cases being viewed as weapons of mass destruction, has since been expanded: contemporary forms of violence have blurred the lines between war, civil war, counterterrorism, policing, and criminal activity, and in doing so highlighted the much wider utility of chemical weapons.

For example, assassination, harassment, incapacitation, economic damage, demoralisation, terrorisation, and allegations and propaganda are some contemporary utilities (re)presenting themselves in the wake of global geopolitics and changing S&T.[12] Indeed, as mentioned, the CWC entirely foresees this. The existence of the GPC indicates that the negotiators were aware that new types of chemical weapons, different utilities of chemical weapons, and evolving contexts in which they would be used, would emerge. This is equally true for biological weapons.

Within the contexts of evolving and less distinct forms and levels of violence, and within contexts clearly marked by rapid advances in S&T, new utilities – or more accurately: new ways of achieving and expanding different utilities – will be possible. These risks undermine the CBW Conventions as both state and non-state actors may see a potential *value* in utilising chemistry or biology for particular violent ends – moreover, if this can be achieved with plausible deniability, the incentives grow. In this case, then, AI may play a role in moderating an actor's intent through perceptions of value and achievability.

Therefore, the challenge emerges that if conditions increase opportunities for the inspiration, planning, research, development and potential deployment of CBW to achieve particular objectives through perceived utilities, the risks to the prohibitions increase. Moreover, if such conditions additionally facilitate the means to research, develop and deploy such CBW for specific utilities - in a way that reduces the developmental footprint or cost, or increases the potential for covert or untraceable deployment – that risk accordingly expands. Unscrupulous actors may seize on these opportunities if particular utilities can maximise their strategic advantage while simultaneously reducing the cost of their use. As such, guarding against processes that can incentivise and deliver upon these negative outcomes is crucial.

3.2 Circumvention, proliferation, and acquisition

A significant challenge to the integrity of the prohibitions can arise if actors who wish to develop or use CBW find ways to circumvent the multiple layers of governance that forms the so-called webs of prevention/deterrence that have been

[11] Robinson, J. 'Difficulties facing the Chemical Weapons Convention' (*International Affairs*, 2008) 84(2)
[12] Ilchmann, K. and Revill, J. 'Chemical and Biological Weapons in the "New Wars"' (*Science and Engineering Ethics*, 2013) 20

developed from international through to local levels.[13] Across many arms control and disarmament treaties this challenge is referred to as being about (non-)proliferation. Within the CBW regimes, the term needs some qualification, and may be better rendered – some may argue, myself included – as 'non-acquisition.'

Indeed, the opportunity for actors to acquire what we might think of as 'ready-made' CBW, or significant components to develop these themselves, seems less likely – or, at any rate, of a different nature - than similar fears in the nuclear regime. Perhaps the most obvious reason is that the CBW Conventions seek complete and irreversible disarmament and non-production, in essence, ensuring that there are no such stockpiled CBW weapons left, and no 'haves-and-have-nots.' As such, traditional proliferation concerns at the sharp end of the development process have been greatly reduced as military stockpiles of chemical weapons have been verifiably destroyed and their means for reproduction placed under international verification measures.[14] Concerns may exist, however, for those countries that remain outside of the CBW regimes, and who may have not insignificant stockpiles and infrastructure.

Another qualification relates to definitions. Any agent and precursor, equipment, weapon system, and means of delivery can be classified as CBW if intended for purposes other than those not prohibited by the Conventions. The GPC implies that CBW can be developed and deployed using a myriad of components or equipment and not just in ways or in forms developed (historically) by states. It seems more likely that the elements needed to develop CBW will be acquired based on their legitimate dual-use applications and not sourced openly as, for example, 'a chemical weapon'. Chlorine is a case-in-point: it can be – and has been - used as a chemical weapon, but its sale, transfer, and use is regulated in ways that would not serve at all to prevent its 'proliferation' in a traditional arms control and disarmament context. Our concerns are how chlorine is used in a specific setting, and not in its material existence. Thus, circumvention and (non-)acquisition/ (non-)proliferation of CBW can be understood to be on a spectrum, given the rather large definitional scale of potential CBW development and format possibilities.

At one end of such a spectrum, traditional conceptions of CBW may exist. In relation to chemical weapons, those that are on the CWC's Schedules, for example, and which face routine declarations to verify their non-production provide an example. At this end, there have been concerns about the proliferation of, in particular, knowledge, with Novichok serving as a recent example.[15] The controls implied through the CWC Schedules and verification regime come close to a traditional non-proliferation mechanism as there is a clearly defined and arbitrated *thing* which is controlled.

---

[13] Rappert, B. and McLeish, C. *A Web of Prevention: Biological Weapons, Life Sciences and the Governance of Research* (Earthscan, London: 2007)

[14] Notable exceptions recognised, for example ongoing discrepancies and uncertainties in regard to Russia and Syria.

[15] For example, Tucker, J. B. and Vogel, K. M 'Preventing the proliferation of chemical and biological weapon materials and know-how' (*The Nonproliferation Review*, 2000) 7(1); Guthrie, R. *The Salisbury Poisonings: Behind the Headlines* (forthcoming)

Moving along the spectrum, however, the hard and defined chemical weapon (and CBW more generally) related controls become more diffused into the wider webs of prevention. These are supposed to be developed and maintained through national implementation of the Convention(s) and through other mechanisms, for example forms of international cooperation such as protection and assistance. Here, the materiality of CBW is less clear, based primarily on the intended purpose to which chemistry and biology is put. The dual-use dilemma is particularly acute in this part of the spectrum, as depending on the stated *purpose* you may define (or struggle to define) activities, processes, outcomes, and technological artefacts as legitimate or illegitimate.

In all cases, however, there are a range of obstacles that can stand in the way of actors seeking a CBW capability. This can include the difficulty of acquiring agents and precursors to use; hiring people to work on CBW who can feel unshackled by moral opprobrium; the knowledge to develop, store, and deploy such weapons effectively; and remaining entirely undetected in these processes. Perhaps a wider obstacle now is the recognition that such weapons may not actually be as effective as one may hope them to be, especially given the constrains one must overcome to acquire them; the knowledge that modern countermeasures, public health responses, and CBW defences may render the effort fruitless, is itself an obstacle.

Of course, the use of CBW for different 'utilities', as discussed above, clearly complicates this matter. If mass destruction is not the goal, but rather demoralisation or terrorisation, then the nature of the obstacles that exist for such actors may also change.[16] Developing high-grade sarin, for example, is a very different set of processes to repurposing commercial chlorine. Nonetheless, to overcome these elements and complexities, actors require an inherent *intention* to pursue CBW and that intention must be maintained by motivation. Additional elements, such as resource, luck, creativity, and novelty, may well be important in sustaining that motivation; likewise, the lack of such elements may reduce motivation and alter intentions.

Indeed, CBW disarmament and non-acquisition efforts will be plagued with the persistent dual-use problem. As such, the challenge of circumvention and acquisition is connected to utilities: for actors to acquire CBW for particular ends there is a need to overcome traditional and new obstacles, and national and international efforts at non-acquisition. However, as advancing S&T changes the landscape of what is *possible* – and for what ends – this series of challenges remain significant and evolving.[17] To complicate matters further, the value that AI can bring in enabling an actor to pursue CBW does not imply that an actor will necessarily seek high-tech, high-casualty CBW capabilities. In fact, AI may support actors acquire low-tech, highly dual-use CBW which are already of the 'hard to prevent' kind. As such, AI raises the worrying potential that the low-hanging fruit might become even lower.

[16] 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)
[17] See, for example, Stewart, I. J. 'Generative AI and weapons of Mass Destruction: will AI Lead to Proliferation?' *Medium* 21/12/23 at: https://medium.com/@ian-j-stewart/generative-ai-and-weapons-of-mass-destruction-will-ai-lead-to-proliferation-c4476580bbc6

3.3 Creeping legitimisation

The phenomenon of creeping legitimisation, whereby unintentional corollaries and intentional CBW research, development, acquisition and employment gain acceptance over time, is a salient contemporary challenge. The effect is not linear, but rather the pressures that result in this ebb and flow. However, unchecked this can be seen as a gradual dilution of the prohibitions, through norm probing, norm weathering and ultimately norm erosion, as particular actors either carve out, or find themselves within, spaces in which (intentional or unintentional) CBW-related activity is pursued.[18]

It can frequently involve minor actions that collectively stretch boundaries and enable more permissive reinterpretations over time, but can also threaten from much larger events. Creeping legitimisation may stem from state or non-state actors across military, scientific, and political spheres, and exists as a direct result of the inverse inability to achieve (perhaps hypothetical) complete and total governance and accountability.

Indeed, there are myriad interconnected factors that contribute to pressures of creeping legitimisation. Most alarmingly, the almost 'routine' use of chemical weapons in Syria indicated a creeping legitimisation that State Parties are seeking to push back on. From Syria to Salisbury, taboos against CW employment are in focus as prohibitions *seem* to grow more malleable behind great power protection and plausible deniability. Unfounded Russian allegations about biolabs in Ukraine make light of the prohibitions, demonstrating the value of CBW allegations as a tool of propaganda and political point-scoring.

Moreover, inadequate enforcement and lack of consequences for verified CBW treaty violations could engender a 'compliance apathy' among some sections of the Conventions' States Parties. Ambiguous legal and rhetorical distinctions between acceptable and prohibited activities also facilitate boundary pushing, notable around definitions of riot control agents and means by which they are employed. Selective transparency via omissions in CWC declarations can further obscure potential violations from scrutiny.

Additionally, dual-use research with ostensibly defensive CBW applications risks enabling offensive capabilities; similarly, unchecked allocation of public funds for potentially dual-use research, without sufficient oversight, provides space for S&T developments that may be strategically appealing for unscrupulous actors. Indeed many advances in the life sciences are inherently dual-use and research into, or application of, particular knowledge or technologies can be viewed as efforts – especially those that are intentional – as probing at norms and stretching what might be possible.

Insidiously, each minor boundary-stretching action can lead to the contours of acceptability being redefined. Initially robust, the CBW Conventions may transform into spaces for reflections on diverging interpretations and the nature of

[18] Crowley, M. 'Monitoring and opposing the misuse of incapacitants – exploring the potential role for independent scientists' in Finney, J. L. and Šlaus, I. *Assessing the threat of weapons of mass destruction* (IOS Press; Amsterdam, 2010); Ilchmann, K. and Revill, J. 'Chemical and Biological Weapons in the "New Wars"' (*Science and Engineering Ethics*, 2013) 20; Revill, J. and Jefferson, C. 'Tacit knowledge and the biological weapons regime' (*Science and Public Policy*, 2014) 14

ambiguous grey zones if states find benefit in applying technologies that tow too closely to the edge. Therefore, gradual, almost imperceptible norm weathering and erosion over time is enabled by, and further enables, situations in which new utilities of CBW, opportunities to circumvent obstacles, and the potentials to acquire CBW become possible. As such, the challenge for the regime is monitoring and closing down spaces in which activities that push boundaries can be fostered. Clearly, advancing S&T make this effort difficult, and the dimensions of impact facilitated by AI clearly enable actors to push (intentionally or otherwise) into marginal spaces that allow them to test the prohibitions, nibble at its edges, or exist in grey ambiguity.

3.4 National interests and normative divergence

The integrity of the CBW regimes rely, more or less, on the consistent rejection of these weapons by all States Parties. That consistent rejection requires more than state delegations affirming the underlying principles in a diplomatic session: the machinery of the state at the national level must also believe it and be prepared to implement it. As the previous challenges have drawn attention to, there may indeed be space for national interests and attitudes towards CBW to diverge over time.

Reinterpretation of CBW acceptability could stem from a state's changing strategic outlook and geopolitical position. Doctrinal shifts, new threat perceptions, and the emergence of new capabilities driven by the life sciences and its convergence with other fields, may cause states to view the existing CBW normative framework as imposing unnecessary constraints on securing national interests. Stigma erosion through creeping legitimisation can also enable states to rationalise their covert efforts at stretching the prohibition.

It may then materialise that states pursue a particular capability that is fundamentally at odds with the collective norms enshrined within the CBW Conventions. Clandestine retention or breakout capability of CBW capacities, despite treaty commitments, is certainly one manifestation of normative divergence, but so too may be those states we may consider norm bystanders and abstainers.

The pursuit of novel CBW-applicable technologies draws on inherently dual-use S&T, however a tacit acceptance – or ignorance - of the ambiguity regarding ‘peaceful uses’ can reflect a normative relaxation. States may feel the implementation of the prohibitions are excessively restrictive for the development of their S&T prowess, and figure that national interests are better suited by static CBW Conventions in which the need for dynamic implementation of the general purpose criterion is of a lesser concern. Moreover, divergence between states is exacerbated by real and perceived asymmetries in oversight and accountability: double standards applied to different states undercut universality and unity. States may frame disregard for norms as correcting perceived hypocrisy and inequity, and be comfortable in prioritising their research and development opportunities over efforts to actively engender new restrictions and strengthened international oversight. As such, the challenges posed here not only stem from S&T but what actor’s feel enabled to do. Thus, we can perceive how in different contexts and conditions AI may facilitate actors to achieve certain objectives which clearly contravene the norm.

3.5 Remarks

The preceding sub-sections presented four main categories of challenge facing the CBW Conventions. These contain some of the main themes and possible pathways that could result in the effectiveness of the Conventions being undermined. They were briefly summarised and did not seek to articulate every permutation or idea. Yet, as broad categories they are inclusive. Importantly, they are not static or context specific, but rather are spatially and temporally flexible: many case studies about challenges to the CBW Conventions can be understood within one or more of these top-level categories. Inherent within all is the effect of advancing S&T and what that enables actors to achieve. While these do not specify what such advances may do in practical terms, they speak to the potential *effects.* As such, these categories can then be applied in specific cases. The case in the present paper is the effects of AI on the CBW Conventions, and so with these categories we have defined a framing device to revisit AI with these categories in mind, and seek to understand *how* and *why* emerging AI-anxieties may result in one or more of these categories of challenge.

4. Summary

This working paper proposes two framing devices – AI's dimensions of impact, and challenges to the CBW Conventions - as starting points to analyse AI's potential impacts on the chemical and biological weapons prohibition regimes.

The first delineates four dimensions of impact and their enabling capacities: process acceleration, pathway generation, ideational mediation, and transparency modification. These denote common routes through which AI could propagate negative outcomes by enabling actors. The second framing helps us then to categorise what those negative outcomes mean for the CBW Conventions. Four categories of challenge are identified: new utilities for CW; circumvention, proliferation, and acquisition; creeping legitimization; and norm divergence. These encompass major known threats to the CBW Conventions, providing a structure to contextualize and embed emerging AI-anxieties and their influences on actors.

Together, these aim to support more nuanced assessment of AI-anxieties by connecting scenarios to recognised challenges through identifiable impact pathways. This moves from AI specifics, to general dynamics and dimensions of impacts, to implications for the CBW Conventions. The goal is to elucidate tangible effects of seemingly intangible technologies by categorising them into familiar terms. Clearly, the next step is to use these framing devices to identify and design policies in response to the CBW specific perspectives the framings provide.

While not exhaustive, these frames offer initial conceptual platforms to build upon. Additional collaborative research and analysis can evolve the proposed typologies; it is critical that these concepts remain flexible frameworks for thinking and using, and not be thought of as not rigid solutions. They provide perspectives to illuminate connections and structure conversations. The devices offer starting points for policy discussions as well analysis, and can be utilised for structuring vignettes and case studies. While not predicting futures, these framing devices support analysis of the dynamics, impacts and choices we face. If constructive, they may have relevance beyond AI in the wider CBW S&T discourse.

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.