CBWNet

Strengthening the norms against
chemical and biological weapons

WORKING PAPER No. 05, July 2023

Artificial intelligence: possible risks and benefits for BWC and CWC.

Anna Krin, Gunnar Jeremias
Carl Friedrich von Weizsäcker Centre for Science and Peace Research (ZNF), University of Hamburg

Artificial intelligence: possible risks and benefits for BWC and CWC.

Anna Krin, Gunnar Jeremias
Carl Friedrich von Weizsäcker Centre for Science and Peace Research (ZNF), University of Hamburg

Executive Summary

Artificial intelligence (AI) is an emerging technology with a dual-use character. Concerns have been raised that some of its applications in life sciences can be misused by nefarious actors for the development of biological and chemical weapons, prohibited by the Biological Weapons Convention (BWC), and the Chemical Weapons Convention (CWC). Areas of AI applications relevant to the BWC and CWC include rational drug design, retrosynthesis planning, and synthetic biology. Research in such areas might also unintentionally produce knowledge, products, or technologies that could be used by others to cause harm.

Table 1: AI applications and misuse potential in rational drug design, retrosynthesis planning, synthetic biology

	Applications	*Misuse scenarios*	*Current limitations*
Drug design	Determination and analysis of correlations in biological datasets to provide a list of pharmacologically promising biological targets; prediction of the native 3D structure of proteins based on the given amino acid sequence; modeling of potential drug candidates.	May provide insight into the susceptibility of population or sub-population groups to some diseases and identify genetic key elements for a disease manifestation; can be used for a *de novo* design of toxic compounds; can be applied to discover new proteins from random amino acid sequences, which might have toxic effects in the human body.	Incomplete, insufficiently and inconsistently labeled data used for the algorithm training, as well as scarce or missing reporting of negative results in public literature limit the outcome of computational modeling; predicted molecules may be not synthesizable, stable, etc.; characterization of complex, highly dynamic molecular systems on a multilevel basis is not adequately captured by current mechanistic models; the correlations determined by the algorithm may be purely coincidental or erroneous; the software for modeling protein 3D structures is not designed to predict the effects of mutations on the native structure; predictions may fail in cases where proteins can adopt different conformations.
Retrosynthesis	Design of retrosynthesis routes using freely available and commercial AI platforms; automatized synthesis by coupling software to robotic systems.	May be misused to propose alternative retrosynthetic pathways for the compounds belonging to the category of chemical weapons.	Publicly available chemical data is highly heterogeneous (e.g., different representations, structured, vs. unstructured), often incomplete, and sometimes contradictory; reported errors in retrosynthesis routes proposed by AI software include a lack of atom conservation and nonsensical chemical transformations; data on the reaction conditions available for the training of the AI algorithms are often incomplete in published literature.
Synthetic Biology	Can drive the process of designing and fine-tuning the experiment, reducing the number of iterative Design-Build-Test-Learn (DBTL) cycles required; can be leveraged to analyze genomic data and to facilitate an understanding of the functional relationship between genome and phenotype manifestation.	Might foster the design of microbial pathogens with enhanced pathogenicity, expanded host range, altered transmission routes, resistance to the available countermeasures, abilities to evade the immune system response, etc.	AI automatization is in a developing stage due to the lack of standardization of hardware models, data flow, and representation; available datasets are oft incomplete in terms of recorded parameters, context-related information, uncertainty quantification, and evaluation of negative outcomes; standard AI evaluation metrics are inadequate for applications in synthetic biology, due to their incapability to capture the complexity and stochasticity of biological systems.

Most of the current limitations in the AI field will likely be overcome in the near future with the emergence of more efficient algorithms and the increasing amount and accessibility of the reported data. The threat landscape is also shaped by the availability of a great number of open-source tools to develop the respective AI-based computational tools “from scratch”. Therefore, a comprehensive legally binding framework is required to regulate AI in the context of biosecurity. The current solutions such as e.g. “Proposal for a Regulation laying down harmonised rules on artificial intelligence” of the EU (“AI Act”) are not sufficient to adequately address the biosecurity risks posed by some of the AI applications in life sciences. Last but not least, AI itself can play a role in strengthening biosecurity by expediting the development of vaccines and antidotes, introducing and improving detection methods, and supporting the implementation of BWC and CWC.

1. Setting the scope

We live in an era of technological advances and scientific breakthroughs. Improvements in medical diagnostics and treatment, more sustainable autonomous industrial processes, novel solutions for renewable energy, and powerful technical devices are just some of the achievements that can be attributed to the developments in life science and technology progressing at a high pace. However, some of these important milestones and insights have a two-sided character. They can be potentially misused by state or non-state actors for the development and production of biological or chemical weapons.

> *In general, the production, development, and stockpiling of biological and chemical weapons are prohibited by several international agreements, such as the Biological Weapons Convention (BWC 1972), and the Chemical Weapons Convention (CWC, 1993).*

In contrast to the CWC, the BWC regime currently does not have established verification mechanisms, an effective scientific and technical review, or other means to keep the treaty regime abreast of relevant R&D advances. In the CWC regime, the Organization for the Prohibition of Chemical Weapons (OPCW)[1], the implementing body of the CWC, is mandated to verify compliance with the provision of the CWC. The Scientific Advisory Board (SAB), a subsidiary body of OPCW, monitors the developments in scientific and technological fields that are relevant to the Convention.[2]

Concerns have been raised that advances in some areas of science and technology are reducing technical barriers and enabling alternative ways to acquire, manufacture, and disseminate hazardous biological agents and toxic chemicals. Of relevance is foremost scientific research which is conducted for solely peaceful purposes, but can potentially provide a toolbox for the production of warfare agents. In the life-science field, research with a specifically high misuse potential is referred to as dual-use research of concern (DURC).[3] It is conceivable that the growing accessibility of dual-use technologies, the rapid cost reduction for their application, and the decreasing expertise required for their use make them potentially attractive to various groups of actors with malicious intents.[4]

Numerous analytical working papers and academic publications are dedicated to the comprehensive discussion of the semantics (civilian vs military use; peaceful vs non-peaceful purposes), ethics (dilemma for scientists and other stakeholders), and regulatory framework of dual-use research.[4-5] In principle, all disciplines in life sciences possess to some extent a dual-use character. However, some particular research areas are repeatedly cited in the vast majority of DURC-related publications due to their relevance in the context of the BWC and CWC regimes. They include synthetic biology, genetic engineering, nanotechnology, artificial intelligence, and additive manufacturing, to name just a few. These research areas are summarized under the umbrella term "emerging technologies". The present working paper focuses on the discussion of artificial intelligence (AI) in life sciences, its current possibilities and limitations, and its implications for the BWC and the CWC.

2. AI as an emerging technology

There is no standardized definition of AI. According to the terminology adopted in the 2018 European Commission Communication[6] AI refers to systems that display intelligent behavior by analyzing their

[1] OPCW: https://www.opcw.org (accessed 2023).

[2] Scientific Advisory Board: https://www.opcw.org/about/subsidiary-bodies/scientific-advisory-board (accessed 2023).

[3] National Research Council . Biotechnology research in an age of terrorism: Confronting the dual use dilemma. National Academy of Sciences, Washington, DC: The National Academies Press, (2004).

[4] National Academies of Sciences, Engineering, and Medicine, Dual Use Research of Concern in Life Sciences: Current Issues and Controversies. Washington (DC): National Academies Press, (2017).

[5] Miller S., Selgelid M.J. Ethical and philosophical consideration of the dual-use dilemma in the biological sciences. *Sci. Eng. Ethics*, 13(4), 523-580, (2007).

[6] European Commission: Artificial Intelligence for Europe, COM(2018), 237 final, April 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN (accessed 2023).

environment and acting – with some degree of autonomy – to achieve specific goals. The synonymical term is "weak AI" to distinguish it from the so-called "strong AI", which is capable of self-controlled thinking and learning and belongs to the realms of highly speculative future developments.

Currently, the subject of AI dominates many news article headlines, especially due to the growing popularity of some of its applications, such as the chatbot ChatGPT, which was developed by OpenAI and launched in November 2022. These developments might create the impression that this rapidly progressing research area has emerged only recently. However, the term "AI" was coined in 1956 by John McCarthy[7], while the underlying concepts were introduced even earlier. The long development path of AI is characterized by numerous milestones with their contribution to the progression in the field. Essential prerequisites for contemporary successful AI development were the accessibility of large amounts of data (big data) and the availability of computing capacities for processing and storing them. The development in the field of AI is further promoted by diminishing computational costs and the availability of various open-source toolkits and libraries that are constantly emerging (TensorFlow, Scikit, OpenCV, PyTorch, Keras, etc.). The popularity of these software tools is reflected in the constantly growing number of their users.[9]

The importance of high-quality data for efficient AI training cannot be underestimated. It is, in fact, the major bottleneck of the technology. Training and validating data sets need to be sufficiently large and representative[8] to reduce bias and to optimize AI performance. This seemingly trivial requirement cannot always be easily fulfilled in reality: data availability can be restricted due to licensing policies, ethical and security considerations, and proprietary rights. This legitimate limitation has bearing on the central question of what is feasible with AI, and what is borderline.

Since its beginnings in the 1950s, AI has gone through several phases of development marked by different achievements that made this technology indispensable in many areas of our everyday life: from mobile voice assistants and online search engines, to self-driven cars and autonomous vehicles, to sophisticated applications in medical diagnostics, chemical and material science, drug design, and robotics.

The increasing number of AI-related publications and patents demonstrates the high impact of AI technology on different areas.

As highlighted in the Artificial Intelligence Index Report 2022,[9] the number of patents filed in 2021 is more than 30 times higher than in 2015 (a compound annual growth rate of 76.9%), and the total number of AI publications doubled in the last decade, with the overall number of worldwide AI publications exceeding 334.000 (status: 2021).[9] Progress in AI rests on the pillars of various converging disciplines including engineering, mathematics, data processing, neurosciences, linguistics, physiology, medicine, chemistry, biology, etc.

The bibliometric analysis by Karger et al.[10] for the period from 1991 to 2022 revealed that the spectrum of publications in which AI/machine learning/deep learning was relevant, covers over 26 research areas, which demonstrates the influence of these computing systems. Areas, in which AI technology has been extensively applied lately:

- Biochemistry, Genetics, and Molecular Biology
- Computer Science
- Pharmacology, Toxicology, and Pharmaceutics
- Chemistry
- Medicine

[7] Russell S.J, Norvig P. Artificial Intelligence: A Modern Approach. Pearson series in artificial intelligence. Pearson education limited., (2020).

[8] Boucher P., Artificial intelligence: How does it work, why does it matter, and what can we do about it? EPRS: European Parliamentary Research Service. Belgium, (2020).

[9] Zhang D., Maslej N., Brynjolfsson E., et al. The AI Index 2022 Annual Report, AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, March (2022).

[10] Karger E., Kureljusic M. Using Artificial Intelligence for Drug Discovery: A Bibliometric Study and Future Research Agenda. *Pharmaceuticals*, 15(12), 1492, 1-22, (2022).

This indicates the important role of AI in life sciences. The above-mentioned aspects underpin the classification of AI as an emerging technology. The term “emerging technology” has a range of definitions depending on the context, e.g., management, IT, science & technology policy, economics, etc.[11] However, some commonly used criteria for categorizing technology as “emerging” are a strong impact across a range of sectors of the economy and society,[12-13] movement beyond the purely conceptual stage,[12] novelty, and growth.[14] The term “novelty” in the context of AI does not refer to its historical timeline, but rather to its ever-evolving manifold applications.

AI progresses at high speed and starts shaping many areas of research and industry. Given the rapid spreading of this technology, ethical and security concerns related to some of its applications should be analyzed in detail to reduce any possible harm to society or individual vulnerable groups. Here we focus on some selected examples of AI applications in life sciences relevant to the context of the BWC and CWC regimes.

3. Rational drug design

Pharmacology and biochemistry are currently thriving with large datasets: information from genomics, proteomics, and metabolomics (“omics”) together with the results from toxicology and physiology represent an intricate landscape of data on potential drug targets. Nevertheless, identifying a drug target (a biological entity in our body interacting with therapeutics to produce a physiological response) for the development of a high-demand efficient drug based on this wealth of information is as challenging as looking for the proverbial needle in the haystack. Machine learning can be applied to tackle these challenges: algorithms have been developed to determine and predict correlations in biological datasets to provide a list of pharmacologically promising biological targets.

> *More than half of the human proteome has a link to any disease but has not been studied for binding to small molecules, while 38% of the entire proteome remains unstudied.[15] AI technology can steer the forward movement in this information space.*

In general, the rational drug design process consists of several consecutive key steps schematically depicted in Figure 1, each with its own set of challenges and particularities.

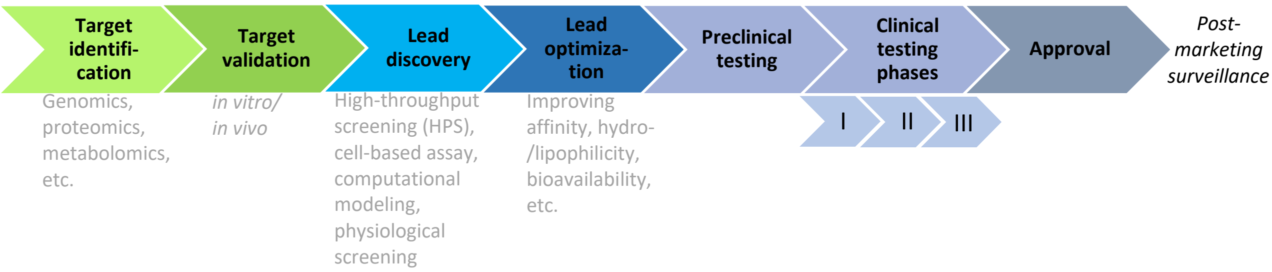
Target identifi-cation
Target validation
Lead discovery
Lead optimiza-tion
Preclinical testing
Clinical testing phases
Approval
Post-marketing surveillance
Genomics, proteomics, metabolomics, etc.
in vitro/ in vivo
High-throughput screening (HPS), cell-based assay, computational modeling, physiological screening
Improving affinity, hydro-/lipophilicity, bioavailability, etc.
I
II
III

Figure 1: Schematic representation of the drug discovery process. The first steps up to clinical testing can be supported by an AI-driven application.

The heart of every drug discovery process is the identification of a molecular component that interacts with the biological target (e.g. a receptor) in a desired manner. This oversimplified description implies a meticulous search for an entity, which would meet a long catalog of criteria (safety, efficiency, bioavailability, etc), progressing from an initial “hit” to the subsequent promising (“lead”) component, which upon further optimization would

[11] Rotolo D., Hicks D., Martin B.R. What is an emerging technology? *Res. Policy*, 44(10), 1827-1843, 2015.

[12] Stahl B.C. What does the future hold? A critical view on emerging information and communication technologies and their social consequences. In Chiasson M., Henfridsson O., Karsten H., DeGross J.I., editors, Researching the Future in Information Systems: IFIP WG 8.2 Working Conference, Future IS 2011, Turku, Finland, Proceedings, 59–76, Springer, Heidelberg (2011).

[13] Martin B.R. Foresight in science and technology. *Technol. Anal. Strateg. Manag.*, 7(2), 139–168, (1995).

[14] Small H., Boyack K.W., Klavans R. Identifying emerging topics in science and technology. *Res. Policy*, 43(8), 1450–1467, (2014).

[15] Doytchinova I. Drug Design-Past, Present, Future. *Molecules*, 27(5):1496, 1-9, (2022).

enter the preclinical/clinical testing phase. For the past two decades, the process of lead discovery was facilitated by the high-throughput screening (HTS) methods with estimated hit rates of 0-0.01%.[16-17] Leveraging AI in this field opens up the possibility of generating new chemical entities from scratch without the necessity of performing HTS.[17]

Not all potential protein targets and protein-based therapeutics have available structural information for the characterization of molecular interactions. AI tools for the prediction of the native protein 3D structure such as nRoseTTAFold[18] can provide the first guess of the respective structure to guide further research. Another deep learning approach, AlphaFold, has attracted substantial attention due to its accurate predictions of protein 3D structures from the given amino acid sequence.[19] The open-access database of predicted protein structure hosted by the AlphaFold developers DeepMind and EMBL-EBI includes as of now over 200 million proteins, “almost every protein known to science”[20]. In comparison, experimentally-derived data collected in the protein database PDB currently encompasses over 200 thousand structures.[21]

The process from the design to the market introduction of the finished product takes over 10-15 years, and the average costs of drug development are assumed to be about $1-2 billion.[22] Undeniably, the leverage of AI in drug discovery is highly beneficial and promising in terms of time and cost reduction.

> *On the other hand, concerns have been raised that this technology can be misused for the targeted development of novel biochemical weapons, given the huge amount of scientific data and computational tools which can be exploited to implement an AI-based prediction model for tailored toxic substances.*

Considering the fast progress and dynamic changes in the field of AI-driven drug discovery, these concerns have a solid foundation and require a scrutinized assessment. Thus, machine learning-based correlation analysis can potentially convey information on hitherto unstudied biological targets in the human organism involved in crucial physiological cascades, which can be efficiently deregulated to cause severe or even lethal effects. Additionally, genomic data mining might provide insight into vulnerabilities of population or sub-population groups in the form of susceptibility to some diseases, and identify genetic key elements for a disease manifestation (a goal usually pursued in the search for novel therapies, which can be misused for malicious purposes). However, the predicted biochemical relationships will still have to be verified experimentally.

It is a ubiquitous phenomenon that complex technologies are also associated with risks from possible misuse for military or criminal purposes. A recent misuse scenario for AI-guided drug design has raised concerns among the general public and scientific communities. As demonstrated by Urbina et al., AI technology can be exploited for the *de novo* design of highly toxic chemicals.[23] In a proof-of-principle study, Urbina et al. altered their commercial machine-learning software for the modeling of toxic entities, exploring the chemical space around the known nerve agent VX.[23] This computational approach resulted in a list of 40,000 molecules, including novel compounds with a predicted *in vivo* toxicity higher than of the known chemical warfare

[16] Bender A., Bojanik D., Davies J.W., et al. Which aspects of HTS are empirically correlated with downstream success? *Curr. Opin. Drug Discov. Devel.,* 11, 327–337, (2008).

[17] Schneider P., Walters W.P., Plowright A.T., et al. Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.*, 19(5), 353-364, (2020).

[18] Baek M., DiMaio F., Anishchenko I., et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557), 871-876, (2021).

[19] Jumper J., Evans R., Pritzel A., et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589 (2021).

[20] Heikkilä M. MIT Technology Review : DeepMind has predicted the structure of almost every protein known to science https://www.technologyreview.com/2022/07/28/1056510/deepmind-predicted-the-structure-of-almost-every-protein-known-to-science/ July, 2022 (accessed 2023).

[21] Protein Data Bank (PDB): https://www.rcsb.org/ (accessed 2023).

[22] Sun D., Gao W., Hu H.,et al. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B*. 12(7), 3049-3062, (2022).

[23] Urbina F., Lentzos F., Invernizzi C., et al. Dual use of artificial-intelligence-powered drug discovery. *Nat. Mach. Intell.* 4, 189–191 (2022).

agents. This illustrates that AI can be potentially applied by actors with malicious intents to explore the area of new highly potent biochemical weapons.

AI-based software can be used not only for a de novo design of small molecules but also for macromolecules, potentially enlarging the arsenal of biochemical substances which can be explored by nefarious actors.

In principle, the novelty of such chemicals might slow down their identification and the initiation of the required countermeasures. Furthermore, AI technology can be misused not only for the modeling of novel lethal warfare agents but also for the design of their subsequent synthesis. Some of the capabilities and limitations of machine learning-driven retrosynthesis platforms are discussed in Section 4 and will not be analyzed here.

Further misuse potential of AI in drug discovery is related to computational structure predictions. If the structural information on the biological target or the interacting molecular entity cannot be retrieved from e.g. PDB database, it can be predicted using efficient AI modeling tools, such as the freely available AlphaFold or RoseTTAFold, provided that the amino acid of the protein is known. This software can even be applied to discover entirely new proteins as shown by Anishchenko et al.[24], who used deep neural networks (category of machine learning algorithms) to design ("hallucinate") new proteins starting from a random sequence of amino acids. A portion of 129 "hallucinated" proteins was expressed in bacteria *E. coli,* resulting in 27 proteins with recorded experimental spectra consistent with the modeled structures. The experimental structure determination of the three of these species revealed a close match with the prediction. This highlights the current predictive power of AI-based computational tools which is expected to increase further, once more high-quality research data flows into the algorithm training.

A critical validation of the achievements and limitations of AI applications is required to set the scope for current biosecurity issues and to identify the key points, which will gain relevance in the near future. Here, we will address the current limitations of computer-aided drug design, bearing in mind that the same issues are of relevance in the discussion of AI misuse potential for a *de-novo* synthesis of hazardous molecular compounds. It's still early days for this target-identification technique.[25] One of the challenges is e.g. the analysis of omics data. Many of the established AI algorithms are not robust enough for the analysis of these datasets due to the inherent difference between such data and other types of information (e.g. text data), for which these algorithms are designed and applied.[26] The main difference is the context dependence of the omics data, which makes it challenging to extract meaningful information via data extrapolation.[26] The characterization of complex, highly dynamic molecular systems on a multilevel basis (e.g. genome vs proteome) is not adequately captured by current mechanistic models. Predictive modeling is further hampered by the tendency of algorithms to determine correlations, some of which may be purely coincidental or erroneous.[27] Furthermore, the collected omics data often lack important information such as proper labeling of recorded parameters, the context of the sampling, and time resolution necessary to comprehend the dynamic changes in the living. As demonstrated by Buescher et al., the integration of multiple omics datasets to understand their crosstalk is more than the sum of the individual experiments.[28] This significantly affects the applicability of the AI algorithms and their reliability, when it comes to constructing models and predicting properties based on the data extracted from such datasets.

[24] Anishchenko I., Pellock S.J., Chidyausiku T.M., et al. De novo protein design by deep network hallucination. *Nature*, 600, 547–552 (2021).

[25] Heaven W.D. MIT Technology Review: AI is dreaming up drugs that no one has ever seen. Now we've got to see if they work. February 2023 https://www.technologyreview.com/2023/02/15/1067904/ai-automation-drug-development/ (accessed 2023).

[26] Eslami M., Adler A., Caceres R.S., et al. Artificial intelligence for synthetic biology. *Commun. ACM.* 65(5). 88-97, (2022).

[27] Yeo H.C., Selvarajoo K. Machine learning alternative to systems biology should not solely depend on data. *Brief Bioinform.,* 23(6):bbac436, 1-6, (2022).

[28] Buescher J.M., Driggers E.M. Integration of omics: more than the sum of its parts. *Cancer Metab*., 4(4), 1-8, (2016).

Limitations are also known concerning the performance of machine learning software for predicting protein 3D structures. Thus, the AlphaFold algorithm is not designed to predict the effects of mutations on the native structure. Prediction is also made on the assumption of a protein "in a vacuum": a molecular entity not interacting with other complex-building proteins (although some progress in this aspect has been made recently[29]) or with compounds ("ligands"), which upon binding induce a conformational change in the respective protein. Giving accurate predictions for proteins with a single well-defined 3D structure, the algorithm may fail in cases where proteins "can adopt different structures in different conformations"[30]. Examples of structural predictions by the software, which could not be verified experimentally have been reported.[30] In one of these cases the modeled structures of the members of the so-called G-protein coupled receptors, which are important for signal transduction into the cell, were incorrect according to the experimental data despite the high confidence of the algorithm in the accuracy of the prediction.[30] This demonstrates that the results obtained with the artificial neural networks cannot be viewed as a substitute for experimental work and expert knowledge, but rather as an approximation complementing the existing experimental methods.

The major limitation in the field of AI-based drug discovery remains the quality of available chemical datasets for algorithm training, the key aspect of AI efficacy.

The available amount of results from the *in vivo* experiments is limited. Training data sets therefore predominantly contain *in vitro* results. Since the *in vitro* conditions do not entirely resemble the *in vivo* situation, the prediction efficiency and the feasibility of the results obtained are reduced. Also, the animal model data has limited transferability regarding drug metabolism in the human body.

Further limitations concerning the available datasets are:

- incomplete, insufficiently, and inconsistently labeled data
- scarce or no reporting of negative results
- limited number and low heterogeneity of the molecules to create a predictive model.[17]

This observation has implications for whether AI would indeed be at the current stage the game-changer for actors with malicious intent.

Computational modeling of a structure and its chemical and physiological properties alone do not necessarily imply that the component is synthesizable and will interact in the body in a predicted manner. Synthesizability is one of the issues known in the context of AI-based *de novo* molecular design.[31] Also, considerations concerning its stability and the transportation routes within the body will not be rendered obsolete by AI. These aspects still require deep expertise in the field. As stated by the CEO of one of the California-based drug companies: "If somebody tells you they can perfectly predict which drug molecule can get through the gut or not get broken up by the liver, things like that, they probably also have land to sell you on Mars"[25]. This is maybe the reason why AI has not yet led to an expected breakthrough in pharmacology. A survey of selected drug discovery companies using AI shows that only a fraction of publicly disclosed drug candidates progressed into clinical trials, despite a large number of programs in the preclinical stage in 2010-2021.[32] That said, this aspect should be viewed critically in terms of the misuse potential of AI technology since the requirements for a compound developed for non-peaceful purposes differ from those of the pharma industry. A high toxicity profile, one of the reasons for the failure of the drug candidates in early clinical testing, is a property desirable for a novel chemical weapon component. An AI algorithm can be designed (or modified) to search for the components with the highest predicted toxicity, as shown in the previously mentioned study by Urbina et al.[23] Still, as emphasized above, the question remains, whether the predicted molecules will be synthesizable, stable, volatile, etc., which were not further assessed by Urbina et al.

[29] Evans R., O'Neill M., Pritzel A., et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 10.04.463034, 1-25, (2021).

[30] Callaway E. What's next for AlphaFold and the AI protein-folding revolution. *Nature*, 604, 234-238, (2022).

[31] Gao W., Coley C. W. The Synthesizability of Molecules Proposed by Generative Models. *J. Chem. Inf. Model.*, 60, 15714–5723, (2020).

[32] Jayatunga M.K.P., Xie W., Ruder L., et al. AI in small-molecule drug discovery: a coming wave? *Nat. Rev. Drug Discov.,* 21(3), 175-176, (2022).

4. Retrosynthesis planning

Once the molecule with the desired properties has been determined, the next crucial step is to plan the optimal retrosynthesis strategy, i.e. to find recursively the synthetic pathway to obtain the compound of interest from readily available consumables (Figure 2). This is not a straightforward task, in particular for complex multi-step reactions. Recently, AI-based solutions have been increasingly gaining attention as a tool to automatize this intricate process.

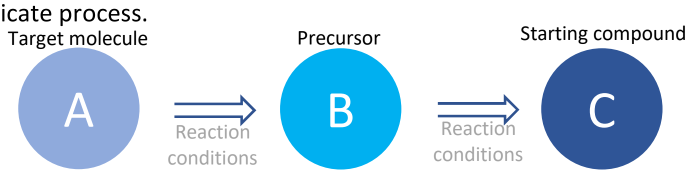

Figure 2: Simplified representation of a retrosynthesis pathway.

The first attempts in the area of computer-aided synthesis planning have been undertaken in the 1960s.[33] However, the field has flourished only recently, partly due to the aforementioned improvements in data storage and processing and the growing amount of scientific data. This data has been collected in publicly available and commercial databases. Some of them are summarized in Table 2, also the list is far from exhaustive. Further open-access chemical reaction databases are currently under development.[34-35] Commercial AI implementations for retrosynthesis design and planning have largely found their way into daily chemical laboratory practice. Reaxys synthesis planning[36], CAS SciFinder[37], and ChemAIRS[38] are some of the commonly used ones. In addition to potential synthesis routes, information on required reaction conditions and even on the pricing of the necessary compounds may be provided by the software. Results of a recently conducted Turing-like test (to access if a computer responds in a human-like manner[39]) show that some of the retrosynthesis routes proposed by AI are largely indistinguishable from those that would be designed by human experts.[40]

Many freely available and commercial platforms have been developed for tailored retrosynthesis planning. Some solutions for coupling this software to robotic systems for a fully automatized synthesis process have also been presented.

One such example is the open-source software ASKCOS, trained on the data from USPTO (US Patent and Trademark Office)[41] and Reaxys. The synthetic route proposed by the algorithm is subsequently validated by a chemist, who also configures the required operations for the robot arm to perform the synthesis.[42] Another popular open-source software is AiZynthFinder[43], which utilizes different multi-step search algorithms to increase efficiency. Also, the AI-driven tool RXN developed by IBN is an online platform for both forward reactions and retrosynthesis planning. The obtained results can be used in combination with another part of the toolset, RoboRXN, “the first remotely accessible, autonomous chemical laboratory”[44].

[33] Corey E.J., Wipke W.T. Computer-Assisted Design of Complex Organic Syntheses. *Science*, 166(3902), 178–192 (1969).

[34] Tavakoli M., Chiu Y.T.T, Baldi P., et al. RMechDB: A Public Database of Elementary Radical Reaction Steps. *J. Chem. Inf. Model.* 63(4), 1114-1123, (2023).

[35] Kearnes S.M., Maser M.R., Wleklinski M., et al. The Open Reaction Database. *J. Am. Chem. Soc.* 143(45), 18820–18826, (2021).

[36] Reaxys https://www.elsevier.com/solutions/reaxys (accessed 2023).

[37] CAS SciFinder https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder (accessed 2023).

[38] ChemAIRS https://chemairs.chemical.ai/ (accessed 2023).

[39] Turing, A.M. I.—computing machinery and intelligence. *Mind*, LIX(236), 433–460, (1950).

[40] Mikulak-Klucznik B., Gołębiowska P., Bayly A.A., et al. Computational planning of the synthesis of complex natural products. *Nature,* 588, 83–88 (2020).

[41] United States Patent and Trademark Office https://www.uspto.gov/ (accessed 2023).

[42] Coley, C.W., Thomas D.A., Lummiss J.A.M., et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* 365 (6453), 1-9, (2019).

[43] Genheden S., Thakkar A., Chadimová V., et al. AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. *J. Cheminformatics.* 12(1), 1-9, (2020).

[44] RNX for chemistry https://rxn.res.ibm.com/ (accessed 2023).

Table 2: Some of the open-source and commercial databases for chemical reactions and compounds.

Open-source databases	
Name	Data amount
United States Patent and Trademark Office (USPTO)-derived data	>1.9 million reactions (1976-2016)[45-46]
ChemSpyder &ChemSpyder synthetic pages	>100 million structures[47]
ZINC	>230 million commercially available compounds[48]
ChEMBL	2.4 million distinct components[49]
PubChem	115 million unique chemical structures[50]
Commercial databases	
Reaxys®	>57 million chemical reaction entries[36] > 170 million compounds
CAS® SciFinder	> 150 million single- and multistep reactions[37]
Pistachio	13.3 million reactions[51-52]

The question, which arises in this context, is biosecurity-related: can these beneficial tools be misused to propose retrosynthetic routes for the compounds belonging to the category of chemical weapons? The respective substances together with their precursors and derivatives are listed in Schedules 1-3 of the "Annex on Chemicals"[53] provided by the OPCW. Facilities that produce (1-3), process (2), or consume (2) these scheduled chemicals must be declared and can be subjected to regular inspections conducted by the OPCW according to the Parts VI-VIII of the "Verification Annex".[53] On a national level, law enforcement units ensure that no illegal purchase, production, or stockpiling of such compounds occurs within the country following the provisions of the CWC. There is a concern that technologies such as AI-driven retrosynthesis tools may be used to circumvent the implemented security measures and determine alternative synthetic routes without involving regulated and monitored chemicals.

Undoubtfully, as discussed above, modern AI-based software has great potential. Given a large amount of available data from various sources, such software can develop synthesis routes not published in literature based on pattern recognition and correlation analysis.

Even if currently available software (free or commercial) would be equipped with security-related restrictions to undermine its misuse by somebody with malicious intent, it does not diminish the risk, since such software can be programmed by everybody with sufficient expert knowledge.

The last aspect is highlighted by the availability of the aforementioned open-source libraries for the development of such programs (Section 2). Nevertheless, a careful risk assessment requires not only the characterization of possibilities of the considered technology but also the evaluation of its limitations.[45] As in the case of AI used in drug design, the restricting factors in the AI performance for retrosynthesis remain the currently available datasets used for the model training. This aspect can be seen as a generally valid point that makes the difference between the hype around AI and the reality of now. Arguably, a large body of literature has been accumulated over the years on different reaction mechanisms and synthetic pathways which can be used to retrieve information on the synthesis and the components involved and to train the machine learning algorithms.

[45] Lowe, D.M. Extraction of chemicalstructures and reactions from the literature. Ph.D. thesis, University of Cambridge, (2012).
[46] Zhong Z., Song J., Feng Z., et al. Recent advances in artificial intelligence for retrosynthesis. 10.48550/arXiv.2301.05864 (2023).
[47] ChemSpyder http://www.chemspider.com/Default.aspx (accessed 2023).
[48] ZINC https://zinc15.docking.org/ (accessed 2023).
[49] ChEMBL https://www.ebi.ac.uk/chembl/ (accessed 2023).
[50] PubChem https://pubchem.ncbi.nlm.nih.gov/ (accessed 2023).
[51] Mayfield J., Lagerstedt I., Sayle R. Pistachio Fantastic reactions and how to use them. NIH Virtual Workshop on Reaction Informatics, May (2021).
[52] NextWove Software. Pistaccio https://www.nextmovesoftware.com/pistachio.html (accessed 2023).
[53] OPCW. Annexes https://www.opcw.org/chemical-weapons-convention/annexes (accessed 2023).

However, publicly available chemical data is highly heterogeneous (e.g., different representations, structured, vs. unstructured), often incomplete, and sometimes contradictory. A recent analysis of over 125.000 pharmaceutical patents from 1976–2015 revealed a lack of essential information including e.g. reaction types or obtained product yields in a large portion of the documents examined.[54-55] Furthermore, a survey of over 1 million applied reactions showed little diversity in the reaction mechanisms, with a biased preference for some particular methods.[55] This limits the scope of chemical data, on which an AI algorithm can be trained to propose a novel (alternative) synthesis route. Datasets usually contain only successful reactions, excluding the failed ones, which are also required for efficient machine learning. All these aspects can lead to poor performance of an AI-based retrosynthesis planning platform, in particular for very specific reactions. Reported errors in AI-guided retrosynthesis include a lack of atom conservation and nonsensical chemical transformations.[56] These challenges pose limitations to both the AI algorithms used for benign and non-peaceful purposes.

In principle, breaking down the regulated components until non-regulated commercially available reagents remain for the synthesis implies creating large and cumbersome synthetic routes. The longer such predicted routes, the lower the prediction confidence level, which might result in chemically implausible and unfeasible pathways. Moreover, the recipe for a successful retrosynthesis consists of more than listing the required synthesis steps. Additional information on reaction conditions, solvents, catalysts, and concentrations is indispensable. As mentioned above, some software currently available does provide information on these parameters. Still, data on the reaction conditions are often incomplete in published literature, limiting also their predictability by the software. Although examples of the accurate prediction of reaction conditions with AI have been reported[57-58], a critical study evaluates some of these results as "an overoptimistic interpretation".[59]

> *The study by Beker et al. shows that an abundance of carefully curated literature data may be insufficient for accurate models of chemical reactivity.[59] Based on the selected example of cross-coupling reactions, they demonstrate that, despite the large database used for machine learning, no meaningful prediction of optimal reaction conditions could be obtained.[59]*

These are just a few caveats in the chemical context of the question at hand. As a consequence, no study so far has reported a novel synthetic route provided entirely by AI and synthesized in the laboratory or industrial setting.[46]

5. Synthetic biology

Synthetic biology is a synergy of biology and engineering principles that nowadays transforms a vast number of sectors including medicine, drug discovery, food industry, energy research, and material science. Advanced applications of synthetic biology enable the building of molecular blocks and circuits from standardized biological parts[60]; the construction of artificial biological systems from synthetic genomes[61]; the

[54] Schneider N., Lowe D. M., Sayle R.A., et al. Big data from pharmaceutical patents: A computational analysis of medicinal chemists' bread and butter. *J. Med. Chem*., 59(9), 4385–4402 (2016).

[55] Almeida A.F., Moreira R., Rodrigues T., Synthetic organic chemistry driven by artificial intelligence. *Nature Rev. Chem.*, 3, 589-604 (2019).

[56] Borrelli W., Schrier J. Evaluating the Performance of a Transformer-based Organic Reaction Prediction Model. *ChemRxiv.*, 3nqv9, (2021).

[57] Gao H., Struble T.J., Coley C.W., et al. Using Machine Learning to Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci*., 4(11), 1465–1476, (2018).

[58] Maser M.R., Cui A.Y., Ryo, S., et al. Multilabel Classification Models for the Prediction of Cross-Coupling Reaction Conditions. *J. Chem. Inf. Model*., 61(1), 156–166, (2021).

[59] Beker W., Roszak R., Wołos A., et al. Machine learning may sometimes simply capture literature popularity trends: A case study of heterocyclic Suzuki-Miyaura coupling. *J. Am. Chem. Soc.,* 144 (11), 4819-4827 (2022).

[60] Knight T. Idempotent Vector Design for Standard Assembly of BioBricks. MIT Artificial Intelligence Laboratory; MIT Synthetic Biology Working Group, (2003).

[61] Venter J.C., Glass J.I., Hutchison C.A., et al. Synthetic chromosomes, genomes, viruses, and cells. *Cell* 185(15). 2708-2724, (2022).

engineering of microorganisms for the production of desired compounds[62]; the manufacturing of novel vaccines, diagnostics, and therapeutics[63]; the expansion of the genetic code to reprogram the native cell translation machinery[64], etc. Precise DNA manipulation is possible with the CRISPR/Cas9 systems, the so-called "genetic scissors".[65] This method has been applied to edit the genomes of various organisms, including bacteria and viruses. Many controversial applications of synthetic biology have sparked debates about dual use and biosecurity, which are beyond the scope of this manuscript.

The experimental procedure in synthetic biology is a sequential iterative process that undergoes the phases of design, implementation, testing, and review of results and failures, referred to as the Design-Build-Test-Learn (DBTL) cycle. AI applications can drive the process of designing and fine-tuning the experiment, reducing the number of iterative cycles required.[26] Neural network models have been e.g. used to design new biological constructs[66], determine plasmid expression, optimize nutrition and fermentation conditions, and predict CRISPR guide efficacy.[26] Additionally, they can be leveraged to analyze genomic data and to facilitate an understanding of the functional relationship between genome and phenotype manifestation.

> *Despite their promising applications at the frontiers of molecular biology, such AI-driven approaches raise biosecurity concerns. They might foster the design of microbial pathogens with enhanced pathogenicity, expanded host range, altered transmission routes, resistance to the available countermeasures, abilities to evade the immune system response, etc.*[67-68]

Machine learning has been e.g. used to improve the production fitness of adeno-associated virus (AAV), a vector used in gene therapy[69], detect novel pathogens from the next-generation sequencing data[70], predict pathogenic potentials for unknown, unrecognized, and novel (e.g. synthetic) DNA sequences.[71] Also, AI-based structure predictions such as e.g. AlphaFold 2 or RoseTTAFold can be exploited in the effort to generate infectious viruses from synthetic DNA or enhance known pathogens.[72] According to the developers of AlphaFold 2, the viral proteins have been excluded from the openly available version of AlphaFold.[73] However, the availability of the source code and respective datasets for the training of the algorithm questions the effectiveness of such precautions.

Computational modeling, however advanced, does not replace experimental endeavor and expertise still required to implement and test the model. Nonetheless, the ongoing automatization of DBTL cycles promoted by the convergence of AI and robotics can lower the know-how threshold and make the technology more accessible, also to nefarious actors.

[62] Chubukov V., Mukhopadhyay A., Petzold C.J., et al. Synthetic and systems biology for microbial production of commodity chemicals. *npj Syst. Biol. Appl.* 2(16009), 1-11, (2016).

[63] Tan X., Letendre J.H., Collins J.J.,et al. Synthetic biology in the clinic: engineering vaccines, diagnostics, and therapeutics. Cell, 184(4), 881-898, (2021).

[64] Shandell M.A., Tan Z., Cornish V.W. Genetic Code Expansion: A Brief History and Perspective. *Biochemistry*, 60(46), 3455-3469, (2021).

[65] Doudna J.A., Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213), (2014).

[66] Eastman P., Shi J., Ramsundar B., et al. Solving the RNA design problem with reinforcement learning. *PLoS Comput. Biol.,* 14(6), 1-15 (2018).

[67] Brockmann K, Bauer S, Boulanin V. BIO PLUS X: Arms Control and the Convergence of Biology and Emerging Technologies. Solna, Sweden: Stockholm International Peace Research Institute, (2019).

[68] O'Brien J.T., Nelson C. Assessing the Risks Posed by the Convergence of Artificial Intelligence and Biotechnology. *Health Secur.,* 18(3), 219-227, (2020).

[69] Mikos G., Chen W., Suh J. Machine learning identification of capsid mutations to improve AAV Production Fitness. *bioRxiv,*. 1-10, (2021).

[70] Deneke C., Rentzsch R., Renard B.Y. PaPrBaG: a machine learning approach for the detection of novel pathogens from NGS data. *Sci Rep.*, 7(39194), 1-13, (2017).

[71] Bartoszewicz J.M., Seidel A., Rentzsch R., et al. DeePaC: predicting pathogenic potential of novel DNA with reversecomplement neural networks. *Bioinformatics*, 36(1), 81-89, (2020).

[72] Sandbrink, J.B., Alley, E.C., Watson, M.C. et al. Insidious Insights: Implications of viral vector engineering for pathogen enhancement. *Gene Ther.*, 30, 407-410, (2022).

[73] Perrigo B. Google's AI Lab, DeepMind, Offers 'Gift to Humanity' with Protein Structure Solution, Time, July 2022, https://time.com/6201423/deepmind-alphafold-proteins/ (accessed 2023).

Experts in the field envision the development of fully-automated self-driven labs to process data, formulate hypotheses and theories and verify them experimentally.[74] An already existing closed-loop robotic system can for instance design and perform experiments to determine gene functions.[75] The technology is currently in a developing stage due to the lack of standardization of hardware models, data flow and representation, and intelligent experiment-selection algorithms.[76]

> *Further advances in the field should be monitored concerning biosecurity, especially with the increasing emergence of the so-called cloud labs. These remote automated workstations promise to improve experimental reproducibility and provide affordable access to sophisticated equipment but can also open a new avenue for misuse.[77] Currently, no active measures are implemented by the providers of such platforms to guard against exploitation for non-peaceful purposes.[78]*

Careful monitoring of the technology readiness level is required to assess possible biosecurity threats from AI applications in synthetic biology. Some general drawbacks, such as the completeness of available datasets in terms of recorded parameters, context-related information, uncertainty quantification, reliability, evaluation of negative outcomes, etc., listed in previous sections also pose limitations on the modeling capabilities of AI in the field of synthetic biology. The limitations of current algorithms in the interpretability of genetic data (omics) were briefly addressed in Section 3. Also, the difficulty of codifying expert/tacit knowledge paramount for a successful experiment in life sciences widens the gap between the computational prediction and the experimental result.

Another technological challenge with a high impact on computational results for both peaceful science and biosecurity-violating projects are the AI evaluation metrics. Standard AI evaluation metrics are inadequate for applications in synthetic biology, due to their incapability to capture the complexity and stochasticity of biological systems.[26] Nevertheless, these obstacles may be overcome in the near future through the development of more sophisticated algorithms and evaluation metrics, driven by huge investments in the synthetic biology sector. These considerations urge the need for a biosecurity framework for AI and robotics and an open dialogue and awareness raising among academia and industry stakeholders.

6. AI and disinformation

So far, we have focused on a few selected AI applications in life sciences relevant to the BWC and CWC without covering the full spectrum of this broad field. However, AI is rapidly gaining access in different civilian and military sectors. Other research areas can also have implications for these arms controls and beyond. Developments in AI technology relevant to the context of cybersecurity, autonomous weapons, and drones will not be further explored within the scope of this paper. Nevertheless, it should be noted that they significantly expand the threat landscape.

A less-noticed realm where AI can be misused for hazardous purposes is disinformation spreading. Intentional disinformation campaigns can have deteriorating effects on the norms against chemical and biological weapons. As reflected in one of the previous project publications by Jakob et al., "false allegations of development, possession, and use of weapons of mass destruction can create perceptions that the taboos against biological, chemical and nuclear weapons no longer hold"[79]. Disinformation campaigns can be launched to trigger propaganda, false flag operations, or tarnish/damage the reputation of institutions such as the OPCW.

[74] Martin H.G., Radivojevic T., Zucker J., et al. Perspectives for self-driving labs in synthetic biology, *Curr. Opin. Biotechnol.*, 79, 1-15, (2023).

[75] King, R., Whelan, K.E., Jones, F.M., et al. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* 427, 247–252 (2004).

[76] Abolhasani M., Kumacheva E. The rise of self-driving labs in chemical and materials sciences. *Nat. Synth.,* 2, 483-492, (2023).

[77] Lentzos F., Invernizzi C. Laboratories in the cloud. *Bull. At. Sci*. 2019 (accessed 2023).

[78] Arnold C. Cloud labs: where robots do the research. *Nature* 606, 612-613 (2022).

[79] Jakob U., Jeremias G., Kelle A., et al. PRIF BLOG: Russian allegations of biological weapons activities in Ukraine. Mai 2022, https://blog.prif.org/2022/03/22/russian-allegations-of-biological-weapons-activities-in-ukraine/(accessed 2023).

AI-based algorithms have been exploited to create or manipulate various forms of media types and text passages known as “deep fakes” or “synthetic media” that can be circulated for disinformation purposes. Counteracting computational methods to detect “deep fakes” are under development,[80] heralding the race between risk-posing and mitigating AI-powered strategies. However, there is no denying that currently, AI advances are radically transforming the ways information and disinformation are disseminated.

In a recently published illustrative example, AI chatbots were successfully challenged to produce a fabricated article about the chemical attacks in Douma in 2018 not attributed to the Syrian government, but staged or orchestrated by the U.S. and other actors, uncovering a considerable breach in the ethical and security restriction mechanisms of these commercial platforms.[81] Interestingly, in the EU law on artificial intelligence, just passed by the European Parliament (AI Act), “only minimum transparency obligations are proposed, in particular when chatbots or ‘deep fakes’ are used”[82]. Artificially generated or manipulated content should be merely labeled as such.[83] This last requirement is waved where the use is “[…]necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties”[83]. More general issues and risks underlying deep fakes are not covered in the current version of the document.

7. Strengthening biosecurity with AI

> *The adoption of AI applications in the life sciences is of course not only associated with biosecurity threats. On the other side of the scale is the notion that AI technology can strengthen biosecurity by facilitating the development of vaccines and antidotes, introducing and improving detection methods, etc.*

The beneficial potential of AI has been for instance illustrated in the recent successful efforts to contain the Covid-19 pandemic. Various AI-driven algorithms have been applied for disease surveillance, patent screening and diagnostics, viral genome sequencing, development of drugs and vaccines, and predicting possible viral mutation landscapes.[84-85] Several machine learning tools have been proposed to support and expand the existing biosecurity measures. A few examples include:

- early warning system for biothreats such as anthrax[86] and potential high-risk Sars-CoV-2 variants;[87]
- model for forensic attribution of biological attacks with the ability to predict both the nation-state-of origin and the ancestor lab;[88]
- a conceptual framework for accurate screening of commercial nucleic acid/peptide synthesis orders.[89]

[80] Malolan B., Parekh A. Kazi F. Explainable Deep-Fake Detection Using Visual Interpretability Methods, *3rd International Conference on Information and Computer Technologies (ICICT)*, San Jose, CA, USA, 289-293, (2020).

[81] Arvanitis L., Sadeghi M. ,Brewster J. Despite OpenAI’s Promises, the Company’s New AI Tool Produces Misinformation More Frequently, and More Persuasively, than its Predecessor. NewsGuard, March 2023 https://www.newsguardtech.com/misinformation-monitor/march-2023/ (accessed 2023).

[82] Proposal for a Regulation laying down harmonised rules on artificial intelligence. COM(2021) 206 final , Brussels, April 2021 https://artificialintelligenceact.eu/the-act/ (accessed 2023). p.3

[83] Ibid. p.70

[84] Bagabir S.A., Ibrahim N.K., Abubaker Bagabir H.A., et al. Covid-19 and Artificial Intelligence: Genome sequencing, drug development and vaccine discovery. *J Infect Public Health*, 15(8), 289-296, (2022).

[85] Arora G., Joshi J., Mandal R.S., et al. Artificial Intelligence in Surveillance, Diagnosis, Drug Discovery and Vaccine Development against COVID-19. *Pathogens*., 10(8), 1-21 (2021).

[86] Jo Y., Park S., Jung J., et al. Holographic deep learning for rapid optical screening of anthrax spores. *Sci Adv.*, 3(8):e1700606, 1-9, (2017).

[87] Beguir K, Skwark M.J., Fu Y., et al. Early computational detection of potential high-risk SARS-CoV-2 variants. *Comput Biol Med.*, 155(106618), 1-9, (2023).

[88] Alley E.C., Turpin M., Liu A.B., et al. A machine learning toolkit for genetic engineering attribution to facilitate biosecurity. *Nat Commun.* 11(6293), 1-12, (2020).

[89] Lee Y-C.J., Cowan A., Tankard A. Peptide Toxins as Biothreats and the Potential for AI Systems to Enhance Biosecurity. *Front. Bioeng. Biotechnol*. 10, 1-6, (2022).

These approaches still have technical limitations and require solid proof of principle, but they represent a promising development in the technological field to mitigate future biothreats.

AI technology also offers some applications to strengthen the framework and implementation of BWC and CWC regulations. For instance, the winners of the 2022 Next Generation for Biosecurity Competition[90] propose in their modular-incremental approach for the potential establishment of a BWC verification regime an AI-based method to support the submission of the so-called Confidence Building Measures (CBM) reports by the States parties. CBMs are a pivotal instrument of the BWC to ensure transparency and to improve international cooperation in the field of peaceful biological activities, which should contain information on research centers and laboratories, vaccine production facilities, national biodefence programs, infectious disease outbreaks, occurrences caused by toxins, etc.[91]

However, the cumbersome collection of relevant information impeded by the lack of consistent reporting standards represents a substantial burden in the CBM submission process.[92] An AI-based approach can be a mainstay in data harmonization and pave the way toward a universal CBM submission.[92] Additionally AI can be used to analyze the CBMs to gain insights and uncover any suspicious inconsistencies or activities that might remain unnoticed by manual screening.[92]

The benefits of AI technologies for the purposes relevant to the CWC were also recognized by the SAB in its fourth report on the developments in science and technology prior to the Review Conference 2018.

As stated in the document: "Advances in fields such as remote sensing, data mining and the analysis of very large amounts of data, artificial intelligence, forensic science, and automated and autonomous systems can be utilized to increase the OPCW's capability to verify compliance"[93]. One of the applications in line with the contemplations of SAB is the project under development conducted at the Tallinn University of Technology. The project is part of the European Defence Fund and aims at leveraging drones, AI, and deep learning to identify chemical and biological weapons in real-time through on-site inspections.[94]

In its report, SAB notices further that the integration of information and communication technologies with other data streams "has potential application for chemical security including recognising unexpected or unusual (bio)chemical change in the environment"[95]. Particular section in the report is devoted to AI. CWC-relevant applications of this technology could be the identification of munitions or the detection of laboratory equipment from on-site photographs.[96] Moreover, AI can facilitate the analysis of big data in the process of data mining, and recognition of "unusual features in information […], especially for threat assessment for counter terrorism"[96]. Based on these recommendations, further AI-based applications were proposed for routine and non-routine verification in accordance with the CWC provisions.[97]

[90] NTI: Winners of 2022 Next Generation for Biosecurity Competition Announced. https://www.nti.org/news/winners-of-2022-next-generation-for-biosecurity-competition-announced/ (accessed 2023).

[91] United Nations. Office of Disarmament Affairs: Confidence Building Measures. https://www.un.org/disarmament/biological-weapons/confidence-building-measures/ (accessed 2023).

[92] Cropper N., Rath S., Teo R. Creating a Verification Protocol for the Biological Weapons Convention: a modular-incremental approach. June 2022 https://www.nti.org/wp-content/uploads/2022/06/Creating-a-Verification-Protocol_FINAL_June2022.pdf (accessed 2023).

[93] OPCW: Report of the Scientific Advisory Board on developments in science and technology for the fourth special session of the Conference of the States parties to review the operation of the Chemical Weapons Convention. April 2018 p. 4-5 (accessed 2023).

[94] Oidermaa J.-J. Scientists seek quicker ways to identify chemical and biological weapons, March 2023 https://news.err.ee/1608928193/scientists-seek-quicker-ways-to-identify-chemical-and-biological-weapons (accessed 2023).

[95] Ibid. OPCW: Report of the Scientific Advisory Board, April 2018, p.27

[96] Ibid. p.28

[97] Kelle, A., Forman, J.E. Verifying the Prohibition of Chemical Weapons in a Digitalized World. In: Reinhold, T., Schörnig, N. (eds) Armament, Arms Control and Artificial Intelligence. Studies in Peace and Security. Springer, Cham., 73-89, (2022).

Possible applications encompass e.g. an AI-based analysis for the verification of declarations of the States parties submitted to the OPCW; an automated tool for effective screening of chemical compounds against the scheduled chemicals and their vast number of derivatives, isotopic species, isomers; recognition of early symptoms of exposure to toxic compounds, etc.[97] Notably, a cheminformatics prototype tool called Nonproliferation Cheminformatics Compliance Tool (NCCT) that automates the task of assessing whether a chemical is part of a CW-control list, has been presented by Costanzi et al.[98] The underlying algorithm is not based on machine learning but represents a database-management system with an embedded database containing generic structures that describe the entries relative to families of chemicals.[98]

This non-exhaustive overview of possible computational applications to strengthen biosecurity at different levels (e.g., medical mitigation measures, forensics, legislative implementations) underscores the "double-sided" nature of technology, possessing both threats and virtues, depending on the purpose for which it is used.

8. AI Governance

Nowadays, the subject of AI receives a lot of attention. In the two polarizing views, this technology is visualized as "revolutionizing the world" or as a "doomsday machine". The reality lies somewhere in the grey area in between. AI applications in drug design, retrosynthesis planning, and synthetic biology are very promising but also harbor a threat of misuse. Publicly available large amounts of chemical data together with the open-source tools to design the respective software with machine learning architecture shape the threat landscape. However, current limitations in this research field, partly related to the quality of available datasets but also to the well-known discrepancies between theory/experiment, *in vitro*/*in vivo* situations, etc., do not only affect the efficiency of AI in peaceful science but also its misuse potential. Broadly speaking, AI is not a panacea. It cannot replace the scientist "in the loop", who still has to evaluate the computational results and validate them experimentally. Albeit AI can facilitate and speed up the process of e.g. screening for potential toxic components (Section 3), it does not eliminate the necessity of further theoretical and experimental work, regarding for instance toxin delivery *in vivo*. This leads to the conclusion that AI alone might currently not be the game-changer in the process of biochemical weapon development by a nefarious actor.

The situation can change drastically given the rapid developments in computer science in general and in AI and robotics technology in particular, which may obtain an extra spin from large investments in Industry 4.0. Once AI-powered automation in life sciences passes infancy and reaches the advanced stage, the current threat landscape will undergo a significant change.

Moreover, the sharing of knowledge and data is subject to a constant drift toward openness and accessibility. The number of publications in open-source literature increases rapidly, as is the proportion of open-access journals worldwide.[99] This trend laudably serves the promotion of transparency in research. Nevertheless, some recorded data can be vulnerable to misuse. This argument applies to all areas of life sciences. The general recommendation for raising awareness among stakeholders contained in the vast majority of biosecurity guidelines is also relevant in the present context. The scientific community should critically review all collected data regarding its possible implications for biosecurity, before disclosing the results in publicly available resources. The principle of awareness-raising also includes the requirement for extensive training in biosecurity and ethics for employees in academia and the private sector working in computational and life sciences. As stated by Urbina et al.: "We are not trained to consider [technology misuse potential], and it is not even required for machine learning research"[23].

[98] Costanzi S., Slavick C., Abides J., et al. Supporting the fight against the proliferation of chemical weapons through cheminformatics. *Pure Appl. Chem*., 94(7), 783-798, (2022).

[99] Adoption of open access is rising – but so too are its costs, LSE, January 2018 https://blogs.lse.ac.uk/impactofsocialsciences/2018/01/22/adoption-of-open-access-is-rising-but-so-too-are-its-costs/, (accessed 2023).

The commercial and non-profit developers of AI-based platforms for biomedical applications and research should also contemplate ethical and biosecurity aspects and e.g. implement biosecurity-related restrictions as a part of their licensing policy. Some companies have committed to "responsible AI", the AI technology "that is fair and non-biased, transparent and explainable, secure and safe, privacy-proof, accountable, and to the benefit of mankind"[100]. Additionally, stakeholders from academia, civil society, industry, and media established a non-profit "Partnership to AI" with currently 103 members.[101] One of the goals of this community is the development of best practices to avert possible misuse of AI and to mitigate the arising risks in both the near and long term. Despite that, the implemented measures are very heterogeneous. Half of the companies committed to the responsible AI principle have not yet undertaken any concrete steps toward achieving this goal.[100] Moreover, open-source algorithms and tools make it possible to circumvent restrictions implemented in commercial products and to develop the required software "from scratch".

Also, commercial providers make some of their software openly available. Thus, the source code of AlphaFold was disclosed[102] due to the consideration that "the entities which could be risky are likely to be a very small handful"[73]. According to the results of consulting with "more than 30 experts in bioethics and security"[73], the benefits of making such software available under open source license "far outweigh any risks"[73]. Although this tool does have significant advantages for science, its dual-use character should not be neglected. Given that the number of unknowns in biochemistry remains large (e.g. unknown genome-phenotype relationships, proteins with unknown functions, unknown regulators of biochemical pathways, etc.), such steps require careful consideration and a broad open discussion in the community.

> *When it comes to the governance of risk-associated technologies, first thoughts might address possible legislative regulations. The regulation of technology risks is widespread and often comes in the form of safety regulations, e.g. for the operation of dangerous facilities, or the handling of harmful substances. For the much more democratized AI systems which also lack the spatial dimension that other risk-associated technologies have, preconditions differ very much.*

The challenges for AI governance lie in the multifaceted character of AI technology that cannot be reduced to a few applications and subsequently regulated. We can therefore not present an elaborate catalog of recommendations for regulation, but rather give an introductory overview of problems and pitfalls. In this case, complexity is multidimensional: it is already difficult to define the subject matter of a possible regulation. Should the code be regulated or rather the way it is used? And how could risk or non-compliance be defined? Complexity arises, among other things, from the need for legislators to deal with a profound problem in mitigating risks from new and emerging technologies, namely that both legislators and users of a potential legal standard have to deal with ignorance. Without the empirical experience of materialized risks from the application of technology, it is only possible to make more or less well-founded assumptions about unknown or at least unclear consequences. A common way of dealing with such uncertain risks is to develop standards based on the precautionary principle. According to the European Commission, "the precautionary principle may be invoked when a phenomenon, product or process may have a dangerous effect" that has been established by a scientific and objective assessment when this assessment does not allow the risk to be assessed with sufficient certainty to determine safety.[103] However, an effective regime for applying the precautionary principle requires a functioning risk assessment mechanism. In other cases, such as in the "EU Directive on the deliberate release into the environment of genetically modified organisms"[104] detailed provisions for an environmental risk assessment were stipulated which must be followed before a release is licensed.

[100] de Laat P.B. Companies Committed to Responsible AI: From Principles towards Implementation and Regulation?. *Philos. Technol.* 34, 1135–1193 (2021).

[101] About Us https://partnershiponai.org/about/ (accessed 2023).

[102] AlphaFold v2.3.1 https://github.com/deepmind/alphafold (accessed 2023).

[103] EU Commission's Communication on the precautionary principle, EU COM/2000/0001, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52000DC0001 (accessed 2023).

[104] EU Directive on the deliberate release into the environment of genetically modified organisms,
OJ L 106, March, 2001 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32001L0018 (accessed 2023).

In the field of AI, we see yet worldwide only extremely sporadic and fragmented legislation. According to the analysis conducted by the AI Index Report 2023, of the 127 countries monitored 31 have passed at least one AI-related bill from 2016 to 2022.[105] The EU law on artificial intelligence, the AI Act, is the first piece of legislation to regulate AI systems by establishing "harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence ("AI systems") in the Union" (Art. 1(a)).[106] At first sight, the AI Act appears to encompass the regulation of AI systems of all kinds of applications in all sectors (Art. 2)[106] except the military and scientific-technical activities before entry into the European market. In this capacity, it would either prohibit or restrict the use of the technologies in question under certain criteria to reduce the risk of use to an acceptable level.

On closer inspection, the legal norm relates exclusively to AI-related risks in the area of data privacy.

The positive lists in Annex III do not refer to other AI-associated risks and conversely exclude them from the scope of the Act. In its systematics, the law divides risks from the use of AI systems into four different classes: low or minimal risk, high risk, and unacceptable risk. While AI systems that are considered to produce unacceptable risks are being excluded from the market (Art. 5)[107], high-risk systems shall become subject to a risk management system (Art. 9)[108] In contrast to the above-mentioned directive on the release of GMOs there are, however, no provisions specifying methodology and criteria for risk assessment and risk management defined in the document. We doubt that meaningful risk classes for AI applications in life sciences can be defined clearly enough. Boundaries between the classes are likely to stay arbitrary instead. For this reason, the question of whether the AI Act can serve as a blueprint for a law that specifically regulates the risks to chemical and biological safety associated with AI must be critically posed.

The German Federal government's AI strategy[109] supports some beneficial developments in AI technology to strengthen biosecurity. Thus, it advocates the necessity to expand the "computational life sciences funding measure focusing on AI for digital infection epidemiology"[110]. The document also requests a rigorous examination of "whether existing legislation adequately addresses the risks and requirements of AI applications and enables effective enforcement"[111] in line with the proposal of the EU AI Act.

It is not clear, how the needed single-case risk assessments could be structured and organized and who would be the involved stakeholders. Especially when AI systems are based upon open source software or applied under unclear spatial dimensions, ownership and accountability of AI systems might become ambiguous and hence be an obstacle for legal regulation. In fact, the unclear spatial dimensions of certain AI systems are another challenge, as AI can hardly be reduced to the territory of one state. Hence, global mandatory rules would be needed to make up a substantial legal system against AI risks. It is not necessary to describe in detail why it is unlikely to see the successful making of such norms. Taking a look into the attempts for DURC regulation might give a picture of the low chances for successful norm development. This is not to say that we do not believe in the need to effectively reduce risks, but instead of publishing an inadequate plan for regulation, we might better induce a debate about possible legal tools for risk mitigation. Although voluntary commitments and other forms of soft law are not the strongest shields against the misuse of technologies, these might hence be the most promising avenue for risk reduction in the field for the time being.

[105] Maslej N., Fattorini L., Brynjolfsson E., et al. The AI Index 2023 Annual Report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, (2023).
[106] Ibid. Proposal for a Regulation laying down harmonised rules on artificial intelligence p.38
[107] Ibid. p.43
[108] Ibid. p.46
[109]Artificial Intelligence Strategy of the German Federal Government, December 2020 https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf (accessed 2023).
[110] Ibid p.27
[111] Ibid p.30

The CBW network for the comprehensive strengthening of norms against chemical and biological weapons (CBWNet)

The research project CBWNet is carried out jointly by the Berlin office of the Institute for Peace Research and Security Policy at the University of Hamburg (IFSH), the Chair for Public Law and International Law at the University of Gießen, the Peace Research Institute Frankfurt (PRIF) and the Carl Friedrich Weizsäcker-Centre for Science and Peace Research (ZNF) at the University of Hamburg. The joint project aims to identify options to comprehensively strengthen the norms against chemical and biological weapons (CBW).

These norms have increasingly been challenged in recent years, *inter alia* by the repeated use of chemical weapons in Syria. The project scrutinizes the forms and consequences of norm contestations within the CBW prohibition regimes from an interdisciplinary perspective. This includes a comprehensive analysis of the normative order of the regimes as well as an investigation of the possible consequences which technological developments, international security dynamics or terrorist threats might yield for the CBW prohibition regimes. Wherever research results point to challenges for or a weakening of CBW norms, the project partners will develop options and proposals to uphold or strengthen these norms and to enhance their resilience.

The joint research project is being funded by the Federal Ministry of Education and Research for four years (April 2022 until March 2026).

Authors information

Dr. **Anna Krin** is Research Associate at the Carl Friedrich von Weizsäcker Centre for Science and Peace Research at Hamburg University (ZNF). She analyses recent developments and convergences in the field of life sciences and technology, which are of relevance in the context of the Chemical and Biological Weapons Conventions.

Dr. **Gunnar Jeremias** heads the Interdisciplinary Research Group for the Analysis of Biological Risks (INFABRI) at the Carl Friedrich von Weizsäcker Centre for Science and Peace Research at Hamburg University (ZNF).

Contact
Responsible for Content §55, Abs.2 RStV:
Dr. Alexander Kelle, IFSH
Reinhardtstraße 7, 10117 Berlin
E-Mail: CBWNet@ifsh.de

www.cbwnet.org
twitter.com/CBWNet

ISSN (Online): 2751-4501

SPONSORED BY THE

Federal Ministry
of Education
and Research