ASSESSING THE RISKS POSED BY THE CONVERGENCE OF ARTIFICIAL INTELLIGENCE AND BIOTECHNOLOGY

John T. O'Brien and Cassidy Nelson

Rapid developments are currently taking place in the fields of artificial intelligence (AI) and biotechnology, and applications arising from the convergence of these 2 fields are likely to offer immense opportunities that could greatly benefit human health and biosecurity. The combination of AI and biotechnology could potentially lead to breakthroughs in precision medicine, improved biosurveillance, and discovery of novel medical countermeasures as well as facilitate a more effective public health emergency response. However, as is the case with many preceding transformative technologies, new opportunities often present new risks in parallel. Understanding the current and emerging risks at the intersection of AI and biotechnology is crucial for health security specialists and unlikely to be achieved by examining either field in isolation. Uncertainties multiply as technologies merge, showcasing the need to identify robust assessment frameworks that could adequately analyze the risk landscape emerging at the convergence of these 2 domains. This paper explores the criteria needed to assess risks associated with AI and biotechnology and evaluates 3 previously published risk assessment frameworks. After highlighting their strengths and limitations and applying to relevant AI and biotechnology examples, the authors suggest a hybrid framework with recommendations for future approaches to risk assessment for convergent technologies.

Keywords: Risk assessment, Artificial intelligence, Biotechnology, Biosecurity

INTRODUCTION

Technological Progress

SINCE THE EXPLOSION OF THE INFORMATION AGE, the collection of data has been increasing. About 2.5 quintillion (10^{18}) bytes of data are generated each day, and 90% of all data have been generated in the past few years alone.¹ Although valuable information is often hidden in these vast datasets, due to their size and complexity, a human analyst could take weeks, months, or even years to discover them, if they are able to at all. If one can analyze a

large enough set of data, it is possible to infer corollary relationships or patterns that may be of salience. In the past, it was not conceivable for extremely large datasets, or big data, to have been meaningfully analyzed due to the daunting task of doing so manually. Partially as a result of the increasing prevalence of big data requiring new tools for analysis and increased computing power enabling their development, the use of artificial intelligence (AI) in a variety of fields, including the biological sciences, has gained popularity in the past few years.

AI can be defined as a system's ability to correctly interpret external data, to learn from such data, and to use

John T. O'Brien, MS, is a Research Associate, Bipartisan Commission on Biodefense, Washington, DC. Cassidy Nelson, MBBS, MPH, is a Research Scholar, Future of Humanity Institute, University of Oxford, Oxford, UK.

those insights to achieve specific goals and tasks through flexible adaptation.² Machine learning, a technique that has enabled rapid advancements in the field of AI, is a subset of AI that is heavily based on pattern recognition and describes methods that help computers learn without explicitly being programmed to do so.² Deep learning, in turn, is a machine learning technique that arose in the early 2010s and uses computational models that exhibit similar characteristics to the hierarchical information processing in the human brain.³ This technique is partially responsible for the increased power potential of machine learning algorithms and has been a large contributor to the recent surge of AI research.⁴

Alongside AI, developments in biotechnology and the life sciences have progressed rapidly in the past several years. In fact, some biotechnological advancements may even be outpacing Moore's Law for computers.⁵ Critical advances in biology such as next-generation DNA sequencing, *de novo* DNA synthesis, gene editing, genomics, and bioinformatics may result in the life sciences becoming a principal component of the next technological revolution. Given recent progress and projected trajectories, AI and biotechnology have both been described as key pillars of the fourth industrial revolution, with a growing number of technologies and applications arising at their point of convergence every year.⁶⁻⁹

Technological Convergence

In tandem, AI and biotechnology increase the potential capability of the life sciences and lower the tacit knowledge required to perform previously tedious laboratory tasks. Widening the pool of actors will almost certainly accelerate technological developments, but this could also pose security risks. With increased access to these technologies, so comes the potential for a resulting high-consequence biological event, some of which may have such large consequences that they constitute a global catastrophic biological risk.¹⁰ The risks posed by the convergence of AI with the biological sciences are plentiful and varied, ranging from AI-assisted identification of virulence factors to *in silico* design of novel pathogens, which could increase the probability of a global catastrophic biological risk event if accidentally, carelessly, or deliberately released.

For the purposes of this paper, convergence is defined as the technological commingling between the life sciences and AI such that the power of their interaction is greater than the sum of their individual disciplines. That is, the convergence of these 2 fields has a multiplicative effect. This interaction mostly manifests as AI-developments influencing advances, as well as new risks, in the life sciences and biotechnology rather than the other way around. A more comprehensive understanding of such emerging threats is highly important for the field of health security but demands rigorous examination and a systematic approach.

In this paper, risk assessment for converging technologies will be discussed and evaluation of 3 existing biosecurity

risk assessment frameworks undertaken. After analyzing their strengths and limitations, suggestions will be made regarding how such frameworks could be adjusted to be applicable to the risks arising at the convergence of AI and biotechnology. Some illustrative examples from the converging risk landscape will be discussed, followed by a proposed hybridized framework and key recommendations to begin addressing gaps in current health security risk assessment for converging technologies.

RISK ASSESSMENTS FOR CONVERGING TECHNOLOGIES

Challenges Assessing Emerging and Converging Risks

Risk assessment for emerging technologies is an inherently difficult task and general frameworks are needed for evaluating the risks associated with technological convergence, including AI and biotechnology. Many frameworks are overly specific to a particular scenario or technology (such as narrowly focusing on synthetic biology), contain many assumptions and guesswork, and are not easily applied to other converging or emerging technologies. This is largely because different frameworks serve different purposes, so expectations that they can be implemented more generally are limited. However, the primary strength of a more generic framework for biosecurity risk assessments would be the ability to compare across different technologies or scenarios. Decision makers could then use such frameworks to be more informed as to which risks may be of higher priority.

Compared to assessments of singular domains, analyzing risks of converging technologies presents an even greater challenge due to the added layers of complexity and the presence of multidisciplinary components. Converging risks such as those described herein are uniquely challenging to assess because their uncertainties multiply such that one must consider all possible outcomes of AI development in tandem with all possible outcomes of developments in the life sciences. For example, the evaluation criteria in a scenario-based risk assessment for virtual laboratories or 'cloud labs' that are too focused on the biological components may miss important bottlenecks, risks, or gaps that exist within the underlying AI components of the technology. Full assessment may not always be possible for some unpredictable technological advancements. Challenges also arise due to limited access to data and the inability to quantitatively assess emerging risks.

Many frameworks include the potential adversary as a component to calculate risk.^{11,12} However, the adversary has a large range of potential attributes that are difficult to capture in their entirety. The adversary could be a nation, a nonstate actor, or an individual. When evaluating the likelihood that one of these adversaries could carry out

an attack, assumptions must be made and variables considered on a spectrum. For example, a nation-state could be a low-resourced or well-resourced country. An individual could be anyone from an undergraduate biology student to a senior research professor. Such approaches fail to capture the range of possibilities and thus suffer from decreased granularity. Along these lines, without access to classified information, it is impossible to appreciate a full picture of adversary capability and intent.

Furthermore, frameworks are sometimes focused on a technology's ability to aid in a specific element in the development process of a biological agent, which then makes it difficult to assess the risks of a full scenario in which it was implemented successfully. Conversely, assessments that solely focus on a full scenario fail to accurately capture the broad range of potential misapplications of a technology and consequently suffer from excessive specificity.

After searching published literature, we identified 3 previously employed frameworks that inform the development of an assessment for converging technologies. These frameworks are examined in this paper in detail, with their strengths and limitations explored.

Overview of Existing Frameworks

The AAAS-FBI-UNICRI Framework

The first framework was published in the 2014 report, *National and Transnational Security Implications of Big Data in the Life Sciences*, and was developed by a joint team consisting of members from the American Association for the Advancement of Science (AAAS) Center for Science, Technology, and Security Policy, the Biological Countermeasures Unit of the Federal Bureau of Investigation (FBI) Weapons of Mass Destruction Directorate, and the United Nations Interregional Crime and Justice Research Institute (UNICRI).¹¹ The main elements of the AAAS-FBI-UNICRI framework are highlighted in Table 1. A strong component of this framework is the inclusion of security specialists, in addition to experts from multiple disciplines to evaluate the risks.

While being one of the only attempts at risk assessment of a converging technology, the framework is limited in generality because it uses a scenario process to assess risks

and its evaluation criteria are heavily skewed to big data and do not neatly map on to other converging technologies such as AI, despite many commonalities between them. The lack of generality is best illustrated by the framework table itself in which the "Needed scientific expertise and skills" block is divided into 2 sections to cover all 3 scenarios evaluated.¹¹ This may be due to the fact that the working group developed the risk scenarios prior to developing the framework itself, resulting in a framework with a narrow scope.

When attempting to implement a scenario where a malicious actor uses AI-enabled cloud labs to synthesize a pathogen *in silico*, the narrowness of this framework becomes more evident. The parameter, "vulnerabilities in data repositories, software, and/or underlying cyber infrastructure"¹¹ does not map onto this scenario since none of those 3 vulnerabilities are necessarily present in the case of cloud labs accessed through legitimate channels, albeit for nefarious purposes.

The NASEM Framework

The second framework emerged from the National Academies of Science, Engineering, and Medicine (NASEM) 2018 report, *Biodefense in the Age of Synthetic Biology*.¹³ The NASEM framework (Table 2) is strong in that it actively considers the synergy with other technologies as a factor to calculate level of concern. The report also considers machine learning and automation as potential technologies that could facilitate the development of a biological agent. This framework, however, excludes actors who may be beneficial, considering the broad range of potential adversaries.

The limitation in using this framework for converging AI and biotechnology risks is that the scope is too specific and restricted to risks from synthetic biology and technologies relevant to the design-build-test cycle. The authors of the report are transparent about these limitations and recognize that every factor may not map on to a specific technology. If one wanted to apply this framework to evaluate the level of concern posed by risks from the convergence of AI and biotechnology, they could face difficulties since the questions used to evaluate the factors are heavily skewed toward synthetic biology. However, since the framework was developed to assess capabilities of a technology to enable harm more

Table 1. Main AAAS-FBI-UNICRI Framework Elements to Evaluate the Risks of Big Data¹¹

<i>Risk Assessment</i>					
<i>Probability</i>		<i>Consequences</i>			
Adversary	Vulnerabilities in data repositories, software, and/or underlying cyber infrastructure	Needed scientific expertise and skills		Severe consequences to economics, political system, society, health, environment, and/or agriculture	Sufficient existing countermeasures
		To exploit vulnerabilities in the system	To use big data analytics to design harmful biological agents		

Table 2. Main NASEM Framework Elements to Evaluate Level of Concern of Technologies in Synthetic Biology¹⁴

<i>Level of Concern About the Capability</i>			
<i>Usability of the Technology</i>	<i>Usability as a Weapon</i>	<i>Requirements of Actors</i>	<i>Potential for Mitigation</i>
<ul style="list-style-type: none"> • Ease of use • Rate of development • Barriers to use • Synergy with other technologies 	<ul style="list-style-type: none"> • Production and delivery • Scope of casualty • Predictability of results 	<ul style="list-style-type: none"> • Access to expertise • Access to resources • Organizational footprint requirements 	<ul style="list-style-type: none"> • Deterrence and prevention capabilities • Capability to recognize an attack • Attribution capabilities • Consequence management capabilities

generally, it is likely that the base framework could easily be adapted to include questions that address concerns such as those resulting from convergence of AI and biotechnology.

The Tucker Framework

The third framework was developed by Jonathan B. Tucker in his 2012 book, *Innovation, Dual Use, and Security: Managing the Risks of Emerging Biological and Chemical Technologies*.¹² This framework (Table 3) benefits from including a governability assessment that is used if the average value of the factors from the risk assessment portion are “medium” or “high.” While assessing governability for approaches to governance has unquestionable value, it can also be an indicator for probability. If a technology is less governable, it is likely to have a higher chance for misuse and resulting risk. A potential weakness in Tucker’s framework is the use of an ordinal scale with just 3 values (high, medium, or low). A spectrum spanning a range of only 3 could result in lost information and decreased specificity, which may have negative downstream effects in determining which technologies pose the greatest or lowest risk. Another potential shortcoming is that the framework considers the imminence of the technology to be of equal weight to the other factors in calculating risk. As a result, it could potentially fail to accurately categorize a long-term, high-consequence risk as one worth governing.

In a 2017 working paper, Koblenz et al expanded on Tucker’s framework, proposing the Tucker-Koblenz framework, which includes the range of applications for misuse as an indicator of risk.¹⁴ This is particularly useful in the context of the converging technologies since they tend to have synergistic effects that increase the range of appli-

cations. The authors also noted how risk assessments are implicitly impacted by the anticipated benefits of a technology since such technologies are more likely to be intensely pursued by researchers and private companies, which results in accelerated innovation and diffusion of knowledge. This perspective is useful in the context of the convergence of AI and biotechnology since the range of beneficial applications is rather large.

THE CONVERGING RISK LANDSCAPE

The overlap between AI and biotechnology is already considerable, with increasing integration of AI into various facets of the biological sciences likely to continue for the foreseeable future.⁹ This poses a challenge for the health security risk assessment frameworks previously described and requires further understanding of the risks that need to be prioritized for examination. This section provides an overview of some of the intersections of AI and biotechnology that could be considered for future risk assessment exercises. These sections are presented as illustrative examples and are not meant to propose a taxonomy of risks.

Synthetic Biology

Increasingly, engineering principles are being applied to biology for research and manufacturing purposes. Synthesis and design of existing and novel microorganisms have many legitimate and beneficial applications; however, these same technologies could be used to engineer pathogens for malicious use. AI and machine learning have the potential

Table 3. Main Tucker Decision Framework Elements to Assess Misuse and Governability¹³

<i>Risk of Misuse</i>	<i>Governability</i>
<ul style="list-style-type: none"> • Accessibility • Ease of misuse • Magnitude of potential harm • Imminence of potential misuse 	<ul style="list-style-type: none"> • Embodiment • Maturity • Convergence • Rate of advance • International diffusion

Note: The elements were determined by averaging the values (high, medium, or low) yielded by the parameters. Governability was determined if risk of misuse is medium or high.

to greatly aid in the design-build-test stages of the synthetic biology process.¹⁵ Current DNA synthesis methods and gene editing technologies, such as the CRISPR-Cas9 system, still pose a high technical barrier for viable pathogen creation. Nonetheless, several recent controversial experiments have showcased the ability to *de novo* synthesize live viruses, including horsepox and the 1918 flu.¹⁶ Research on emerging techniques, such as enzymatic synthesis of DNA, aims to revolutionize the field of synthetic biology in the next decade, with the end goal being benchtop printers able to synthesize gene-length sequences of DNA to completion.¹⁷ Advances in this space are not only increasing the potential power and capability of synthetic biology but are also lowering the technical barriers and tacit knowledge required for its widespread use.¹⁸

Within the realm of synthetic biology, AI could potentially lower some of the barriers for a malicious actor to design dangerous pathogens with custom features. It is worth noting, however, that AI will have a differential effect on technological hurdles, with some bottlenecks remaining difficult to traverse. For a more detailed examination of how the difficulty of design steps can affect the probability of a successful biological attack, we recommend the paper by Sandberg and Nelson also published in this issue.¹⁹ AI has enabled a substantial reduction in the time required to design and test functionality in fields such as protein design.²⁰ It may also lower barriers in the design stages of the design-build-test cycle and could be used to identify optimal mutations of a pathogen that enhance virulence factors, expand host range, alter routes of transmission, encode resistance to medical countermeasures, and intensify mortality.²¹ Genetic changes could be identified and made to a virus that is traditionally benign in nature, allowing it to bypass screening procedures currently used by mail-order DNA synthesis services. Similarly, AI could be leveraged to circumvent field detection technologies by identifying code sequences that could be altered without changing function.¹¹

Deep learning could potentially identify genetic functions that code for vulnerabilities and interconnections between the immune system and microbiome. It may be possible that malicious actors could leverage this information to engineer "precision maladies," or pathogens that target important mechanisms the immune system or the microbiome of specific subpopulations.²² It is important to note, however, that simply identifying certain vulnerabilities does not mean actors can then freely make changes, significant bottlenecks would still have to be traversed in the overall development process and the feasibility of targeting specific subpopulations would remain an unknown.

Moreover, in the past few years, several studies have been published on predicting the pathogenic potential of natural and synthetic DNA using AI. Although improving biosecurity is one of the motives for this research, the results of this work may have the reverse effect and pose information

hazards if the methods and algorithms are published. PaPrBaG and DeePaC are examples of machine learning algorithms predicting pathogenic potential that have made all source code available online.^{23,24} These algorithms could potentially aid a bad actor with the intent of designing a pathogen that can cause maximal harm.

AI can not only assist in the design step of the process, but it can also, to a lesser degree, support the building and testing phases. The rise of AI-enabled cloud labs to automate procedures that used to require hands-on, tacit knowledge to perform tasks could further widen the pool of actors who could perform certain biology experiments. Cloud labs could allow a malicious actor to perform synthesis and carry out subsequent testing undetected and with greater ease.²⁵ The next evolution of cloud labs may further enable AI scientists capable of carrying out experiments in entirety by developing and testing hypotheses, interpreting results, and amending hypotheses for retesting, thus lowering the technical barrier even further.²⁶ Furthermore, directed evolution of pathogens, which spans all 3 phases of the design-build-test cycle, is another route a malicious actor could take in the development of a customized pathogen.¹³ AI techniques have already been leveraged to accelerate this process and guide directed evolution in protein engineering.²⁷

Overall, in the future, it may be possible for a bad actor to perform complete hands-off, *in silico* design, building, and testing of a novel or recreated pathogen by leveraging the technologies and advancements outlined above.^{25,28} The potential democratization and synergy of these technologies pose potentially high-consequence risks that could elevate to a globally catastrophic level if not recognized and mitigated.

Vulnerabilities and AI Systems

As technologies in the biological sciences continue to be developed, AI-enabled offensive cyber-attacks may become an increasing biosecurity threat. For example, cybersecurity vulnerabilities in machines such as DNA synthesizers, could allow the introduction of malware to corrupt the design or audio-record printed DNA sequences, disrupt laboratory biosecurity, or allow intruder access to sensitive information.^{29,30} In addition, AI system vulnerabilities could also have downstream effects on human health. If healthcare becomes increasingly automated with the use of AI, so does increased risk for cyber-intrusion and subsequent manipulation of medical devices, which could result in harm for the end-user. TrapX Security Labs notes in their latest MED-JACK report³¹ that many medical devices and equipment in use today, such as infusion pumps, medical lasers, LASIK surgical machines, and life support equipment are vulnerable to hijacking by malicious actors. Wireless connectivity is becoming increasingly common in implantable medical devices, opening up the risk of malicious actors mounting attacks from anywhere in the world.³²

Beyond the risks associated with medical device exploitation, it is possible that in the future computer systems will be integrated with human physiology and, therefore, pose novel vulnerabilities. Brain-computer interfaces (BCIs), traditionally used in medicine for motor-neurological disorders, are AI systems that allow for direct communication between the brain and an external computer.³³ BCIs allow for a bidirectional flow of information, meaning the brain can receive signals from an external source and vice versa.

The neurotechnology company, Neuralink, has recently claimed that a monkey was able to control a computer using one of their implants.³⁴ This concept may seem farfetched, but in 2004 a paralyzed man with an implanted BCI was able to play computer games and check email using only his mind.³⁵ Other studies have shown a “brain-brain” interface between mammals is possible.^{36,37} In 2013, one researcher at the University of Washington was able to send a brain signal captured by electroencephalography over the internet to control the hand movements of another by way of transcranial magnetic stimulation.³⁸ Advances are occurring at a rapid pace and many previous technical bottlenecks that have prevented BCIs from widespread implementation are beginning to be overcome.³⁹⁻⁴¹

Research and development of BCIs have accelerated quickly in the past decade.⁴² Future directions seek to achieve a symbiosis of AI and the human brain for cognitive enhancement and rapid transfer of information between individuals or computer systems.³⁴ Rather than having to spend time looking up a subject, performing a calculation, or even speaking to another individual, the transfer of information could be nearly instantaneous. There have already been numerous studies conducted researching the use of BCIs for cognitive enhancement in domains such as learning and memory, perception, attention, and risk aversion (one being able to incite riskier behavior).⁴³ Additionally, studies have explored the military applications of BCIs, and the field receives a bulk of its funding from US Department of Defense sources such as the Defense Advanced Research Projects Agency.⁴⁴

While the commercial implementation of BCIs may not occur until well into the future, it is still valuable to consider the risks that could arise in order to highlight the need for security-by-design thinking and avoid path dependency, which could result in vulnerabilities—like those seen with current medical devices—persisting in future implementations. Cyber vulnerabilities in current BCIs have already been identified, including those that could cause physical harm to the user and influence behavior.⁴⁵⁻⁴⁸ In a future where BCIs are commonplace alongside advanced understandings of neuroscience, it may be possible for a bad actor to achieve limited influence over the behavior of a population or cause potential harm to users. This issue highlights the need to have robust risk assessment prior to widespread technological adoption, allowing for regulation, governance, and security measures to take identified concerns into account.

COMBINED ILLUSTRATIVE FRAMEWORK

Given the rapid developments in AI and biotechnology and the convergence of these 2 areas, existing risk assessment frameworks could be adapted to properly examine emerging threats. Table 3 is an illustrative framework that adapts the original AAAS-FBI-UNICRI framework by removing some elements while adding some from other frameworks to improve its capability to assess converging risks, such as those posed by AI and biotechnology. Convergent risk scenarios from this paper are included as examples of how this framework might be implemented; the completed boxes presented in the table are illustrative only and not meant to be conclusive, as there was no large-scale expert review and consensus.

In Table 4, the adversary spans the same range for each element, which may suggest it is a variable that might not be worth including or at least reconsidered. However, a comprehensive assessment by multidisciplinary experts with multiple scenarios would need to be conducted to determine this. In this framework, rather than using an ordinal scale of only 3 (low, moderate, high) to assess each parameter, an ordinal scale of 5 (very low, low, moderate, high, very high) was used to increase specificity and better delineate between scenarios. Each parameter is meant to be evaluated based on the projected timeline. For instance, in the case of BCIs, it is assumed the parameter will be highly accessible and democratized in the long-term. Governability was included as a component contributing to probability since a technology convergence that is less governable may be more likely to be maliciously used. Democratization, called “accessibility” in other frameworks, was included because a technology that is more accessible to a wide range of individuals is assumed to have higher potential for misuse. The remaining evaluation criteria (vulnerabilities, needed skills and expertise, magnitude of potential consequences, and existing countermeasures) were selected because they are common to many of the frameworks reviewed herein and we believe are integral to any risk assessment framework. “Vulnerabilities” is meant to describe and illuminate any gaps that could be exploited to cause harm. “Needed skills and expertise” is meant to describe the scope of potential adversaries and help identify potential bottlenecks. “Magnitude of potential consequences” and “existing countermeasures” are meant to contextualize the scope of potential harm and balance that with any mitigating factors in place.

As one might expect, this combined framework has many limitations. For example, depending on the type of pathogen designed or synthesized, the high variance changes the amount of expertise needed and potential consequences over time, as does the nature and sophistication of the type of BCI exploitation. The framework also includes uncertainties related to the timeline and assessing governability, which highlights the high degree of assumptions required to complete and use it. Moreover, the framework lacks a standardized mechanism for evaluating factors in a way that calculates risk. The hope is that explicitly identifying

Table 4. Illustrative Framework Adapted from AAAS-FBI-UNICI, NASEM, and Tucker Frameworks¹²⁻¹⁵

			Convergent Risk Scenarios		
			In Silico Design of a Pathogen	In Silico Synthesis of a Pathogen	Brain-Computer Interface Exploitation
Risk Assessment	Probability	Adversary	Nation state, nonstate group, or individual	Nation state, nonstate group, or individual	Nation state, nonstate group or individual
		Timeline	Near term (0 to 5 years)	Mid term (6 to 10 years)	Long term (11+ years)
		Democratization	Moderate	High	Very High
		Vulnerabilities	Open access data, open source software	Limited customer verification	Wireless connectivity, cybersecurity
		Needed scientific expertise and skills	Synthetic biology, genomics, bioinformatics	Synthetic biology, bioinformatics	Neuroscience, computer science, electronics
		Governability	Low	Moderate	Moderate
	Consequences	Magnitude of potential consequences to economies, political systems, society, health, environment, and agriculture	Moderate	High	Very High
		Sufficient existing countermeasures	None	None	None
Risk			Moderate	Moderate	Very High

these limitations can help illustrate the gaps and challenges in developing a comprehensive framework for converging technologies that may pose high-consequence risks.

RECOMMENDATIONS

Risk Assessment Framework Approaches

A risk assessment framework is needed that is applicable to converging technologies in the life sciences that balances generality to capture a broad scope but also maintains enough specificity to capture nuances. It is paramount that such frameworks be developed and implemented by a multidisciplinary team of experts that includes both security officials and scientists. Such a strategy allows for a more comprehensive and, subsequently, more accurate depiction of the potential risks. To further increase robustness, existing frameworks should be iteratively built upon and improved, such as the combined framework presented in this paper.

A general framework intended to apply to a diverse range of converging or emerging technologies should be developed prior to determining the scenario to be evaluated. It can then be revised as scenarios or technologies are assessed. This would benefit the applicability of such a framework to other scenarios or technologies rather than narrowly limiting it to a specific issue.

The role that time plays in calculating risk for an emerging or converging technology should be weighed. While a technology that is assumed to not be relevant until the far future may not pose an immediate risk, it should still be considered, as it could pose a significant threat once the technology materializes. Additionally, projected timelines for emerging technologies are inherently uncertain and technologies may develop more rapidly than expected. Therefore, it is worth considering projected timelines as a framing element rather than a factor to calculate risk.

Revisiting the AAAS-FBI-UNICRI Framework

As 5 years have passed since publication of the initial AAAS-FBI-UNICRI report, we recommend this working group reconvene to reassess the risks outlined in this paper and update their framework to be more inclusive of other converging threats. As noted in their last report, "if scientists and security experts can assess the risks and benefits together and on a routine basis, the potential risk that biological [big data] would be used by adversaries decreases." We believe that this team is best equipped to develop an iterative set of frameworks related to this topic because the workgroup includes knowledgeable experts from a wide range of disciplines and the issues related to big data and AI are similar in nature. With perspectives from experts in law

enforcement, the life sciences, and AI, the team is uniquely positioned to take on such a task. While their initial project focused narrowly on big data, it would be of great value to extend that scope to AI in the life sciences and provide relevant stakeholders in academia, government, and the private sector with an informational risk assessment that will help them make informed research, policy, and security decisions on these topics.

This working group can leverage frameworks—such as the NASEM and Tucker-Koblentz frameworks—that have been developed since their initial report to improve upon their own. The Koblentz et al¹⁴ paper could provide them with an insightful review of the strengths and weaknesses of many previous frameworks and would be a useful resource as the working group develops the next iteration. Components from the NASEM framework could be leveraged to increase the generality and broadness of their framework. The illustrative framework presented in this paper (Table 4) may serve as a useful starting point for the joint team.

CONCLUSION

AI and biotechnology are both rapidly growing fields. The biosecurity risks arising at their points of convergence are vast, highlighting the need for comprehensive mapping and thorough, robust risk assessment. Current frameworks used for risk assessment in health security settings may fail to capture the nuanced areas of these converging fields and could benefit from considering generality and timeline for technology adoption. We recommend that a multidisciplinary working group comprised of AAAS, the FBI, UNICRI, and other specialists gather to formulate a framework to comprehensively assess risks the results from converging technologies such as AI and biotechnology as well as others that may emerge in the future. The combined framework presented in this paper could serve as a starting point for the development of more robust biosecurity risk assessments for converging technologies like AI and biotechnology. While this new framework does not solve all of the identified challenges, it explicitly elucidates the gaps in our ability to accurately assess risks associated with emerging and converging technologies and will hopefully facilitate adaption of previous frameworks in order to develop a more inclusive health security risk assessment that properly balances generality and specificity and is regularly and iteratively revisited.

ACKNOWLEDGMENTS

We thank Gregory Lewis and Jacob Swett for reviewing and commenting on early versions of this manuscript. For this manuscript preparation, funding was received from the Open Philanthropy Project. The authors declare no competing conflicts of interest.

REFERENCES

1. Khan N, Yaqoob I, Hashem IAT, et al. Big data: survey, technologies, opportunities, and challenges. *ScientificWorld-Journal*. 2014;2014:712826.
2. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Bus Horiz*. 2019;62(1):15-25.
3. Arel I, Rose DC, Karnowski TP. Deep machine learning - a new frontier in artificial intelligence research [research frontier]. *IEEE Comput Intell Mag*. 2010;5(4):13-18.
4. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85-117.
5. Wetterstrand KA. DNA sequencing costs: data. National Institutes of Health National Human Genome Research Institute website. Accessed September 6, 2019. <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>
6. World Economic Forum (WEF). *Health and Healthcare in the Fourth Industrial Revolution Global Future Council on the Future of Health and Healthcare 2016-2018*. Geneva: WEF; 2019. Accessed September 24, 2019. http://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Health_Council_Report.pdf
7. Dalmia V, Sharma K. The moral dilemmas of the Fourth Industrial Revolution. World Economic Forum website. Published February 13, 2017. Accessed September 24, 2019. <https://www.weforum.org/agenda/2017/02/ethics-2-0-how-the-brave-new-world-needs-a-moral-compass>
8. Mesko B. The role of artificial intelligence in precision medicine. *Expert Rev Precis Med Drug Dev*. 2017;2(5):239-241.
9. Oliveira AL. Biotechnology, big data and artificial intelligence. *Biotechnol J*. 2019;14(8):e1800613.
10. Schoch-Spana M, Cicero A, Adalja A, et al. Global catastrophic biological risks: toward a working definition. *Health Secur*. 2017;15(4):323-328.
11. American Association for the Advancement of Science (AAAS), Federal Bureau of Investigation, United Nations Interregional Crime and Justice Research Institute. *National and Transnational Security Implications of Big Data in the Life Sciences*. Washington, DC: AAAS; 2014. Accessed September 16, 2019. https://www.aaas.org/sites/default/files/AAAS-FBI-UNICRI_Big_Data_Report_111014.pdf
12. Tucker JB. *Innovation, Dual Use, and Security: Managing the Risks of Emerging Biological and Chemical Technologies*. Cambridge, MA: MIT Press; 2012.
13. National Academies of Sciences Engineering and Medicine. *Biodefense in the Age of Synthetic Biology*. Washington, DC: National Academies Press; 2018. Accessed September 15, 2019. <https://doi.org/10.17226/24890>
14. Koblentz GD, Kirkpatrick J, Palmer MJ, Denton SW, Tiu B, Gloss K. *Biotechnology Risk Assessment: State of the Field*. Editing Biosecurity Working Paper No 1. Arlington, VA: George Mason University; 2017. Accessed September 16, 2019. https://static1.squarespace.com/static/58eaf565197aea9c401e0497/t/5c051efb88251b7718b58f0b/1543839485066/Biotech+Risk+Assessment_WP1.pdf
15. Carbonell P, Radivojevic T, García Martín H. Opportunities at the intersection of synthetic biology, machine learning, and automation. *ACS Synth Biol*. 2019;8(7):1474-1477.
16. Thiel V. Synthetic viruses—anything new? *PLoS Pathog*. 2018;14(10):e1007019.

17. Fernández CR. DNA script raises €34M to bring rapid DNA synthesis to the lab bench. Labiotech website. Published May 16, 2019. Accessed September 6, 2019. <https://labiotech.eu/industrial/dna-script-fundraising-dna-synthesis/>
18. Edwards B. *Insecurity and Emerging Biotechnology: Governing Misuse Potential*. Cham, Switzerland: Springer Nature Switzerland; 2019.
19. Sandberg A, Nelson C. Who should we fear more: biohackers, disgruntled postdocs, or bad governments? *Health Secur.* 2020;18(3):155–163.
20. Hutson M. AI protein-folding algorithms solve structures faster than ever: deep learning makes its mark on protein-structure prediction. Nature website. Published July 22, 2019. Accessed May 29, 2020. <https://www.nature.com/articles/d41586-019-01357-6>
21. 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 Accessed August 14, 2019. https://www.sipri.org/sites/default/files/2019-03/sipri2019_bioplusx_0.pdf
22. Pauwels E. *The New Geopolitics of Converging Risks: The UN and Prevention in the Era of AI*. New York: United Nations University Centre for Policy Research; 2019. Accessed September 4, 2019. <https://i.unu.edu/media/cpr.unu.edu/attachment/3472/PauwelsAIGeopolitics.pdf>
23. Deneke C, Rentzsch R, Renard BY. PaPrBaG: a machine learning approach for the detection of novel pathogens from NGS data. *Sci Rep.* 2017;7:39194.
24. Bartoszewicz JM, Seidel A, Rentzsch R, Renard BY. DeePaC: predicting pathogenic potential of novel DNA with reverse-complement neural networks. *Bioinformatics.* 2020;36(1):81–89.
25. Dunlap G, Pauwels E. *The Intelligent and Connected Bio-Labs of the Future: Promise and Peril in the Fourth Industrial Revolution*. Washington, DC: Wilson Center; 2017. Accessed August 14, 2019. https://www.wilsoncenter.org/sites/default/files/dunlap_pauwels_intelligent_connected_biolabs_of_future.pdf
26. Bilsland E, van Vliet L, Williams K, et al. *Plasmodium* dihydrofolate reductase is a second enzyme target for the antimalarial action of triclosan. *Sci Rep.* 2018;8(1):1038.
27. Yang KK, Wu Z, Arnold FH. Machine-learning-guided directed evolution for protein engineering. *Nat Methods.* 2019;16(8):687–694.
28. Schabacker DS, Levy L-A, Evans NJ, Fowler JM, Dickey EA. Assessing cyberbiosecurity vulnerabilities and infrastructure resilience. *Front Bioeng Biotechnol.* 2019;7:61.
29. Peccoud J, Gallegos JE, Murch R, Buchholz WG, Raman S. Cyberbiosecurity: from naive trust to risk awareness. *Trends Biotechnol.* 2018;36(1):4–7.
30. Faezi S, Chhetri R, Malawade AV, et al. Oligo-Snoop: a non-invasive side channel attack against DNA synthesis machines. In: *26th Annual Network and Distributed System Security Symposium, MDSS 2019*. The Internet Society; 2019. Accessed August 14, 2019. <https://dx.doi.org/10.14722/ndss.2019.23544>
31. TrapX Research Labs. *MEDJACK.4 Medical Device Hijacking*. San Jose, CA: TrapX Security; 2018 Accessed September 4, 2019. <https://trapx.com/wp-content/uploads/2018/04/MedJack.4.pdf>
32. Pycroft L, Aziz TZ. Security of implantable medical devices with wireless connections: the dangers of cyber-attacks. *Expert Rev Med Devices.* 2018;15(6):403–406.
33. Nicolas-Alonso LF, Gomez-Gil J. Brain computer interfaces, a review. *Sensors (Basel).* 2012;12(2):1211–1279.
34. Bowman M, Koebler J. Elon Musk announces plan to ‘merge’ human brains with AI. Vice. Published July 17, 2019. Accessed September 6, 2019. https://www.vice.com/en_uk/article/7xgnxd/elon-musk-announces-plan-to-merge-human-brains-with-ai
35. Khamsi R. Paralyzed man sends e-mail by thought. Nature. Published October 13, 2004. Accessed September 6, 2019. <http://www.nature.com/articles/news041011-9>
36. Pais-Vieira M, Lebedev M, Kunicki C, Wang J, Nicoletis MAL. A brain-to-brain interface for real-time sharing of sensorimotor information. *Sci Rep.* 2013;3:1319.
37. Pais-Vieira M, Chiufta G, Lebedev M, Yadav A, Nicoletis MAL. Building an organic computing device with multiple interconnected brains. *Sci Rep.* 2015;5:11869.
38. Rao RPN, Stocco A, Bryan M, et al. A direct brain-to-brain interface in humans. *PLoS One.* 2014;9(11):e11332.
39. Patel SR, Lieber CM. Precision electronic medicine in the brain. *Nat Biotechnol.* 2019;37(9):1007–1012.
40. Musk E, Neuralink. An integrated brain-machine interface platform with thousands of channels. *J Med Internet Res.* 2019;21(10):e16194.
41. Martins NRB, Angelica A, Chakravarthy K, et al. Human brain/cloud interface. *Front Neurosci.* 2019;13:112.
42. Kögel J, Schmid JR, Jox RJ, Friedrich O. Using brain-computer interfaces: a scoping review of studies employing social research methods. *BMC Med Ethics.* 2019;20(1):18.
43. Cinel C, Valeriani D, Poli R. Neurotechnologies for human cognitive augmentation: current state of the art and future prospects. *Front Hum Neurosci.* 2019;13:13.
44. Munyon CN. Neuroethics of non-primary brain computer interface: focus on potential military applications. *Front Neurosci.* 2018;12:696.
45. Ienca M, Haselager P. Hacking the brain: brain–computer interfacing technology and the ethics of neurosecurity. *Ethics Inf Technol.* 2016;18(2):117–129.
46. Marin E, Singelée D, Yang B, et al. Securing wireless neurostimulators. In: *CODASPY '18: Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy*. New York: ACM; 2018:287–298. Accessed September 6, 2019. <https://doi.org/10.1145/3176258.3176310>
47. Bernal SL, Celdrán AH, Pérez GM, Barros MT, Balasubramanian S. Cybersecurity in brain-computer interfaces: state-of-the-art, opportunities, and future challenges. *arXiv Cryptogr Secur.* 2019;arXiv:1908.03536 [cr.CR]. Accessed September 6, 2019. <http://arxiv.org/abs/1908.03536>
48. Pycroft L, Boccard SG, Owen SLF, et al. Brainjacking: implant security issues in invasive neuromodulation. *World Neurosurg.* 2016 Aug;92:454–462. <https://linkinghub.elsevier.com/retrieve/pii/S1878875016302728>

Manuscript received October 14, 2019;
revision returned March 4, 2020;
accepted for publication April 29, 2020.

Address correspondence to:
Cassidy Nelson
Research Scholar
Future of Humanity Institute
University of Oxford
16-17 St Ebbe's Street
Oxford OX1 1PT, UK

Email: cassidy.nelson@gtc.ox.ac.uk