The Security Governance Challenge of Emerging Technologies

August 2022

Robert F. Trager

Abstract: *In recent decades, governments have been ineffectual at regulating dangerous emerging technologies like lethal autonomous weapons and synthetic biology. In today's era of great power competition, changing course is difficult but imperative. Several areas of technological development illustrate the generic problem of responding to innovation: the complexity of the analysis required to determine the best governance responses, the difficulties of building a consensus for action, and the urgent need for institutional frameworks that facilitate anticipatory rather than reactive regulation. The international community needs new, standing international governance frameworks to address this growing challenge.*

In December 2021 at the United Nations in Geneva, government and civil society representatives met to set an agenda for regulating lethal autonomous weapons. Most nations favor adopting rules to govern the development and use of these technologies. Yet, following eight years of discussions in the context of the Convention on Certain Conventional Weapons (CCW), the governments of the world found virtually no common ground. In the end, they barely managed to agree on ten more days of discussion in the coming year. Unfortunately, governments historically have experienced similar difficulties implementing collective governance solutions for other potentially dangerous technologies.

In stark contrast to governmental sclerosis, the pace of technological development appears set to increase. Arguably, we are experiencing a second industrial revolution. In the first, machines harnessed new energy sources to supplement human labor. In the second, artificial intelligence is harnessing non-human energy sources to supplement human thought. The implications of this current change are still unknown, but an increase in the rate of

doi: 10.1016/j.orbis.2022.08.008

discoveries is certainly a possibility.[1] Such progress could open possibilities of extraordinary futures of human flourishing, and yet powerful technologies can also cause harm. There is no guarantee that technological solutions will arrive in time to solve the problems that technology creates. Sometimes, a governance solution is available when a technical one is not.

This article contends that, even without such a change in the rate of progress, technologies are already here that require urgent, new regulatory actions by the international community. It suggests governance paths that may be politically possible. Four emerging-technology security-governance challenges of today are described, but this list of challenges is just a start. They are examples of the challenges the world community should expect to face with increasing regularity as technological insights compound and the rate of discovery increases. They illustrate the generic problem of responding to innovation—the complexity of the analysis required to determine the best governance responses, the difficulties of building a consensus for action, and the urgent need for new institutional frameworks that facilitate anticipatory, rather than reactive, regulation.

Each of these emerging technologies demands a tailored solution. But there are also commonalities across them in terms of actor interests and effective institutional responses. As new technologies emerge, we can undoubtedly expect the same. This reality suggests the need for general solutions to the problem of controlling emerging technology rather than *de novo* responses in each case. The international community needs new, standing international governance frameworks to address this growing challenge.

Autonomous Weapons Technologies

The technology governance challenge is unique to the modern age. To be sure, actors in earlier eras saw the benefits of regulating particular security technologies. In the twelfth century, the Catholic Church attempted to ban the use of projectile weapons by and against Christians.[2] The Hague Convention of 1899 prohibited the use of certain technologies in war: poison gas, balloons to deliver bombs, and bullets designed to "expand or flatten" in the body.[3] Yet, never before has there been a wide recognition of the need to protect human

[1] For a wide-ranging treatment of these issues, see, Henry A. Kissinger, Eric Schmidt, and Daniel Huttenlocher, *The Age of A.I.: And Our Human Future* (New York, NY: Little Brown and Company, 2021).
[2] Martin Van Creveld, *Technology and War: From 2000 BC to the Present* (New York, NY: Simon and Schuster, 2010), p. 71.
[3] James B. Scott, ed., *The Hague Conventions and Declarations of 1899 and 1907: Accompanied by Tables of Signatures, Ratifications and Adhesions of the Various Powers, and Texts of Reservations* (Oxford, UK: Oxford University Press, 1915).

security by regulating not just the use of technologies in wartime but even the spread of various technologies in peacetime.

The new era began with the development of nuclear weapons in 1945.[4] The destructive potential of the technology immediately prompted an attempt to bring it under international control through the United Nations in what became known as the Baruch Plan. While this effort failed, test-ban and non-proliferation regimes eventually were established.[5] Protocols were accepted during the Cold War alongside shared safety technologies to make accidental or unauthorized use less likely.

The era continued with the regulation of biological and chemical weapons technologies. These conventions prohibit development, production, acquisition, transfer, stockpiling, and use of these weapons, as well as their precursors. The conventions illustrate that arms control and governance regimes are important even when they are violated. Only one state, Syria, has recently claimed to possess biological weapons, for instance, and members of the scientific community could not admit to working on them.[6] The impact of these technologies on world affairs is thus probably greatly reduced, despite the large-scale Soviet violation of the Biological Weapons Convention.[7]

[4] See earlier examples of attempts to limit certain types of arms in peacetime. The Washington Naval Treaty briefly limited the size of ships, for instance. In the first half of the twentieth century, there were proposals for collective security through the internationalization of airpower. Winston Churchill advocated for this approach to airpower in the interwar period. See Winston Churchill, *The Aftermath: The World Crisis, 1918–1928* (New York, NY: Charles Scribner's Sons, 1929), p. 27. This is one of the few instances of a broad-based call for limiting the spread of a technology because the technology was considered dangerous in itself. Such proposals gained broader interest with the invention of nuclear weapons. See Waqar H. Zaidi, *Technological Internationalism and World Order* (Cambridge, UK: Cambridge University Press, 2021).

[5] While most scholars and policymakers have agreed on the desirability of some non-proliferation regime, there is a robust debate about whether the existing regime organized around the International Atomic Energy Agency has done more harm than good. Critics contend that the policy of sharing peaceful nuclear technology in return for countries refraining from developing a weapons program has led states to develop nuclear weapons when they would not otherwise have done so. See, for instance, Christoph Bluth, et al., "Civilian Nuclear Cooperation and the Proliferation of Nuclear Weapons," *International Security*, Summer 2010.

[6] "Chemical and Biological Weapons Status at a Glance," Fact Sheet, Arms Control Association, 2022, https://www.armscontrol.org/factsheets/cbwprolif.

[7] It is not possible to have a high degree of certainty about what biological weapons would have been developed, stockpiled, and used in the absence of a convention.

Today, many countries around the world recognize the need to regulate Lethal Autonomous Weapons (LAWS). Several factors make regulation a priority. The first is that autonomy implies scalability. When weapons systems no longer require human operators, they can be produced at scale and trained to coordinate in so-called "swarms." They could be used to do large scale damage, but perhaps more importantly, they could be used to influence political outcomes through large-scale, but finely targeted, applications of violence and the threat of violence. This threat has the potential to influence state-society relations, including facilitating tighter authoritarian control. More broadly, it is a technology that, in a more developed form, could eventually enable the rapid deployment of extreme violence on the level of weapons of mass destruction (WMD) capabilities.

The second reason autonomous capabilities pose a security challenge is the difficulty of knowing how they will behave when deployed.[8] The use of autonomous and machine learning capabilities in real-world settings often leads to unintended consequences. This risk is particularly high in an adversarial context in which systems would be designed to thwart each other. Such systems would need to be unpredictable to frustrate adversaries. This possibility implies that they would be unpredictable even to their own designers—and this generates risks.[9] Further, in the absence of regulation, security dilemmas may cause states to deploy technologies without adequate safety testing to compete with adversaries.

Relatedly, the idea that conflicts could remain *only* between autonomous systems is probably a chimera. If one side were weaker in a contest between autonomous systems, it might turn to, or threaten, violence against human targets to deter its destruction. It is also likely that offensive capabilities (the capability to do violence) would be dominant over defensive capabilities (the capability to prevent it) involving civilian targeting by LAWS.[10] This scenario

However, the willingness of some leaders to take actions to shock and intimidate rival countries and their own populations suggests that these weapons would have been used more.

[8] Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York, NY: WW Norton & Company, 2018); Michael C. Horowitz, "When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability," *Journal of Strategic Studies,* vol. 42, no. 6 (2019) pp. 764–88.

[9] Human warfighters also necessarily exhibit a certain amount of tactical unpredictability, but human commanders have a better chance of understanding what human fighters might do in different contexts. Artificial intelligence is often valuable precisely because it considers possibilities that humans do not.

[10] For considerations on how the offense-defense balance may shift as autonomous technology improves, see Benjamin Garfinkel, and Allan Dafoe, "How Does the

was true in 1932 when British Prime Minister Stanley Baldwin stated that "the bomber will always get through" and it is likely to remain so in the near-term.[11] Thus, there is no *a priori* reason to expect bloodless autonomous weapons conflicts and there is the potential for radical escalations to extreme violence.[12]

A third danger of LAWs is that they may facilitate violence without attribution. Combined with their ability to evade traditional defenses in carrying out targeted and large-scale violence, LAWs have the potential to reshape the political order. If the technology is unregulated, as it spreads, attribution of attacks will become ever more difficult. Conceivably, political leaders and others, such as human rights campaigners, could be targeted without the ability of the targeted states to identify the perpetrators. Certainly, this scenario has the potential to lead to new cycles of violence as targeted groups lash out against potential, rather than actual, perpetrators.[13]

For these reasons, and others,[14] there is a case for governance solutions along the lines of the nuclear, chemical, and biological regulatory regimes.[15] While it may not be possible to prevent the spread of simple, small-scale applications of AI technology in drones, larger-scale, sophisticated applications

Offense-Defense Balance Scale?" *Journal of Strategic Studies*, vol. 42, no. 6 (2019), pp. 736–63.

[11] In his speech to the British Parliament entitled "Fear for the Future," Baldwin argued that wars would require the sides to "kill more women and children more quickly than the enemy if you want to save yourselves."

[12] E. Gartzke, "Blood and Robots: How Remotely Piloted Vehicles and Related Technologies Affect the Politics of Violence," *Journal of Strategic Studies*, 2019, pp.1–31.

[13] Sandeep Baliga, Ethan Bueno De Mesquita, and Alexander Wolitzky, "Deterrence with Imperfect Attribution," *American Political Science Review* vol. 114, no. 4 (2020), pp. 1155–78.

[14] For instance, scalable LAWS technologies may fill in gaps in the escalatory ladder between "conventional" and "nuclear," potentially encouraging higher levels of escalation on average. They may also make decisions so quickly that even those aspects of crises that remain under human control become more difficult to manage. Alternatively, they pose new threats against sensors with the potential to lead to crisis decision making in lower information environments.

[15] Note that LAWS could also have positive effects in some cases, such as reducing collateral damage from attacks and the need to escalate in response to attacks. See Ronald Arkin, "Lethal Autonomous Systems and the Plight of the Non-combatant," *AISB Quarterly,* July 2013. Note also other models of arms control agreements that could be applied to this case including: the Ottawa Convention to Ban Landmines and the Treaty on the Prohibition of Nuclear Weapons, even though many of the most powerful countries have not signed them.

of lethal autonomous technologies may require the resources of larger actors.[16] This is particularly the case for building LAWS hardware on the scale of swarms. LAWS software, by contrast, spreads easily through copying, but there are nevertheless opportunities for restricting its development and spread. Regulatory regimes with the capacity to punish actors can greatly reduce the spread of both hardware and software.

These objectives cannot be achieved unilaterally, and there may be a window of opportunity for coordinated action.[17] If a powerful actor enables the spread of these technologies, they will spread widely. Once they do, it may be much more difficult to destroy existing capabilities and prevent the use of existing know-how than to prevent the capabilities from spreading to certain actors in the first place.

Yet, it appears likely that some powerful actors will not forgo development of these technologies. The failure to regulate these technologies in the CCW strongly suggests this reality. Here, the Chemical and Biological Conventions were an easier case for arms control. They were facilitated because some powerful countries viewed these weapons as less effective than others that were available to them. Monitoring compliance was therefore less essential.[18] LAWS—in part because they may eventually allow precise targeting and shield their operators from harm—are seen as eventually becoming more useful. Yet, some states are unlikely to agree to the invasive monitoring regime that would be required to guarantee compliance—just as the Soviet Union would not agree to a strict monitoring and sanctions regime at the time of the Baruch Plan. Thus, as in the case of the nuclear regime, there is an argument for authorizing a small number of actors to develop and even stockpile the technology. These actors would then be incentivized to uphold the non-proliferation regime. Some actors could view a non-proliferation regime as threatening, but security guarantees for actors who decline to develop and deploy the technology could alleviate these concerns, as they have in the nuclear non-proliferation regime. Alternatively, there may be opportunities to ban less

16 Thousands of AI and robotics researchers have signed an open letter advocating a ban of LAWS. UN Secretary General António Guterres has also said the world should "ensure that autonomous machines are never given lethal capacity outside human judgment or control." See the Secretary-General's remarks to the UN Human Rights Council: "The Highest Aspiration: A Call to Action for Human Rights," Feb. 24, 2020.

17 The United States took unilateral action in promising to maintain broad command and control of autonomous technologies in Department of Defense Directive 3000.09. This directive does nothing to restrict the spread of the technologies, however.

18 The chemical, but not the biological, convention has a monitoring component because the former weapons were seen as more effective on the battlefield.

sophisticated applications of the technology that are available to larger numbers of actors. This approach would provide incentive for the few more capable actors to maintain their special status by declining to share the technology and policing its development and use.

There is also a role for the development and policing of technical safety standards. For instance, it will be important to prevent commercially available AI technologies from being reprogrammed to weaponize them. This, too, follows the playbook from the Cold War when the United States developed Permissive Action Link (PAL) technology and shared it with the Soviet Union to prevent unauthorized arming of a nuclear weapon. Analogous technologies may be developed to aid AI regulatory regimes once global standards are set.[19]

In general, arms control does not happen when it limits the power of the powerful. The full idealism of a ban on the technology may not be necessary to safeguard global security, however. Currently, a non-proliferation regime that includes security guarantees for actors that would otherwise be disadvantaged by it appears to be the most promising approach.

AI Misaligned with Human Values

AI researchers agree that misaligned AI systems pose dangers to human security, but they disagree about the extent of the dangers. Clearly, however, AI systems that seek to optimize one set of variables will often have unexpected, pernicious effects on other variables. Social media algorithms designed to maximize viewers' attention, for instance, may end up driving political polarization as a byproduct.[20] Such effects can occur despite the best intentions of software designers.[21]

As AI systems become ever more capable, the magnitude of potential pernicious effects increases. In driving political polarization, an unintended

[19] National Security Commission on Artificial Intelligence, Final Report, 2021, https://www.nscai.gov/2021-final-report/, p. 106.

[20] H. Allcott, L. Braghieri, S. Eichmeyer, and M. Gentzkow, "The Welfare Effects of Social Media," *American Economic Review*, vol. 110, no.3 (2020), pp. 629–76; M. H. Ribeiro, R. Ottoni, R. West, V. A, Almeida, and W. Meira Jr, "Auditing Radicalization Pathways on YouTube," *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 131–41.

[21] Prominent AI research scientist Stuart Russell argues that instead of designing systems with hard-coded objectives, we should design systems that use reverse reinforcement learning. These systems attempt to infer human goals from human behavior. This may well solve one set of problems, but it is not clear that this approach resolves all the concerns.

consequence of AI systems deployed to date may well have been to undermine democratic processes. But these systems are in their infancy compared to systems that are currently in development. If democracy is undermined by the infant version, what might more mature technologies do? These effects are likely to cross national boundaries. Some believe that powerful, misaligned AI systems may pose catastrophic risks.[22]

A key factor driving risk appears to be the tradeoff between AI safety and performance. This tradeoff exists across the emerging technology landscape and often relates to competitive contexts where speed of development and deployment is of the essence. The Chernobyl RBMK reactor, for instance, seems to have been designed with efficiency over safety in mind.[23] Technology companies often deploy new products as soon as they are workable —before safety measures to secure individuals' data have been put in place.[24] The AI field has its own examples. The company formerly known as Facebook, for instance, knowingly employed algorithms that promoted anger-provoking content, potentially furthering societal polarization, because these algorithms were the most effective at protecting declining market share.[25] Thus, AI safety involves both a variety of difficult-to-measure risks and incentives to cut corners on safety. This raises the specter of races to the bottom—a classic argument for regulatory interventions.

Considerations like these have caused prominent AI developers and technologists to advocate for global regulatory regimes. In fact, many are tracking the discussions of the proposed European Union AI Act. Advocates hope to identify the riskiest areas of development and put safety incentives in place. There are credible suggestions for domestic legislation to mitigate risks,[26]

[22] Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (New York, NY: Hachette Books, 2020); Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (New York, NY: Penguin, 2019). The "alignment problem" is often considered in terms of a single principal (sometimes "humanity") and a single AI agent. See Andrew Critch and David Krueger, "AI Research Considerations for Human Existential Safety (ARCHES)," *arXiv preprint*:2006.04948. They point out that the problem is far more complex: we need to worry about safety in worlds of multiple principals and multiple agents.

[23] Thomas Filburn and Stephan Gregory Bullard, *Three Mile Island, Chernobyl and Fukushima* (New York, NY: Springer, 2016).

[24] Nicole Perlroth, *This Is How They Tell Me the World Ends: The Cyberweapons Arms Race* (London, UK: Bloomsbury Publishing, 2021).

[25] See, "The Facebook Files: A Wall Street Journal Investigation," https://www.wsj.com/articles/the-facebook-files-11631713039.

[26] M. Brundage, et al., 2020, "Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims," *arXiv preprint*: 2004.07213.

but internationally, the outlines of the most effective governance solutions are difficult to determine. Further consideration of these problems is important.

The Future of Mutually Assured Destruction

The dawn of the nuclear era produced a burst of concern about whether humanity would long survive. This question was exemplified in philosopher Bertrand Russell's essay "Will Man Survive?" and with physicist Robert Oppenheimer quoting Hindu scripture as he witnessed the first nuclear detonation: "Now I am become Death, the destroyer of worlds." Even in the early 1960s, US President John F. Kennedy warned the public of vast increases in the number of nuclear weapons states within a decade; many thought this would dramatically increase the risk of a nuclear conflict.[27]

As a result of these concerns, leaders around the world searched for governance solutions. The possibility of creating aspects of a world government was taken seriously by some global elites and publics.[28] In a world of rival states possessing the capacity to destroy advanced life, the costs of anarchy were seen as too high to bear. When world government and UN control of nuclear weapons both failed, the Kennedy administration sought another governance solution to slow proliferation: a partial test ban treaty. The administration saw this treaty as "a necessary, but not a sufficient, condition for keeping the number of nuclear countries small."[29] The administration also initiated the fundamental bargain of détente by promising to restrain Germany from acquiring nuclear weapons of its own.[30] After Kennedy's death, the Lyndon Johnson administration negotiated the Non-Proliferation Treaty with more state signatories than any other arms control treaty in history.

These actions shaped the world of today. It is possible that our civilization would not exist without them. Instead of dozens of nuclear

[27] *Public Papers of the Presidents of the United States: John F. Kennedy,* Press Conference, Mar. 21, 1963 (Washington, DC, US Government Printing Office, 1964), p. 280; and *New York Times*, Mar. 23, 1963.

[28] Waqar H. Zaidi, *Technological Internationalism and World Order: Aviation, Atomic Energy, and the Search for International Peace, 1920–1950.* (Cambridge, UK: Cambridge University Press, 2021).

[29] McNamara to Kennedy, "The Diffusion of Nuclear Weapons with and without a Test Ban Agreement," cited in Peter Lavoy, "Predicting Nuclear Proliferation: A Declassified Documentary Record," *Strategic Insights*, vol. III, no. 1 (Jan. 2004), p. 3.

[30] Marc Trachtenberg, *A Constructed Peace: The Making of the European Settlement, 1945–1963* (Princeton, NY: Princeton University Press, 1999), Ch. 9.

weapons states, there are nine. In almost eight decades since World War II ended, these weapons have not been used.

Yet, this success has also ushered in a complacency that is not justified by the facts. Nuclear weapons states continue to employ the strategy of mutually assured destruction. Statistically, avoiding nuclear destruction until the present produces weak confidence that the probability of destruction in, say, a given decade is extremely low.[31] And even a limited nuclear exchange between smaller powers might produce a catastrophic nuclear winter, blotting out the sun and starving much of humanity.[32]

Unfortunately, technological developments may increase the risk of nuclear weapons use. The idea of delegating launch decisions to autonomous systems is anathema to most nuclear security experts today.[33] Yet it is uncertain whether this restraint will remain the case.[34] Technological developments will increase the speed of delivery systems, could potentially enhance the ability to target all arms of the nuclear triad, and may undermine command and control. These developments would release pressures for greater autonomy. They could also diminish second-strike capabilities, particularly if states do not respond to the first signs of an attack. In such situations, *the demands of deterrence* can imply the need to commit to action before it is certain that adversaries are even attacking. Such policies were adopted during the Cold War—so-called "Launch on Warning." This policy can assist in deterring against a first strike, but it raises the risk of accidental conflict.

This is not to say that nuclear weapons use is likely or will become so soon. But the question we should ask is not whether it is likely but whether it is very unlikely. If a mutually assured destruction equilibrium is to persist over a long period of time, the yearly risk must be extremely low for the probability of long-term survival to be high.[35]

Thus, governance solutions are sorely needed in this space as well. Aside from the bilateral agreements between the United States and Russia, the

[31] Helen Caldicott, ed., *Sleepwalking to Armageddon: The Threat of Nuclear Annihilation* (New York: The New Press, 2017).
[32] Alan Robock, and Owen Brian Toon, "Local Nuclear War, Global Suffering," *Scientific American,* vol. 302 no. 1 (2010), pp. 74-81; and Alan Robock, et al., "Climatic Consequences of Regional Nuclear Conflicts," *Atmospheric Chemistry and Physics,* vol. 7, no. 8 (2007), pp. 2003–12.
[33] Vincent Boulanin, et. al., "Artificial Intelligence, Strategic Stability and Nuclear Risk," Stockholm International Peace Research Institute, 2020, p. 138.
[34] Indeed, Russia already has a so-called "doomsday" system, although many believe that it is usually turned off. See, Nicholas Thompson, "Inside the Apocalyptic Soviet Doomsday Machine," *Wired,* vol. 17, no. 10, Sep. 21, 2009.
[35] If there is a 1 percent probability of destruction in any given year, for instance, the probability of destruction over a thousand years is over 99.99 percent.

core of the nuclear arms control regime is a half century old. We should not expect that the institutions and agreements that met the needs—or seemed to—of that time will meet the needs of today. We require governance solutions on a similar and even greater scale at present. New technologies must be met with new arms limitation treaties. In particular, countries should agree on sets of technologies that destabilize the nuclear balance that they will not pursue. This agreement may be possible because developing such technologies can undermine the interests of all, especially when adversaries would then feel forced into "launch on warning" postures and developing such technologies themselves. When they cannot agree on technologies to foreswear, they should turn again to non-proliferation and confidence-building measures.[36]

Biological Engineering—Dispersing the Capacity to Cause Catastrophe

Biologists have played the leading role in regulating gene editing themselves, beginning with the Asilomar Conference in 1975. But the tools of genetic modification are becoming ever more widely dispersed among biologists and students of biology. This dispersion gives ever-increasing numbers of individuals the means to cause catastrophic events. To see why, consider the achievement of Dutch researchers in 2012. They demonstrated that the deadly H5N1 virus could be made far more contagious using relatively common laboratory techniques.[37] The modifications made the virus less deadly, but its release could kill large numbers around the world, potentially disrupting supply chains and causing further catastrophe.

Today, the tools of genetic modification are even more sophisticated and less expensive than a decade ago.[38] CRISPR-Cas9 techniques of gene editing are now widely used around the world and even taught in some high schools.[39] Most, but not all, companies that "print" DNA sequences to order follow voluntary industry guidelines to screen customers for nefarious motives

[36] Michael Horowitz and Paul Scharre, "AI and International Stability: Risks and Confidence-Building Measures," *Center for a New American Security Report,* 2021.
[37] Sander Herfst, "Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets," Science, vol. 336, no. 6088 (2012), pp. 1534–41.
[38] Jennifer Doudna and Samuel H Sternberg, *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution,* (New York, NY: Houghton Mifflin Harcourt, 2017).
[39] See, for instance, "CUT! How Does CRISPR Work?" Science Buddies, https://www.sciencebuddies.org/teacher-resources/lesson-plans/how-does-CRISPR-work.

and requested sequences for known pathogens.[40] In the future, however, it is likely that the printing technology will become less expensive and more widely available. Thus, even more individuals will have the opportunity to create a catastrophe by creating a new, deadly pathogen or just using one of the known pathogens whose genomes have been published, such as smallpox, which killed more than 300 million people in the twentieth century.[41] A complicating factor, of course, is that these same technologies can also be of enormous benefit to humanity.[42]

These technologies exemplify a broader trend: the increasing power of technology to be used for great good or great harm by ever larger numbers of individuals acting alone or in ever-smaller groups. One might argue that during the Cold War, there were only a few individuals who could cause global-scale catastrophes: the leaders of the nuclear states. Today, because of the diffusion of biological capabilities, there are many individuals with that potential.[43] But cyber capabilities and other technologies also appear to enable smaller numbers of people to do greater harm.

These dynamics require governance solutions too, and in an increasingly interconnected world, internationally coordinated solutions will be more effective. In the case of biological engineering, a place to start is with a mandatory global regulatory regime for biological engineering as a first layer of defense. If most companies comply with US government guidance voluntarily,[44] but not all do, nefarious actors can simply funnel projects through the few that do not. This loophole will require capacity building and coordination with local authorities around the world. A next layer of defense is testing for pathogens in populations and putting rapid quarantine and healthcare responses in place. Rapid responses will require both governance

[40] R. A. Leo Elworth et al., "Synthetic DNA and Biosecurity: Nuances of Predicting Pathogenicity and the Impetus for Novel Computational Approaches for Screening Oligonucleotides," *PLoS Pathog*, vol. 16, no. 8 (2020).
[41] Laura Spinney, "Smallpox and Other Viruses Plagued Humans Much Earlier than Suspected," *Nature*, vol. 584, (2020), pp. 30–32.
[42] Science & Tech Spotlight: Genomic Sequencing of Infectious Pathogens, *Government Accountability Office Report,* March 30, 2021.
[43] Rob Reid, "How synthetic biology could wipe out humanity -- and how we can stop it, TED Talk 2019, https://www.ted.com/talks/rob_reid_how_synthetic_biology_could_wipe_out_humanity_and_how_we_can_stop_it/reading-list?referrer=playlist-itunes_podcasts_health.
[44] Voluntary US Government guidance for companies involved in gene synthesis includes screening sequences in requestors' orders for known pathogens and screening requestors for malicious motives. See, Public Health Emergency, PHE.Gov, https://www.phe.gov/preparedness/legal/guidance/syndna/documents/syndna-guidance.pdf and https://genesynthesisconsortium.org/.

and technology solutions. Efforts in this direction are already under way—for instance, through the Broad Institute's Sentinel program, which is currently deployed in West Africa.[45] These efforts need to be developed and scaled quickly. Once they are, public health responses must be coordinated. Together, these steps hold out hope for severely limiting both manmade and naturally occurring pandemics.[46] Accomplishing these goals will not be easy, requiring public-private partnerships, overcoming underinvestment due to collective action problems, finding solutions to privacy concerns, and necessitating cross-national agreements.

Looking Ahead

At the dawn of the new era, nuclear weapons seemed to pose a unique challenge. The scale of their destructiveness demanded radical thinking about governance solutions. But if we could only manage that problem, it seemed, humanity's future would be bright. Unfortunately, it is becoming ever clearer that the invention of nuclear weapons was only the beginning. Technologies with similar capacities for destruction and for other profoundly negative, and positive, effects on societies are springing up. We must expect them to continue to appear at an increasing rate. The governance challenges presented by nuclear weapons were not the hurdle to be overcome but the beginning of a dangerous and impactful period in the history of mankind.

The philosopher Derek Parfit argued that we are coming to the "hinge of history."[47] He meant that we are at a moment where mistakes lead to destruction and the right choices lead to grand futures. But this is uncertain and appears less likely today.[48] While the challenges we face are great, we have little reason to expect smooth sailing once they are overcome. There are other clouds on the horizon.

[45] David Cameron, "Scientific Coalition Developing Surveillance System for Detecting Emerging Pandemics in Real-Time." https://www.broadinstitute.org/news/scientific-coalition-developing-surveillance-system-detecting-emerging-pandemics-real-time, 2020.
[46] For an insightful discussion of these issues, see the 80,000 Hours Podcast interview with Jaime Yassif, https://80000hours.org/podcast/episodes/jaime-yassif-safeguarding-bioscience/.
[47] Derek Parfit, *On What Matters*, vol. 2 (Oxford, UK: Oxford University Press, 2011), p. 616.
[48] William MacAskill, "Are We Living at the Hinge of History?" *GPI Working Paper,* no. 12 (2020).

Thus, what we require today is not a single solution for a single existing technology, or even a set of solutions for a set of technologies. Rather, we must reorient our expectations about the governance solutions required by technological change from those of the early nuclear era.

We must develop the means of quickly addressing the governance challenges of new technologies arising one after another with no expectation that a current challenge will be the last. We must constitute highly capable, resourced, standing international bodies to monitor emerging technologies and propose governance responses that operate hand-in-glove with new technological responses.

In debates about the response to innovation, we should not fall into the trap of two sides calling each other complacent and alarmist. Both framings ask which scenario is more likely, but this is usually the wrong question. Debates over the governance response to lethal autonomous weapons, for instance, are sometimes framed in these terms. Proponents of aggressive attempts to slow their spread argue that they will revolutionize political order. Proponents of a more measured approach to governance note that governments will develop countermeasures that stand a strong chance of preventing this. But a better question to ask is how confident we can be that countermeasures would be effective—that lethal autonomous technologies would not upend the political order. Even if countermeasures are likely to be effective, there is still a strong argument for restricting the spread of technology until we know for sure. In this context and in others, we need to ask not just how likely a scenario is, but whether the risks are sufficient to justify actions to mitigate them. Often the answer will be yes, even when the likelihoods are relatively small. Uncertainty about the effects of technologies should often lead to a bias for the status quo, especially when a horse, genie, or devil, once summoned, cannot be returned from whence it came. Otherwise, we may say with environmentalist Rachel Carson: "All this has been risked — for what?" As she noted, "Future historians may well be amazed by our distorted sense of proportion."[49]

The approach to governance must be tailored to the technology. Some can be banned outright. As with chemical and biological weapons, bans can be useful even when they are violated. Others can be restricted to smaller clubs, as nuclear weapons have been. In some cases, it may be important to anticipate technologies—to regulate their development even before they exist. This anticipation applies to the most transformative and potentially destructive technologies, such as diffused gene printing. Sometimes, international regulation and agreements will be most effective; sometimes reciprocal domestic legislation will be. At times, combinations of institutions and norms

[49] Rachel Carson, Rachel. *Silent Spring.* (Boston, MA: Houghton Mifflin, 2002), p. 8.

will be required. It took decades for a global norm of no-first-use of nuclear weapons to emerge.[50] Such processes will need to be accelerated through careful consideration of the best governance solutions in each case and institutions with the credibility to demand the attention of the international community at the highest levels.

In the early days of the nuclear revolution, world leaders discussed significant changes to world governance structures in response to challenges posed by technology, including limitations on sovereignty and strict weapons development monitoring in collective security regimes. Our world has gotten by without these for seventy years, but challenges are multiplying and interacting. It may be that discussions that approach that scale should begin again.

Robert F. Trager is a Professor of International Relations at the University of California at Los Angeles and a Senior Researcher at the Centre for the Governance of Artificial Intelligence.

[50] Richard Hanania, "Tracing the Development of the Nuclear Taboo: The Eisenhower Administration and Four Crises in East Asia," *Journal of Cold War Studies,* vol. 19, no. 2 (2017), 43–83.