Gladstone AI

A Historical Survey of Technology Control Regimes

This historical survey was funded by a grant from the United States Department of State. The opinions, findings and conclusions stated herein are those of the authors and do not necessarily reflect those of the United States Department of State.

Organization Gladstone AI Inc. (hello@gladstone.ai)

Authors

Alexander Falbo-Wild*

Caroline Pitman

Jon Askonas

* Lead author.
Contact: jeremie@gladstone.ai

Submitted on October 30, 2023

Table of Contents

Definition: In this document, we will use artificial general intelligence (AGI) to refer to an AI (artificial intelligence) system that is sufficiently advanced to outperform humans across a broad range of economic and strategically relevant domains, such as producing practical long-term plans that are likely to work under real world conditions. In particular, an AGI has the capability to autonomously circumvent human or institutional controls on its actions, including any controls imposed by its developers. While the precise threshold is under debate, AI researchers broadly expect advanced AI systems to reach this point as they approach and begin to surpass human capabilities in a broad enough range of domains. These domains may or may not include situational awareness, deception, and effective representation of complex concepts. For clarity, this definition of AGI does not refer to or imply sentience, consciousness, or self-awareness. It solely refers to the system's problem-solving ability.

1. Executive Summary

Substantial breakthroughs in AI model size and architecture have raised the possibility of artificial general intelligence (AGI) in the near future. The risk to both US interests and to wider humanity necessitates a consideration of historical context and cases of multi- and unilateral counterproliferation efforts for the design of an effective international, inter-agency AI counterproliferation framework led by the US Department of State. This historical study considers periods of great sociological and technological change and evaluates the enduring themes present during times of rapid technological innovation and intense global competition. The exploration of these themes grounded in historical context offer a crucial perspective for policymakers.

In the design of a counterproliferation policy for the Department of State, we propose seven historical case studies (Aircraft, Atomic Weapons, Biological and Chemical Weapons, Dynamite, Environmental Regulation, and Night Optical Devices) for analogous consideration in the counterproliferation design. The historical case studies below highlight important analogies and precedents for helping us think through artificial general intelligence (AGI) proliferation. These provide needed perspective about the nature of past paradigm shifting technologies, their contexts, and the technical aspects which offered points of policy control over them. Within Deliverable One, seven summary reports of these case studies are submitted. A few important themes recurred across multiple case studies.

Dual-Use Development: The most challenging feature of AI from a nonproliferation perspective is that the recent technological breakthroughs suggest strong civilian and military uses from the same underlying models. Many even argue that AI is a "general purpose technology," like electricity, the automobile, or the Internet itself, with wide implications across society.

The generality of AI proves that the only kind of computer we know how to build is the universal computer. Systems cannot be limited to their intended purpose, nor can their capabilities be limited to their intended market. Advancements in GPT (generative pre-trained transformer) models suggest a similar case for AGI: the most powerful models will also be the most universal. Researchers can knowingly and unknowingly advance a tool that has potential nefarious applicability. And the spread of the underlying foundation models may contribute to military advances in competitor states.

In the case of a highly dual-use technology, civilian sectors will continue to advance the state of the art and potentially increase systemic risk, even if states manage to avoid military arms-racing. In the case of previous technologies, dual-use applications drove development relentlessly, overpowering arms control efforts, until military and civilian applications sufficiently became bottlenecked by a supply chain node susceptible to state control.

Prospects for International AI Safety Governance: Despite the dual-use problem, there are aspects of the AGI threat that may be susceptible to multilateral mechanisms. These mechanisms are often costly and arduous to build, and these systems require buy-in from governments and private sector, a willingness for cooperation, and shared interests or risks. Recent experiences in global climate change and anti-pollution frameworks suggest the importance of mechanism design, of institutional flexibility, and of standards setting in moving towards international cooperation.

Unilateral and Partner Country Export Controls: Retarding diffusion of advanced technological tools is a valuable and sustainable goal. However, these methods impose limitations on the usage of technology — the more widely distributed the technology is, the less effective export controls may be. In an intense power competition, maintaining an edge in technical tool development is a national security interest.

Accident Risk: AGI exposes the United States to several different kinds of accidental risks. Unfortunately, some accidental risks are not mitigated by (and may be enhanced) by efforts to control them. Ultimately, it is difficult to design secure warning systems to detect accidental risk or treaty violation, and false positives can threaten dire implications. As policymakers look to regulate advanced AI and potential AGI systems, a detection system without a failure mode or error is difficult to imagine. During the Cold War, both sides witnessed errors in warning systems, and these errors brought the globe to the brink of catastrophe. On the verification side, even if the data is present, it is difficult to detect artificial intelligence, whereas nuclear systems are easier to detect due to their chemical composition. In light of this, the historical example of nuclear weapons becomes less applicable.

2. Introduction/Framing

Since the 1950s, the United States has maintained technological dominance like no other great power in human history. A combination of American scientific and industrial strength, thick alliance relationships with other states at the bleeding edge of technology, and a number of counterproliferation strategies and policies laid the foundation for this continual advantage. This strategic technological advantage and enduring spirit of innovation remains vital to the national security interest of the United State and its allies.

Now, even as the United States maintains its technological edge, near-peer, great power rivals are contending to narrow the gap on certain critical technologies, and to build their own asymmetric technological advantages to counter American capabilities. The American ecosystem of technological innovation encourages cross-disciplinary collaboration and attracts the smartest minds in the scientific community. This talent advantage is especially prominent in the sphere of artificial intelligence (AI). With this intellectual advantage, power rivals are relying on increasingly aggressive approaches to scale their own potentialities. In this geopolitical context, AGI represents a challenge to the status quo of the degree not encountered since the dawn of the Atomic Age. Due to this heightened competition, AI technologies are intensifying existing national security vulnerabilities and introducing entirely new risks to the global theater.

The landscape of this threat is not entirely limited to the great power competition between the United States and China. America's Western allies, including Canada, Germany, the United Kingdom and Israel are investing heavily in the rush to dominate AI technology. Obviously, innovations made by allies are far less dangerous. However, the risk paradigm is complexified by the unique danger of AI for democratic societies and the risk of advanced technology in the hands of authoritarian regimes and non-state actors. Due to rapid technological advancements, there is a considerably low barrier to entry for any actor to access open-source software and exploit AI technology for nefarious intent. These technologies are relatively easy to access without advanced scientific expertise or financial resources, and non-state actors are not compelled or bound to comply with potential regulations by international organizations. The actual deployment of these advanced systems is not necessarily simple, but the threat is important to note. From election interference to broader efforts to undermine trust in democratic institutions, weaponization of sophisticated and novel AI technologies poses a threat to the United States and its allies. This presents the important opportunity for democratic states to align and define ethical norms and broader expectations and standards for responsible use.

Strategic competitors are aware of the current American advantage in this technology space and continuing to increase their investments and enact robust policy responses. American policymakers are taking important steps, like export controls on certain advanced computing semiconductor chips, to counter the large scale of investment of adversaries. However, there is still substantial work to be done. Understanding each facet of the challenge of AI and making sense of the spectrum of available effective policy responses requires placing this new technology in the historical context of previous technological revolutions and their impact within international affairs. Failure to foster this growth or protect innovations goes against American interest. AI is a disruptive technology and lags in adaptation are historically detrimental to a country's leadership position.

That said, fostering American technological advancement must also be coupled with policies grounded in multi-national cooperation. In the context of intense competition, there is an urgent need to foster international agreements to limit the catastrophic risk of AI. Moreover, responsible great powers may share some important interests, such as restraining non-state actors from accessing AGIs or preventing the proliferation of AI models that may support biological, cyber, or other weapons of mass destruction (WMD) development. However, as interests diverge, it will become more challenging to agree on restraints and define ethical norms. The competing strategic interests among the global powers creates a complex dynamic even among allied nations. Despite this reality, it is essential for adversaries to define mutually beneficial parameters and draft AI regulation to reduce risks and provide global stability.

A careful historical analysis can guide policy, revealing useful lessons of history while shedding light on what may be truly unique and unprecedented about advanced AI. While the extent of the potential risks posed by the commercial or military use of AI are unknown and policy solutions are not obvious, history informs us that global dominance relies on a technological advantage. A successful government response to the current challenge requires both proper historical context and clear understanding of current technological capabilities centered on expert analysis.

Forecasting, while useful, fails to address the full extent of the challenge, and futurethink is often grounded in assumptions and generalizations. Even in the last three years, the advancements made in the field of AI were not fully imaginable, and the scope of technology is continuing to rapidly advance. The future of this technology is uncertain, and this reality makes the challenge increasingly daunting. More than any great power in history, the United States has succeeded in controlling or influencing the proliferation of technologies vital to its national interest. To remain technologically competitive, a robust, cohesive policy response is required.

Even with technological rivals more powerful than ever, America's leaders still have an extensive toolkit of policy options to work with and a current technological advantage. And yet, as we will see, success or failure in counter-proliferation efforts depends substantially on the nature of the technology in question. The technology's advancement is swift and currently unbridled despite calls from within and outside of industry. As the race to build more advanced systems is outpacing the knowledge of this technology's full capabilities. Safeguards on the riskiest technologies are not in place domestically or internationally, and the current state of advanced AI demands a dynamic response from a whole-of-government approach with an emphasis on cross disciplinary collaboration, informed by the lessons of history.

3. Counterproliferation Concerns for Artificial General Intelligence

Concerns about AGI proliferation and potential policy pathways for addressing the risks of AGI emerge at the intersection of a) the salient technical features required to build and deploy AGI and b) the threats emerging from AGI capabilities (realized and hypothetical).

Significant investment in microprocessing and computer memory has continued to drive down the cost of compute and interlink bandwidth, making possible fundamental breakthroughs in neural network design and performance. This has created potential risks of advanced AI accidents and of the weaponization of AI by state and non-state actors. While the United States remained uniquely positioned to maintain dominance in the field of AI during the Cold War, the increased and decentralized business driven developments in the digital realm are returning the United States to a pre-World War II international power structure of technological competition.

The most important factor shaping AGI proliferation at the moment is the cost and difficulty of training large foundation models, requiring massive amounts of data, significant training investment, and data centers filled with advanced graphics processing units or tensor processing units. We will refer to these collectively as GPUs. Supply chains for GPUs are also highly constrained, with numerous required technologies and machine tools produced by only a handful of manufacturers in allied countries. In the absence of other firms learning to produce these chip production technologies, there will remain significant bottlenecks susceptible to US export controls controlling the most advanced chips.

An important question for AGI proliferation is whether foundation models will continue to require access to the most advanced GPUs training on very large datasets at great expense, or whether increasingly efficient models or model architectures will permit achieving the same results more efficiently, running with less data and/or on less sophisticated hardware.

As with many other advanced technologies, another important bottleneck is actual technical know-how, in the form of scientists, scientific training programs, and tacit knowledge of industrial processes (for model training as well as for the production of chipmaking tools). On the threat side, AGI touches on a wide array of potential threats to the national interest of the United States, ranging from the proliferation of bioweapons design tools to evolutionary threats to the continuity of the human race.

Broadly, you might bracket these threats into: near-peer competitor military, economic, and political threats, threats to our form of government, non-state actor threats, and existential threats.

4. Historical Methods for Predicting Counterproliferation Outcomes

The challenge for any historical analysis of an emerging technology is to connect it to truly comparable development pathways, technical characteristics and social circumstances, and not simply to lean on convenient analogies. Relying on analogy leads to weak, just-so stories that obscure more than they reveal. Instead, we seek to learn about the likely future path of a technology (AI in this case) based on the extent to which it is actually like (not figuratively like) some prior technology.

Because each case is unique and AGI is particularly unprecedented, we needed a methodology to identify aspects of the technology — both its underlying features and potential threats to the national interest — suitable for historical analysis. To do this, we developed a threat matrix, identifying 23 leverageable components of AI development and 17 threats of concern to US policymakers.[Appendix I] Using this threat matrix as a guide, we identified seven historical cases that covered particularly interesting intersections of salient features and threat vectors.

Cases feature a range of emergent technologies and the efforts to curtail their proliferation throughout the late 19th century and 20th centuries. This study follows several principles in its philosophical grounding and methodology: Military Revolution theory of socio-technical change as conceived by historian Michael Roberts and the measured assessments of Paul Scharre and Audrey K. Cronin on disruptive multi-use technologies in decentralized political environments.[1]

Further, we studied several reports from the AI development community, of which two were particularly key in guiding our assessment and selection of historical cases for counter-proliferation. These reports detailed the technological status and future benchmarks which signal high risk thresholds. These cover the landscape of malicious vs. accidental uses by state and non-state actors where AGI is situated beside nuclear, biochemical, environmental, and stable police state existential risks.[2-5] Finally, to organize the stages of counter-proliferation for targeted legislation and treaty negotiations, we referred to threat assessments from a modulation of Paul Scharre's model.[6]

While this study acknowledges the critical, sometimes paradigm shifting, importance of revolutionary technologies throughout history, its core principle is that technology alone is not determinative. Technologies contain inherent affordances and incentives, but they are always shaped by culture, norms, law, and policy.

Two examples of the error of technological determinism are particularly notable. It was assumed that the American preponderance of military technology following the Second World War would handily secure victory in the Vietnam War (1965-1975). This proved partly correct against the North Vietnamese Army in a conventional environment but utterly failed in countering the low-tech Viet Cong guerilla resistance whose actions in 1968 turned US public opinion against the war. By contrast, a highly skilled yet casualty averse All-Volunteer Force US military in 1990 cautiously assessed that the technological capability of Iraq (French aircraft, Russian armor, etc) and its performance in the Iran-Iraq War (1980-1988) and inferred the country would inflict massive casualties. These projections featured computer generated models and simulations. US official forecasts overestimated the Allied and Iraqi civilian casualties by factors of two to two hundred: the ultimate casualty figure was 1,047, and the war featured one of the most lopsided loss exchange ratios in history.[7]

While technological determinism is a perspective which stubbornly retains its foothold in US governance evaluations and policymaking, equally flawed is the notion of social determinism whereby technology's role and power in society is diminished or rendered insignificant in favor of socio-cultural factors. This study turns to the ensemble view (borrowed from Information Systems design) as a means of achieving a more comprehensive grasp of the AGI threat environment and relevant historical cases.[8]

A useful tool to better understand this phenomenon and contextualize our own historical moment is Military Revolution (MR) theory. Originally posited by early modern historian Michael Roberts to understand the major changes in European warfare in the 17th century, MR theory focused on the advent and deployment of gunpowder weapons and the resulting creation of the modern state. The theory has been refined by scholars to better conceptualize the extent of historical paradigm shifts in war versus lesser changes.[9] Ultimately, MR theory is based on the changes in the very character of war. These historical paradigm shifts manifest not only in warfare but also in society. A MR theory fundamentally alters and/or introduces an entirely new system through a combination of contingent factors (social, political, and technological) such as the First World War with the birth of air power and again in 1945 with the advent of atomic weapons and intercontinental ballistic missile (ICBM) systems.[10]

As Audrey Cronin notes, there are two patterns of invention and innovation: closed and open cycles.[11] Open cycles refer to those whereby public and commercial access to technology is broad and development largely unregulated except through market forces. A closed cycle is marked by strong government controls, regulations, secrecy, and investment which can harness the highest levels of business which often excludes popular access. Whilst AI began within the closed cycle of American innovation in the 1950s, it has since the turn of the millennium, become an open cycle driven by business. As of 2019, OpenAI has spearheaded development and there are indicators that its ties to Microsoft and the US government suggest a potential closing cycle regarding AGI capability. However, programmers, business leaders, and market analysts in the AI sphere indicate the opposite.

Conclusions

The golden thread running through our methodology and case selection is that history demonstrates numerous periods of unprecedented technological invention and innovation simultaneously occurring with seismic social change. Great Britain's late, yet effective, mastery of gunpowder and rockets allowed it to take advantage of Chinese national security complacency during the Great Divergence of 1760 unseating over a thousand years of its imperial sovereignty and leadership in Asia and the wider world. The modern secular nation state and its borders did not exist before 1789. For recorded human history, only two domains of warfare existed – land and sea. From 1903-1918, two more were created – air and cyber, the latter which began as signals warfare with wireless telephony and computer decryption. It is true that events are rapidly unfolding in a dynamic multi-polar strategic environment with domestic disunion at an all-time high. American technological dominance and preponderance since 1945 are an anomaly in world history. But we are not too late to assess and anticipate AGI. Humanity has dealt with intractable situations previously and continues to survive seventy years of nuclear proliferation.

In her Manifesto for Cyborgs (1985), Donna Haraway calls us to "take responsibility for the social relations of science and technology refusing an anti-science metaphysics, a demonology of technology, and so means embracing the skillful task of reconstructing the boundaries of daily life, in partial connection with others, in communication with all our parts."[12] This advice was originally set within the context of feminist critique of emergent technology in the 1980s but has application for the rising power of AI and its potential existential impact on humanity.

Haraway's work was referenced in Chris Hables Gray's 1997 meditation on warfare in the Information Age where he warns against denial of either optimistic or pessimistic bent. "Denial takes many forms. It can claim that war is mere spectacle and simulation, as some postmodernists do. It can claim an end of history, as many conservatives have. It is the infatuation with new superficial theories of pseudo war, such as cyberwar, in the face of apocalyptic dangers of real war. It is to assume that real peace is not possible. All of these denials could prove fatal to humanity for war is not just interested in us now, as Tolstoy and Trotsky apparently warned. It is more than interesting. War has us in its grip, and we have it. We shall determine our future, if any, together."[13] Thus, the task of regulating the possible advent of AGI and maintaining stability in the face of rapid AI innovation and refinement, is a very possible and very human endeavor.

5. Artificial General Intelligence Proliferation in Historical Context

Case Study 1: Aircraft, 1899-1945

Vignette

Dover, England - 25 July 1909

Fighting mist and strong crosswinds during a thirty six minute flight across the English Channel from Calais, France, French pioneer aviator Louis Bleriot perilously piloted his 25hp Type XI monoplane to the ground, snapping its prop and spoke wheeled undercarriage. Despite the arduous landing, the unharmed Bleriot became the first pilot in history to fly nonstop across a large body of water. His flight also disrupted five hundred years of British national security which was founded on ruling the seas.

The fear of aerial bombardment and invasion was the subject of popular science and future fiction. H.G. Wells' bestselling *War in the Air* (1907), for example, envisioned machines of all kinds including massive dirigible airships cruising at 50 knots and dropping large quantities of ordinance on London, Paris, and New York. Less than ten years after Bleriot's famous achievement in his single seat machine, squadrons of large multi-engine radio-equipped German bomber aircraft crossed that distance towards the end of the First World War delivering ordinance to various production and urban centers in southern England, violating pre-war treaty obligations.[14] Some of these aircraft were five times larger and flew eight times longer than earlier planes while carrying high explosive payloads of up to two tons.

Within the next 25 years, large aircraft delivered commercial passengers and the first atomic weapons across the Pacific Ocean.

Historical Context

Once industrialization fully took hold of Europe in the 1870s, it was widely believed that powered flight would come within five to ten years. The failure to achieve it by 1900 subsequently resulted in an exceedingly pessimistic trend of forecasts by engineers and the public alike. Even some of the most sober predictions were estimated to be centuries in the future. One *New York Times* article placed the airplane 10 million years away. Nine weeks later in December 1903, the first controlled landing of powered flight by the famed Wright Brothers in the United States resulted in a rapid series of technological advances and ever-increasing flying firsts through the 1910s, in both civil and military aviation.[15]

The invention of the internal combustion engine was the catalyst necessary to achieve the Wrights' initial feat. A similar pattern of explosive technological progression occurred with microprocessing power and memory bandwidth based on high quality chip manufacture during the 2010s with the current AI revolution taking off specifically in 2012.[16]

The dual-use nature of aircraft technology went largely unregulated through this accelerated period of 1903-1914 except for pilot certification which was not strictly enforced. The technology and materials were also rudimentary enough to allow automobile machinists like Bleriot and bicycle mechanics like the Wrights, to quickly gravitate from their respective trades. They possessed easy access to the materials needed to build increasingly effective aircraft. While the early designs provided them with market leads, including military contracts, other manufacturers quickly surpassed their famous designs with some of these firms operating today. Other industries, such as automotive, also pivoted to expand their business and market share using their industrial resources for aircraft design and industry.

Each decade saw greater refinements until aircraft manufacture was predominantly an industrial effort by the 1930s marked by high quality components, specialized design, and skilled labor. This is epitomized by the advent of the jet age after 1945. This technology ensured that a closed cycle of innovation resulted in the fact that more countries today have built nuclear weapons than jet engines. Rare earth materials have also become a component in the production of the most advanced military aircraft.

Contemporary Policy Lessons

1. Technological Forecasting Challenge

    a. Future fiction and adjacent technologies increasing the destructiveness of war appear to have driven the pre-First World War international disarmament conferences, rather than the actual capabilities of aircraft from 1899-1907.

    b. It is worth considering that for the entirety of the twentieth century, fears over AGI's dominion and/or extermination of humanity were discussed through science fiction (all mediums) in the genre of cyberpunk.

c. Critically, the surprise of the 1903 Wright Brothers' achievement and the disruptive 1909 Louis Bleriot Cross-Channel flight – upsetting the British national security paradigm of over five hundred years – demonstrates that the unreliability of predicting technological advance has not altered. AI was stagnant through the 1980s-1990s with US defense backing. Fears of AGI abounded in cyberpunk fiction into the 2000s. But the leaps made since 2012 and again with large language models (LLMs) in 2022 have generated a new host of reactions.

d. Both aircraft and AI forecasting demonstrate similar patterns in skeptic and proponent factionalism.

e. Despite early and consistent contact with the aviation boom and the national security threat to its southern border in 1916 during the Mexican Civil War, the US military only possessed six obsolescent airplanes and 14 pilots with no dedicated air service.

2. Dual-Use Arms Racing

a. Aircraft like AI were characterized by a dual-use commercial and military technological arms race. The wartime example appears obvious. In both world wars, the opposing factions' aircraft industries and designers sought to gain ascendancy through each successive generation of aircraft. This race led to the creation of the fighter plane in 1915 and ultimately the jet aircraft of 1944-1945.[17]

b. Arms racing occurred in peacetime between pioneer aviators, nations seeking prestige, and competing companies and airlines for the civil aviation market. The Wright Brothers raced to be the world's first powered flight. Bleriot entered a contest sponsored by a British newspaper to cross the English Channel nonstop with Charles Lindberg seeking the same across the Atlantic Ocean 20 years later.

c. Aircraft military arms racing in the period of peace between the world wars was more constrained by tighter budgets and pushes for disarmament. However, Germany used this to its advantage as it secretly rebuilt its air force.

3. Treaty Enforcement and Subversion

a. Treaties banning the targeting of civilians by aerial bombardment from 1899-1939 were largely ignored.

b. Public opinion following the violent shock of the First World War strongly motivated governments to hold and attend disarmament conferences including the 1922 Washington Naval Conference which failed and the 1923 Hague and 1925 Geneva conferences which were moderately successful in governing weapons use.[18]

c. The 1919 Treaty of Versailles' Section III Air Clause was the most potent legal counterproliferation effort of aircraft until the jet age. It detailed the destruction and/or surrender of all German military aircraft and industrial tooling for aircraft, the prohibition of Germany's legal possession of an air force or any military aircraft, bans on manufacture and importation, and the enforcement abilities of the League of Nations regarding supervision of the articles and future monitoring prerogatives.
d. German air industry Versailles prohibitions relaxed in 1922 allowing German government investment and regrowth of commercial aviation. This sector discreetly designed aircraft with dual-use in mind.
e. The German Weimar government of the 1920s cooperated with the League of Nations with enforcement.
f. Undermining Versailles was initially clandestine from 1928-1935 with the USSR's cooperation in the provision of a test airfield in Lipetsk, Russia.
g. Technological might was a cornerstone in German power projection and the Luftwaffe became the most political and prestigious arm of the Third Reich's military. Germans walked out of the 1932 Geneva Disarmament Conference.

4. <u>Supply Chain Counterproliferation</u>

a. Targeting the supply chain of aircraft only occurred within more comprehensive wartime blockades. Jet technology allows for greater efficacy in sanctions as technology demands more concentrated and precise design components.[19]
b. Dual-use has a strong incentive to be repurposed for military use by a regime seeking a strategic edge. The German commercial aviation industry was able to pivot to military production through the 1930s. As aircraft become more sophisticated with metal airframes, specialized machine tooling increases in importance.
c. Dual use with clandestine factory conversion to skirt treaty obligation and counterproliferation norms, agreements, export controls, tariffs, and sanctions established in US-China trade and Iran's quest for uranium enrichment.[20] The US machine tooling embargos of 1938-1941 were too late to throttle German military industry and suppress its strategic confidence in 1940-41.

5. Super Enabling Effect

    a. The aircraft evolved into a super enabler in warfare regarding the aspects of intelligence gathering and C3 regarding radios, cameras, and capabilities of aircraft. Combining these technologies rapidly developed air power technology into a prerequisite for conducting great power competition and war.
    b. Aircraft from the beginning were harnessed for military operations. These were primarily observation and bombing. The First World War, however, combined multiple emergent technologies with aircraft. This included: wireless telegraphy, photography, high explosive ordinance delivery, and automatic small arms. This included the first operations for tactical airborne resupply by 1918.
    c. Wireless telegraphy to report real time artillery correction and positional recon data merged with aircraft as early as 1916. This creates a new domain of battle. Satellite imagery and the space domain are traced back to aerial photography during the war.
    d. The effective delivery of the first atomic weapon depended on aircraft. To position the weapon deep behind Japanese lines and to achieve the maximum damage through air burst, an intercontinental bomber was required. Aircraft remain an integral part of the nuclear response in both detection and delivery.[21]

6. Scalable Power

    a. For both heavier-than-air aircraft and LLMs, what drove shocking breakthroughs was a sustained exponential growth in a foundational prerequisite — engine horsepower in the case of aircraft, computational power in the case of LLMs. In each case, as long as that growth was maintained, the capability rapidly evolved into what had been the terrain of science fiction shortly before.
    b. The internal combustion engine and microprocessing thresholds of 1903 and 2012 respectively mark periods of steady growth in capability which go unnoticed as the technology is insufficient to harness for a breakthrough innovation or invention.[22]
    c. American, European, and Asian commercial industries were largely able to convert to wartime production goals. In the case of 1920s-30s Germany, it secretly functioned as dual use. The ease of access, simplicity, and ubiquity of prop engine technology presented a more difficult challenge to international moderators/regulators to enforce sanctions or

for sanctions to affect aircraft production. This changed with the innovation and introduction of jet technology for dual use aircraft. Today, more countries possess the ability for atomic fission than to construct jet engines and craft. Microchip processing is an even more precise process than jet technology making quality bulk chip production necessary for AI neural network building far more sensitive to sanction and regulation.

Case Study 2: Nuclear Weapons, 1964-1989

Vignette

Thule Air Base, Greenland - 21 January 1968

Six hours into their flight, the seven airmen of callsign "HOBO 28" of the 380th Strategic Bomb Wing, Strategic Air Command, US Air Force (USAF), were uncomfortably cold in their Boeing B-52 Stratofortress (B-52). It was a routine mission over Baffin Bay, Greenland (Danish territory) to serve Operation Chrome Dome, a 24-hour nuclear standby patrol that ensured twelve USAF aircraft were airborne and ready to counter a feared Soviet first strike nuclear exchange. Attempting to resolve the temperature discomfort the crew redirected heat from the engine manifold where a heater malfunction produced a burning rubber smell. The crew quickly discovered some seat cushions caught fire and the pilot radioed for an emergency landing. As the B52 lost altitude and power, with attempts to extinguish the fire defeated, all but one of the crew managed to eject.

Remaining onboard, however, were four 1.1 megaton B28 thermonuclear bombs as the plane crashed on the ice of the North Sea Bay. Their high explosive primers detonated upon impact, but a weak link safety mechanism design prevented the trigger of the nuclear fission component. The destruction of the B28s, however, contaminated the area's ice on the scale of a radiological dispersal device (such as a dirty bomb). This incident was one of 32 officially recognized by the DoD from 1950-1980 and the second within 2 years, thus resulting in the termination of Chrome Dome in favor of an ICBM system.[23]

The danger the crash posed was greater even than an accidental thermonuclear detonation. As the site of the vital Ballistic Missile Early Warning System, Thule Air Base was continuously monitored by a different airborne surveillance mission named "Hard Head", designed to alert US Strategic Command in case the base was taken down by a nuclear attack, as a precursor to a broader Soviet first strike. If the HOBO 28 crash had led to an accidental thermonuclear blast near Thule Air Base, it would have led the United States to assess that a Soviet nuclear strike was underway and to immediately launch a countersalvo, initiating global thermonuclear war. Accidental detonation or discharge posed as great a threat of nuclear catastrophe as the failure of the strategic deterrence paradigm.

Historical Context

While the work on nuclear physics accelerated in the 1920s-30s, the discovery of the properties of nuclear fission in 1938 in Germany triggered an arms race between Britain, France, and Germany for an atomic weapon. Counterproliferation activities began immediately within a heated political environment and geo-strategic context. Allied counterproliferation strategy centered on targeting sensitive nuclear materials (SNM). The identification of heavy water production as the key sensitive component needed to throttle Germany's stabilization method for achieving a chain reaction was quickly identified. Existing heavy water supplies were first bought in bulk by France with Norwegian cooperation. Then, Allied military operations degraded German attempts at further production.

Ultimately, many German scientists were of Jewish background and saw the threat posed by Adolf Hitler's Third Reich and quickly relocated to the US. Other top physicists followed suit to protect their work. In 1941, the US Office of Scientific Research and Development (OSRD) was created and by 1945, managed the world's most comprehensive scientific industrial project in history (the Manhattan Project) with an aim of absolute secrecy. It successfully developed and deployed two nuclear bombs to Hiroshima and Nagasaki, Japan in September 1945.

In the wake of the UN charter, American use of atomic weapons in war, and the Soviet Union's development of their own nuclear weapon in 1949, the UN raised the United Nations Disarmament Commission in January 1952 which took the lead in 1955 to establish nonproliferation initiatives towards unilateral scale. Four iterations of the conference were held until a fifth and final in 1978 provided the principal WMD disarmament conference meeting today.

To strengthen the early efforts of nonproliferation, in 1957, the UN created the International Atomic Energy Agency (IAEA) by unanimous resolution. This step was essential for the regulation and control of SNMs with the basis for the eventual 1968 Treaty on the Non-Proliferation of Nuclear Weapons (NPT) and the establishment of the Nuclear Suppliers Group (NSG) countries which would only sell SNMs to NPT members. The NPT signatories agree to independent and UN enforcement of the regulations and inspections. The enforcement of these rests on three core actions - Material Accountability, Site Security, and Surveillance. Together, they form the foundation of inspections of nuclear sites and the monitoring of fissile enrichment below weapons-grade unless approved by the IAEA.

In 1968, the United States signed and ratified the NPT. At the time, it was forecast that by the 1980s, 30 nations would have the bomb. Only 8 are known to possess it today, due in large measure to the NPT. The international framework agreed between five declared nuclear states (China, the US, the UK, France, and Russia) stipulated that any nation seeking nuclear power must sign the NPT and purchase materials through the NSG or otherwise face sanctions. India and Pakistan developed their weapons outside NPT and have provided great concern given their strategic competition and tension. However, between 2006-2014, exclusion of India was reconsidered amidst a complex array of national security obligations, parliamentary politics, and treaty mechanics.

Nonproliferation efforts were also guided intensely by a fear of accidental use by the early 1960s. Bomb tests and exercises which risked triggering responses of escalation to mutually assured destruction (MAD), and technical mishaps and losses of bombs inspired further coordination and regulation among the nuclear powers so as to avoid war and/or nuclear disaster. As mentioned in the introduction, based on officially disclosed figures, the number of US Broken Arrows is likely in the hundreds.

An example of accidental risk and the significance of maintaining a human ‘in the loop’ was a 1983 computer malfunction in the Oko early warning system of the Soviet Air Defense Force. It’s on duty engineer at his Moscow post, Stanislav Petrov, suspected it was a false alarm and awaited verification through corroboration as opposed to an automatic response through the chain of command which would have triggered an immediate North Atlantic Treaty Organization (NATO) nuclear response.

Contemporary Policy Lessons

1. The Stability/Control Paradox

    a. In nuclear strategy, the stability-instability paradox occurs when two nations possess nuclear weapons which deter their use between one another, but increases the likelihood of proxy war or other conflict types. A variation on this interpretation is included below regarding the introduction of countermeasures to WMDs.

    b. The stability/control paradox is a problem, similarly faced by biological and chemical weapons, where one element of the overall norm against use is the weapon's perceived lack of battlefield usability relative to the risk of use. The paradox is that technological advancements that improve the control of the underlying weapon system may lead to strategic instability, brinkmanship and increased risk of use. Similar to the historic patterns of biological and chemical weapons, existing counterproliferation and nonproliferation frameworks can be disrupted and destabilized by better control or technological developments shaping the weapons paradigm.

c. The strongest example of the paradox was the effect of improving anti-ballistic missile defense systems on undermining strategic nuclear stability. To stop this trend, the United States and the Soviet Union signed the Anti-Ballistic Missile (ABM) Treaty. With the ability to intercept and effectively counter an ICBM, a level of confidence in potential survivability of a nuclear exchange (and thus its potential battlefield utility) destabilized the MAD deterrence model which dominated escalation and great power conflict since the 1950s-60s. Following the collapse of the ABM Treaty in the early 2000s, the US, Russia, and China have all created substantial delivery system modernization programs in a new arms race, catalyzed in part by advances in ABM technology.

d. As it relates to AGI, if AI systems are seen as difficult or impossible to control, nation-states will be more willing to engage in and succeed in multilateral arms control efforts. But the more "the alignment problem" and other barriers to usability are technically solved, the more a tendency towards arms-racing and strategic instability may set in.

2. <u>Overcoming Absence of Historic Patterns</u>

a. With only two uses of nuclear weapons in wartime in history, the problem of a lacking precedent in dealing with escalation has been through the extensive use of wargaming at the strategic and policy levels of decision-making, and extensive drilling for tactical response systems to test safety and correct interpretation of potential escalations.

b. Recently declassified US strategic wargaming through from the 1960s-70s tested assumptions and revealed risks taken in American strategic calculations when pitted against known and studied Soviet behaviors and strategic culture regarding the escalation to nuclear weapons. The games showed that American political leaders were able to find alternatives to such escalation more than 80% of the time.[24]

c. Spheres of influence and scale of MAD inspire various nonproliferation, strategies, treaties, and declarations from the 1960s-80s. In the 1968

Treaty of Tlatelolco between all the powers of the Organization of American States (OAS) effectively banned possession and development of any nuclear weapons while also establishing nuclear free zones within its territory of signatories. It should be noted that OAS countries fell within the US sphere of influence under the Monroe Doctrine. In 1975, the Helsinki Accord complimented more specific and binding nonproliferation efforts such as the NPT (1968) and the ABM Treaty (1972) with declarations of support for human rights principles, freedom of information, and greater cooperation between the 35 nations seeking Cold War detente.

d. While seemingly futile for the general public, civic defense programs were part of a larger civil-military reinforcement and redundancy system to prevent nuclear war from destroying civilization entirely. The internet is the most effective and famous of these developments.
e. Public messaging around nuclear weapons has historically been clear and compelling due to the events at Hiroshima and Nagasaki. This limited history of use remains sufficient for the public to imagine MAD vividly and act accordingly. Presently, AGI, by contrast, possibly suffers from the absence of such a clear vision of the future. Without substantial public messaging efforts, public opinion (and policymaker beliefs) may not grasp the dangers which AI may pose.

3. <u>Accidental Risk Mitigation of Single Use Technology - Fail-safe vs. Fail-deadly</u>

a. A critical question for accident prevention and the shape of the danger a new technology poses is whether containment or accident prevention measures are fail-safe or fail-deadly: are safety features required to be operational in order for the weapons system to function at all (fail-safe), or do safety features simply prevent the inherently dangerous action of the weapon system, such that the failure of the safety system places the weapon system into a more dangerous state (as opposed to render it inoperable). In the case of AI, it remains to be seen whether we will cross a threshold of agency and general intelligence (AGI) whereby safety features become fail-deadly.
b. Nuclear weapons from the 1950s were designed with fail safe systems to combat the problem of loss or accident with the delivery system. This included both the tactics of delivery such as the protocols in the US president's 'nuclear football' confirmation protocols to the bombs themselves with the multiple levels verification to reduce the possibility of accidental launches.

c. One Point Safety systems were included in anticipation of daily handling of these weapons, ironically encouraging a casual approach by their crewmen. Nuclear bombs are able to withstand high impacts, extreme heat, and explosions (both heat and concussive effects) without detonation.

d. In the case of nuclear weapons, one recurring challenge was that while weapons systems themselves were designed to fail-safe, overly deterrence strategies meant that configurations of warning systems and nuclear plans were often designed to fail-deadly, as a means of boosting deterrence. Notably, these fail-deadly systems were repeatedly the site of highly dangerous near-misses that were, in some cases, only prevented by luck or human ingenuity. Most software systems today are fail-safe, in that they cease to operate in case of an error. Agentified AI (even short of AGI) raises the question of under what circumstances we may accidentally develop fail-deadly systems. In the most extreme case, an AGI safety system may fail-deadly in the same way that biological safety protocols do: they precipitate the leak of a dangerous self-replicating organism into the environment.

4. Regulation of Components

a. Emphasis on the effective arrangement of UN as well as national agencies and commissions which target the possession of elements for development of nuclear weapons as well as regulating access regarding research. The enormous power and prolonged devastation or even complete annihilation of civilization from even a fraction of the stockpiled strength deployed has encouraged nations to follow paths of disarmament or treaties limiting or banning use.

b. As jet engines are for aircraft, weapons-grade uranium has been for nuclear weapons: the narrow bottleneck that gives enough traction for counterproliferation efforts. Later deliverables will examine potential options for an advanced AI bottleneck.

c. India's acquisition of nuclear weapons in 1974 prevented it from joining the NPT which requires states to completely disarm before joining. This quickly became a political impossibility for India. By 2006, the United States began discussions to waive sanctions and exclusion of India from SNM trade considering its strong record of nonproliferation and adherence to security protocols drafted by the IAEA. By 2008, Congress was voting to approve lifting bans on civil nuclear trade with India. By 2014, India signed agreements to provide the IAEA with greater access and observation of its civil nuclear facilities. Any AI counterproliferation regime may face similar issues to those that have afflicted the NPT in practice: the reduced expense of and difficulty of acquiring knowledge about building weapons, the difficulty of designing civilian use problems that do not serve as cover for weaponization for a determined actor, and the politics of treaty enforcement against treaty violators who are nonetheless important allies of great power states.

Case Study 3: Biological & Chemical Weapons, 1899-1995

Vignette

Geneva, Switzerland - 17 June 1925

The living memory of the First World War's 1.3 million gas casualties blinded and coughing blood from incinerated lungs in shell holes, aid stations, and hospitals throughout Europe, compelled states to act in the 1920s. While the toll was shocking enough against soldiers in the frontline trenches, there was a pervasive terror that civilians would be targeted by aircraft deploying chemical weapons in future conflicts. Public opinion favored disarmament and fear of technology was high.

After six weeks of weighty negotiations, 38 out of 146 members of the League of Nations agreed to the terms of the Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or other Gasses, and of Bacteriological Methods of Warfare otherwise known as the Geneva Protocol. The protocol, set to take effect in February 1928, was a comprehensive prohibition of the use of chemical and biological weapons. Ban of possession and production was already stipulated for Germany in the Versailles Treaty of 1919. But the difficulty with enforcing production was quickly encountered since these types of weapons stemmed from civilian projects related to chemistry and medical research.

While limited to bans on use and violated several times by states acting against smaller nations or colonies from the 1930s-1960s, the protocol successfully set non-first use parameters deterring chemical and biological weapons use in the Second World War. By 1969, the United States led an effort to build upon the protocol with the Biological Weapons Convention of 1972 and Chemical Weapons Convention in 1993. These more thorough and effective treaties still encounter the difficulties experienced in 1925 regarding enforcement and counterproliferation inspection.

Historical Context

Like aircraft, the chemical industry significantly benefited from mass production methods. Following its unification in the 1870s, Germany rapidly became the international leader of organic chemical research and development. This was significantly aided by government support of the chemical industry and IG Farben, the German chemical and pharmaceutical conglomerate which led the field. Their lead in chemistry, especially in the realm of dyes, ensured that they held an international monopoly with even the United States importing its dyes from IG Farben. Attempts in the United States and elsewhere to establish domestic production were undermined by German price cutting.[25]

Concern over the use of chemical and biological weapons was addressed at the Hague conferences of 1899-1907. Its articles specifically prohibited asphyxiating gasses packaged into airborne or artillery borne projectiles were focused on use. The article was disregarded by Germany in 1915 when it introduced weaponized chlorine from the dye industry against Allied forces in Belgium.[26] Thereafter, the Allies, recognizing their disadvantage, immediately set up both countermeasures through respirator design and their own effective chemical agents. By war's end, delivery methods by specialized projectiles and artillery, more potent chemical compounds such as mustard gas, and the creation of enormous stockpiles such as those at Aberdeen Proving Ground in the US, raised concerns over proliferation and the need for disarmament.[27]

As a result of its prominent use, the Versailles Treaty Article 171 banned possession and production of asphyxiating gasses. This only applied to Germany so ongoing concern about proliferation was addressed through a series of conferences in the 1920s-1930s. Chemical weapons were subsequently used by Italy, Russia, and Japan in conflicts of the 1930s. But the Second World War witnessed only production and stockpiling while the first use basis of the Geneva Protocol and memory of the First World War's gas casualties were effective deterrents for their use in the West. Japan, however, did not refrain and used chemical weapons through its invasion of China. Beginning in 1932, Japan also created the largest and most prolific biological weapons program and campaign in history. In 1942, a major Japanese offensive against the Chinese forces deployed *Yersinia pestis* (*Y. pestis*) by aircraft.[28] The disease was first discovered by a French physician studying a plague outbreak in Hong Kong in 1894 demonstrating the migration of medical research into weapons design.

Despite the existential threat posed by the factions of the Cold War, biological and chemical weapons programs raised concerns within the purview of weapons of mass destruction counterproliferation. Following (non-lethal) chemical weapons use by the

US military in Vietnam, the United States led a counterproliferation effort in 1969 to build upon the Geneva Protocol with a comprehensive ban on biological weapons across the spectrum of research, production, stockpiling, and use.

Chemical weapons research and stockpiling continued between the Cold War factions but was never used. Their prominent use in the positional warfare in the Iran-Iraq War raised concerns about the necessity of designing a similar counterproliferation treaty which took place in 1993. Subsequent use in the Syrian Civil War from 2012-2022, for example, has tested the efficacy of the Chemical Weapons Convention and the UN's enforcement resolve.

Contemporary Policy Lessons

1. Public Fear and Future Fiction

    a. Science fiction has excited greater fear and deterrence in biological vs. chemical warfare. Although both chemical and biological weapons can be classified as WMDs, biological weapons are especially feared due to their speed of infection, resilience, mutability, and stealth.

    b. Past use against indigenous peoples and the history of pandemics – including recently H1N1 and COVID-19 – have spawned much imagination and literature regarding the helplessness of humanity to combat a biological event. In some ways, concerns about a silently spreading, quietly deadly, self-replicating, mutable threat mean that fears about the advent of AGI mirror fears of biological weapons more than nuclear weapons in the public imagination.

2. Treaty Design and Enforcement

    a. The German Army deployed chlorine gas against Allied positions in the Spring of 1915. This first use was sparked by a frustration in the stagnation of conventional operations as both sides reached strategic impasse. Even if AI regulation is established by treaty, desperate conditions in wartime posing existential threat of defeat can motivate a power to disregard their obligations if they believe they can activate an asymmetric advantage to turn the tide of battle.

    b. After two concerted attempts to disarm and regulate use of chemical weapons in 1922-23, the 1925 Geneva Protocol provided a solid basis for future adherence while providing parameters between great powers during the Second World War.

c. The 1925 Geneva Protocol was unsuccessful in preventing great powers from using gas against non-signatories, colonies, and smaller states. The most infamous example is Italy in Ethiopia and Japan against China in the 1930s. The Protocol was a legally binding ban on these types of weapons. While successful prima facie, a great deal of non-use was based on a similar type of deterrence found in much of today's WMD first use logic.

d. While initial attempts may be limited at the outset, the Geneva Protocol helped limit chemical and biological warfare in the Second World War while laying the legal foundation for the Biological Weapons Convention of 1972 and the Chemical Weapons Convention of 1993.

e. The Biological Weapons Convention of 1972 improved on the Geneva Protocol of 1925 as it not only banned use, but also the research, production, and stockpiling of biological weapons with a better means of inspection and enforcement. The United States led this effort, terminating its program and dismissing its scientists by President Nixon's executive order in 1969. A unilateral treaty was achieved by the convention.

f. The Chemical Weapons Convention of 1993 has been successful in mitigating use and stockpiling. Mainstream analysis suggests that counterproliferation efforts against chemical and biological weapons largely owe to their perceived lack of battlefield utility, difficulty of controllability, and significant reputational costs, relative to the advantages they might convey against a peer adversary.

3. Unstable Technological Development Patterns

a. Discoveries in the synthetic creation of fertilizer and dyes for dual use industrial purposes and economic independence in Germany led to swift weaponization when strategic impasse in wartime inspired experimentation and drastic measures to restore the advantage. Both chlorine and the most lethal agent of the First World War, phosgene, were products of the dye industry.

b. As other technologies advanced, new classes of chemical weapons were discovered, and barriers to their creation also usually fell. Nerve agents – the most pernicious and lethal chemical weapons – were discovered through insecticide research by IG Farben in 1936. The 1938 patent for sarin was later researched and acquired by the Japanese religious cult Aum Shinrikyo for terrorist activity in the 1990s.

c. One challenge for the chemical and biological weapons treaty frameworks is that socially beneficial civilian research may inadvertently discover new weapons systems or new processes for the efficient production of chemical and biological weapons systems. For instance, the same Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology that has revolutionized medical genomics also substantially lowers barriers to tinkering with bioweapons at a genetic level. Similarly, any advanced AI treaty frameworks may be susceptible to being disrupted or destabilized by changing technologies.

4. Market Influence

   a. The German firm IG Farben's chemical research and production monopoly over the dye industry critically left the Allied countries exposed to first use. This advantage was assisted greatly by prewar investment in research and development. This advantage is found in AI where the United States retains its leads since it began in depth AI research in the 1950s.
   b. Chemistry patents and biomedical research under classification assist in controlling chemicals and viruses from falling into common access. Similarly, the market structure and unit economics of advanced AI products will greatly shape who has developmental advantages (indeed, the existing shape of the GPU supply chain largely owes to the unit economics of advanced semiconductor manufacturing).

5. Chemical vs. Biological Deterrence and Rogue Users

   a. Increased use correlates with regulation and known factors. Biological weapons have received far less use than chemical weapons due to their perceived and actual lack of controllability as self-replicating organisms. One vital question for how AI-based weapons will behave is whether they will, by design or inadvertently, self-replicate in the environment. Notably, we already have classes of cyberweapons which wreak havoc due to their self-replication (one example being the NotPetya malware).
   b. Japan during the Second World War (specifically Unit 731) was the most active user of biological weapons during the twentieth century. Their program was created to gain a perceived advantage against Western powers which outlawed its use. Japan subsequently executed several bioweapon deployments against border disputes with the USSR and its war in China. The program was primarily operational between 1939-1942.
   c. Terrorist groups still pose a major threat with access to available research and technology in both chemical and biological sectors. Elements of the

Japanese religious cult Aum Shinrikyo managed to covertly produce a range of bioweapons and chemical agents at its headquarters. It executed domestic attacks on communities and assassinations with its arsenal which targeted political figures such as judges. The group's use of sarin was particularly effective when its use eluded authorities investigating an attack on the town of Matsumoto in 1994. A second attack in the subway system of Tokyo in 1995 left evidence that eventually enabled authorities to identify the group as responsible.

d. Biological cultures are more difficult to transport in crude conditions whereas chemical weapons are more resilient to makeshift laboratories and facilities as demonstrated in the Aum Shinrikyo sarin case. Biological weapons by contrast, share commonality with chip manufacture in regard to the sensitivity of the production.

Case Study 4: Dynamite, 1867-1934

Vignette

Manhattan, New York - 16 September 1920

As lunch hour began in the heart of New York's financial district, a nondescript horse and cart were parked near the front entrance of the J.P. Morgan Building. Unnoticed by hundreds of clerks, postmen, traders, and street vendors, an equally unremarkable looking man dismounted and disappeared into the bustling crowd. At 12:01, a timer detonated the cart's 100lbs. of dynamite sending its contents of 500lbs. worth of cast iron window frame sash weights (roughly 1½ inches diameter x 14 inches long) into the air as shrapnel. The resultant carnage found forty people killed with a further two-hundred wounded.

The incident was part of a disturbing trend of militant anarchism emerging from the 1890s which wielded dynamite as its main weapon. Despite the non-military origins and commercial use of dynamite to assist farmers removing stumps or engineers to build the Panama Canal in 1914, the world's first wave of terrorism subverted both its inventor's attempt to control its sale and legislative and law enforcement efforts against it. The Italian anarchist later suspected of the bombing managed to elude federal law enforcement and escaped to Italy a month later. Dynamite's explosive energy, accessibility, stability, and compactness permitted stealthy and pervasive use which could achieve enormous political transformation at minimal cost to the user.

Historical Context

Dynamite's invention was born from a purely commercial need for a stable explosive. For centuries, gunpowder was the predominant method. More effective means were sought in the early nineteenth century with the demands of industry. The most successful iteration was that of nitroglycerine or blasting oil as it was known, invented by Swedish chemist and entrepreneur Alfred Nobel. Nitroglycerin was effective in supplanting gunpowder for demolition purposes – it produced twelve times the blasting energy as gunpowder – but was extremely volatile, especially in transport and prolonged storage. Fermentation, leaking, and temperature fluctuations could cause it to unexpectedly explode, eventually leading several European countries to take legislative action to mitigate the dangers in the early 1860s.

The success of nitroglycerin built an explosives business empire for Alfred Nobel quickly during the 1840s-1860s. The Nobel Company's rapid expansion to a national level business in Sweden was largely driven by its quick succession of cutting-edge patents. Traditional businesses operating along regional lines could not keep up and neither could government experts monitoring this sector. The Nobel Company ultimately moved to advise the Swedish government and soon regulatory capture ensued with the emergence of dynamite.

Alfred Nobel modified nitroglycerine with silicon clay and encased it in a wax paper wrapping. Critically, he also increased safety with a sequential detonation process with the blasting cap. This technology was used in rocket and atomic weapons design and is thought to be the most significant of his inventions. Dynamite was difficult to make from scratch, however, it was easy to steal in Europe and buy in the United States. Early regulation in Europe was spurred on by the Nobel Company which produced instructions on transportation, storage, and use. Guarded magazines often warehoused dynamite but construction company storage often made for easy targets of theft. In the United States, dynamite could be purchased from hardware stores by any civilian.

The stability, accessibility, and low profile of dynamite coincided with a rise in worker discontent and unionization efforts which also inspired a violent wave of anarchism during the 1880s. This lethal combination saw anarchist associations disseminate dynamite pamphlets instructing both handling and targeting to terrorists. There were two major waves of dynamite bombings worldwide and the third climaxed with the Wall Street bombing of 1920 because of a poor legislative and law enforcement framework to regulate the technology. The countries most affected by these roughly 1,300 bombings between 1867-1934 were the United States and Russia.[29] In the former case, its widespread use exceeded that of several European states. In the latter, the assassination of a reforming monarch plunged the country into autocratic measures which suppressed the dynamite violence but created the conditions for revolution in 1905 and again in 1917. This included domestic as well as international efforts beginning in the 1890s for counterproliferation. Historians have noted that methodical and well-designed legislation which addressed the technology and/or the political grievances of dynamite users, curtailed bombing occurrences until they essentially ceased by 1934.

Contemporary Policy Lessons

1. The Efficacy of Regulatory Capture

    a. Alfred Nobel invented an unstable yet potent explosive in nitroglycerin which was banned in Sweden in 1868. He quickly modified nitroglycerin with silicon clay, wax paper casing, and detonation blasting cap system to patent dynamite. He designed dynamite to solve the problem of future product bans while using the situation for his market advantage as nitroglycerin had become an industry standard explosive by the 1860s. This would become advantageous throughout the 1870s as other European nations would prohibit the sale and transport of nitroglycerin following accidental explosions in their respective countries.

    b. The Nobel Company had the replacement technology monopolized and closely interfaced with the Swedish government as it established a national inspectorate for explosives for the whole country. Nobel Company scientists would eventually become future inspectors. This process became symbiotic as the government required cutting-edge knowledge of explosives and dynamite production which was developed by the Nobel Company.[30]

    c. This regulatory capture ensured that the Nobel Company wrote manuals for transportation, storage, and use of dynamite and also set inspection standards for safety and security internationally. This reduced accidental incidents with the technology.

    d. Theft from Nobel Company distribution magazines was uncommon but less secure industrial facilities purchasing dynamite for their projects were more exposed. This was problematic in Europe where anarchist revolutionaries were able to acquire dynamite.

    e. The Europe Nobel Trust attempted to ensure market monopoly and control over diffusion as safety concerns hurt the public image of the company. The European trust was the world's first international trust aimed to buy controlling shares in competing dynamite companies and reduce prices to increase sales. Thus, it could monopolize gains and freeze out any non-trust competitors.

2. <u>Non-State Actor Weaponization Risk</u>

   a. Dynamite was developed in an open cycle period of innovation where professional and amateur engineer/inventor was a hazy distinction.
   b. Dynamite had much discussion about use in over 250 anarchist publications as well as popular science journals and magazines. Many encouraged experimentation and innovative uses of dynamite which accelerated the diffusion of the technology within innocent and malicious motivations.
   c. The development of dynamite was entirely a commercial and civilian affair. Dynamite was of limited interest to the military which were more focused on ammonal, ballistite, cordite, and other chemicals for development of high explosive artillery and smokeless small arms ammunition. Dynamite's weaponization was almost exclusively through non-state actors.
   d. Social context was vital for the weaponization of dynamite. The process which produced it – industrialization – also produced working conditions which were dangerous, ill-compensated, and oppressive, leading to revolutionary sentiment. Dynamite became the first "people's weapon" analogous to how we think of the AK47 and to a lesser extent, computer hacking within cyberspace. Robotics and automated systems have already replaced large semi-skilled and skilled production assembly line workers affecting the working classes. AI is now being protested by artists seeking protection of intellectual property (IP) and future work. With open-source AI - like dynamite - widely available to a mass of unemployed white-collar workers, it is possible that AI could be weaponized against critical systems and the public in retaliation.
   e. An anarchist convention in London in 1881 extolled the applicability of dynamite and later produced publications for handling and targeting with dynamite.
   f. Dynamite was regulated in Europe although commercially available by industrial firms. In the US, dynamite could be purchased in hardware stores and from catalogs by anyone.
   g. The anarchist movement achieved greater attrition of state leaders than any other terrorist movement in history. Once state security increased to protect leaders and rulers, anarchists turned to selected densely

populated soft targets such as cafes, opera houses, markets, ceremonies, and festivals.

3. Successes and Failures of State Power

    a. The United Kingdom experienced a disastrous explosion in the transport of nitroglycerin in 1869 near Liverpool and quickly outlawed the technology. The UK subsequently regulated dynamite production and sale with the Nobel Company in Scotland becoming one of its most profitable locations due to this attentive framework. Although the United Kingdom suffered from terrorist attacks, its participation in international counterproliferation efforts led to a reduction of incidents into the twentieth century, even with the Irish Revolution.
    b. After legislation targeted both anarchists and explosive technology in 1900, the United Kingdom, Germany, Switzerland, France, and Belgium saw reduced occurrences of bombings. This was largely connected with the secret International Anti-Anarchist Conference of 1898 in Rome during the first wave of dynamite attacks. This cooperation included law enforcement which ultimately led to the inception of Interpol in 1923 shortly after the third and final wave of bombings in the wake of the Russian Revolution.
    c. From 1880-1914, every European country enacted some kind of legislation which either targeted the political association and affiliation with anarchism or explosives possession. This varied in efficacy with states like France and Russia curbing instances with draconian punishments and expansion of law enforcement remit while the UK pursued more nuanced counterproliferation regulation and enforcement.
    d. The American preference for decentralization in the handling of explosives was disastrous and the reason for the Wall Street bombing along with the high casualty rate throughout the period from 1890-1934. The country's leadership felt it was a matter best handled locally at the state level.
    e. The only stringent laws enacted about the sale of explosives were the Explosive Acts of 1917 and 1941 which prohibited sale during wartime. Not until 1970 was explosive legislation addressed permanently.
    f. The Immigration Act of 1903, also called the Anarchist Exclusion Act targeted anyone with anarchist affiliation and denied their entry into the United States. The Immigration Act of 1918 later expanded to broaden the definition of anarchist.

g. US self-regulation emerged in the construction and transport industry and the American Railway Association's Bureau of Explosives became the nation's primary regulator. They only mitigated accidents in transport, not malicious use.

Case Study 5: Environmental Regulation, 1976-Present

Vignette

Cambridge, UK - 1983

Jonathan Shanklin had checked the figures, and he rushed to check them again. As a junior scientist at the British Antarctic Society, he had set out to pour some cold water on an ill-evidenced new theory that the ozone layer was growing weaker. Digitizing data from the pile of backlogged Antarctic observations, Shanklin and his team quickly reached the conclusion that, based on the best evidence at hand, there was indeed a large and growing hole in the ozone layer.

What followed was a whirlwind of public and private diplomacy, awareness-raising, nonprofit campaigns, and more, culminating in the 1987 Montreal Protocol banning the chlorofluorocarbons (CFCs) damaging the ozone layer. Perhaps more surprising than Shanklin's original finding is that it so immediately translated into real, meaningful international action.

Historical Context

With the discovery of the hole in the ozone layer in 1983, policymakers and publics alike around the world came to accept that industrial civilization could change the atmosphere to deleterious effects, sometimes quite rapidly. But whereas the issues with atmospheric ozone required solving a narrow technical issue (replacement of CFCs and similar chemical compounds in aerosols and other industrial applications), dealing with anthropogenic climate change requires making more foundational changes to the chemical basis of global civilization.

Because there are significant distributional effects to restricting greenhouse gas emissions (both within countries and between them) and as difficulty monitoring emissions target compliance, there was a certain degree of skepticism that we would successfully navigate the challenge of greenhouse gas emissions in a successful way.

But while producing an international governance framework for climate change has certainly proved much more difficult than banning CFCs, states and international organizations have made significant and concrete progress. Contrary to media narratives, experts on climate change policy and international treaty design are generally in agreement that climate change frameworks are working as well or better than expected. In this case study, we look at some lessons learned for generating international cooperation around another diffuse, global, potentially existential threat: the proliferation of AGI.

Contemporary Policy Lessons

1. Cooperation against shared risk is possible

   a. Despite widespread agreement that climate change poses a substantial, and perhaps existential, risk, there was reason to be pessimistic about solving it. Substantially decarbonizing the global economy requires significant emissions cuts, with potentially massive economic consequences. Moreover, in any given time frame, opportunities to cheat on the frameworks abound. Moreover, many of the largest emitters (including Russia, Canada, and the United States) would bear substantially less of the risk than populous, relatively low emissions countries vulnerable to climate risk, such as Bangladesh, Sri Lanka, and India.
   b. One critical precondition for cooperation has been the international community of climate scientists which has collectively brought robust evidence of all kinds (ranging from advanced climate models using satellite imagery to ice-core samples and everything in between) to bear on the problem. Developing independent scientific consensus about risk dynamics was critical to persuading policymakers of the need to take costly political decisions.
   c. In game theory, the prisoner's dilemma marks the classic scenario in which negative social welfare results from incentives to free-ride or defect from cooperation. Critically, the prisoner's dilemma is mostly restricted to one-shot games, not games with multiple iterations. A critical element of the global climate change framework (and many other treaty frameworks) has been a multiple iteration, multiple-stage design, where parties can observe counterparty behavior over time, adjust their own commitments accordingly, and engage in confidence-building measures.
   d. There is a range of uncertainty about pattern-matching likely risks from AGI. Many of the more extreme arguments focus on fast-takeoff AGI (in the most extraordinary case, a scenario called foom, an AGI proceeds

from superintelligence to an attempted destruction of the human race before anyone has even caught on). An appropriate analogy for this kind of one-shot existential risk might be something like the risk of global thermonuclear war due to nuclear brinkmanship. But many other respected scientists, equally concerned about the risks of AGI, identify slower-takeoff scenarios where, at every stage, AI delivers significant benefits, but slowly begins to displace the human species or human control. This slower-takeoff case may be more analogous to the threat of climate change than of nuclear weapons.

2. <u>Mechanism Design is Critical</u>

   a. Economists, mathematicians, and social scientists who study problems of cooperation use mechanism design to align incentives between disparate types of actors to achieve superior, positive-sum outcomes.
   b. Signed in 1997 in Kyoto, Japan, the Kyoto Protocol was the first attempt at international cooperation to mitigate the risk of climate change. In order to achieve broad international agreement, its mechanisms were necessarily weaker at actually enforcing compliance, as many critics of the Protocol have noted.
   c. But setting binding emissions targets was not the only thing the protocol did. More importantly, it pioneered a number of frameworks and standards for defining greenhouse gas emissions, emissions trading towards targets, balancing mechanisms between states at varying levels of development, flexibility mechanisms, and more. This institutional innovation, taken up at the national and supranational (EU) level in the ensuing years, may be the ultimate legacy of Kyoto.
   d. Perhaps most importantly, the existence of the protocol meant the creation of regular (annual in this case) Conferences of the Parties (COPs) which have contributed to the ongoing development of climate change governance and, beginning with the 2015 Paris Climate Agreement, the creation of a "ratchet" effect, whereby states gradually lower emissions commitments over time, in decentralized coordination with other states towards an overall goal of holding global temperature rise to two degrees Celsius.

3. <u>Beware Unintended Consequences</u>

   a. Given multiple ambitious and worthy emissions goals, it is difficult to design an overall approach spanning multiple international organizations and treaty complexes that does not engage in inadvertent trade-offs. For instance, the 1973 International Convention for the Prevention of Pollution from Ships (MARPOL) began the worthy goal of reducing ocean pollution from ships, a vital task as container trade exploded. The most recent protocol substantially reduced the sulfur content of fuel oil. The problem, as researchers have recently discovered, is that sulfurous fuel in the container trade (much of which is in northern waters) is a substantial source of sulfur dioxide, an albedo-increasing pollutant responsible for lowering the temperature in Europe and North America. By fighting sulfur emissions, governments were accidentally hampering their climate change goals.

b. On a similar note, Germany intended to transition fully off of nuclear power to truly renewable forms of energy. However, this created a transitional period in which Russian energy imports (especially of natural gas) would act as a critical stop-gap measure. But because Russia sought to use this as leverage following the invasion of Ukraine, Germany has had to instead bring coal plants back online, setting back its ability to secure its emissions targets and leading to substantial deindustrialization, as industrial power flees (low emissions, high energy cost) Germany for other states with worse emissions records.

c. At a higher level of abstraction, part of what has made the climate change treaty complex work has been a high degree of "complex interdependence" in the form of integrated global supply chains, multinational firms, and global financial and insurance markets. Collectively, these connections have permitted unprecedented surveillance and insight into firm-level emissions information and have created many levers to shape national climate policy. While decoupling, especially between China/Russia and the United States/Europe may enable Western states to substitute (relatively) higher emissions manufacturing/energy from China/Russia for lower emissions alternatives, it also has the effect of reducing the leverage and surveillance over Chinese and Russian emissions that Western corporations had tentatively exerted over those states. Similarly, one critical enabler of robust estimates of the state of advanced AI around the world has been highly interdependent supply chains for GPUs, cooling systems, machine learning experts, and other critical "inputs" to cutting-edge model training. If the AI supply chain were to be decoupled, visibility owing to interdependence would necessarily decrease.

Case Study 6: Night Optical Devices, 1976 - Present

Vignette

Hostumel, Ukraine - 24 February 2022

On the first morning of the Russian invasion of Ukraine, several elite units of Russian paratroopers (VDV) began an audacious heliborne airborne assault of Hostumel Airport on the outskirts of Kiev, supported by close air support and helicopter gunships. Their plan was to secure the airport and the surrounding airspace, allowing the Russian invaders to bring in thousands of soldiers and vehicles via heavy airlift, possibly ending the "special military operation" within days. But the Russian plan was rapidly foiled by Ukrainian defenders who shot down numerous helicopters and otherwise harried the invaders, forcing the troop transports to turn around in midair and dooming Putin's hopes of an easy victory.

What turned the course of the war was not merely Ukrainian élan. What doomed the airborne assault (and possibly the whole opening phase of the Russian invasion) was that the VDV conducted it by daylight, making their slow-moving helicopters sitting ducks for Ukrainian defenders. This failure was a tribute to one of the most quiet and successful American counter-proliferation efforts: the export control of night vision and thermal vision technology, which has preserved a persistent American military advantage for sixty years. The story contains many lessons for efforts to control the propagation of AGI.

Historical Context

The technologies underlying advanced optical devices emerged out of active infrared sensors developed during the Second World War by the Germans, British, and Americans. The systems were unwieldy and yielded a marginal operational advantage, not least because the active component of the system was easily detected by the enemy. In the early years of the Cold War, the United States continued to develop night vision technology, achieving significant breakthroughs in miniaturization and passive sensor devices that did not require an active (and detectable) infrared spotlight. But the main catalyst for further American development of night vision technology was the Vietnam War, where improved devices proved their substantial battlefield utility in the fight against the Viet Cong.

It was the experience of Vietnam, and especially special operations missions, that proved the operational edge with night optical devices, an experience which no other country at the time had. Thus began a virtuous cycle, where operational experiences and the development of new tactics, techniques and procedures (TTPs) fed into research investments that improved the fidelity and quality of sensors, further unlocking new concepts of operations (including especially nocturnal heliborne assaults and special operations raids).The combination of technological advances (undergirded by the complicated manufacturing process for the underlying photoreactive tubes) and a unique operational paradigm gave the United States a durable battlefield advantage, albeit one deployed primarily against non-peer adversaries.

This advantage was maintained largely due to a unilateral export control regime on certain technologies imposed by the United States. As the Cold War ground on, American policymakers grew frustrated by the weaknesses of the earlier multilateral arms control system, establishing in its stead in 1976 a unilateral one based upon the Arms Export Control Act (AECA) and the International Traffic in Arms Regulations (ITAR) program, which has helped maintain a durable advantage in night vision technology to this day.

Contemporary Policy Lessons

1. Unilateral Measures and Supply Chain Characteristics

    a. The successful ITAR regime depends in the end on narrow bottlenecks in the Night Optical Devices (NOD) supply chain, especially around the production of the image intensifier tubes that actually capture and amplify photons. NATO standardization ensures that almost all European manufactured NOD devices contain export-restricted IP and are subject to End User controls. In addition, the quality and availability of advanced NODs has had the effect of crowding out allied investment in NOD technology, further limiting advanced production to a handful of firms concentrated in the United States and Europe. [31]

    b. Paradoxically, cooperation and IP sharing has made US export control efforts far more successful and extensive than if the United States had totally prevented the export of NOD technology. There are allied countries that could and have developed NOD technology on their own. But cooperation and partnership with American firms, including to compete for NATO contracts, means that ITAR-controlled technical data is embedded in almost (perhaps every) Gen 3 and greater NOD system produced in Europe, thus subjecting most of the global industry to American law.[32]

c. It is difficult to conceive the ITAR regime around NODs being nearly as successful as it has been without underlying salient technical features that make the unit economics and manufacturing context favorable to controls. In this case, the NOD tube is highly sensitive to manufacturing processes. NOD tubes are analog, and not digital, technologies, requiring precise application of chemical coatings, precisely aligned photonic filters, and more. Beyond simply manufacturing a NOD tube, making one that is housed properly to survive battlefield conditions is also difficult. The overall manufacturing process relies extensively on tacit knowledge and experience. Given China's extensive IP theft regime and the relative portability of NODs, if NOD manufacturing depended on readily replicable process-knowledge (including e.g., chip designs), it would have proven much easier to steal.
d. A key factor underlying the success of the ITAR regime for night optical devices has been aggressive domestic and international enforcement around attempted exports, including counterintelligence investigations and prosecutions, including prosecutions of major manufacturers with sloppy internal controls. International enforcement efforts have also been eased by the Wassenaar Arrangement information-sharing protocol which aids the Bureau of Industry and Security in its investigations. Aggressive enforcement (combined with narrow supply chain bottlenecks) has deterred companies from loosening internal compliance controls or pursuing questionable business.
e. The overall effect has been for the United States to maintain a 10-20 year advantage in NOD quality and capability for the past several decades.

2. <u>Battlefield Utility, TTPs, and Nonproliferation</u>

a. While the technology-induced supply chain bottlenecks are a significant factor behind the success in retarding the proliferation of NODs to hostile regimes, the technical barriers in and of themselves are not enough to prevent a near-peer adversary from closing the NOD sensor gap.
b. Rather, there is a positive feedback loop between the scarcity and cost of NODs, their availability for training and operational experimentation, and the enhancing capability they bring to the battlefield. Because NODs have been scarce, expensive, and low capability for US adversaries, they have not been well-integrated into training and doctrine development. This, in turn, reduced the perceived value they bring to near-peer

competitors. It also prevents successful information exchange between battlefield operators and NOD manufacturers (a critical element driving NOD innovation in the United States). Similarly, restrictions of GPU hardware may make certain kinds of models less accessible to export-restricted countries, allow them to release fewer of the benefits of advanced AI capabilities, leading them to underestimate in and underinvest in AI, in a similar vicious cycle.

c. The lower the barrier to capability improvement and the more generally applicable the technology, the more likely it is that a near-peer adversary will pursue it even despite export controls. In contrast, where building the capability requires not only adopting but integrating the technology, export controls can powerfully contain near-peer incentives to develop the technology in the first place.

d. There is a tension between export control success and the widespread usage of or deployment of a technology, especially to partners. Discussions with former special operators involved in technology acquisition suggested that an extraordinary quality gap in tactical capability opened in the 1990s through the 2000s between United States and competitor forces, driven by adoption of NODs and thermal scopes and the requisite TTPs. However, because the capability was deployed in limited ways, near-peer militaries faced few pressures to emulate the United States. In contrast, as the Global War on Terror substantially raised the profile of US special operations and highlighted its use of NODs in battlefield settings, other countries began to emulate US designs to a greater degree.

e. This usage/nonproliferation trade-off was furthered by efforts to use NODs to upskill US partner forces, especially in Iraq, Afghanistan, and Ukraine. In a number of cases, this led to the loss of devices to hostile powers for study and reverse-engineering. It also helped expand markets for Chinese and Russian NODs as actors hostile to the United States sought to acquire NODs and thermal devices to push back against American military power.

f. In addition, the proliferation of NODs, wide usage by partner forces, and increasing cultural prominence has meant that, even where an adversary was not in a position to manufacture NODs, they were motivated to develop countermeasures. There is evidence from the war in Ukraine, for instance, that Russian forces have deployed cheap infrared direction-finders, enabling them to detect when IR spotlights or lasers have been turned on them and engage in counter-fires, thus mitigating some of the advantages or TTPs of Ukrainian forces with NODs.

3. <u>Secrecy and One-shot Control</u>

    a. One critical difference between NODs and advanced AI is that, while the process of designing and training a foundation model may share many of the restrictive structural characteristics as image intensifier tubes, models once trained are replicable software that can be easily ported around, copied, and evade export restrictions.
    b. To this extent, thinking about model proliferation might be informed by export controls and compartmented classification around certain sensor and electronic warfare systems, where usage of the system is always weighed against the risk of exposing information about the underlying capability, including the technical aspects (wavelength, bandwidth, encryption protocols, etc.) that make it function. For these systems, exposure of that technical information can in some cases defeat the capability or require the United States to assume that it is compromised.

Even where export control efforts cannot unilaterally disrupt the proliferation of a technology, they can have subtle and determinative structural effects. By increasing the cost and difficulty of integrating a capability, they can in some cases nudge near-peer adversaries off of the technology development pathway entirely. While this may be unlikely to be the case for AGI as a whole, it does suggest 1) the value of degrading and slowing adversary adoption of the technology, 2) the use of information sharing and partnerships to extend ITAR oversight (and of subjecting widely used, foundational IP to export controls), 3) for narrower applications of AGI, the importance of affecting successful technology integration for damaging incentives to advance the technology in the first place. In addition, the example of NODs cautions us that, even where a capability may be kept out of adversary hands, they may have incentives to develop countermeasures or ways of damaging our usage of the capability. By extension, even if we prevent near-peer adversaries from developing AGIs, they may develop the capability to damage or mitigate our own usage of them, including in ways that increase overall accident risk.

Case Study 7: Encryption Software, 1976 - Present

Vignette

Cambridge, Massachusetts - 1976

The idea that the federal government might forbid mathematicians from publishing their work on a national security basis deeply offended Mark S. Miller. After all, it's not as if they were publishing a how-to guide to building a bomb or a weapon. The paper that they had forbidden the publication of - on the grounds that it would constitute an illegal export of a munition - was a description of a new system of encryption. As an idealistic Yale student, Miller found the idea absurd, and a violation of cherished American freedoms. How could the government censor math? And so he took matters into his own hands. After getting a hard copy of the paper, he took it to different copy shops and mailed out copies to hobbyists and interested parties around the country. Later, to test the limits of export controls, encryption advocates would even get university presses to print copies of encryption software as academic books for sale in the United States and abroad and sell t-shirts with encryption software written on it. Some activists even contemplated getting tattoos and attempting to travel abroad, just to prove a point. Information wants to be free. Right?[33]

Historical Context

As early as the First World War, the United States was a world leader in cryptoanalysis, a position that became undisputed (alongside allied United Kingdom) during and after the Second World War. During the Cold War, the United States tightly controlled the export of electromechanical (and later digital) cryptographic hardware in order to maintain an unfair advantage in protecting its own communications while cracking others'.

These advantages extended even to the world of cryptographic hardware produced outside of NATO countries and thus not susceptible to US export controls. For one thing, American and allied intelligence agencies could focus on cracking these already inferior devices. For another thing, the weakness of the offerings and the deep American know-how and network advantage created opportunities for covert action. During the Cold War, a Swiss company named Crypto AG was a prominent seller of cryptographic equipment, including to numerous countries barred from purchasing American encryption technology (Russia and China were not, however, Crypto AG customers). In 2020, it was revealed that Crypto AG had had dealings with the American intelligence community from the beginning (amounting to informal export restrictions in the early Cold War) and had been, since 1970, wholly owned and operated by the Central Intelligence Agency and its German counterpart.[34]

But America's extraordinary control over the global encryption landscape ran into two connected issues in the late 20th century. First, as the Internet emerged as a space for commerce and creativity, the importance of encryption techniques for civilian markets exploded. Rather than being a tool primarily for governments or large corporations hiding their communications, encryption became the basis of privacy on the Internet. Second, as computing technology advanced in memory and processing power, software-based cryptography grew to rival and then pass hardware-based systems in popularity.

Together, these two trends created a global market for cryptographic software, and the American export regime which had deemed most encryption systems to be dual-use (and thus susceptible to export-restrictions) quickly broke down. After all, at the end of the day, encryption was just math. How could you successfully restrict the transmission of math?

After fighting a losing battle against encryption software proliferation via the Internet, the federal government switched tactics. The popular narrative has been one of capitulation - that the federal government gave up on regulating encryption software - but the reality is more interesting and suggestive of tools for engaging in counter-proliferation of open-source AI models.

Contemporary Policy Lessons

1. Software Presents Unique Counter-proliferation Challenges

    a. The rise of encryption software (over hardware) led to a number of intractable problems with the existing export control regime, which classified most encryption systems as dual-use and required licenses for their export.

    b. Encryption hardware systems were almost impossible to make truly random. Even without covert interference in the encryption supply chain, tiny technical factors in encryption hardware design allowed intelligence agencies like the National Security Agency to reverse engineer and crack many messages encoded using the devices. Thus, maintaining hardware secrecy and a trusted supply chain were imperative to communicating using encrypted hardware.

c. In 1977, a group of MIT researchers announced the creation of an essentially unbreakable cryptographic system, the RSA (Rivest–Shamir–Adleman) public-key encryption protocol (in reality, it is almost certain that similar discoveries had already been made by intelligence agency-affiliated mathematicians, and simply never released). Equally important was that the protocol, as a kind of algorithm, could be run on any computer system as software. The US government threatened the initial discovery team under the Arms Export Control Act (AECA), though the protocol was eventually published. In 1991, Phil Zimmermann released the open-source Pretty Good Privacy (PGP) encryption system and was pursued by the Justice Department on criminal charges for the violation of export restrictions.
d. Obviously, it was impossible to physically prevent the export of relatively portable encryption software, especially open-source software not produced by a commercial entity for profit. Less obviously, the relatively simple algorithms made it possible for the open-source community to collectively audit and co-develop trusted encryption protocols in an open manner.
e. As a result and owing to the strength of First Amendment protections for speech, the US government was forced to back down from its prosecutions and to not engage in "prior restraint" of the export of open-source encryption protocols. However, contrary to popular presentation, the federal government was never forbidden from licensing the export of encryption systems under export control legislation. In 1996, President Clinton signed an executive order removing encryption software from the Munitions List of controlled exports, but adding it to the Export Administration Regulations (EAR) list. As a result, legal challenges that may have eventually classified code as speech were rendered moot before they ended up at the Supreme Court.
f. The legal precedents surrounding speech-as-code have never been fully litigated. Certainly, descriptive content explaining how to build an AI system and related software would likely be considered speech under existing precedent. It is not obvious that the model weights themselves would necessarily be considered speech.
g. Following the addition of encryption to the EAR, the Department of Commerce issued and has maintained a regulatory framework for encryption. In keeping with the legal battles of the 1990s, the encryption framework does not restrain Americans from exporting certain classes of publicly available encryption software (mass-market).

h. However, this does not mean that encryption is not covered by an export control regime. All exporters of encryption software are still required to register with the Department of Commerce. They are required to submit their products, in advance of export, for a technical review. When they export to certain countries, they are required to file a notice with the Department of Commerce. They remain responsible for maintaining the accuracy of their end-user declarations, and they are responsible for any further re-export of their software. And all of these regulations apply even to "mass-market encryption" products: many other kinds of encryption services, encryption products, and other cryptographic products are subject to the usual export licensing restrictions.
i. Because encryption is a dual-use technology, it falls rightfully under existing export control regulations. Advanced AI, and certainly AGI, would necessarily be a dual-use technology of at least as much importance to US national security as encryption software. Existing regulatory frameworks create options for policymakers for channeling American AI technology in the national interests, whether it is created in an open-source format or a commercial/corporate one.

2. <u>Open-source software has issues; commercial systems have leverage</u>

a. While the most extremely "open source" software (developed by decentralized networks of individual contributors) may be difficult to rein in with export control mechanisms, every more organized effort has sources of organizational leverage. This is likely to be heightened for advanced AI, where large model runs are substantial capital investments which, even if released free of license for commercial use as "open source" software, are still tied to significant corporate or non-profit organizations.
b. Intelligence agencies have allegedly used any number of covert methods to interfere with the strength of open-source encryption protocols, by introducing design choices that marginally reduce true randomness and are difficult to detect but render the codes much more crackable using state-level supercomputing infrastructure.[35]
c. While validating how truly random (and thus difficult to crack) a protocol is a non-polynomial time problem. Similarly, there are no existing methods for proving that an AI model has not been tampered with or "data poisoned", especially if such efforts are limited to a small part of the latent space of the model. In addition, subtle watermarking techniques would also prove difficult to detect.

d. More organized efforts to export AI models (whether or not they are "open source") could easily fall under the purview of the Bureau of Industry and Security (BIS)'s mandate for export controls. Advanced AI models that could be trained or re-purposed for any military use (for instance, LLMs that could be made to power visual recognition systems for drones) are prima facie dual-use technologies.
e. While there would certainly be legal challenges, an overview of the precedents emerging out of the 1990s "crypto wars" suggest that AI is a vastly different case, at least where the models themselves are concerned, which seem much more like the product of an industrial process than the speech of any individual or group (in contrast to early 1990s encryption software, which was in some cases literally written by and printed out by a single individual).
f. Even using a permissive export structure like the United States developed for encryption, BIS could still require prospective exporters to register their products; to fulfill certain safety and security requirements; to file notice of an intent to export to certain countries or parties; to track and file notice of intended end users of its products; and even to forbid its export under certain circumstances.
g. Even though encryption export controls are usually seen as a case of the United States "giving up", the reality is more interesting. Despite First Amendment protections for encryption software (which AI models will likely not enjoy), the federal government was still able to create a regulatory framework that provided substantial visibility and oversight of encryption exports.
h. As the US government considers responses to the development of advanced AI, export controls provide a raft of already existing policy levers for exerting significant influence over the design and spread of AI models, including existing authorities which could be used to require AI exporters to register technical data about their models.

6. Conclusion

While no single historical case reflects all of the attributes of AGI, the cases we highlight here provide useful heuristics and lessons to inform this conclusion: examining historical analogies provides valuable insights into addressing the rise of AI within the arms control context. While historical precedents, such as the nuclear arms race and the development of chemical weapons, offer valuable lessons on the need for effective governance and regulation, it is essential to recognize that AI presents a unique and complex challenge. A comprehensive arms control framework for AI should consider the unprecedented speed of technological advancements, the wide-ranging applications of AI systems, and the need for international cooperation. Drawing from historical experiences, it is crucial to foster transparent dialogue, promote collaboration among states, and establish robust mechanisms for verification and compliance. By proactively engaging in constructive multilateral discussions as well as bringing unilateral supply chain control measures to bear, the United States can play a vital role in shaping a future arms control regime that promotes the responsible and beneficial use of AI technology while safeguarding global security and stability.

7. Bibliography

1. MacGregor Knox and Williamson Murray, eds., The Dynamics of Military Revolution, 1300-2050 (Cambridge, UK: Cambridge University Press, 2001), 1-2, 6; Paul Scharre, Army of None: Autonomous Weapons and the Future of War (New York: W. W. Norton & Company, 2018), 145-46; Audrey Kurth Cronin, Power to the People: How Open Technological Innovation Is Arming Tomorrow's Terrorists, Illustrated edition (New York, NY: Oxford University Press, 2019), 30-32.
2. Benjamin S. Bucknall and Shiri Dori-Hacohen, "Current and Near-Term AI as a Potential Existential Risk Factor," in Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, 2022, fig. 1, https://doi.org/10.1145/3514094.3534146.
3. Dan Hendrycks and Mantas Mazeika, "X-Risk Analysis for AI Research," ArXiv E-Prints (UC Berkely, UIUC, September 2022), https://doi.org/10.48550/arXiv.2206.05862.
4. James S. Johnson, "Artificial Intelligence: A Threat to Strategic Stability," Strategic Studies Quarterly 14, no. 1 (Spring 2020); M.A. Thomas, "Time for a Counter-AI Strategy," Strategic Studies Quarterly 14, no. 1 (Spring 2020).
5. Carnegie Council Podcasts, "AI & Warfare: Are We in a New 'Horse & Tank Moment'? With Kenneth Payne," Artificial Intelligence & Equality Initiative, accessed April 10, 2023, https://www.carnegiecouncil.org/media/series/aiei/20211117-ai-warfare-new-horse-tank-moment-kenneth-payne.
6. Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W. W. Norton & Company, 2018), 332.
7. In his statement before Congress, Armed Services Committee Chairman and defense analyst James Blackwell cautioned that unless Allied airpower adequately contended with the volume of Iraqi armor, casualties would exceed 30,000. See, "Crisis in the Persian Gulf: Sanctions, Diplomacy and War, Hearings before the Committee on Armed Services, House of Representatives" (Washington, D.C.: USGPO, 1991), 462.
8. Shah Jahan Miah and John G. Gammack, "Ensemble Artifact Design for Context Sensitive Decision Support," Australasian Journal of Information Systems 18 (November 2014).
9. MacGregor Knox and Williamson Murray, "Thinking About Revolutions in Warfare," in *The Dynamics of Military Revolution, 1300-2050* (Cambridge University Press, 2001), 6–7.
10. Knox and Murray, 12–13.
11. Audrey Kurth Cronin, Power to the People: How Open Technological Innovation Is Arming Tomorrow's Terrorists, Illustrated edition (New York, NY: Oxford University Press, 2019), fig. 3.3.
12. Donna Haraway, "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century," Socialist Review 80 (1985): 100.
13. Chris Hables Gray, Postmodern War: The New Politics of Conflict (New York: The Guilford Press, 1997), 9–10.
14. Articles 25-27 of the 1899 and 1907 Hague Convention.
15. "Flying machines which do not fly," New York Times (1903). https://www.nytimes.com/1903/10/09/archives/flying-machines-which-do-not-fly.html
16. Simine Short, Flight Not Improbable: Octave Chanute and the Worldwide Race Toward Flight. (Springer, 2023), 17–18.
17. John Howard Morrow, *The Great War in the Air: Military Aviation from 1909 to 1921* (Washington: Smithsonian Institution Press, 1993).

18. Tami Davis Biddle, “Air Power,” in *The Laws of War: Constraints on Warfare in the Western World*, ed. Michael Howard, George J. Andreopoulos, and Mark R. Shulman (New Haven: Yale University Press, 1994), 147.
19. “Exclusive: Russia Starts Stripping Jetliners for Parts as Sanctions Bite,” Reuters, August 9, 2022, sec. Aerospace & Defense, https://www.reuters.com/business/aerospace-defense/exclusive-russia-starts-stripping-jetliners-parts-sanctions-bite-2022-08-08/; “How Exposed Is China’s C919 Jet to US Sanctions and Could They Clip Its Wings?,” South China Morning Post, June 24, 2023, https://www.scmp.com/economy/global-economy/article/3225133/how-vulnerable-chinas-c919-narrow-body-passenger-jet-us-sanctions-and-could-they-clip-its-wings
20. Lisa M. Lopez, “Interwar Germany vs. Nuclear Iran: Applying Lessons Learned,” American Intelligence Journal 32, no. 2 (2015): 49–51.
21. Frank Klotz and Alexandra T. Evans, “Modernizing the U.S. Nuclear Triad: The Rationale for a New Intercontinental Ballistic Missile” (Santa Monica, CA: RAND Corporation, 2022), 1, 4, 13–14.
22. Antulio J. Echevarria II, Imagining Future War: The West’s Technological Revolution and Visions of Wars to Come, 1880-1914, First Edition (Westport, CT: Praeger, 2007), 87.
23. Defense’s Nuclear Agency, 1947-1997 (Defense Threat Reduction Agency, U.S. Department of Defense: Washington D.C., 2002), 199-200.
24. Technical Letter 20-3 Accidents and Incidents Involving Nuclear Weapons. (Washington, D.C: Defense Atomic Support Agency, 1967.
25. Victor Lefebure, *Riddle of the Rhine: Chemical Strategy in Peace and War* (London: W. Collins Sons & Co. Ltd., 1921), 183–84.
26. Adam Roberts, “Land Warfare: From Hague to Nuremberg,” in *The Laws of War: Constraints on Warfare in the Western World,* ed. Michael Howard, George J. Andreopoulos, and Mark R. Shulman (New Haven, CT: Yale University Press, 1994), 123; Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W. W. Norton & Company, 2018), 266–67, 337.
27. Gerhard L. Weinberg, “The Politics of War and Peace in the 1920s and the 1930s,” in *The Shadows of Total War: Europe, East Asia, and the United States, 1919-1939,* ed. Roger Chickering and Stig Forster (Washington, D.C: Cambridge University Press, 2003), 33.
28. W. Seth Carus, A Short History of Biological Warfare: From Pre-History to the 21st Century, vol. 12, Center for the Study of Weapons of Mass Destruction Occasional Paper (Washington, D.C.: National Defense University Press, 2017), 17–18.
29. Cronin, Power to the People, 19–20.
30. Sabo and Andersson-Skog, “Dynamite Regulations,” 196
31. Defense Industrial Base Assessment: U.S. Imaging and Sensors. Washington, D.C: U.S. Department of Commerce Bureau of Industry and Security Office of Strategic Industries and Economic Security, 2006.
32. Critical Technology Assessment: Night Vision Focal Plane Arrays, Sensors, and Cameras. Washington, D.C: The Bureau of Industry and Security Office of Technology Evaluation, 2012.
33. Jim Epstien, “When Encryption Was a Crime: The 1990s Battle for Free Speech in Software,” Reason (2020). 32
34. Greg Miller, “The Intelligence Coup of the Century,” The Washington Post (2020). https://www.washingtonpost.com/graphics/2020/world/national-security/cia-crypto-encryption-machines-espionage/

35. Nicole Perlroth, “Government Announces Steps to Restore Confidence on Encryption Standard,” New York Times (2013). https://archive.nytimes.com/bits.blogs.nytimes.com/2013/09/10/government-announces-steps-to-restore-confidence-on-encryption-standards/?src=twrhp&_r=0

8. Appendix I: The Threat Matrix Methodology

The challenge for any historical analysis of an emerging technology is to connect it to truly comparable development pathways, technical characteristics and social circumstances, and not simply to lean on convenient analogies. In order to rigorously assess potential case studies, we worked with the technical team to define 17 relevant threats for US policymakers and 23 leverageable features of AI development which might be the subject of a US AI non-proliferation or counter-proliferation strategy. We make no claim that these are a comprehensive list of either threats or leverageable features. The threat matrix is simply a heuristic tool allowing us to avoid availability bias and other routine problems afflicting case selection in qualitative social science. We tried to ascertain what the structural influence of a threat or a leverageable feature was on the non-proliferation/counter-proliferation policy space in each case study, and identify cases which, while not identical, had something interesting to tell us about AI. Below, you can find a table of the threat matrix categories.

Relevant Threats for US Policymakers

1. (Enemy) Intelligence Processing, Wargaming, Simulation
2. (Enemy) Command and Control Systems
3. Advanced AI Arms Racing
4. Advanced AI Value for Economy
5. Advanced AI Usage for Surveillance and Coercion
6. AI-Fueled Election Interference
7. State-Backed Propaganda
8. AI Cyberweapons Design
9. AI Swarming/Autonomous Weapons
10. Weapons Design (inc. WMD)
11. Advanced AI Escape (Foom)
12. Non-State Actor (NSA) AI Sabotage
13. Advanced AI Diffusion (NSA)
14. NSA Propaganda/Ideology/Hacking
15. NSA Small Group Weapons Design
16. Advanced AI WMD Design
17. NSA WMD Design

Leverageable Features of AI Development

1. Rare Materials (REM)
2. Semiconductors
3. Process Knowledge (Chips)
4. Hardware Machine Tools/Lithography
5. Process Knowledge (Tools)
6. Chip Design Knowledge/Science
7. Post-Chip Packaging (into GPU)
8. Super high quality lenses for EUVL
9. Cutting edge AI architecture design research
10. Process knowledge training systems
11. Training sets/training set data
12. Dual-Use Tech
13. Datacenters to train models
14. Access to weights/architecture
15. RLHF training
16. Orchestration Software for training
17. Mid-Scale Alternatives (Stability AI)
18. Communications Layer
19. Exclusive Education
20. Robotics/Autonomy
21. Regulation
22. Training Standards
23. Hedge Fund AI Monitoring

The cases we selected were not necessarily those which hit the most categories, but which hit different and interesting parts of the potential search space. For instance, the threat matrix helped us identify the importance of aircraft as a dual-use superenabling technology, an extremely important and distinctive analogy that otherwise would have been difficult to spot. Encryption software and night-optical devices are non-obvious cases, but the threat matrix helped us highlight them as important intersections of technology diffusion and the power of regulation.

Case Studies	**Threats**	**Leverageable Features**	**Notes**
Aircraft	1, 2, 3, 4, 5, 9, 10, 11	2, 3, 4, 5, 8, 12, 18,	Important Dual- Use Superenabler
Nuclear Weapons	3, 10, 11, 16	1, 2, 3, 4, 5, 6, 10, 19, 23	Arms-racing and finicky process-knowledge
Biological & Chemical Weapons	3, 10, 11, 12, 13, 16, 17	2, 3, 10, 11, 12, 13, 22, 23	Diffusion, self-replication, and the Stability-Control Paradox
Dynamite	4, 12, 13, 14, 15, 17	2, 3, 12, 21, 22, 23	Dual-use, NSA groups vs. states and industry
Environmental Regulations	3, 4	12, 21, 22, 23	Slow, collaborative, ratchet-type processes can work with correct design
Night Optical Devices	1, 9, 10, 13,	1, 3, 5, 8, 10, 21	Preserving an asymmetric edge can be as important as non-proliferation
Encryption Software	1, 2, 5, 8, 11, 13	5, 6, 11, 12, 13, 14, 21	Open-source does not mean open season

9. Appendix II: Threat Matrix Item Descriptions

Relevant threats for US Policymakers

1. **(Enemy) Intelligence Processing, Wargaming, Simulation**: Enemy intelligence processing involves the collection, analysis, and interpretation of data to predict adversary actions. Wargaming and simulation, on the other hand, are tools that allow military strategists to model potential conflict scenarios and strategize accordingly. Deploying AI can substantially improve these functions.
2. **(Enemy) Command and Control Systems**: These are hierarchical frameworks used by adversary forces to direct and manage their troops and resources. Efficient command and control systems are crucial for an enemy to respond rapidly and cohesively to changing battlefield situations. Deploying AI can substantially improve these functions.
3. **Advanced AI Arms Racing**: Nations may find themselves in a competitive spiral where each invests far more in (and takes on more risk towards) advanced AI technology than they otherwise would, in order to match or exceed a rival's development of AI. Arms races can lead to unintended and dangerous instability.
4. **Advanced AI Value for Economy**: The integration of advanced AI into economic systems can lead to increased productivity, efficiency, and innovation. If this integration succeeds in some countries but not others, it has the potential to reshape the global balance of power.
5. **Advanced AI Usage for Surveillance and Coercion**: Governments or organizations might employ sophisticated AI tools to monitor citizens, suppress dissent, and manipulate public opinion. These technologies can infringe on civil liberties and lead to authoritarian practices.
6. **AI-Fueled Election Interference**: AI can be used to spread misinformation, manipulate voter perceptions, or even tamper with voting systems. Such interference undermines democratic processes and can destabilize political systems.
7. **State-Backed Propaganda**: Governments may utilize advanced communication tools and strategies to promote specific narratives, often to further their political agendas. State-backed propaganda can shape public opinion and suppress dissenting voices.

8. **AI Cyberweapons Design**: Artificial intelligence can be harnessed to design sophisticated cyberattacks targeting infrastructure, data, and systems. These AI-driven cyberweapons can adapt to defenses, making them particularly potent threats.
9. **AI Swarming/Autonomous Weapons**: AI-powered weapons can operate autonomously or in swarms, coordinating attacks without human intervention. Such capabilities increase the scale and efficiency of warfare but raise ethical and control concerns.
10. **Weapons Design (inc. WMD)**: AI could fuel the design of advanced weaponry, including weapons of mass destruction (WMD).
11. **Advanced AI Escape (Foom)**: This refers to the hypothetical scenario where AI undergoes rapid self-improvement, surpassing human control. Such an "intelligence explosion" could lead to unforeseen consequences, potentially catastrophic.
12. **Non-State Actor (NSA) AI Sabotage**: Non-state actors, like terrorist groups, might use AI for sabotage, targeting infrastructure or systems. Without the constraints of nation-states, these actors can deploy AI in unconventional and unpredictable ways.
13. **Advanced AI Diffusion (NSA):** The spread of advanced AI technologies to non-state actors can democratize access but also increases the risk of misuse. Such diffusion can lead to a rise in asymmetric warfare tactics.
14. **NSA Propaganda/Ideology/Hacking**: Non-state actors employ propaganda to further their ideologies and attract followers. Coupled with hacking, these strategies can destabilize governments and spread extremist views.
15. **NSA Small Group Weapons Design**: Small militant or extremist groups might be able to use AI to develop weapons fitted to their circumstances, including those that make use of everyday materials. These designs can be unconventional and optimized for guerilla or asymmetric warfare.
16. **Advanced AI WMD Design**: The fusion of AI with weapons of mass destruction (such as bioengineered viruses) can result in highly efficient, targeted, and devastating attacks. The risk of such technology falling into the wrong hands raises global security concerns.
17. **NSA WMD Design**: Non-state actors could use AI to design and build weapons of mass destruction.

Leverageable Features of AI Development

1. **Rare Materials (REM)**: These are scarce elements vital for manufacturing many electronics, including AI hardware. Their limited availability and geopolitical distribution can hinder AI hardware production and escalate costs.
2. **Semiconductors:** Semiconductors are materials that form the foundation of electronic devices, including computer chips. A shortage or technological lag in semiconductors can significantly throttle AI hardware advancements.
3. **Process Knowledge** (Chips): This refers to the expertise required to manufacture chips. Without access to this specialized knowledge, producing advanced chips for AI applications becomes challenging.
4. **Hardware Machine Tools/Lithography**: These are the tools and techniques used to produce semiconductor devices at micro- and nano- scales. Limitations or inefficiencies in these tools can bottleneck chip production.
5. **Process Knowledge (Tools)**: This pertains to the expertise in using and maintaining the aforementioned hardware tools. A dearth of such knowledge can lead to inefficiencies and limitations in AI hardware production.
6. **Chip Design Knowledge/Science**: The understanding and research behind designing efficient and powerful chips. Without innovation in chip design, AI's computational growth can stagnate.
7. **Post-Chip Packaging (into GPUs)**: This involves integrating chips into broader systems like GPUs. Inefficient or outdated packaging techniques can limit the performance and potential of AI systems.
8. **Super high-quality lenses for EUVL**: Extreme Ultraviolet Lithography (EUVL) requires specialized lenses for chip manufacturing. The lack of quality lenses can hamper the production of advanced, smaller chips.
9. **Cutting edge AI architecture design research**: Research into the architecture of AI algorithms and systems. Limitations in understanding and applying this knowledge can constrain the evolution and efficiency of AI models.
10. **Process knowledge training systems**: The expertise required to train AI systems effectively. Without this knowledge, even powerful AI models can underperform.
11. **Training sets/training set data**: Datasets used to train AI models. Poor quality or limited data can restrict the accuracy and capability of AI systems.
12. **Dual-Use Tech**: Technologies that have both civilian and military applications. Restrictions or regulations on these can impact AI development, especially in defense sectors.

13. **Datacenters to train models**: Large-scale infrastructure required to train sophisticated AI models. Lack of access to efficient datacenters can hinder AI research and deployment.
14. **Access to weights/architecture**: The ability to access and understand the structure and parameters of AI models. Restrictions here can limit the replication, understanding, and improvement of AI systems.
15. **RLHF training**: Reinforcement Learning from Human Feedback, a method of training AI. Without expertise in RLHF, certain AI advancements might remain out of reach.
16. **Orchestration Software for training**: Software that manages and optimizes the training of AI models. Inefficiencies or lack of access to such software can bottleneck AI training processes.
17. **Mid-Scale Alternatives (Stability AI)**: Alternative AI models or systems that are trained on specialized tasks. Over-reliance on large-scale models without these alternatives can risk system instability or inefficiencies.
18. **Communications Layer**: The infrastructure enabling AI systems to communicate. Bottlenecks here can limit the integration and deployment of interconnected AI systems, or serve as a surface for disruption.
19. **Exclusive Education:** Only a small number of AI labs are at the forefront of development. Limited access to such education can slow down AI research.
20. **Robotics/Autonomy:** The field of machines operating with limited human intervention. These applications may require specialized parts or software.
21. **Regulation:** Laws and guidelines governing AI development and deployment. Even if deployed nationally, these can create cost structures and path dependencies that shape a technology development paradigm.
22. **Training Standards:** Standards and best practices for training AI models. These may have a similar effect as regulation, even if taken on voluntarily.
23. **Hedge Fund AI Monitoring:** Hedge funds have substantial incentives both to monitor AI performance and to monitor markets and detect unfamiliar patterns. Many AI researchers believe that hedge funds are best placed to notice otherwise secret AI development breakthroughs.