JACK D. GORDON INSTITUTE
FOR PUBLIC POLICY

ARTIFICIAL INTELLIGENCE AND CYBER POWER

FOREIGN POLICY IMPLICATIONS

AUTHOR
JOE DEVANNY

ACKNOWLEDGEMENTS

I am very grateful to the team at Florida International University for publishing this paper and to the anonymous reviewers for their helpful feedback. I would like to thank the British Academy for the Innovation Fellowship (2022-23) that made this research possible, as well as Professor Russell Buchan and all the officials working on cyber policy at the Foreign, Commonwealth and Development Office, from whom I learned so much about cyber diplomacy in practice during my fellowship. I would also like to thank Tim Stevens and the wider cyber community – researchers and students – in the Department of War Studies, my home at King’s College London, with whom I have regularly explored these issues in conversations, research, and teaching.

MAY 2024

The Jack D. Gordon Institute for Public Policy, part of FIU's Steven J. Green School for International & Public Affairs was founded in 1985 to establish, promote, and advance the study of public policy and national security studies. The Gordon Institute serves as the forefront of public policy discourse by leading, integrating, and delivering exceptional multidisciplinary education and research while improving the competitiveness and diversity of students and professionals entering the workforce. The Gordon Institute is centered on four pillars: Academics, Professional Education, Research, and Community Outreach, each serving the mission of shaping public policy and national security solutions in the 21st century.

Disclaimer: The views expressed in this report are those of the author and do not necessarily reflect the official policy or position of Florida International University or any other affiliated institutions.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Cyber diplomacy has grown steadily for more than a decade in the context of increasingly urgent international debate about the need for responsible state behavior in cyberspace. As our lives are lived increasingly online or facilitated by digital technologies and the internet—from the services we consume to the critical infrastructure underpinning everyday life—states have also had to consider the darker side of these developments, the vulnerabilities they create, and the threat actors that exploit them. Conversely, as states compete for strategic advantage, they recognize that sovereign cyber capabilities can be instruments to pursue national strategic objectives.

Foreign ministries have had to become more "cyber literate" to meet these challenges, participate in global cyber diplomacy, and contribute international expertise to the intra-governmental (and increasingly whole-of-society) process of shaping national cyber strategies. While all foreign ministries face similar challenges, they do so in very different contexts and with significant variations in capacities and resources. While foreign ministries can learn from each other in this ongoing process of institutional adaptation to new technological challenges, each foreign ministry must also consider how best to ensure any adaptation is appropriate to the national context. The one-size-fits-all approach to preparing a foreign ministry for cyber diplomacy is unlikely to exist.

The field of cyber diplomacy has also contended, as we all have, with the increasing turn in global public policy conversation toward artificial intelligence (AI). The global turn to AI predates but has intensified since the November 2022 launch of ChatGPT—the AI chatbot developed by OpenAI. While this conversation did not start with ChatGPT, the global public sphere has been captivated by the potential implicit in the development of AI, particularly with generative AI. This has added urgency to preexisting debates about regulating and controlling the impact of AI research and development. The intensified focus of executives, legislatures, and other stakeholders has been evident throughout 2023 and 2024 in activities ranging from congressional hearings and preparing executive orders, to the hurried convening of multilateral and multistakeholder summits about AI safety and regulation.

These debates are all connected to concerns about AI's geopolitical impact, including its implications for cyber strategy and operations. Just as foreign ministries must adapt to the rising use of cyber diplomacy, they now must contend with the rising prominence of AI diplomacy. Given the overlap between the two, foreign ministries should reflect on how best to align their AI and cyber diplomacy activities and ensure the foreign policy implications of AI are considered when making and implementing national cyber strategies.

AI is a broad field. Recent debates have focused on the implications of advances in machine learning, especially its generative uses. Accordingly, this is the principal focus of this paper. The paper frames its contribution to these debates by focusing on the institutional role of foreign ministries as governments seek to address four major questions:

(1) To what extent is AI interdependent with cyber power (in short, the ability of a state to achieve objectives in and through cyberspace)? (2) Concerning international security, does AI promise more advantages to attackers or defenders in cyberspace? How does it affect the so-called "offense-defense balance" (if that is a helpful concept in cyberspace), (3) How advances in AI will affect the balance of power in cyberspace, and (4) Perhaps the most pertinent institutional question for foreign ministries is if they should adapt to meet the diplomatic and foreign policy challenges posed by the interdependence of AI and cyber power? Each of these questions merits greater engagement than is possible in this paper. The intention is, therefore, to use this paper as a starting point to further understand the role of foreign ministries in whole-of-government efforts to navigate the implications of AI for cyber strategy.

The paper argues that AI and cyber power are indeed interdependent. This interdependence is a striking example of the wider contemporary trend in which the prism of geopolitical competition affects how states perceive economic interdependence and the need to reduce reliance on competitors and adversaries in science and technology. Although states see the potential for AI to tip the balance in strategic competition, including in cyberspace, it is hard to produce a net assessment of the impact of AI on the competition between "attackers" and "defenders" in cyberspace. Advances in AI offer opportunities to both. The paper argues that early innovators will likely achieve temporary advantages, but insufficient evidence exists to forecast more sustained shifts over time. Given the speed of recent advances in AI—particularly in machine learning—and the range of new uses these advances have created, forecasting future scenarios is a very uncertain business.

Regarding AI's impact on the balance of power in cyberspace, the paper argues that fluctuations are expected in a competitive environment. However, AI advances are more likely to entrench than overturn existing power relationships between states. And while the proliferation of AI tools might lower barriers to entry for less skilled cyber operatives, high-end cyber operations will remain the preserve of the most capable states.

Finally, foreign ministries should continue adapting to emerging technologies in diplomacy and foreign policy. How they do so will look different depending on the national context. As foreign ministries build cohorts of expertise in cyber policy, they should ensure that AI and cyber policy are not pursued in silos. They should invest in (and use effectively) supporting functions such as research and analysis, particularly to understand how AI is integrated into the cyber strategies of adversaries. Some foreign ministries can build these capacities within themselves; others will need to find ways of accessing external expertise. Foreign ministries have a key role in mitigating strategic risks and exploiting the opportunities of AI's impact on cyber power, but they must be organized, resourced, and empowered to do so.

INTRODUCTION

Over the last decade, cyber statecraft—how states cooperate, compete, and come into conflict in, through, and regarding cyberspace—has been increasingly recognized as an important component of national strategy. This is true even if states are not listed on the Belfer Center's Cyber Power Index or have not followed the U.K. government in using cyber power as an explicit framing concept and a trope of strategic communication.[1]

It is clear today that Artificial Intelligence (AI) and cyber power will be increasingly interdependent, which has important implications for developing and implementing foreign policy. For example, the United States and the United Kingdom already identify AI as one of several cyber-critical emerging technologies in their respective national cyber strategies. A national edge in these technologies is widely seen as necessary for global competitiveness.[2] Put simply, no understanding of the future of cyber power is possible—and no national cyber strategy truly effective—without understanding the relationship between AI and cyber power and the implications of that relationship in the pursuit of national strategic objectives.

And yet, as AI creates challenges and opportunities for both defenders and attackers, it is particularly hard to forecast AI's future "net effect" on cyber power. National foreign ministries—although they each operate in different systems of government, with corresponding differences in their capacities, missions, and degrees of influence on the direction of national policy—should have an important role in shaping and advancing national strategy. As the lead institutional actor in diplomacy, foreign ministries advocate for national policies with other states and global stakeholders. They might also provide insight into other actors' AI and cyber power developments to inform wider national strategy. The challenges for AI-related cyber diplomacy largely mirror those for non-AI-related cyber diplomacy over the last decade—from the continued fostering of norms of responsible behavior to efforts deterring and responding to irresponsible behavior. This paper reflects on the relevance of these problems for all states. In its analysis of the role of foreign ministries in cyber strategy, the paper uses some U.K.-specific examples, however, the implications of AI for cyber strategy are more widely applicable to other states. The bigger picture should be clear, notwithstanding some parochial British detail.

METHODOLOGY

This paper intends to establish a starting point for further public engagement and research on these issues. It does not try to comprehensively analyze every dimension of AI's impact on cybersecurity, cyber diplomacy, or wider cyber statecraft. Nor is it a technical paper. It focuses on how states should interpret AI's impact on cyber power from a national strategic perspective. It does so by exploring the operational impact of AI on cyberspace, identifying the institutional implications of these developments for governmental actors.

The paper's methodology is a critical literature review focusing on three relevant bodies of literature comprising predominantly academic and think tank outputs: cyber power, AI's impact on cybersecurity, and the strategic implications of AI. The paper uses this critical review approach to survey and evaluate the state of knowledge regarding the consequences of AI advances for the competition between states in and through cyberspace. The paper adopts a synthesizing approach, explicitly recognizing that each section merits a deeper and longer-form engagement than is possible here. The paper aims to highlight the connections between the different issues and suggest a pathway for future research. In so doing, the paper spotlights a somewhat neglected issue across the literature: the implications of AI's impact on cyber power for the institutional role and departmental contribution of foreign ministries in national cyber strategy.

Most of the relevant literature concentrates on institutional implications for defense, intelligence, and security actors, but this paper presents the problem from the perspective of foreign ministries. The defense and intelligence-oriented literature on cybersecurity contains several clear implications for foreign ministries, but these are rarely highlighted in this literature. This paper addresses this gap, connecting the contemporary literature on AI and cyber power with the institutional contributions foreign ministries can make—when organized effectively to engage—in the national cyber policy and strategy decision-making process. A healthy institutional balance is desirable between diplomats, the armed services, intelligence and national security officers in shaping national strategy, to say nothing of opening the strategy-making process to stakeholders outside government. This will not occur spontaneously. Foreign ministries should develop the appropriate capacity to make this contribution. More capable states, alongside a plethora of non-state actors (from civil society and the private sector), can assist by supporting other foreign ministries through cyber capacity-building programs.

The paper draws on examples from the U.K. to ground this argument, considering evidence that the Foreign Commonwealth and Development Office (FCDO) has played an increasingly active role in U.K. cyber strategy and its conception of "cyber power" as an instrument of national strategy.[3] The U.K. example is not presented uncritically, as an exemplar for all to emulate, but rather as a useful case study to introduce the paper's key themes and identify the institutional challenges all foreign ministries will face addressing the interdependence of AI and cyber power. While the challenges are similar, the institutional context in which different national foreign ministries operate will significantly shape their ability to address them. Not all countries will agree with the particular modalities of the U.K.'s adaptation of its foreign ministry to pursue cyber diplomacy, and not all countries will regard emulation of that adaptation as feasible. Context is crucial and should be factored into any attempt to draw lessons from one national case to inform deliberations in another.

AI: A BRIEF EXPLANATION

As mentioned above, this is not a technical paper. However, it is worthwhile to summarize AI. As one researcher has noted, "the concept of AI means different things to different people, partly because its subject matter—intelligence—is hard to define."[4] The U.S. Department of State offers the following definition: "Artificial intelligence may be understood to refer to the ability of machines to perform tasks that would otherwise require human intelligence. This could include recognizing patterns, learning from experience, drawing conclusions, making predictions, or generating recommendations."[5] This and similar definitions are common in national policy papers. Some include additional clauses that emphasize the importance of data, for example. While understanding that AI "can mean a lot of things," the U.K. government's (2021) National AI Strategy provides the following working definition: "Machines that perform tasks normally requiring human intelligence, especially when the machines learn from data how to do those tasks."[6]

The European Commission's 2020 White Paper on AI highlighted key factors behind its recent advances: "AI is a collection of technologies that combine data, algorithms, and computing power. Advances in computing and the increasing availability of data are therefore key drivers of the current upsurge of AI."[7] The U.K. strategy also explicitly connects these "key drivers of [AI] progress, discovery and strategic advantage" (namely: "access to people, data compute and finance") with the geopolitical theme of "global competition."[8]

AI research is multidisciplinary and has a much longer history than is conveyed by references to it as an "emerging" technology.[9] The breadth and historical pedigree of the field are reflected in the contrast between different approaches to AI research, such as, on one hand, symbolic AI and, on the other, machine learning (ML) approaches using neural networks. The former was particularly evident in earlier generations of AI research, relying on logical programming and rules-based systems. More recently, ML-based approaches have come to global prominence due to breakthroughs in deep learning, the exploitation of advances in neural networks and the availability of large quantities of data and computing power. Artificial neural networks—inspired by the structure and function of neurons in the human brain—are combinations of nodes for transmitting and processing data (e.g., text or images); these networks "learn" from that data to perform given functions.[10] One example of applying a neural networks approach is Google's 2016 introduction of its Neural Machine Translation to improve Google Translate.[11] Large Language Models (LLMs) and generative pretrained transformer (GPT) models are examples of neural networks trained on large quantities of text data.[12] A tool based on such models, such as OpenAI's ChatGPT, can be prompted to generate new natural language text (song lyrics, a novel, or a computer program). Advances in generative AI research have led to urgent global policy and regulatory concern about the future implications of 'frontier' of cutting-edge AI developments. Whilst 'frontier AI' is variously defined, the U.K. government uses the term to refer to: 'highly capable general-purpose AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models.'[13]

An accessible example of how different AI approaches have evolved over the past 80 years is the improvement of computer programs for playing chess. These have greatly progressed from their beginnings as relatively simple—and, for human players, beatable—algorithms for playing a single game. The progress in applying more sophisticated algorithms and more powerful computing to chess was evident in IBM Deep Blue's famous 1997 victory over then-World Chess Champion Garry Kasparov. More recently, in 2017, Google DeepMind's AlphaZero used neural networks and reinforcement learning to attain "superhuman levels" of chess performance from random play in 24 hours without the benefit of any game-specific knowledge except the rules.[14]

Of course, if the utility of neural networks and deep learning techniques were restricted to beating humans at playing chess, the intensity and urgency of global public policy debates about regulating AI research would be more muted. These ML techniques produce a wider range of results that imply far greater societal impact. Such results include assisting medical researchers in identifying new treatments and using ML to significantly increase the speed and volume of data processing.[15] There are also a range of defense and security applications, ranging from facial recognition software to integration into systems such as the Patriot Air and Missile Defense System. Such applications, especially those that involve the potential loss of life, raise sharp questions about ethics and the extent of human control in the decision-making process.[16]

Three categories of AI-related concerns can be applied to cyber strategy. First is the cybersecurity of AI systems and the harm that compromising those systems could inflict. Second is the potential for AI techniques to improve (or transform) existing approaches to cybersecurity, enabling defenders to respond to threats quickly and in large volume. The third is the mirror image of what AI offers to defenders, namely the prospect of AI enabling new methods of cyber attack. Each of these is an independent policy issue. States should be increasingly mindful of AI-related challenges when developing and implementing national cyber strategies, as well as during participation in regional, multilateral, and multistakeholder cyber diplomacy discussions regarding the evolution of cyber norms. The capacity of different states to address the challenges and seize the opportunities presented by AI will vary considerably. But a common baseline of understanding of what is at stake is a good place to start.

STRUCTURE OF THE PAPER

The paper is structured to address three related issues sequentially: (1) the origins of the concept of cyber power and its interdependence with AI, (2) a practical focus on the implications of this interdependence, particularly the oft-cited question of whether AI will favor attackers or defenders in cyberspace, and (3) the growing role of cyber and science/technology diplomacy in national policy; how states can and should adapt the institutional role played by foreign ministries in contributing to national AI and cyber strategies.

In the first section, the paper explores and explains the impact of AI advances on the concept and practice of cyber power. In the second part, the paper introduces how AI and cyber power are perceived as instruments of national strategy. It uses the U.K. example because of the prominence of cyber power as a framing concept for its cyber strategy. How countries answer this question will shape how they prioritize investments and lines of effort.

Having articulated the general principle that AI and cyber power are strategic instruments, the third section provides more specificity, with a detailed assessment of AI's impact on cybersecurity and cyber operations. This section explains operational issues, adversarial behaviors, and the capabilities countries should monitor and assess as they develop foreign policy responses to AI's impact on cyber power.

This advances the paper's argument into its fourth and final section, in which the paper reflects on the development and execution of foreign policies and national AI and cyber strategies. It also highlights the rapidly increasing intensity and plurality of events and processes that explore the potential for common ground between states on AI's future regulation, safety, security, and trustworthiness. This includes national initiatives with potential international reach—such as the Biden administration's October 2023 executive order on AI—and various international processes that have commenced in the last three years and in which states are increasingly active. These include the United Nations, G7/G20, Organisation for Economic Co-operation and Development (OECD), and the BRICS group.[17] This section provides practical recommendations for foreign ministries to adapt and address the new challenges created by AI's impact on cyber power and broader technology diplomacy.[18]

FOREIGN MINISTRIES, AI, AND CYBER POWER

The paper argues that the complexity and resource intensity of remaining at the cutting edge of AI research, development, and exploitation will likely reinforce existing asymmetries of power between countries in cyberspace. Recent evidence shows that only a relatively small number of states have the resources to acquire the necessary hardware and recruit the talent required to compete.[19] The paper argues that foreign ministries should adapt to an era of increasingly AI-enabled competition between states and non-state actors in cyberspace. It concludes by clarifying where foreign ministries can play their most effective role in the sometimes busy institutional landscape of cyber-relevant agencies and departments and the increasingly busy schedule of bilateral, multilateral, and multistakeholder events and processes devoted to cyber diplomacy and the safety, security, trustworthiness, and future governance of AI. AI's impact on cyber power is best understood as an ongoing process requiring strategic patience rather than a panicked, winner-takes-all race toward a fixed endpoint.

CLARIFYING THE RELATIONSHIP BETWEEN AI AND CYBER POWER

As noted in an early paper on "Cyber Power" by Joseph S. Nye, Jr., power is difficult to measure but far from meaningless.[20] Both AI and cyber power are prominent themes in discourse about geopolitical competition.[21] Just as states have always tried to use the instruments at their disposal to achieve their national objectives, today's states are exploring how best to integrate cyber capabilities and emerging technologies such as AI in their wider national strategies. Publicly, at least, governments were initially slower than companies to respond to the mid-2000s breakthrough in machine learning.[22] But governments began to move faster toward the end of the last decade. For example, according to one researcher: "Between 2017 and early 2019 at least 17 countries released a national strategy or made a strategic policy announcement on AI."[23] There is a similar story in the development of national cybersecurity strategies. Between 2003, when Norway developed a national cyber strategy, and 2018, 76 countries had developed a cyber strategy. By 2021, that figure had grown to 127.[24]

Though this is impressive growth, the picture is uneven. In Latin America and the Caribbean, for example, while the Organization of American States (OAS) was the first regional body to produce a cyber strategy (2003), there are still 17 countries in the region that lack a national cybersecurity strategy addressing critical infrastructure and resilience, and 14 countries without a national computer incident response team according to the International Telecommunication Union.[25] As the diplomatic agenda on cyber norms and AI-related concerns intensifies, it is clear that national foreign ministries will start from various national baselines as they work out how to best engage with this increasingly busy agenda. To better inform this process of adaptation and engagement, it is useful to consider examples of states that were early adopters or thought leaders in this area.

The U.K. is a good example of a state that has been at the forefront of developing an integrated national approach to cyber strategy, has prioritized the importance of cyber capacity-building assistance as an element of that strategy, and has increasingly recognized the connection between AI and cybersecurity. The 2021 U.K. Integrated Review of Security, Defense, Development and Foreign Policy (hereafter the Integrated Review) framed this as part of a wider "struggle to shape the global digital environment between 'digital freedom' and 'digital authoritarianism,' which will have significant implications for real-world governance."[26] This applies as much to the competition to develop leading AI and cyber capabilities as it does to how actors use these capabilities to achieve their goals in and through cyberspace. Interstate competition for strategic advantage in and through science and technology is wider than AI and technologies vital to cyber power. Yet, AI is regularly cited among the highest strategic science and technology priorities.[27] The U.K. government regards AI as one of five key emerging technologies crucial to its cyber power agenda.[28]

The U.K. is far from the only state publicly articulating the connection between geopolitics and technology. Nor is it alone in highlighting the tension between democracies and authoritarian governments over the future of internet governance and developing a competitive edge in AI and other emerging technologies. These are salient contemporary themes, visible, for example, in the Biden administration's U.S. National Security Strategy.[29] This framing highlights that AI diplomacy and cyber diplomacy are as political as they are technical processes. This also goes for cyber capacity-building efforts, arguably intended to serve developmental and political-strategic, i.e., "influencing" objectives.

However, nation-states are not the only, nor even the most consequential, actors in the AI and cyber fields. In contrast with many twentieth-century computing developments associated with defense contracts, many contemporary AI developments are driven by private sector research with broader commercial, rather than defense, incentives.[30] This demonstrates a stark reversal over time. In 1967, the U.S. government funded 67 percent of research and development in science and technology. By 2020, the private sector accounted for 73 percent of such spending.[31]

A major theme of public policy discourse about AI regulation, therefore, is how governments can effectively position themselves in a field in which corporate, nongovernmental entities are making the biggest strides.[32] The sense of sudden urgency is palpable, with reporting on one recent international summit on AI's existential risks noting, "The lack of government controls on AI has largely left an industry built on profit to self-police the risks and moral implications of a technology capable of next-level disinformation, ruining reputations and careers, even taking human life."[33] This statement reflects the fact that governments collectively feel they are playing catch up with the private sector, rather than necessarily implying that the private sector has no interest in exercising a responsible approach to research, development, and use of new AI systems.[34] The extent of the reversal of power between governments and the private sector is perhaps most evident in U.K. Prime Minister Rishi Sunak's somewhat hyperbolic hailing of a largely voluntary, non-binding agreement in November 2023 by technology companies to share models for testing by a new U.K. AI safety institute as a "landmark achievement."[35]

Notwithstanding the evident proliferation of global initiatives regarding AI regulation, AI is not the first disruptive scientific or technological development to affect international security.[36] Similarly, we should consider that the study of AI's implications for cyberspace already has a history.[37] Describing AI as "emerging technology' arguably stretches the definition of emergence.[38] This applies specifically to the impact of AI on cybersecurity, where machine learning techniques have been integrated into tools and systems for over 20 years.[39]

In the foreseeable short term, AI will likely continue to have an evolutionary rather than revolutionary impact on the competition between defenders and attackers in cyberspace.[40] Successful integration of AI can increase the scale, speed, and sophistication of defensive and offensive cyber operations. Attackers and defenders are highly unlikely to advance simultaneously across operational applications of AI in cyberspace. The process of competitive adaptation will determine this, particularly in how states approach the key inputs of data and expertise. According to one researcher, "It really comes down to data, and whoever has the most data to train the models, whoever has the most educated workforce to spend the time to build these systems. So we're in a national security race right now against our adversaries."[41]

This assumption is hardly unique to AI. It draws on the history of interstate competition to achieve advantage through innovation and adaptation, which is a highly contingent, dynamic, and interactive process.[42] Therefore, as immediate, reciprocal defensive advances in AI for each specific advance in offensive applications are unlikely, developments in AI will continue to affect the fluctuating global shape and character of cyber power, rather than simply cementing the status quo.[43] To date, regardless of the hype and media interest in the specter of AI-enabled cyber weapons, there is, as yet, no agreement about precisely how or how quickly AI will affect competition in cyberspace.[44] The race to innovate and integrate new technologies is not solely a technical enterprise. It is contingent on human factors, institutional cultures, and the skill and strategy with which new technologies are employed.[45] Humans and their institutions still matter.

So, while the status quo has the potential to change, the benefits will likely remain with those who can combine strategic purpose, skilful employment, and the necessary scale of resources and computational power for cutting-edge AI research. This is a field dominated by already-powerful states and companies.[46] Similarly, the increasing commercial availability of AI-enabled capabilities (such as for surveillance) is likely to further entrench rather than overturn states' cyber power in relation to their citizens.[47] The current dominance of a small number of leading players in the development and application of AI is likely to lead to those corporate and national actors trying to maintain this leading position, extending to efforts to safeguard the intellectual property and hardware (such as advanced graphics processing units) that enable it, access to which would allow competitors to catch up.[48] If states came to perceive the costs and risks of such AI competition in cyberspace as unbearable, it is conceivable that collective restraint (on capability development and circumstances of use) might eventually be pursued, but the politics and modalities of reaching agreement would likely be more difficult than the technicalities of verification.[49]

While AI is unlikely to be geopolitically transformative, and current evidence is inconclusive, its future applications would likely affect the balance between offense and defense in cybersecurity for a time.[50] For defense, AI has implications for each of the five operational phases of the well-known Cybersecurity Framework developed by the U.S. National Institute for Standards and Technology (NIST).[51] For attackers, AI presents similar opportunities across the Cyber Threat Framework.[52]

Figure 1. Cyber Defense: The NIST Cybersecurity Framework

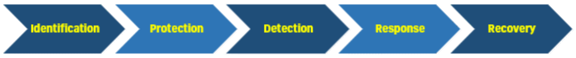
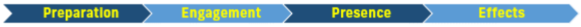


Figure 2. Offensive Cyber: The U.S. government Cyber Threat Framework



Rapid, unmatched acquisition and effective employment by one country of tools that produce a temporary offensive advantage in cyberspace would not transform the underlying structure of international security. It could, however, affect that state's appetite for risk and bias for action, particularly if it was aware of its sudden advantage and recognized a short-term window of opportunity that adversaries would endeavor to close as quickly as possible. As more states incorporate more powerful AI-enabled capabilities into their cyber (and non-cyber) operations, the speed and scale of engagement should be expected to increase, thereby increasing the potential risk of inadvertent escalation.[53] This risk could be exacerbated by reduced predictability and control over AI-enabled systems rushed into service.[54] The strategic imperative for states to compete is obvious, but so is the need to explore ways to agree to norms of responsible uses of AI to mitigate these risks.[55] Overall, the international security implications of AI advances and their impact on cyber power will be challenging for states to manage. Furthermore, they will likely highlight sharp divides between liberal democracy and digital authoritarianism. This will require active and inclusive global diplomacy in which national foreign ministries will play an important role.[56]

AI AND CYBER POWER AS INSTRUMENTS OF NATIONAL STRATEGY

This section will explain how cyber power and AI were conceived as instruments of national strategy and explore the steadily increasing institutional contribution of foreign ministries to the foreign policy and diplomatic aspects of cyber statecraft. The concept of cyber power was first coined in the United States, but soon made the journey across the Atlantic and beyond.[57] The systematic effort to articulate and measure cyber power is more recent. A prominent example is the U.K. government's effort to promote "Responsible and Democratic Cyber Power" as a key feature and framing concept of its strategic communications about cyber strategy.[58] The Integrated Review offers a broad definition of cyber power, situating it as an instrument of national strategy. It encompasses the promotion of national security, the national interest, prosperity, and the effective diffusion of national values. It also includes the importance of having the operational capacity to pursue national interests and achieve "real world" effects through cyberspace.[59] Thus, irrespective of whether the rhetoric of cyber power is effective as a trope of strategic communication, it is arguably too early to judge this definitively. However, it is a useful shorthand for a range of important developments in national security that states should address, exploit, or overcome.

The 2022 U.K. National Cyber Strategy defined AI as "a technology in which a computing system is coded to "think for itself," adapting and operating autonomously. AI is increasingly used in more complex tasks, such as medical diagnosis, drug discovery, and predictive maintenance."[60] Public debates about AI emphasize it represents a qualitative difference to previous trends in technological innovation, with potentially revolutionary impact.[61] In the last decade, the urgency of geopolitical competition for "AI power" has intensified. And in many of the public contributions to these debates, the roll call of risk factors associated with the diffusion of AI power includes the potential for malevolent actors to use generative AI to create cyberweapons.[62] As a starting point, a working definition of AI power is the use of AI to pursue the national interest.

Any form of power entails the ability of one actor to influence another's behavior effectively.[63] This means relationships are at the heart of effective cyber power. Within a government, different agencies and departments must establish relationships to ensure the optimal effectiveness of the governmental approach. But a national cyber strategy must look beyond the silos of government and ensure effective working relationships between government and other domestic stakeholders, such as the private sector, civil society, and academia. Similarly, effective cyber strategy requires cultivating effective relationships at regional and global levels to address transnational challenges. It is easy, therefore, to see why diplomacy plays an important role in contemporary cyber strategy, but the diplomatic skills of relationship management are not confined to the international domain. They also apply to managing relationships with key domestic stakeholders.

The U.K. is a good example of a state that has recognized and publicly articulated the importance of this in successive iterations of national cyber strategy. It has also tried to export this strategic insight through cyber capacity building, for example, through long-term funding for the Oxford Global Cyber Security Capacity Centre and the dissemination of its influential Cybersecurity Capacity Maturity Model for Nations. The U.K. strategic approach to cyberspace realizes that state power and influence is enhanced by a comprehensive whole-of-cyber/whole-of-society approach to national strategy, embracing all relevant stakeholders, including nongovernmental ones.[64]

Another trend over time is cyber transparency. Some states are talking more openly about the broad potential utility of offensive cyber operations as a tool of national strategy.[65] This is an initiative that should be encouraged. It can be perceived as a confidence-building measure. As such, it arguably represents the achievable, relatively low-hanging fruit of cyber arms control diplomacy.[66]

The Belfer Center's National Cyber Power Index is a prominent example of efforts to measure cyber power. The first iteration in 2020 designated the United States, China, and the United Kingdom as the world's top three cyber powers. The second in 2022 saw Russia overtake the U.K., with the United States and China retaining the top rankings.[67] Brazil was the only Latin American country included in Belfer's analysis that ranked outside the Top 10 in both iterations. The Belfer approach says, "there is no single measure of cyber power," and measurement needs to be sensitive to "the context of a country's national objectives."[68]

Initiatives like the Belfer Index are welcome efforts to provide structure to measuring cyber power, but suffer from a lack of robust, publicly-available data, resulting in some incongruous conclusions, like its justification for Russia supplanting the U.K. in the most recent edition.[69] Effective cyber power must be about more than simply conducting publicly-known cyber operations. The U.K. is a good example of how, over a decade, a state has tried to coordinate within and beyond the governmental sphere to produce a coherent national approach to building and effectively—and responsibly—employing cyber power.

The institutional role of a foreign ministry differs according to national context. Many national foreign ministries face similar challenges. This is true internally, for example, in the increasingly proactive involvement of heads of government in direct diplomatic engagements with their counterparts, facilitated by modern telecommunications and air travel. It is also true internationally, given that states cooperate in various regional and global forums—that a significant asymmetry in resources and capacity between different national foreign ministries exists, reflecting the asymmetries of power and wealth in the international system.

This paper presents the case of the U.K. Foreign Ministry and its involvement in national cyber strategy as a specific country example. In the U.K., the Foreign Commonwealth and Development Office (FCDO) plays the institutional role of the foreign ministry.[70] The FCDO leads the diplomatic and foreign policy-related efforts under the different pillars of the

National Cyber Strategy. One interpretation of the evolution of U.K. cyber strategy across its successive iterations is that the FCDO has come to exert more influence over the public presentation and articulation of the national view of cyber power and what it entails.[71] This manifests in the FCDO's lead advancing the national objective to exercise global cyber leadership, including in such initiatives as coordinated public attribution statements, sanctions, incident response, and capacity- and confidence-building projects. Cyber and a wider emerging technology diplomacy shape the governance arrangements, norms, and regulatory environment in which future developments in emerging technologies vital to cyber power will occur. Conversely, as diplomacy and wider public policy can lag behind the pace of technological advances, new trends in emerging technology are likely to shape the cyber and technology diplomacy agenda.[72]

While the FCDO is the lead institution for diplomacy and foreign policy, the domestic and international elements of cyber strategy overlap, meaning the policy equities of other institutions are also in play. As such, several U.K. government agencies and departments have stakes in the international elements of national cyber strategy. These include the digital ministry (DCMS/Department for Science, Innovation and Technology), the Ministry of Defence, and the U.K. intelligence community – especially its cyber and signals intelligence agency, Government Communications Headquarters (GCHQ). An institutional landscape as busy as this can potentially strengthen a national strategy. At its best, it can generate creative tension between different cultures and expertise as part of a coordinated, whole-of-government effort. But this is not a given. Without effective coordination, multiple institutional actors can lead to disorganization and unresolved disagreement.[73]

The inter-institutional challenges of coordinating an effective governmental response to cyber strategy depend on the given national framework. For example, some states will have fewer actors than those listed above or exhibit different power dynamics and differentials between institutional actors. In the United States, for example, there have been recent reforms within the Department of State (creating a new Bureau of Cyberspace and Digital Policy, headed by Ambassador Nathaniel C. Fick) and in the White House (a new Office of the National Cyber Director with a deputy national security adviser for cyber and emerging technology). These reforms highlight that governments are increasingly trying to address how to reshape their institutions to embrace the opportunities and overcome the challenges of cyber and emerging technology. And, where national governmental context means authority is more personalized and informally executed by the head of government, particularly when characterized by relatively low institutional maturity in cyber-relevant sectors, national approaches are likely to reflect swift decision making and severe implementation challenges.

In the example of the U.K., the FCDO has a clear role in national cyber strategy. It informs national strategy by providing insight into foreign governments and other global actors. It uses its diplomatic expertise and understanding of regional and multilateral forums to coordinate national strategy with partners, persuade non-likeminded states of the merits of U.K. views, and effectively counter the efforts of adversaries and competitor states. Whatever the inter-institutional dynamics and governmental context, foreign ministries will be increasingly expected to play these roles. This turn in cyber diplomacy encapsulates much of what the current U.S. Central Intelligence Agency Director (and former career diplomat), William J. Burns, describes as the "quiet power" of diplomacy. This description, more broadly intended, applies easily to AI and cyber diplomacy: "the largely invisible work of tending alliances, twisting arms, tempering disputes, and making long-term investments in relationships and societies."[74]

THE IMPACT OF AI ON CYBER POWER

Significant advances in AI research have underpinned and accelerated the global interest in AI. Yet, advances in AI have not resulted in a clear advantage for either defense or offense in cybersecurity. Since states have differing cyber capabilities, each will experience the impact of an AI-enabled attack differently. As noted above, this is not solely a technical issue. An actor must also bring human factors to bear to achieve the successful employment of an advantage notionally afforded by technical superiority.

If recent spending and rapid advances in available computational power continue, capabilities may improve sharply beyond what is currently available.[75] Alternatively, obstacles such as reports of the rapidly dwindling quantity of high-quality text data could slow advances using large-language models.[76] This uncertain outlook makes forecasting difficult, but the implications of forecasts are too important to be ignored. For example, those that posit scenarios of steady or accelerating gains in capability pose challenges for policymakers and regulators—particularly given the attendant risks accelerating AI capability poses to fairness and safety, as well as the impact of malicious use.[77]

One challenge in assessing the impact of AI on cyber power is determining whether AI will prove predominantly positive or negative for it. The expanding Internet of Things and Smart Cities are two challenges that ubiquitous AI poses for cybersecurity.[78] The proliferation of internet-connected devices causes a sharp increase in the size of the "attack surface" of vulnerable devices needing to be secured. This is a potentially negative outcome of AI. AI-enabled devices not only increase the size of the attack surface but also introduce new methods of compromise, such as contaminating or poisoning the training data.[79]

Conversely, the sharp increase in available data, computational power, and connected devices offers opportunities to improve cybersecurity practices, potentially mitigating AI's negative effects. Recent assessments of the overall impact of machine learning on cybersecurity concluded that a focus on specific cases rather than a top-level view was needed. "Policymakers and practitioners alike need to think about how machine learning can alter specific tasks within cybersecurity, rather than talking in general terms about how machine learning can alter cybersecurity as a whole."[80] This grounded approach—which, logically, should apply to both generative and non-generative AI tools—has the potential over time to inform a more comprehensive net assessment. The challenge is to integrate different discrete elements into such an assessment. A task-specific focus is a crucial part of such an effort, but governments ultimately need to take a broad view for strategic prioritization.

NET ASSESSMENT OF AI'S IMPACT ON CYBER POWER

While restricting the analysis to specific cases should present a more manageable challenge, governments should still adopt a broader view to guide strategic prioritization. The net assessment of AI's impact on cyber power should weigh a series of specific impacts to inform a comprehensive judgment of AI's impact on the totality of cyber power. Issues to consider would include the likelihood of sustained progress along established trend lines[81] in defensive and offensive applications of AI in cyberspace, recognition of the potentially transformative impact of advances that represent a steep change in the development and application of AI-enabled cyber capabilities, and the geopolitical implications of these trends (partly accounting) for states' differing AI and cybersecurity abilities. Depending on a state or head of state's risk appetite, its ability to deliberate effectively, and its strategic orientation, its behavior may change as a result of perceived changes in its relational power. The more opportunistic a state's leadership, the greater the chance it will seek to exploit a perceived advantage.

The increasing use of AI-powered applications to write code has a broader impact on cybersecurity. As with human programmers, current AI has significant potential to produce insecure code. One recent study found that 40 percent of AI-produced code contained exploitable vulnerabilities.[82] Furthermore, the recent release of OpenAI's ChatGPT—a chatbot enabling users to query a large language model (LLM)—led a popular programming knowledge-sharing website to prohibit uploads generated by ChatGPT, due to the large volume of flawed (but superficially plausible) code it had generated.[83] There were also reported instances of users effectively disabling safeguards built into ChatGPT and other LLM-based tools.[84] Notwithstanding these incidents, there is a clear use case for AI integration into normal business, which is seen in Microsoft's development of AI copiloting for information security.[85] These episodes highlight not just the risks posed by subversion of AI for unintended uses, but also the opportunities created by the speed and volume of AI-assisted activity.[86] Over time, the wider availability of increasingly sophisticated AI is likely to add further plausibility to the tenet that the cyber domain can have a "low barrier to entry."[87]

As a step toward net assessment, we can explore some specific implications of AI developments for cybersecurity and cyber operations. Over the last 20 years, AI has been increasingly integrated into most elements of cybersecurity software.[88] The market has expanded as capabilities developed, though the sector is not immune from vendor hype outstripping current performance.[89] The logic of using AI to improve cybersecurity is that the automation of cyber defenses increase the scale and speed of threat identification, detection, and response.[90]

One specific example of this use of AI has been the incorporation of machine learning into cybersecurity to automate the detection of spam, intrusion, and malware.[91] The gradual incorporation of machine learning into more sophisticated anti-spam tools improved the accuracy of spam detection but is considered an evolutionary rather than transformative application of AI to cybersecurity. Machine learning has also been applied in the two main methods for detecting intrusions, misuse-based and anomaly-based detection, being similarly incremental rather than revolutionary.[92] [93] However, malware developers responded, increasing the sophistication of their tools to evade AI-enabled detection. One estimate is that more than 90 percent of malware in 2018 exhibited polymorphic features (e.g., the existence of multiple variants or encryption methods to avoid certain types of AI-enabled detection).[94]

Consequently, while machine learning offers incremental benefits, current evidence suggests these techniques have not delivered a knock-out blow in their impact on the performance of malware-detection systems.[95] This should not be surprising. Advances in cybersecurity tools are unlikely ever to enable network defenders to declare a permanent victory over their adversaries. Such advances will, instead, spur competitive innovation and adaptation.

It is similar to other fields of cybersecurity research focused on AI. Academic and private sector researchers are increasingly using machine learning techniques (e.g., deep-learning neural networks) to improve software security testing in "neural fuzzing."[96] AI can also enhance the ability of cybersecurity tools to learn to detect new cases of previously unseen malware with no records/indicators of compromise. Similarly, reinforcement-learning-based penetration-testing (pen-testing) techniques could ultimately improve the speed and scale of existing automated approaches. However, the industry's ability to exploit this potential is limited by the computational intensity required to test larger, more complex systems. Furthermore, evidence suggests these techniques do not produce significantly better results than existing, non-reinforcement-learning-based techniques. As with more traditional pen-testing tools that attackers have appropriated, significant advances in this field might also be misappropriated to enable attackers as much as defenders.[97]

Just as AI improves cyber defenses, it can also assist malicious cyber capabilities. Two private-sector cybersecurity executives observed in 2019 that "AI-powered cyberattacks are not a hypothetical future concept. All the required building blocks for the use of offensive AI already exist."[98] They point to the already highly-sophisticated quality of some malware, the availability of capable, open-source AI programs, and the perennial existence of motivation for malicious use. The promise of AI for malicious cyber actors essentially mirrors its promise for defenders, namely that malware becomes better at deceiving intended targets, harder to detect, smarter and faster at spreading across networks and exploiting its access to exfiltrate data, or to disrupt, degrade, or destroy.[99]

Possible future advances in AI-enabled

malicious cyber activities sound formidable, but these claims of (not yet demonstrated) potential can also be made about AI-enabled cyber defenses. A recent report lists several current and possible future defensive applications of machine learning, but none constitute a revolution in existing practice.[100] However, concerted application of machine learning to these "active defense" practices would increase operational friction for cyber attackers, wasting their time, misdirecting them, and enabling defenders to improve their understanding of their adversaries' tools, techniques, and procedures.

The use of AI creates new attack surfaces and novel manifestations of traditional security risks, particularly in terms of the cybersecurity of AI systems.[101] Established scholarly debates about the utility of cyber operations have started focusing on cyber threats to new AI developments. One example is the challenge of ensuring the cybersecurity of lethal autonomous weapons systems (LAWS).[102] This is a "new wine in old bottles" problem, as AI-enabled platforms and systems are likely, in principle, vulnerable to the intrinsically subversive potential of cyber operations.[103] In the context of LAWS, such sabotage could focus on data-poisoning, affecting the training process, or directly altering algorithms.

These cyber threats are not unique to LAWS. They apply to all AI systems. But the possibility of rogue autonomous weapons has a greater impact on public opinion, even when primed by fiction. Corollaries exist in the field of AI-enabled cybersecurity tools. The increasing use of machine learning to enhance cybersecurity likely incentivizes attackers to compromise those tools during development.[104] Such cyber operations, while challenging to execute and with success difficult to measure, are hypothetical ways adversaries could undermine each other's progress in pursuit of strategic advantage.

To conclude this section, debates about the impact of AI on cyberspace reflect uncertainty about the net effect of these advances in defending and penetrating computer networks.[105] AI is highly likely to increase scale and speed on both sides of the offense-defense divide. However, others contend advances in AI might lead to more direct confrontation between state actors in cyberspace in a manner that increases opportunities for "tacit bargaining" rather than escalation risks and instability.[106] This theory suggests state behavior in cyberspace is largely exploitative rather than coercive. It contends that the operational activity of state actors in cyberspace is continuous and best conceived as campaigns of activity rather than a series of disconnected, one-off actions, meaning operational gains are cumulative. It also maintains that the dynamic of operational interaction is more competitive than escalatory, with states aiming to achieve fait accomplis without their adversaries' knowledge or leaving adversaries unable or unwilling to respond.[107]

The above account should not instill complacency about the future impact of more sophisticated AI-enabled cyber operations, particularly regarding escalation risks and concerns about whether the risk appetite of some states might lead them to deploy an AI-enabled capability without robust controls. The media has long highlighted unintended escalation in cyberspace as a major risk, which is arguably heightened by the potential for AI to accelerate the speed of decision-making in cyber and counter-cyber operations.[108] A recent study suggests there is some basis for the fear that future AI advances may have an escalatory impact on competition in cyberspace.[109] Even precautions, such as keeping humans in the decision-making loop rather than ceding full autonomy could be undermined by the cognitive impact of the accelerated pace of activity and the potential inscrutability of AI-derived network effects.[110]

IMPLICATIONS FOR FOREIGN POLICY AND NATIONAL STRATEGY

Regardless of whether cyber power or other appellations, such as the U.K.'s "responsible and democratic cyber power," are invoked in a given state's strategic communications, the capabilities and practices under its conceptual umbrella are an integral part of contemporary national security strategy. Advocates for the prominence and utility of cyber power will also need to advocate for guardrails to promote and shape responsible state behavior in cyberspace. They will find both initiatives enhanced by the effective integration of AI and other technologies vital to cyber power within the mainstream of their respective national cyber strategies. This will include exploring the need to revise the institutional mission—and perhaps also the organizational design—of national foreign ministries.

Foreign ministries increasingly face a busy agenda of cyber and emerging technology diplomacy. Over the last decade, global cyber diplomacy has become more formally inclusive. This is apparent, for example, at the United Nations, where we have seen the transition from the restricted membership of successive Groups of Governmental Experts to the more open process of the Open-Ended Working Group. The prospective Programme of Action and the cyber-relevant dimensions of the UN Secretary-General's Global Digital Compact and Summit of the Future are also more inclusive.

This turn toward greater inclusivity and participation must be seen as a positive development, but it also places larger burdens on smaller states and foreign ministries with less capacity to engage fully across a busy international agenda. Although this was already true in cyber diplomacy, it has become even more true in the last three years of AI diplomatic developments. Besides creating increased demands on the cohort of existing cyber and emerging technology diplomats, it makes it crucial to increase the resources, workforce, and external support available to foreign ministries to increase their engagement with the growing global agenda of AI and cyber diplomacy.

THE GROWING GLOBAL AGENDA ON AI

The field of global AI diplomacy has become increasingly crowded. According to one recent study, "more than 50 active international AI governance initiatives, related to nearly as many forums, bodies, or actors driving them. Most of these are anchored within the established international cooperation system, with nearly a quarter originating within the UN system itself."[111] While states will not participate in all forums or participate to the same extent in every forum, there is a proliferation of initiatives and processes that states should monitor and, where necessary, engage to ensure their national viewpoint and interests are reflected in global deliberations.

At the national level, there are dozens of pieces of AI-related legislation worldwide, but the most prominent example of a significantly strong effort to legislate standards for AI is the development of the European Union's (E.U.) AI Act.[112] But legislation is not everything. As this paper has argued, given the transnational nature of challenges related to AI and cybersecurity, states must cultivate effective relationships with relevant stakeholders, both state and non-state actors. This international relationship management should be seen as an important complement to domestic efforts. For example, in addition to developing its AI Act, the E.U. also collaborates with allies, both bilaterally through initiatives such as the E.U.-U.S. Trade and Technology Council and the E.U.-Japan Digital Partnership, as well as multilaterally through the G7 and other processes. As one observer of E.U. efforts noted recently, this is an "already-crowded regime complex of international organizations, standards, principles, and codes of conduct."[113] For example, the G7 has developed, under its Hiroshima AI process, International Guiding Principles for developing advanced AI systems. Moreover, the Biden administration is also pursuing an increasingly active and assertive agenda (see below) through executive orders and international engagements (such as Vice President Kamala Harris's U.K. visit in November 2023), to shape the global landscape of AI safety, security, and trustworthiness.[114]

A range of models have been proposed for regulating artificial intelligence. Maas and Villalobos have categorized these models as focusing on scientific consensus-building, political consensus-building and norm-setting, coordinating policy and regulation, enforcing standards and restrictions, stabilization and emergency response, joint international research, and distributing access and benefits.[115] These models are not mutually exclusive. In recent years, language that invoked several models was seen in a series of summit communiques.

At the global multilateral level, the UN Secretary-

Advisory Body on Artificial Intelligence.[116] This effort is aligned with the UN's wider, multistakeholder initiative to negotiate a Global Digital Compact, culminating in the 2024 Summit for the Future. The compact aims to "outline shared principles for an open, free and secure digital future for all"—including efforts to counter discrimination and disinformation, as well as to uphold human rights and expand access to the internet.[117] The compact is also focused on exploring the "regulation of artificial intelligence to ensure that this is aligned with shared global values."[118] The UN University's Centre for Policy Research has explored the scope of the UN's use of its "convening power and moral authority" to shape the global policy and regulatory debates about frontier AI. Given the UN's lack of technical expertise, it perceives its role as focusing on improving the global diversity of representation and participation in these debates, as well as the global equity of "benefit-sharing" and the building of a strong normative consensus as a prelude to the longer term development of "an effective international regime complex for AI."[119]

Regardless of the model being pursued, there is a need to look beneath surface similarities to identify points of consequential difference. For example, in 2019, the OECD agreed to five high-level principles for "responsible stewardship of trustworthy AI" that respect human rights and democratic values. The principles are inclusive growth, sustainable development and well-being, human-centered values and fairness, transparency and explainability, robustness, security and safety, and accountability.[120] The OECD principles formed the basis for a similar declaration in June 2019 by the G20 regarding non-binding principles for a human-centered approach to AI. The OECD also created an AI Policy Observatory to provide a shared baseline of information to support states in making AI policies. Since 2020, the OECD has provided secretariat support for the Global Partnership on AI (GPAI), which promotes research and global knowledge sharing and has 29 member states.

This multilateral approach to fostering common understanding and developing shared principles has continued to the present day. Since an April 2023 summit meeting in Japan, the G7 initiated the Hiroshima AI Process of inclusive dialogue to explore common ground on governance to improve AI safety and trust. A subsequent G7 leaders' statement of October 30, 2023, encapsulates this emerging international agreement on broad, high-level principles of AI safety and trustworthiness and the need to balance the pursuit of AI opportunities (in the process of "closing digital divides and achieving digital inclusion") with mitigation of the risks of doing so.[121] The G7 Hiroshima AI Process has produced a Comprehensive Policy Framework with four pillars: "analysis of priority risks, challenges and opportunities of generative AI … [an] International Guiding Principles for all AI actors in the AI ecosystem … [an] International Code of Conduct for Organizations Developing Advanced AI Systems, [and] project based cooperation in support of the development of responsible AI tools and best practices."[122] Both the Hiroshima Process and the OECD-supported GPAI are soft, non-binding initiatives aimed at improving understanding and exploring the potential for common ground between states on the broad set of issues relating to AI development, use, and governance.

Other recent noteworthy multilateral developments are the Council of Europe's draft Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law,[123] the G20's New Delhi Leaders' Declaration,[124] a new BRICS AI Study Group to establish governance frameworks and facilitate both information exchange and technical cooperation,[125] and a recently announced Global AI Governance Initiative through China's Belt and Road Forum, through which China aims to export its vision for AI governance and safety testing, as well as advocate for an international institution to govern AI.[126] This proliferation of multilateral events, frameworks, and initiatives highlights not only the increasing pace of international activity on AI governance, the high-level framing of common principles regarding a human-centered, rights-respecting, and responsible approach to AI, but also substantive geopolitical disagreement about the best governance and safety testing models to pursue.

Notwithstanding efforts such as those of the OECD to support international consensus and an open, inclusive, and multistakeholder approach to AI legislation, how countries interpret the existing high-level principles about responsible stewardship of AI differ significantly. This is not just a case of two camps divided, such as a disagreement between the like-minded states and China. There is also disagreement between close allies. For example, the E.U. and the United States agree on the broad principles of AI risk management and the need to enhance the safety, security, and trustworthiness of AI systems. This agreement on the principle of risk management is, however, subject to diverging methods of implementation. For example, the United States broadly favors an approach focusing on executive orders,[127] voluntary agreements, and frameworks, whereas the E.U. has pursued a more legislative method to managing AI risks.[128]

THE BIDEN ADMINISTRATION AND AI

In late October 2023, the Biden administration released an executive order (EO) on the safety, security, and trustworthiness of the development and use of AI systems.[129] This order was far-reaching and broad, encompassing national security, privacy protection, countering AI threats to employment, and AI's potential to exacerbate discrimination. It invoked the Defense Production Act to require "that developers of the most powerful AI systems share their safety test results and other critical information with the U.S. government."[130] The National Institute of Standards and Technology is required to develop the standards for such testing, and the Departments of Homeland Security (DHS) and Energy are charged with addressing AI systems' risks to critical infrastructure, with DHS also convening an AI Safety and Security Board. The EO also addresses AI risks in biosecurity, cybersecurity, and disinformation. In addition to its protective aspects, the order also aims to "promote innovation and competition, advance U.S. leadership in AI technologies, and ensure the responsible and effective government use of the technology."[131] Under the heading of "Advancing American Leadership Abroad," the order also recognizes the need to work with partners outside of the United States, in "bilateral, multilateral, and multistakeholder engagements to collaborate on AI."[132]

The recent EO was not the Biden administration's first AI-related initiative. The administration had, for example, previously secured voluntary commitments from leading U.S.-based AI companies and started to build international agreement on the principles underpinning responsible military uses of AI.[133] These are all examples of executive-led initiatives. The EO's initial reception noted the administration's use of executive initiatives to drive policy in the absence of congressional actions.[134] Reporting highlighted the administration's ongoing effort to work closely with industry[135] and the likelihood that legislation would ultimately follow.[136] Pro-enterprise and pro-innovation commentaries criticized the EO as a "red tape wishlist" and a panicked, misguided effort to move faster than Congress and use executive power to constrain AI research, likely stifling corporate innovation and competition.[137] This critical strand of commentary reflected the administration's difficulty in balancing competing priorities—pursuing the opportunities of AI while managing its risks—in the context of uncertainty.

While the AI regulation field is crowded with overlapping and competing initiatives, there is a shared sense of purpose: "Most of these guidelines say more or less the same thing—that all must balance the potential risk of AI systems against the risk of losing the economic and social benefits the new technology can bring."[138] The difficulty of pursuing a legislative route to further regulate AI is regarded as another reason for the series of EOs issued by the Biden administration. The October 2023 EO on safe, secure, and trustworthy AI, for example, "follows an earlier EO from August 2023 that limited U.S. investment in AI with potential military and intelligence uses in China."[139]

Notwithstanding the shared sense of purpose animating many of the AI-related initiatives from the United States, the E.U. and other like-minded states, these initiatives sometimes reflect different nuances of approach (e.g., in the case of the United States and E.U.) and also sometimes reveal the adverse impact of haste on allied coordination, particularly when states are competing to promote their respective national interests in the AI field. This last tendency was arguably on display during the Bletchley Park AI Safety Summit in November 2023.

THE BLETCHLEY PARK AI SAFETY SUMMIT

In March 2023, more than 1,000 scientists and entrepreneurs signed an open letter calling for a pause in research into the most advanced (frontier) generative AI systems, motivated by fear about the potential risks of such research in the absence of better understanding and regulatory guardrails.[140] Shortly after the letter's publication, Prime Minister Sunak announced the U.K. would host a summit meeting on AI safety, focusing on the existential risks of frontier AI.[141]

The U.K. summit took place in early November 2023, hosted at the historic intelligence and computing venue of Bletchley Park, to discuss the safety of frontier AI and its potential impact on international security.[142] It convened 28 countries, the E.U., and a range of nongovernmental attendees drawn mostly from large technology companies.[143] Two proposed outcomes ahead of the summit were a global AI safety institute to research further frontier AI safety and a voluntary global registry of frontier AI models, to enable red team testing. The summit did not wholly secure either of these objectives, reflecting the difficulty of achieving agreement between a relatively small grouping of states, particularly when the summit organizer—the U.K.—appeared out of step with the U.S. government on desirable outcomes.[144]

During preparations for the summit, there was skepticism about what it could realistically hope to achieve.[145] It had been criticized for its rushed design and delivery and its unclear contribution to an already crowded landscape of multilateral AI forums.[146] The exclusivity of its attendance list, especially a perceived failure to be sufficiently inclusive to civil society, was also critiqued.[147] Furthermore, some described its focus on frontier AI as rendering it a "doom" summit[148] that was also destined to mediocrity by the lack of global consensus for significant regulatory proposals—and indeed a lack of consensus on the precise meaning of "frontier AI."[149]

Some described the summit as motivated by the U.K. government's domestic political plight, symptomatic of "government by photo op."[150] Similarly, the U.K.-hosted summit appeared to expose differences in approach between the U.K. and the United States on AI.[151] This was evident, for example, in the themes addressed in Harris's speech on the margins of the summit, in which she announced a new U.S. AI safety institute and highlighted near-term AI risks not considered the perceived focus of the U.K.-hosted summit.[152] Between the lines, the overall impression was that, understandably enough, the U.S. government had resisted U.K. efforts and asserted itself to ensure it played the leading role in developing testing standards and compelling corporate developers to release test results for the most sensitive new systems.[153]

The sequencing of the Biden administration's new executive order, announced at the beginning of the week, and Harris's speech,[154] delivered on the first day of the U.K. conference—but pointedly separate from its proceedings—added to the impression that the U.S. administration was not exactly thrilled by the U.K.'s concept for the summit and its intended impact. In her speech, Harris covered the topics encompassed by the EO, emphasizing the need for action to address current problems, arguably more urgent than the long-term "existential" threats that appeared to have been the intended subject of the U.K.-hosted summit. Without overstating the significance of this apparent misalignment between the U.K. and U.S. governments, it underscores the difficulty of achieving wide international agreement when countries compete to position themselves to achieve national advantage. Arguably, the U.S.-U.K. misalignment perceived by some during the Bletchley summit highlighted the rapid pace and stark publicity of still on-going developments: a subsequent agreement, announced in April 2024, outlined the principles for the U.K. and U.S. AI safety institutes to collaborate.[155]

Given the current geopolitical tensions, it was widely seen as a positive outcome of the U.K.-hosted summit that China, the E.U., and the U.S. government signed the summit communique, the Bletchley Park Declaration. One report described the presence of U.S. and China officials on the same stage as "a major diplomatic coup for the U.K. government."[156] One of the reasons for this "coup" was likely that the brief communique was extremely bland and unobjectionable, "a statement of mission and purpose ... [that] did not contain specifics on how global cooperation could take shape."[157] Notwithstanding the apparent gap between description and reality—and uncertainty about the clarity of communications regarding the extent of China's involvement in the summit,[158] events like the U.K.-hosted summit can have value as a wider contribution to the slow, patient effort to build common understanding between states.

GEOPOLOTICS AND AI DIPLOMACY

It should be emphasized from the above brief survey that the United States, the U.K., and the E.U. and its member states have notably led these early efforts in global discussions about AI safety and governance. However, the broader "Global Majority" has started to complement these initiatives through the G20 and BRICS, for example.[159] These wider initiatives highlight concerns for the inclusive development and governance of AI, as do the multilateral processes under the UN umbrella. It is in this context, therefore, of competition to deliver globally inclusive outcomes—and achieve geopolitical influence—that we should interpret the commitments regarding AI for development, aligned to the Bletchley Park AI Safety Summit. The U.K. government pledged US$48 million to a US$100 million collaborative fund to assist the expansion of safe and responsible AI projects in the developing world, beginning in Africa.[160] These commitments played directly into the competitive geopolitics that form the context of global dialogue about AI safety and governance.

This proliferation of different AI-related forums and initiatives in recent years creates an additional burden on already-stretched foreign ministries. Not only must foreign ministries keep abreast of the specific processes and outcomes of these AI-related initiatives, but they must also identify the points of overlap between the proliferation of AI-related initiatives and the range of cyber-specific diplomatic activities—most notably the UN processes aimed at building norms of responsible state behavior in cyberspace and countering cybercrime. Given the potential impact of AI on cyberspace, foreign ministries need to ensure they can address the overlap in AI and cyber diplomacy. This is especially important when both AI and cyber diplomacy are conducted in the context of geopolitical competition.

HOW SHOULD FOREIGN MINISTRIES RESPOND TO THESE CHALLENGES?

The major challenge for foreign ministries is to organize themselves to participate effectively in this burgeoning global agenda of AI and cyber diplomacy. Effective participation will mean different things in varying national contexts. All states should consider how to achieve practical payoffs in the competition for strategic advantage. Foreign ministries have important roles, but so do several other departments, nongovernmental actors, and foreign partners. Achieving strategic advantage will require active coordination and relationship management within and beyond the national government and persistent, patient effort to achieve clear, realistic objectives.[161] It also requires consideration of where AI and cyber power should be situated in the list of national (and other states') foreign policy priorities. This is something a foreign ministry is uniquely placed to address in a coordinated, whole-of-government approach.

Close collaboration with allies and partners will be key, as are adversary-focused activities. These include managing and mitigating the risks of interdependence, good awareness and insight into what adversaries are doing, and planning on how best to counter them. There are many potential instruments to use, such as the U.S. Creating Helpful Incentives to Produce Semiconductors (CHIPS) and Science Act[162] or President Joseph R. Biden's Executive Order to restrict China's access to sensitive technology,[163] but they will require effective orchestration, within national governments and between states and their allies and partners.

As the Belfer Center has noted, "The asymmetric nature of cyber capabilities means smaller countries can punch above their weight, exerting more influence using cyber means than with traditional tools."[164] This is somewhat true, but rational smaller actors should think twice about punching too far above their weight too often, in case they invite retaliation from stronger, more capable actors. And however asymmetric the domain, the most sophisticated operations remain beyond the capability of weaker cyber powers.[165] The advent of AI-augmented cyber capabilities will, theoretically, increase the potential for smaller actors to achieve bigger effects than traditionally expected, up to a point.[166] The underlying logic of prudence that states should not provoke retaliation they cannot handle will not change. Forecasts, therefore, should consider not only the challenge posed by hostile state actors but also the possible AI-augmented challenges posed by non-state actors, for whom states are likely to need different instruments of deterrence and compulsion.

Building national AI and cyber power requires a thriving ecosystem of knowledge and innovation. This will mean different things in varying national contexts. But, in terms of foreign policy relevance, this will mean strategic investment in education and skills. It will also mean trying to cultivate the country's attractiveness as a magnet for globally mobile talent. Contemporary national cyber strategies increasingly include this. For example, recent U.K. strategies for cyber and AI recognize the need for the state to play an interventionist role in creating conditions for the nation to thrive in global competition (in producing and attracting talent and exploiting the commercial value of research).[167] But it is clear that, at present, the United States is leading the world in AI and cyber, and the Biden administration wants to extend that lead to include the emerging U.S. approach to AI regulation. This reality confronts other states that are trying to identify how best to find a niche for themselves to exploit the benefits and reduce the risks of AI.[168]

As science and technology success depends on a skilled workforce, national foreign ministries have an important role in amplifying the appeal of their national image internationally. Continued success would also depend on the domestic policies that would draw people to live and work in a given state. The efforts of one government to attract globally mobile talent do not exist in a vacuum, as other states would also seek limited global AI talent.[169] The outlook for any state depends to a large extent on whether it can increase the international appeal of its science and technology institutions.

Talent pipelines and a thriving ecosystem should co-exist with a rigorous focus on producing a detailed understanding of strategic competitors' relevant AI and cyber power activities. The specific question for foreign ministries is whether their existing open-source monitoring and research/analytical capacities would merit uplift and reconfiguration to contribute systematically to such a task and complement wider national government and allied efforts. Foreign Ministries must be nimble when internal capacity is limited and adept at accessing resources and expertise outside of government.

The geopolitics of strategic competition over emerging technologies, such as AI, is driving a shift, particularly in the United States but also elsewhere, toward risk management and reducing dependence on competitors and adversaries, chiefly China.[170] Concerns arise from the national security implications of foreign investment, from dependencies on supply chains, and from restricting access of strategic competitors to advanced technology and opportunities for research collaboration. This shift is particularly evident in the United States. For example, Microsoft President Brad Smith and co-author Carol Ann Browne recently counseled restraint, highlighting Microsoft's experience of the benefits of a global approach to technology development.[171]

It is difficult to forecast the implications of this shift on future advances in AI. It certainly appears that several Western, like-minded states are reevaluating the extent of science and technology collaboration with Chinese scientists and institutions.[172] Determining the net effect of decoupling from China on research is difficult. Isolating China may slow down its advances, but it has already generated reported retaliatory measures[173] and could conceivably create a context in which further innovation emerges from adversity.[174] Conversely, by reappraising bilateral research collaboration with China, states could find that this policy exerts a negative impact not only on China but also on its domestic research outcomes. While some collaborations might be too sensitive to pursue, policymakers should not lose sight of the bigger picture, namely that there are benefits as well as costs to such collaborations, and a careful risk management approach would be needed. And for those states that currently pursue a pragmatic policy of engaging with China and like-minded states, there is a constant need to reflect on their room to maneuver and the implications of their situation in the context of such geopolitical competition. Foreign ministries are likely crucial institutional contributors to that process.

CONCLUSION

A debate about AI policies and regulatory frameworks is now firmly on the global agenda. It is imperative that foreign ministries identify the consequential overlap between the proliferation of AI-related diplomatic processes and the parallel track of global cyber diplomacy. At times, this paper has presented the U.K. as an example of a government striving to accelerate its engagement with AI issues, most recently in hosting the AI Safety Summit in November 2023, but also in domestic debates about the extent of its ambition to improve national computing resources.[175] The U.K. example is presented critically with due regard for its limits as a case for emulation.

In some ways, the U.K. government's recent foray into global thought leadership on AI highlights the limits of what one state can realistically achieve. One way of interpreting the outcomes of the Bletchley Park summit, for example, is that there is more to gain from working within the already established, inclusive, and multistakeholder processes under the auspices of the UN and other multilateral organizations. Reducing duplication of effort is also more likely to conserve executive bandwidth and limited foreign ministry resources. With fewer resources to allocate to their foreign ministries, smaller states should advocate for the most efficient, streamlined, and globally inclusive approach to AI and cyber diplomacy.

Policy debates about AI's impact are broad and multifaceted. This paper has contributed to that wider debate by focusing on the implications of AI's impact on cyber power, particularly from the perspective of foreign ministries as institutional actors in the national strategy. With the rising salience of cyber diplomacy in recent years, this should be a prominent theme in contemporary arguments about the geopolitical impact of AI and the balance of power in cyberspace.

The paper addressed the connections between four crucial questions in the contemporary debate. First, it argues that AI and cyber power are interdependent. No one wants to be left behind in the race to secure a strategic advantage in AI or cyberspace.[176] Second, cautioning about the difficulty of producing a net assessment of AI's impact on the offense-defense balance in cyberspace, the paper argues that AI will be increasingly integrated into both cyber defensive and offensive operations.[177] Third, the paper argues that advances in AI are more likely to entrench than overturn existing asymmetries of cyber power between states. Temporary advantages are likely obtainable, and the dynamics of competition may increase the temptation for some states to adopt a greater appetite for risk when using AI-enabled capabilities. This has potential implications for stability and the risk of (unintended) escalation.

Finally, a major argument of this paper is that foreign ministries should continue to adapt to address the diplomatic and foreign policy challenges implicit in the interdependence of AI and cyber power. The challenges for AI-related cyber diplomacy largely mirror those of non-AI-related cyber diplomacy, from the continued fostering of norms of responsible behavior to the collaborative improvement of a broader capacity of efforts deterring irresponsible conduct. This is a crowded agenda. For example, the U.K. hosted the Bletchley Park AI Safety Summit, which was one of several global initiatives addressing the impact of AI.[178] Moving forward, states should pursue cooperation more effectively, avoiding unnecessary duplication of effort and mixed or contradictory messages. Foreign ministries should ensure that the proliferation of AI-related initiatives does not undermine the coherence between the separate but interdependent agendas of AI diplomacy and cyber diplomacy. Foreign ministries should organize themselves to effectively contribute to this effort. The following section presents recommendations for foreign ministries to address the implications of AI's impact on cyber power.[179]

POLICY RECOMMENDATIONS

Four policy recommendations flow from the analysis presented in this paper. First, as Foreign Ministries lead in managing the foreign policy implications of AI's impact on cyber power, they should ensure they are internally well-configured to play this role. This means breaking down silos between policy teams– as well as those that might exist between policy teams and other relevant functions, such as research and analysis. Growing investment in institutional capacity to develop AI/cyber policy should be complemented by a proportionate increase in supporting capacities, whether through direct recruitment or accessing external expertise. When resource and workforce constraints make this infeasible, foreign ministries should find ways to access external knowledge or capacity-building sources, for example, through engagement with relevant regional organizations.

Second, as foreign ministries are not the only institutional actors with a stake in the foreign policy implications of AI's impact on cyber power, they should ensure the national strategy allows all relevant stakeholders to participate in a coherent process. This is arguably easier for smaller states to achieve. The larger the cast of institutional actors, the greater the need for active leadership from the government center. However, small foreign ministries might also struggle to devote the necessary resources to what is essentially domestic relationship management.

Third, as effective diplomacy and foreign policy require a sound understanding of the context in which they are implemented, states should invest in a systematic process to monitor the cyber applications of AI, particularly by adversary states. By better understanding where the most relevant threats lie, states can prioritize effectively in crafting their responses. This paper does not prescribe one blueprint over others. Depending on the national or regional context, a joint unit, assessment center, or another format might work best. However, regardless of the bureaucratic form adopted, it is unlikely to succeed without sufficient prioritization and coordinated effort. Building on existing regional organizations with a track record of cyber diplomacy and capacity building—such as the OAS—would provide a sensible route to pooling effort, sharing the burden, and maximizing outcomes for all states.

Fourth, no individual state can address these issues alone. Solutions must be global, but many of the most plausible incremental gains in international cooperation are more likely to proceed at a lower level, away from the limelight and glitzy summit meetings. As noted in this paper, the United States, the U.K., and other states have already been developing strategic blueprints for integrating AI into national cyber diplomacy. It is increasingly necessary for all states to mainstream these considerations in the different strands of cyber diplomacy, including the global debate about norms and the capacity-building agenda. When capacity constraints make this difficult for some states, there is a clear role for cyber and AI capacity-building investments to enable more equitable and representative participation in global deliberations.

This paper has focused on the practice of state actors. The benefits of AI-enabled cyber capabilities are indeed available to state and non-state actors. For the latter, AI may lower the bar of what can be achieved with limited resources and expertise. Nonetheless, high-end cyber operations will remain the preserve of elite state actors. Responding to this competition to secure strategic advantage is a whole-of-society effort, with a clear imperative for governments to coordinate with both state and non-state actors.[180] It will also require strategic patience, as demonstrated by China's long-term state intervention to boost higher education in the Science, Technology, Engineering, and Mathematics disciplines and its apparent impact on the quantity and quality of China's AI and cyber-related research.[181] This patient approach to long-term investment in education and skills should apply beyond the hard sciences; effective regulation and responsible use of AI-enabled capabilities will require better understanding and multidisciplinary perspectives from the humanities and social sciences, for example. All this underlines that states must integrate domestic and foreign policy effectively in pursuing national strategies regarding AI and cyber power.

The most plausible near-future implications of AI for cyber power involve evolutionary changes in the potential scale and speed of cyber attacks and their identification and remediation. State and non-state actors will actively exploit these stages, provoking efforts to overturn or offset perceived advantages. Some states will advance faster than others. If these advances are not addressed swiftly, they will incrementally alter the balance of cyber power. Early adopters of AI advances can exploit these advantages, strengthening their position until adversaries catch up or counterbalance that advantage.[182] As such, it is best to view the competition for AI and cyber power advantage not as a one-off, "winner takes all" tournament but as an ongoing process with no foreseeable endpoint. Foreign ministries should plan for the long haul.

ABOUT THE AUTHOR

JOE DEVANNY

Dr. Joe Devanny is a Lecturer in the Department of War Studies at King's College London. He is also deputy director of the Centre for Defence Studies at King's. In 2022-23 he was a British Academy Innovation Fellow at the U.K. Government's Foreign, Commonwealth and Development Office, where he researched issues related to cyber diplomacy.

He is also a 2023-25 Project Fellow with the Research Institute for Sociotechnical Cyber Security. His research focuses on cyber operations and cyber diplomacy as instruments of statecraft, and more broadly on the relationship between technology and national security.

END NOTES

[1] The U.K. government uses a broad definition of cyber power as an instrument of national strategy. It encompasses the promotion of national security, the national interest, prosperity, and the effective diffusion of national values. It also includes the importance of having the operational capacity to pursue national interests in cyberspace and to achieve "real world" effects through cyberspace. Government of the United Kingdom, Global Britain in a competitive age: The Integrated Review of Security, Defence, Development and Foreign Policy (CP403) (2021): 40.

[2] Government of the United States, National Cybersecurity Strategy (2023), 23; Government of the United Kingdom, National Cyber Strategy: Pioneering a cyber future with the whole of the UK (2022), 78-89.

[3] For an extended elaboration of this argument, see: Joe Devanny and Andrew Dwyer, 'From Cyber Security to Cyber Power: Appraising the Emergence of "Responsible, Democratic Cyber Power" in UK Strategy,' in T. Jančárková, D. Giovannelli, K. Podiņš, & I. Winther (Eds.), 15th International Conference on Cyber Conflict Meeting Reality (International Conference on Cyber Conflict (CyCon)) (NATO Cooperative Cyber Defence Centre of Excellence (CCDCoE), 2023).

[4] Vincent Boulanin, "Artificial Intelligence: a primer," in The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk. Volume 1: Euro-Atlantic Perspectives, Vincent Boulanin, ed. (Stockholm: Stockholm International Peace Research Institute, 2019), 13, https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf.

[5] Government of the United States, "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy," Department of State, February 2, 2023, https://www.state.gov/wp-content/uploads/2023/10/Latest-Version-Political-Declaration-on-Responsible-Military-Use-of-AI-and-Autonomy.pdf.

[6] Government of the United Kingdom, National AI Strategy, September 2021, 16, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf.

[7] European Commission, White Paper: On Artificial Intelligence - A European approach to excellence and trust, February 19, 2020, 2, https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[8] Government of the United Kingdom, National AI Strategy, 10. For a theoretical and historical contextualization of the relationship between technological change and the competition for strategic advantage, see Manuel Castells, The Rise of the Networked Society 2E (Oxford: Wiley-Blackwell, 2010), 7.

[9] Nils Nilsson, The Quest for Artificial Intelligence: A History of Ideas and Achievements (Cambridge: Cambridge University Press, 2010).

[10] IBM (no date), "What is a neural network?," https://www.ibm.com/topics/neural-networks, accessed February 20, 2024.

[11] Matt Burgess, "Google's AI just created its own universal 'language'," Wired, November 23, 2016, https://www.wired.co.uk/article/google-ai-language-create.

[12] Models trained on different data types, such as text and images, are called large multimodal models. OpenAI's GPT-4 is an example. See "OpenAI, GPT-4," https://openai.com/research/gpt-4, accessed February 22, 2024.

[13] Government of the United Kingdom, Future Risks of Frontier AI (Government Office for Science, 2023), 2.

[14] Joanna Goodrich, "How IBM's Deep Blue Beat World Champion Chess Player Garry Kasparov," IEEE Spectrum, January 25, 2021, https://spectrum.ieee.org/how-ibms-deep-blue-beat-world-champion-chess-player-garry-kasparov; and Google DeepMind, "AlphaZero: Shedding new light o

15. James Gallagher, "New superbug-killing antibiotic discovered using AI," BBC News, May 25, 2023, https://www.bbc.co.uk/news/health-65709834.

16. Paul Scharre, Army of None: Autonomous Weapons and the Future of War (New York: W.W. Norton, 2019); and Kenneth Payne, I, Warbot: The Dawn of Artificially Intelligent Conflict (London: Hurst, 2021).

17. For a good summary of international efforts shaping the AI agenda, see Lea Kaspar, Maria Paz Canales, and Michaela Nakayama Shapiro, "Navigating the Global Governance Landscape," Global Partners Digital (October 31, 2023), https://www.gp-digital.org/navigating-the-global-ai-governance-landscape/.

18. For a truncated version of this argument, see Joe Devanny, "Commentary: Foreign Ministries and Cyber Power: Implications of Artificial Intelligence," Royal United Services Institute (July 21, 2023), https://www.rusi.org/explore-our-research/publications/commentary/foreign-ministries-and-cyber-power-implications-artificial-intelligence.

19. Madhumita Murgia, Andrew England, Qianer Liu, Eleanor Olcott, and Samir Al-Atrush, "Saudi Arabia and UAE race to buy Nvidia chips to power AI ambitions," Financial Times, August 14, 2023, https://www.ft.com/content/c93d2a76-16f3-4585-af61-86667c5090ba.

20. Joseph S. Nye, Jr., Cyber Power (Boston: Belfer Center for Science and International Affairs, 2010), 2.

21. André Barrinha and George Christou, "Speaking sovereignty: the EU in the cyber domain," European Security 31(3) (2022): 356-376; and Government of the United Kingdom, National Cyber Strategy, 11.

22. Geoffrey E. Hinton, Simon Osithero, and Yee-Whye Teh, "A fast learning algorithm for deep belief nets," Neural Computation 18 (7), (2006): 1527–54.

23. Boulanin, "Artificial Intelligence: a primer," 18.

24. International Telecommunication Union, Guide to Developing a National Cybersecurity Strategy 2E (2021): vii.

25. Louise Hurel and Joe Devanny, "Raising the Political Priority of Cybersecurity in Latin America," Council on Foreign Relations (March 16, 2023), https://www.cfr.org/blog/raising-political-priority-cybersecurity-latin-america.

26. Government of the United Kingdom, Global Britain in a Competitive Age: The Integrated Review of Security, Defence, Development and Foreign Policy (CP 403) (2021): 29.

27. One example is the United States. The National Artificial Intelligence Initiative, https://www.ai.gov/, accessed February 26, 2024. Another is the EU. Raluca Csernatoni and Katerina Mavrona, "The Artificial Intelligence and Cybersecurity Nexus: Taking Stock of the European Union's Approach," Carnegie Europe (2022), https://carnegieeurope.eu/2022/09/15/artificial-intelligence-and-cybersecurity-nexus-taking-stock-of-european-union-s-approach-pub-87886.

28. Government of the United Kingdom, Integrated Review Refresh 2023: Responding to a more contested and volatile world, March 13, 2023, https://www.gov.uk/government/publications/integrated-review-refresh-2023-responding-to-a-more-contested-and-volatile-world, 56.

29. Government of the United States, National Security Strategy, The White House, October 2022, 8, 21, 32-33, https://www.whitehouse.gov/wp-content/uploads/2022/10/Biden-Harris-Administrations-National-Security-Strategy-10.2022.pdf.

30. Historical trends are evident in Margaret O'Mara, The Code Silicon Valley and the Remaking of America (London: Penguin, 2019); and Nils Nilsson, The Quest for Artificial Intelligence: A History of Ideas and Achievements (Cambridge: Cambridge University Press, 2010). See also Janet Abbate, Inventing the Internet (Cambridge, MA: MIT Press, 1999). The Chinese model is different, seeing closer direction by the party-state. It is explored specifically from the perspective of AI in a recent book drawing on the wider research conducted by the U.S. think tank, Center for Security

anderging Technology, Chinese Power and Artificial Intelligence: Perspectives and Challenges, William C. Hannas and Huey-Meei Chang, eds. (Abingdon: Routledge, 2022).

31. Anthony Faiola and Cat Zakrzewski, “Governments used to lead innovation. On AI, they’re falling behind,” Washington Post, November 2, 2023, https://www.washingtonpost.com/technology/2023/11/02/ai-regulation-bletchley-park/.

32. Ian Bremmer and Mustafa Suleyman, “The AI Power Paradox: Can States Learn to Govern Artificial Intelligence—Before It’s Too Late?,” Foreign Affairs, August 16, 2023, https://reader.foreignaffairs.com/2023/08/16/the-ai-power-paradox/content.html; and Jason Matheny, “Opinion: Here’s a simple way to regulate powerful AI models,” Washington Post, August 16, 2023, https://www.washingtonpost.com/opinions/2023/08/16/ai-danger-regulation-united-states/.

33. Faiola and Zakrzewski, “Governments used to lead innovation.”

34. The private sector has a prudential self-interest and incentive to improve the security of its AI systems. For example, some companies fund vulnerability rewards programs (or “bug bounties”) to assist in discovery and remediation. Megan Crouse, “Google Offers Bug Bounties for Generative AI Security Vulnerabilities,” TechRepublic, October 31, 2023, https://www.techrepublic.com/article/google-ai-bug-bounty-program/.

35. Faiola and Zakrzewski, “Governments used to lead innovation.”

36. Does Technology Drive History?: The Dilemma of Technological Determinism, Merritt Roe Smith and Leo Marx, eds. (Cambridge, MA: MIT Press, 1999); and George Basalla, The Evolution of Technology (Cambridge: Cambridge University Press, 2010).

37. O’Mara, The Code; and Paul Edwards, The Closed World: Computers and the Politics of Discourse in Cold War America (Cambridge, MA: MIT Press, 1996).

38. Nilsson, The Quest for Artificial Intelligence.

39. Micah Musser and Ashton Garriott, Machine Learning and Cybersecurity: Hype and Reality (Washington, D.C.: Center for Security and Emerging Technology, 2021). Although AI is not a new phenomenon in cybersecurity, there are novel aspects to applying increasingly powerful machine learning: “Whereas the automatization of defensive cyber actions is hardly new, AI/ML are, in the sense of technology which produces an output for a given input without allowing reconstruction of the digital reasoning process or the line of thought of the machine or software that led to a specific decision. This creates situations in which the code produces decisions that are no longer deducible and thus prevent humans from intervening based on reasoning. When such AI/ML-enabled measures are used for offensive actions, this creates serious problems in connection with the necessary human integration and interaction.” See Thomas Reinhold and Christain Reuter, “Cyber Weapons and Artificial Intelligence: Impact, Influence and the Challenges for Arms Control,” in Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm, T. Reinhold and N. Schornig, eds. (New York: Springer, 2022), 150.

40. Musser and Garriott, Machine Learning and Cybersecurity.

41. Ben Schreckinger, “The future of AIs fighting AIs is already here,” Politico Digital Future Daily, October 31, 2023, https://www.politico.com/newsletters/digital-future-daily/2023/10/31/the-future-of-ais-fighting-ais-is-already-here-00124568.

42. Williamson Murray, Military Adaptation in War (Cambridge: Cambridge University Press, 2011).

43. Paul Scharre, Four Battlegrounds: Power in the Age of Artificial Intelligence (New York: W.W. Norton & Company, 2023).

44. In the recent words of three writers particularly influential in the recent turn toward “Defend Forward” and “Persistent Engagement” in the U.S. Department of Defense cyber strategy: “At this time, all that can be offered regarding potential AI cyberspace futures is conjecture.” Michael P. Fischerkeller, Emily O. Goldman, and Richard J. Harknett, Cyber Persistence Theory: Redefining National Security in Cyberspace (Oxford: Oxford University Press, 2022), 54.

45. This point has been argued persuasively as a critique "offense-defense balance" in warfare. Stephen Biddle, "Rebuilding the Foundations of Offense-Defense Theory," The Journal of Politics 63, no. 3 (August 2001): 741-774. See also: Stephen Biddle, Military Power: Explaining Victory and Defeat in Modern Battle (Princeton: Princeton University Press, 2004). For a critical assessment of the applicability of the concept of offense-defense balance to cyberspace, see Rebecca Slayton, "What Is the Cyber Offense-Defense Balance?," International Security 41, no. 3 (Winter 2016): 72-109. Note also Avi Goldfarb and Jon Lindsay's argument: "It is reasonable to expect organizational and strategic context to condition the performance of automated systems, as with any other information technology." Avi Goldfarb and Jon R. Lindsay, "Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War," International Security 46, no. 3 (Winter 2021): 10.

46. Although big tech companies maintain an advantage, there has been a significant investment in AI start-ups in recent years. Joe McKendrick, "Pace Of Artificial Intelligence Investments Slows, But AI Is Still Hotter Than Ever," Forbes, October 15, 2022, https://www.forbes.com/sites/joemckendrick/2022/10/15/pace-of-artificial-intelligence-investments-slows-but-ai-is-still-hotter-than-ever/?sh=6a8dc5f94c76.

47. Kate Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence (New Haven: Yale University Press, 2021), 181-209.

48. Thomas Reinhold, "Arms Control for Artificial Intelligence," in Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm, T. Reinhold and N. Schornig, eds. (New York: Springer, 2022), 222-225.

49. For a recent exploration of nuclear arms control analogies, see Mauricio Baker, "Nuclear Arms Control Verification and Lessons for AI Treaties," Arxiv.org (April 8, 2023), https://arxiv.org/abs/2304.04123. For a broader historical and theoretical perspective on the impact of context and contingency on the evolution of cyber norms, see Brian M. Mazanec, The Evolution of Cyber War: International Norms for Emerging-Technology Weapons (Washington, D.C.: Potomac Books, 2015).

50. Brandon Valeriano, "The Failure of Offense/Defense Balance in Cyber Security," Cyber Defense Review 7, no. 3 (2022): 96.

51. The five phases are: identification, protection, detection, response, and recovery. See Government of the United States, "Cybersecurity Framework," U.S. National Institute for Standards and Technology, https://www.nist.gov/cyberframework. The NIST framework is widely used, but other conceptualisations exist that focus more on the perspective of the adversary. See, for example: Jorge Orchilles, "Cyber Kill Chain, MITRE ATT&CK, and Purple Team," SANS Institute, March 24, 2022, https://www.sans.org/blog/cyber-kill-chain-mitre-attack-purple-team/.

52. Government of the United States, "A Common Cyber Threat Framework: A Foundation for Communication," Director of National Intelligence, accessed April 2, 2024, https://nsarchive.gwu.edu/document/22929-office-director-national-intelligence. See also Daniel Moore, Offensive Cyber Operations: Understanding Intangible Warfare (London: Hurst, 2022), 75.

53. Fischerkeller, Goldman, and Harknett, Cyber Persistence Theory, 86. For these authors, operational activities can effectively establish tacit bounds of agreed competition between states in cyberspace. The dynamic of action and response would form a kind of communication between actors, clarifying the boundaries of acceptability. The problem with such tacit guardrails is the implicit risk of miscommunication, which is arguably likely exacerbated in a future context of more automated operations.

54. For an interesting discussion of the scenarios in which humans could lose control of AI, see Karl Von Wendt, "Paths to failure," Less Wrong (April 25, 2023), https://www.lesswrong.com/posts/yv4xAnkEyWvpXNBte/paths-to-failure. The risks posed by the premature use of insufficiently understood AI tools are not unique to cyberspace. Regarding the financial risks of AI, see Felix Salmon, "AI will be at the center of the next financial crisis, SEC chair warns," Axios, August 12, 2023, https://www.axios.com/2023/08/12/artificial-intelligent-stock-market-algorithms.

55. Michele Flournoy, “AI is Already at War: How Artificial Intelligence Will Transform the Military,” Foreign Affairs (November/December 2023), https://reader.foreignaffairs.com/2023/10/24/ai-is-already-at-war/content.html.

56. Nicholas Wright, “How Artificial Intelligence Will Reshape the Global Order: The Coming Competition Between Digital Authoritarianism and Liberal Democracy,” Foreign Affairs, 2018, https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order.

57. There is a reference to cyber power in a 2009 House of Lords debate on nuclear deterrence. It was related to the cyber vulnerabilities of the deterrent’s command and control infrastructure. It was made by the late Baroness Shirley Williams, and was very possibly a case of direct Transatlantic transmission from Joseph Nye, given Williams’s long association with the Harvard Kennedy School. See Baroness Williams, “Nuclear Proliferation,” Hansard, vol. 709, col. 790 (March 2009).

58. See Andrew Dwyer, Workshop Report: Crafting a democratic and responsible cyber power? Offensive Cyber Working Group, August 22, 2022; and Joe Devanny and Andrew Dwyer, “From Cyber Security to Cyber Power: Appraising the Emergence of ‘Responsible, Democratic Cyber Power’ in UK Strategy,” in 15th International Conference on Cyber Conflict Meeting Reality (International Conference on Cyber Conflict), T. Jančárková, D. Giovannelli, K. Podiņš, and I. Winther, eds., NATO Cooperative Cyber Defence Centre of Excellence, 2023.

59. Government of the United Kingdom, Global Britain in a competitive age, 40.

60. Government of the United Kingdom, National Cyber Strategy 2022, 125.

61. Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher, The Age of AI and our human future (London: John Murray, 2022), 50.

62. Bremmer and Suleyman, “The AI Power Paradox: Can States Learn to Govern Artificial Intelligence.

63. Robert A. Dahl, “The concept of power,” Behavioral Science 2, vol. 3 (1957): 201–215.

64. Joe Devanny, “The Review and Responsible, Democratic Cyber Power,” in The Integrated Review in Context: Defence and Security in Focus (London: Centre for Defence Studies, 2021): 62-64; and Nick Beecroft, “The UK’s Cyber Strategy Is No Longer Just About Security,” Carnegie Endowment for International Peace, December 17, 2021.

65. A recent example from the U.K. is the effort by the National Cyber Force to improve public understanding of its activities while advocating for the U.K. government’s strategic narrative about responsible cyber power. National Cyber Force, The National Cyber Force: Responsible Cyber Power in Practice (April 4, 2023), https://www.gov.uk/government/publications/responsible-cyber-power-in-practice/responsible-cyber-power-in-practice-html.

66. For more context on this argument, see Reinhold, “Arms Control for Artificial Intelligence,” 211.

67. Julia Voo, Irfan Hemani, Simon Jones, Winnona DeSombre, Dan Cassidy, and Anina Schwarzenbach, National Cyber Power Index 2020 (Boston: Belfer Center for Science and International Affairs, 2020); Julia Voo, Irfan Hemani, and Daniel Cassidy, National Cyber Power Index 2022 (Boston: Belfer Center for Science and International Affairs, 2022).

68. Voo et al. National Cyber Power Index 2020, 1.

69. Russia was elevated ahead of the U.K. in the 2022 Index, apparently on the (in the author’s view) basis that the increase in observable malicious activity attributed to Russia during the index’s assessment period justified the move. Alternative views, such as that of Marcus Willetts, contend that Russian cyber operations during the 2022 conflict with Ukraine should lead to a critical reappraisal of assumptions about Russian prowess in military cyber operations. See Marcus Willetts, “The Cyber Dimension of the Russia–Ukraine War,” International Institute for Strategic Studies (October 6, 2022), https://www.iiss.org/blogs/survival-blog/2022/10/the-cyber-dimension-of-the-russia-ukraine-war.

70. This awkward name results from a (still) controversial merger in 2020 of the Foreign and

Commonwealth Office with the Department for International Development. Historically, it is attributable to a decision to highlight the prominence of the Commonwealth in U.K. foreign policy, which was itself the result of an earlier merger of the then-separate Foreign Office and Commonwealth Office in 1968.

71. This argument is explained more fully in Devanny and Dwyer, "From Cyber Security to Cyber Power: Appraising the Emergence of 'Responsible, Democratic Cyber Power' in UK Strategy."

72. Markus Anderljung and Paul Scharre, "How to Prevent AI Catastrophe: Society Must Get Ready for Very Powerful Artificial Intelligence," Foreign Affairs, August 14, 2023, https://www.foreignaffairs.com/world/how-prevent-ai-catastrophe-artificial-intelligence; and Tom Bristow, Gian Volpicelli, and Laurie Clark, "Rishi Sunak promised big on his AI summit. He's running out of time to deliver," Politico, August 15, 2023, https://www.politico.eu/article/uk-pm-rishi-sunak-promise-big-ai-summit-but-running-out-of-time/.

73. This is a particular risk given that (1) Various agencies and departments will understandably start from different viewpoints about specific aspects of national policy, and (2) The more operational actors will have sensitive equities to defend. This simply underscores the need for a coordinated approach.

74. William J. Burns, The Back Channel: American Diplomacy in a Disordered World (London: Hurst, 2021), 406.

75. Between 2012 and 2018, the computing used in AI training grew exponentially (Open AI, "AI and Compute," May 16, 2018, https://openai.com/blog/ai-and-compute/). At this time, private investment flooded into AI start-ups, with global AI funding growing from US$670 million (2011) to US$69.5 billion (2020) to US$111.4 billion (2021), with 2022 falling between 2020 and 2021 levels. See also Joe McKendrick, "Pace Of Artificial Intelligence Investments Slows, But AI Is Still Hotter Than Ever," Forbes, October 15, 2022, https://www.forbes.com/sites/joemckendrick/2022/10/15/pace-of-artificial-intelligence-investments-slows-but-ai-is-still-hotter-than-ever/?sh=6a8dc5f94c76).

76. "Battle of the Boffins: the race of the AI labs heats up," The Economist, January 30, 2023, https://www.economist.com/business/2023/01/30/the-race-of-the-ai-labs-heats-up; and Pablo Villalobos, Jaime Sevilla, Lennart Heim, Tamay Besiroglu, Marius Hobbhahn, and Anson Ho, "Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning," Epoch, 2022, https://arxiv.org/pdf/2211.04325.pdf.

77. Christian Vasquez, "Top US cyber official warns AI may be the 'most powerful weapon of our time'," Cyberscoop, May 5, 2023, https://cyberscoop.com/easterly-warning-weapons-artificial-intelligence-chatgpt/; Gerrit De Vynck, "AI leaders warn Congress that AI could be used to create bioweapons," Washington Post, July 25, 2023, https://www.washingtonpost.com/technology/2023/07/25/ai-bengio-anthropic-senate-hearing/; and Gerrit DeVynck, "Google's AI ambassador walks a fine line between hype and doom," Washington Post, August 9, 2023, https://www.washingtonpost.com/technology/2023/08/09/google-james-manyika-ai-existential-threat/.

78. Mark Muro, Julian Jacobs, and Sifan Liu, "Building AI cities: How to spread the benefits of an emerging technology across more of America," Brookings, July 20, 2023, https://www.brookings.edu/articles/building-ai-cities-how-to-spread-the-benefits-of-an-emerging-technology-across-more-of-america/. AI has the potential to benefit people everywhere, but the corporate boom in AI-related research has disproportionately benefited a relatively small number of places with high-tech clusters. See Ryan Heath, "AI boom's big winners are all in four states," Axios, August 10, 2023, https://www.axios.com/2023/07/24/ai-goldrush-concentrated-4-states.

79. Alexander Wan, Eric Wallace, Sheng Shen, and Dan Klein, "Poisoning Language Models During Instruction Tuning," Proceedings of the 40th International Conference on Machine Learning, 2023, https://arxiv.org/pdf/2305.00944.pdf; Florian Tramèr, Reza Shokri, Ayrton San Joaquin, Hoang Le, Matthew Jagielski, Sanghyun Hong, and Nicholas Carlini, "Truth Serum: Poisoning Machine Learning Models to Reveal Their Secrets," Arvix, 2022, https://arxiv.org/pdf/2204.00032.pdf; and "It

doesn't take much to make machine-learning algorithms go awry," The Economist, April 5, 2023, https://www.economist.com/science-and-technology/2023/04/05/it-doesnt-take-much-to-make-machine-learning-algorithms-go-awry.

80. Musser and Garriott, Machine Learning and Cybersecurity, 36.

81. To the extent that measuring evolutionary performance is possible in this competition. Jose David Mireles, Eric Ficke, Jin-Hee Cho, Patrick Hurley, and Shouhuai Xu, "Metrics Towards Measuring Cyber Agility," Arvix, June 12, 2019, https://arxiv.org/pdf/1906.05395.pdf.

82. Hammond Pearce, Baleegh Ahmad, Benjamin Tan, Brendan Dolan-Gavitt, and Ramesh Karri, Asleep at the Keyboard? Assessing the Security of GitHub Copilot's Code Contributions, Arvix, December 16, 2021, https://arxiv.org/pdf/2108.09293.pdf.

83. Elias Groll, "ChatGPT shows promise of using AI to write malware," CyberScoop News, December 6, 2022, https://www.cyberscoop.com/chatgpt-ai-malware/.

84. Will Oremus, "The clever trick that turns ChatGPT into its evil twin," Washington Post, February 14, 2023, https://www.washingtonpost.com/technology/2023/02/14/chatgpt-dan-jailbreak/.

85. Tom Warren, "Microsoft Security Copilot is a new GPT-4 AI assistant for cybersecurity," The Verge, March 28, 2023, https://www.theverge.com/2023/3/28/23659711/microsoft-security-copilot-gpt-4-ai-tool-features.

86. For instances of subverting existing tools and creating chatbots explicitly to assist cybercrime, see Will Knight, "A New Attack Impacts Major AI Chatbots—and No One Knows How to Stop It," Wired, August 1, 2023, https://www.wired.com/story/ai-adversarial-attacks/; Matt Burgess, "Criminals Have Created Their Own ChatGPT Clones," Wired, August 7, 2023, https://www.wired.co.uk/article/chatgpt-scams-fraudgpt-wormgpt-crime; Andy Zou, Zifan Wang, J. Zico Kolter, and Matt Fredrikson, "Universal and Transferable Adversarial Attacks on Aligned Language Models," Arxiv, July 27, 2023, https://arxiv.org/abs/2307.15043; Chenta Lee, "Unmasking hypnotized AI: The hidden risks of large language models," Security Intelligence, August 8, 2023, https://securityintelligence.com/posts/unmasking-hypnotized-ai-hidden-risks-large-language-models; and Will Oremus, "The 'red team' race to make AI go rogue," Washington Post, August 8, 2023, https://www.washingtonpost.com/technology/2023/08/08/ai-red-team-defcon/.

87. There has long been a low barrier to conducting relatively low-impact operations, but this should not be confused with the obstacles to achieving significant effects in cyberspace. Sophisticated operations such as Stuxnet most likely required a substantial investment of time, resources, and expertise. It is not unreasonable to expect future advances in AI to "democratize" the task of writing sophisticated malware, but this still leaves many stages of the cyber "kill chain." High-end cyber operations are likely to remain an elite activity.

88. Government of the United Kingdom, National Cyber Strategy (2022): 102.

89. Hannah Murphy, "Machine-on-machine cyber defence edges closer," Financial Times, November 7, 2022, https://www.ft.com/content/a978b8a6-ebc5-4929-ac63-6be7acb7d738.

90. Government of the United Kingdom, Pioneering a New National Security: The Ethics of Artificial Intelligence, Government Communications Headquarters, 2021, 6.

91. Musser and Garriott, Machine Learning and Cybersecurity, 4.

92. Musser and Garriott, Machine Learning and Cybersecurity, 8.

93. Musser and Garriott, Machine Learning and Cybersecurity, 7-9; and Yang Xin et al., "Machine Learning and Deep Learning Methods for Cybersecurity," IEEE Access 6 (2018): 35365–81.

94. Musser and Garriott, Machine Learning and Cybersecurity, 10.

95. Musser and Garriott, Machine Learning and Cybersecurity, 10.

96. William Blum, "Neural fuzzing: applying DNN to software security testing," Microsoft Research Blog, November 13, 2017, https://www.microsoft.com/en-us/research/blog/neural-fuzzing/; Nicole

Nichols, Mark Raugas, Robert Jasper, and Nathan Hilliard, "Faster Fuzzing: Reinitialization with Deep Neural Models," Working Paper (November 8, 2017), https://arxiv.org/pdf/1711.02807.pdf; and Yunchao Wang et al., "NeuFuzz: Efficient Fuzzing With Deep Neural Network," IEEE Access 7, 36340–52.

97. Musser and Garriott, Machine Learning and Cybersecurity, 17.

98. William Dixon and Nicole Eagan, "3 ways AI will change the nature of cyber attacks," World Economic Forum (June 19, 2019), https://www.weforum.org/agenda/2019/06/ai-is-powering-a-new-generation-of-cyberattack-its-also-our-best-defence/.

99. Blessing Guembe, Ambrose Azeta, Sanjay Misra, Victor Chukwudi Osamor, Luis Fernandez-Sanz, and Vera Pospelova, "The Emerging Threat of AI-driven Cyber Attacks: A Review," Applied Artificial Intelligence 36, no. 1 (2022), https://doi.org/10.1080/08839514.2022.2037254.

100. Musser and Garriott, Machine Learning and Cybersecurity, 25.

101. Csernatoni and Mavrona, "The Artificial Intelligence and Cybersecurity Nexus."

102. Lennart Maschmeyer, "Subverting Skynet: The Strategic Promise of Lethal Autonomous Weapons and the Perils of Exploitation," in 14th International Conference on Cyber Conflict: Keep Moving, T. Jancarkova, G. Visky, and I. Winther, eds. (Tallinn: NATO CCD COE, 2022), 155-171.

103. Forrest E. Morgan, Benjamin Boudreaux, Andrew J. Lohn, Mark Ashby, Christian Curriden, Kelly Klima, and Derek Grossman, Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World (Santa Monica: RAND, 2020); and James Johnson, "Automating the OODA loop in the age of intelligent machines: reaffirming the role of humans in command-and-control decision-making in the digital age," Defence Studies 23, no. 1, (2023): 43-67, DOI: 10.1080/14702436.2022.2102486.

104. Wyatt Hoffman, AI and the Future of Cyber Competiton (Washington, D.C.: Center for Security and Emerging Technology, 2021): 2.

105. Liu Yangyue, "The role of artificial intelligence in cyber-deterrence," in The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume II East Asian Perspectives, Lora Saalman, ed. (Stockholm: Stockholm International Peace Research Institute, 2019): 20-23.

106. Fischerkeller, Goldman, and Harknett, Cyber Persistence Theory, 53-55.

107. Fischerkeller, Goldman, and Harknett, Cyber Persistence Theory, 36, 61.

108. Robert Jervis, "Some thoughts on deterrence in the cyber era," Journal of Information Warfare 15, no. 2 (2016): 66-73. An alternative perspective is offered in Avi Goldfarb and Jon R. Lindsay, "Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War," International Security 46, no. 3 (Winter 2021): 7–50.

109. Payne, I, Warbot: The Dawn of Artificially Intelligent Conflict, 187-188. For the RAND report, see Yuna Huh Wong, John Yurchak, Robert W. Button, Aaron Frank, Burgess Laird, Osonde A. Osoba, Randall Steeb, Benjamin N. Harris, and Sébastian Joon Bae, Deterrence in the Age of Thinking Machines (Santa Monica: RAND Corporation, 2020).

110. Paul Scharre, Army of None: Autonomous Weapons and the Future of War (New York: W.W. Norton, 2019).

111. Lea Kaspar, Maria Paz Canales, and Michaela Nakayama Shapiro, "Navigating the Global Governance Landscape," Global Partners Digital (October 31, 2023), https://www.gp-digital.org/navigating-the-global-ai-governance-landscape/.

112. Melissa Heikkilä, 'The AI Act is done. Here's what will (and won't) change,' MIT Technology Review (March 19 2024), https://www.technologyreview.com/2024/03/19/1089919/the-ai-act-is-done-heres-what-will-and-wont-change/.

113. Raluca Csernatoni, "Generative AI Poses Challenges for Europe," Carnegie Europe (October 19, 2023), https://www.carnegieeurope.eu/strategiceurope/90803.

114. Cat Zakrzewski, “Vice President Harris to unveil AI safety plans in U.K. speech,” Washington Post, November 1, 2023, https://www.washingtonpost.com/technology/2023/11/01/ai-kamala-harris-summit/.

115. Matthijs M. Maas and Jose Jaime Villalobos, “International AI Institutions: A Literature Review of Models, Examples and Proposals,” AI Foundations Report 1 (October 25, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4579773.

116. United Nations, “Secretary-General Announces Creation of New Artificial Intelligence Advisory Board,” October 26, 2023, https://press.un.org/en/2023/sga2236.doc.htm.

117. United Nations, “Our Common Agenda: Report of the Secretary-General,” 2021: 63, https://www.un.org/en/content/common-agenda-report/assets/pdf/Common_Agenda_Report_English.pdf.

118. United Nations, Our Common Agenda, 63.

119. Jason Hausenloy and Claire Dennis, “Towards a UN Role in Governing Foundation Artificial Intelligence Models,” UN University Centre for Policy Research (September 9, 2023): 3-4, https://unu.edu/cpr/working-paper/towards-un-role-governing-foundation-artificial-intelligence-models.

120. Organisation of Economic Co-operation and Development, “Recommendation of the Council on Artificial Intelligence,” OECD/LEGAL/0449 (May 22, 2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

121. Group of 7/Group of 20 Documents Database, ‘G7 Leaders’ Statement on the Hiroshima AI Process’ (30 October 2023), https://g7g20-documents.org/database/document/2023-g7-japan-leaders-leaders-language-g7-leaders-statement-on-the-hiroshima-ai-process.

122. European Commission, ‘G7 Leaders’ Statement on the Hiroshima AI Process’ (30 October 2023), https://digital-strategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-ai-process.

123. Council of Europe, “Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law,” Committee on AI, 2023, https://rm.coe.int/cai-2023-01-revised-zero-draft-framework-convention-public/1680aa193f.

124. In sections of the communique entitled “Harnessing AI responsibly for good and for all” and “G20 New Delhi Leaders’ Declaration,” G20 (September 9-10, 2023): 24, https://www.g20.org/content/dam/gtwenty/gtwenty_new/document/G20-New-Delhi-Leaders-Declaration.pdf.

125. Ephraim Modise, “BREAKING: BRICS nations to form study group to monitor AI,” Tech Cabal (August 23, 2023), https://techcabal.com/2023/08/23/ai-study-group-brics/.

126. Government of the People’s Republic of China, “Foreign Ministry Spokesperson’s Remarks on the Global AI Governance Initiative,” Ministry of Foreign Affairs, October 18, 2023, https://www.fmprc.gov.cn/eng/xwfw_665399/s2510_665401/202310/t20231018_11162874.html.

127. Arguably of necessity, pending evidence of legislative momentum on these issues.

128. Alex Engler, “The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment,” Brookings (April 25, 2023), https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/.

129. Government of the United States, “FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence,” The White House, October 30, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/; and Carlos Lima and Cat Zakrzewski, “Biden signs AI executive order, the most expansive regulatory attempt yet,” Washington Post, October 30, 2023, https://www.washingtonpost.com/technology/2023/10/30/biden-artificial-intelligence-executive-order/.

130. Government of the United States, “FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.”

131. Emilia David, “Biden releases AI executive order directing agencies to develop safety guidelines,”

The Verge, October 30, 2023, https://www.theverge.com/2023/10/30/23914507/biden-ai-executive-order-regulation-standards.

132. Government of the United States, "FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence."

133. Government of the United States, "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy," U.S. Department of State, November 13, 2023, https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/.

134. James A. Lewis, Emily Benson, and Michael Frank, "The Biden Administration's Executive Order on Artificial Intelligence," Center for Strategic and International Studies (October 31, 2023), https://www.csis.org/analysis/biden-administrations-executive-order-artificial-intelligence.

135. Josh Tyrangiel, "Opinion: Gina Raimondo is the secret MVP of Biden's sweeping new executive order," Washington Post, October 31, 2023, https://www.washingtonpost.com/opinions/2023/10/31/ai-gina-raimondo-is-steph-curry/.

136. Megan Crouse, "White House Executive Order on AI Provides Guidelines for AI Privacy and Safety," TechRepublic, October 30, 2023, https://www.techrepublic.com/article/white-house-executive-order-ai-privacy/.

137. Krista Chavez, "Biden Releases AI Red Tape Wishlist in New Executive Order," NetChoice, October 30, 2023, https://netchoice.org/biden-releases-ai-red-tape-wishlist-in-new-executive-order/; and Simon Sinofsky, "Regulating AI by Executive Order is the Real AI Risk," Hardcore Software Substack (November 1, 2023), https://hardcoresoftware.learningbyshipping.com/p/211-regulating-ai-by-executive-order.

138. Lewis, Benson, and Frank, "The Biden Administration's Executive Order on Artificial Intelligence."

139. Lewis, Benson, and Frank, "The Biden Administration's Executive Order on Artificial Intelligence."

140. Cade Metz and Gregory Schmidt, "Elon Musk and Others Call for Pause on A.I., Citing 'Profound Risks to Society'," The New York Times, March 29, 2023, https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html.

141. Esther Webber, "UK to host major AI summit of 'like-minded' countries," Politico, June 7, 2023, https://www.politico.eu/article/u-k-to-host-major-ai-summit-of-like-minded-countries/.

142. Vincent Manancourt and Gian Volpicelli, "UK to pitch new 'AI Safety Institute' to allies," Politico, October 3, 2023, https://www.politico.eu/article/uk-pitch-ai-safety-institute-rishi-sunak/.

143. Government of the United Kingdom, "The Bletchley Declaration by Countries Attending the AI Safety Summit," November 1-2, 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

144. "The Guardian view on AI regulation: the threat is too grave for Sunak's light-touch approach," The Guardian, November 1, 2023, https://www.theguardian.com/commentisfree/2023/nov/01/the-guardian-view-on-ai-regulation-the-threat-is-too-grave-for-sunaks-light-touch-approach; Kiran Stacey and Dan Milmo, "U.K., U.S., E.U. and China sign declaration of AI's 'catastrophic' danger," The Guardian, November 1, 2023, https://www.theguardian.com/technology/2023/nov/01/uk-us-eu-and-china-sign-declaration-of-ais-catastrophic-danger.

145. Devanny, "Foreign Ministries and Cyber Power; and Kenneth Payne, "Thoughts ahead of Bletchley," Ken's Substack (October 24, 2023), https://www.kennethpayne.uk/p/thoughts-ahead-of-bletchley.

146. Peter Guest, "Britain's Big AI Summit Is a Doom-Obsessed Mess," Wired, October 23, 2023, https://www.wired.co.uk/article/britains-ai-summit-doom-obsessed-mess.

147. Kaspar, Canales, and Shapiro, "Navigating the Global AI Governance Landscape."

148. Cat Zakrzewski, Anthony Faiola, and Gerard De Vynck, "World leaders are gathering at the U.K.'s AI

Summit. Doom is on the agenda," Washington Post, October 31, 2023, https://www.washingtonpost.com/technology/2023/10/31/uk-ai-safety-summit-rishi-sunak-elon-musk/.

149. Guest, "Britain's Big AI Summit Is a Doom-Obsessed Mess."

150. Guest, "Britain's Big AI Summit Is a Doom-Obsessed Mess."

151. Zakrzewski, Faiola, and De Vynck, "World leaders are gathering at the U.K.'s AI Summit."

152. Cat Zakrzewski, "Vice President Harris to unveil AI safety plans in U.K. speech," Washington Post, November 1, 2023, https://www.washingtonpost.com/technology/2023/11/01/ai-kamala-harris-summit/; Vincent Manancourt, Eugene Daniels, and Brendan Bordelon, "Existential to who?' US VP Kamala Harris urges focus on near-term AI risks," Politico, November 1, 2023, https://www.politico.eu/article/existential-to-who-us-vp-kamala-harris-urges-focus-on-near-term-ai-risks/; and Vincent Manancourt and Eugene Daniels, "Kamala Harris seizes agenda as Rishi Sunak's AI summit kicks off," Politico, November 1, 2023, https://www.politico.eu/article/us-vp-kamala-harris-seizes-agenda-as-rishi-sunaks-ai-summit-kicks-off/.

153. Manancourt and Daniels, "Kamala Harris seizes agenda as Rishi Sunak's AI summit kicks off;" and Vincent Manancourt, Eugene Daniels, and Annabelle Dickson, "U.K., U.S. slated to announce AI safety partnership," Politico, November 1, 2023, https://www.politico.eu/article/uk-us-slated-to-announce-ai-safety-partnership/.

154. "Vice President Harris Remarks in London on Artificial Intelligence," C-SPAN, November 1, 2023, https://www.c-span.org/video/?531550-1/vice-president-harris-remarks-london-artificial-intelligence.

155. Government of the United States, 'U.S. and UK Announce Partnership on Science of AI Safety,' Department of Commerce (1 April 2024), https://www.commerce.gov/news/press-releases/2024/04/us-and-uk-announce-partnership-science-ai-safety.

156. Mark Scott, Tom Bristow, and Gian Volpicelli, "US and China join global leaders to lay out need for AI rulemaking," Politico, November 1, 2023, https://www.politico.eu/article/artificial-intelligence-safety-summit-rulemaking-us-china/.

157. Cat Zakrzewski and Anthony Faiola, "Global adversaries, allies reach first agreement on containing AI risks," Washington Post, November 1, 2023, https://www.washingtonpost.com/technology/2023/11/01/ai-kamala-harris-summit/.

158. Brenda Goh and Paul Sandle, "Exclusive: China took part in leaders' AI meeting even though UK did not acknowledge," Reuters, November 3, 2023, https://www.reuters.com/technology/china-took-part-leaders-ai-meeting-even-though-uk-did-not-acknowledge-2023-11-03/.

159. Kaspar, Canales, and Shapiro, "Navigating the Global Governance Landscape."

160. Government of the United Kingdom, "UK unites with global partners to accelerate development using AI," Foreign, Commonwealth and Development Office, November 1, 2023, https://www.gov.uk/government/news/uk-unites-with-global-partners-to-accelerate-development-using-ai.

161. Devanny, "The Review and Responsible, Democratic Cyber Power," 64.

162. Government of the United States, FACT SHEET: "CHIPS and Science Act Will Lower Costs, Create Jobs, Strengthen Supply Chains, and Counter China," The White House, August 9, 2022, https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/.

163. "Biden orders restrictions on U.S. investments in Chinese technology," Associated Press, August 9, 2023, https://www.npr.org/2023/08/09/1193013362/biden-executive-order-restricts-investments-china-tech.

164. Voo et al., Reconceptualizing Cyber Power, 1.

165. Rebecca Slayton, "What Is the Cyber Offense-Defense Balance?," International Security 41, no. 3 (Winter 2016): 72-109.

166. This fear is evident in recent commentary about the urgent need for states to regulate AI. See, for example: Bremmer and Suleyman: "The AI Power Paradox: Can States Learn to Govern Artificial Intelligence.

167. Government of the United Kingdom, National AI Strategy, 2021, 10.

168. Cat Zakrzewski, "Vice President Harris to unveil AI safety plans in U.K. speech," Washington Post, November 1, 2023, https://www.washingtonpost.com/technology/2023/11/01/ai-kamala-harris-summit/; and Cat Zakrzewski and Anthony Faiola, "Global adversaries, allies reach first agreement on containing AI risks," Washington Post, November 1, 2023, https://www.washingtonpost.com/technology/2023/11/01/ai-kamala-harris-summit/.

169. Emily Weinstein and Jeffrey Stoff, "China's Quest for AI talent," in Chinese Power and Artificial Intelligence: Perspectives and Challenges, William C. Hannas and Huey-Meei Chang, eds. (Abingdon: Routledge, 2022), 57.

170. Hans Nichols, "Biden's real target on Chinese investment restrictions," Axios, August 9, 2023, https://www.axios.com/2023/08/09/biden-us-china-investments; and Karen Hao and Sha Hua, "The U.S. is Turning Away From its Biggest Scientific Partner at a Precarious Time," The Wall Street Journal, August 16, 2023, https://www.wsj.com/world/china/the-u-s-is-turning-away-from-its-biggest-scientific-partner-at-a-precarious-time-9fb9adaa.

171. Brad Smith and Carol Ann Browne, Tools and Weapons: The Promise and Peril of the Digital Age (London: Hodder and Staughton, 2022), 324. Also see the more recent: Eleanor Olcott, Qianer Liu, and Ryan McMorrow, "Microsoft to move top AI experts from China to new lab in Canada," Financial Times, June 10, 2023, https://www.ft.com/content/d21d2f85-7531-4536-bcce-8ca38620fe55.

172. Peter Carlyon, "Science and Security: Setting the Direction for the UK's Research Relationship with China," The RAND Blog (June 30, 2022), https://www.rand.org/blog/2022/06/science-and-security-setting-the-direction-for-the.html; and Karen Hao and Sha Hua, "The U.S. Is Turning Away From Its Biggest Scientific Partner at a Precarious Time," The Wall Street Journal, August 16, 2023, https://www.wsj.com/world/china/the-u-s-is-turning-away-from-its-biggest-scientific-partner-at-a-precarious-time-9fb9adaa.

173. Saleem H. Ali, "Mining For Gallium: The Next Step Towards American Semiconductor Supremacy," Forbes, August 15, 2023, https://www.forbes.com/sites/saleemali/2023/08/15/sourcing-gallium-for-american-semiconductor-supremacy/.

174. The twentieth century provides examples of geopolitical necessity driving scientific innovation, particularly during wartime as states were forced to develop domestic alternatives to products or materials unavailable due to the impact of conflict on world trade. Daniel Immerwahr, How to Hide an Empire: A History of the Greater United States (London: Vintage, 2020).

175. George Parker, "Britain to host first global AI regulation summit in autumn," Financial Times, June 7, 2023, https://www.ft.com/content/3929908e-0f6a-4223-9c1c-5cd68d82a828; George Parker and Tim Bradshaw, "Money for UK supercomputer plan 'needs to be in billions,' says Javid," Financial Times, June 13, 2023, https://www.ft.com/content/6e3938b2-1952-4ae9-b52e-c2f2ddc1812b.

176. Payne, I, Warbot: The Dawn of Artificially Intelligent Conflict, 7.

177. Berry Zwets, "AI advances cybersecurity, but also offers hackers opportunities," Techzine, September 9, 2022.

178. Devanny, "Commentary: Foreign Ministries and Cyber Power;" Anu Bradford, "The Race to Regulate Artificial Intelligence: Why Europe Has an Edge Over America and China," Foreign Affairs (June 27, 2023), https://www.foreignaffairs.com/united-states/race-regulate-artificial-intelligence; Tom Bristow, Gian Volpicelli and Clark, "Rishi Sunak promised big on his AI summit. He's running out of time to deliver;" and Elias Groll, "White House is fast-tracking executive order on artificial intelligence," Cyberscoop, August 15, 2023, https://cyberscoop.com/white-house-executive-order-artificial-intelligence/.

179. The policy recommendations presented in this paper focus on the institutional functions of foreign ministries and the roles they should play in effectively shaping and implementing policy. The paper does not consider the potential to integrate AI tools in the normal business of foreign ministries. This is a productive area for further study, along similar lines to existing work on the affordances offered by AI tools to intelligence analysts. See: Defence Science and Technology Laboratory, Human-centered ways of working with AI in intelligence analysis (July 26 2023), https://www.gov.uk/government/publications/human-centred-ways-of-working-with-ai-in-intelligence-analysis/human-centred-ways-of-working-with-ai-in-intelligence-analysis; https://cetas.turing.ac.uk/publications/large-language-models-and-intelligence-analysis; Adam C and Richard Carter, Large Language Models and Intelligence Analysis Expert Analysis, Centre for Emerging Technology and Security (2023), https://cetas.turing.ac.uk/publications/large-language-models-and-intelligence-analysis; Kenneth Payne, 'AI Spy,' Ken's Substack (August 16 2023), https://www.kennethpayne.uk/p/ai-spy; and: Joe Devanny, 'On AI Spies: The value of AI tools for cyber operations and intelligence analysis,' Cyber Policy & Strategy (August 16 2023), https://open.substack.com/pub/joedevanny/p/on-ai-spies?utm_campaign=post&utm_medium=web.

180. A point made persuasively in Bremmer and Suleyman, "The AI Power Paradox: Can States Learn to Govern Artificial Intelligence."

181. Weinstein and Stoff, "China's Quest for AI talent," 61.

182. Arquilla, Bitskrieg, 82

FIU | Jack D. Gordon
Institute for Public Policy

MAY 2024