ONE

Introduction

THE CENTAUR'S DILEMMA

FROM SPRING TO SUMMER

The development of artificial intelligence (AI) is often described with seasonal reference. Since its advent, arguably in 1956, AI is said to have gone through a series of AI winters, pockets of time with little support and even less to show for that support. AI is here now. This is not winter; perhaps it is not even spring. It feels like summer. Most experts think AI is about to bloom in exponential ways. It is already embedded in our everyday lives, in how we shop, listen to music, and navigate while driving. AI has great promise to solve problems. It might help cure cancer, address climate change, and alleviate hunger. The cybernetics movement dreams of immortality. AI will bring wealth to some and unemployment to others; it already has. It will also transform national security practice. It will enable intelligence applications, augment human decisionmaking, and offer autonomous logistics and weapons systems. The question is how, not whether, this will happen.

AI will also power, and empower, the research and development of other emerging technologies, each with potential transformative effects. Quantum computing seeks to harness the physics of subatomic particles to make computers exponentially faster than current classical computers. Such speed could allow quantum computers to break existing encryption keys using brute calculating force alone—unless, of course, quantum keys are designed and implemented first. To be sure, quantum computing depends on mastering and controlling the quantum physical properties known as superposition and entanglement and then finding a means to measure quantum bits that does not otherwise disrupt those properties. AI will not solve these challenges, but it may *help* solve them by optimizing the speed and calculating capacity of classical computers and by modeling outcomes.

AI applications will also help identify and model complex molecular structures as well as rapidly calculate and adjust alternative outcomes. Synthetic biology involves the application of engineering principles to create new biological substances and components as well as edit and change existing biological components. Synthetic biology can be used to make new biological and chemical agents. It can also be used to drive genetic evolution in purposeful directions and, in the case of species with short lifespans, such as mosquitoes, do so rapidly. And it can be used to create bio-enhanced equipment and implants, like climbing gloves modeled on the molecular structure of a gecko's feet, or microchip implants to store data and track personnel. Synthetic biology has beneficial purposes, such as the eradication of diseases—for example, malaria—and the creation of drought-resistant crops. However, these same processes can be used to create new biological weapons and destroy an adversary's crops.

No wonder AI is described as a defining technology of the twenty-first century. The government of China thinks so. In July 2017, China's state council announced a US$150 billion centralized program to develop AI and become the world's leader in AI by 2030.[1] When the AI-driven computer AlphaGo defeated Lee Sedol in the Chinese game of Go in 2016, 60 million Chinese were watching live on television or social media. Vladimir Putin declared "whoever controls [AI] will be the ruler of the world." Hyperbole, no doubt, but, significantly, Putin's other comments suggest he is being briefed on the security applications of AI. For its part, the U.S. Department of Defense (DOD) has made AI a central component of what was initially referred to as the Third Offset, the use of technology to offset the perceived or real advantages of a potential adversary in manpower or geography. Governments, however, are chasing the private sector, which is leading the way in developing AI for commercial advantage—or, in the case of authoritarian states, social and political control.

Scholars use a variety of metrics to measure research and development potential and progress, including: (1) research funding; (2) patents issued; (3) scholarly papers produced and cited; and (4) the number of PhDs and undergraduate degrees awarded in specific fields. Contextual metrics are also used, such as the strength and number of supercomputers within a country, the number of start-ups and venture capital spending in a field, as well as the volume and nature of IP theft. The numbers are impressive. Although estimates of the number of AI researchers and engineers vary widely, they are measured in the hundreds of thousands, and LinkedIn reports "AI specialist" and "robotics engineer" as the two fastest-growing jobs globally as well as in the United States.[2] A June 2017 McKinsey & Company report estimated global corporate spending on AI between US$26 billion and US$39 billion in 2016, with 90 percent of this amount going toward research and development.[3] In contrast, the Manhattan Project employed 130,000 people at its height and cost approximately US$2 billion, the equivalent of US$23 billion in 2007 dollars.[4]

Exact numbers are hard to establish, because AI incorporates multiple subfields. "AI is not a single piece of hardware or software, but rather a constellation of technologies."[5] For now, let us define AI as a process of machine optimization relying on algorithms, data, and calculation that gives "a computer system the ability to solve problems and to perform tasks that would otherwise require human intelligence."[6] Current AI is particularly good at correlating, connecting, and classifying data; recognizing patterns; and weighing probabilities, which is why current AI is good, and getting better, at tasks like facial recognition, image compression and identification, and voice recognition. Metrics are also hard to establish, because some of the research is secret, for proprietary commercial reasons as well as for security reasons. But there is no doubt the twenty-first century is an AI century in commerce, academia, and government. The capacity exists. The motive exists. And there is money to drive the process. There is also security incentive.

AI offers significant security applications and advantages. However, there are also significant security implications. The philosophical questions posed by the potential advent of superintelligent artificial intelligence—a hypothetical era when machines become generally smarter than humans—are interesting and theoretically existential. But there are more immediate and pragmatic risks. Security risk will come first, as states—and perhaps other actors—race to develop and defend against the advantages of AI-enabled intelligence, weapons, and decisionmaking. Security advantage will be found in machine automation, autonomy, augmentation, and speed. But such advantage comes with risk. Where there is parity, states will seek the marginal advantages of additional speed, additional data, and additional autonomy, and they will take shortcuts in time and safety to do so. That is what happens in technology races when security is at stake. Risk also comes from the way humans interface with technology. Operators may not understand the technology they are using, including its limitations, its strengths, or its faults. They may not rely on it enough or they may rely on it too much. Chapter 4 considers instances when technology failed to work as intended, or just as often worked as intended but was misunderstood or applied by human operators with disastrous impact.

The secretary of defense stated in 2016 that the Defense Department would not employ AI-enabled weapons, known as lethal autonomous weapons systems (LAWS), without a human in the loop or on the loop, at least not unless another state does so first. Autonomous and semiautonomous weapons systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force. In theory, this means that a human will make the decision to fire a weapon or can override a weapon system once the shooting begins. The Defense Department calls this a centaur model of employment, but instead of being part-human and part-horse, this centaur is part-human and part-machine, the machine augmenting human capacity with the human seeking to understand and control the machine's capabilities while adding human judgment, strategy, and intuition to the machine's operation. As a result, the employment of AI presents two immediate decisional issues: (1) when to rely on the machine alone and when to rely on machine augmentation of human capacity; and (2) how to assert positive control over a technology that operates at machine speed and, in some cases, without transparent explanation.

Indeed, machine speed is an AI strength. AI-driven machines can sense, calculate, and predict instantaneously—and thus track targets, plan logistics, and link data—in ways and at speeds humans cannot. Slow this process down to ensure human input and control and you may lose some or all the advantage. This is the Centaur's Dilemma: how to reap the benefit of AI for national security purposes without losing control of the consequences.

These factors will magnify the risks already associated with using AI to enable security applications of the sort described in the next three chapters. One of these risks is that we are not ready for an AI race. AI—technology, doctrine, and law—is not yet a core national security subject. AI foundations and think tanks are studying the subject, but it is not yet fully embedded in the work of think tanks or war colleges, as nuclear doctrine once was and cybersecurity and grand strategy are today. It should be. As a result, there is no corresponding national, policy, and legislative dialogue that cuts across constituencies to frame issues, set expectations, and resolve disputes. We do not yet have an ethical framework, legal framework, or associated operational doctrine commensurate with the security benefits and risks of AI.

Neither do we have an effective governance structure. To be sure, there are lead agencies and processes. But there is no national agency, or lead agency with responsibility for formulating *national* policy. Moreover, unlike national security problems and issues in the past, AI challenges come with new actors and incentives, from industry, academia, and government. Because most AI R & D is occurring in the commercial and academic sectors, any effort to regulate AI for national security purposes needs to effectively address these sectors as well as government conduct. This also means that at present, by default, policy is set by the individual decisions and individual actors engaged in AI research, development, and deployment based on individual motives and incentives, just as social media policy is driven by industry actions. This is lowest common denominator decisionmaking, not *national* security policy. If we do not make conscious policy decisions—and, where necessary, embed them in law—we will end up with piecemeal policy through litigation and during moments of crisis, with all the resulting pathologies associated with these modes of decision. In doing so, we will magnify the risks and minimize the advantages of AI.

PURPOSE OF THIS BOOK

This book is intended to make AI and the law accessible to national security policy and legal generalists so that they can make wise and strategic decisions about regulating the security uses of AI. In short, if I were asked to brief a senior official about AI, and was given the time, this is what I would say. These chapters are not intended to tell specialists what they already know about AI; they are intended to tell generalists what they should know in order to perform their policy and legal tasks. It is also intended to be read as needed, from the index. Toward this goal, the book identifies legal issues that might arise in the context of using AI for security purposes. However, the reader should not mistake a discussion of how a law *might*, or *could*, be used for a suggestion that the law *should* be used in this manner or without challenge or controversy. Such a discussion, however, might inspire policymakers to consider clarifying the law or finding a better way to address legal needs.

This book intends to encourage key actors—in industry, academia, and, most of all, government, where the responsibility for national security ultimately resides—to make informed, purposeful, and accountable decisions about the security uses and governance of AI. This is hard for generalists to do if they do not understand the essential policy, legal, or technical elements of the debate. Thus, beyond identifying the law and how it might be used or interpreted, this book identifies law, or principles of law, that might, do, or should apply to AI by implication or analogy. In the absence of specific law, for example, policymakers might look to the law of armed conflict or arms control concepts for principles to regulate AI. The absence of specific law will also elevate the importance of constitutional law in regulating AI, including and especially the First, Fourth, and Fifth Amendments. Thus, these amendments are addressed at the policy level and in plain English.

This book also makes the obvious point that law never keeps up with technology. Moore's Law is always faster than statutory law and case law. But if this is an obvious lesson, it is a lesson we repeatedly learn. One answer is to focus on laws that require good process and procedural checkpoints, rather than laws that seek to dictate substantive results or that address specific technologies in the present tense when we know AI will evolve and change in unanticipated ways with unanticipated uses. Law, if well designed and wisely used, can help decisionmakers maximize benefits and minimize risks by defining boundaries and requiring accountable process before risks are assumed.

If there are good ideas in this book, they are intended to inspire as a point of departure. If there are bad ideas, they should be used to pivot more quickly to better ideas. As Yale's president A. Whitney Griswold said, "The only sure weapon against bad ideas is better ideas."

ROADMAP

This book is divided into two parts. Part I describes AI, its security uses and risks. Specifically, chapter 2 introduces the reader to AI, its history, its components, and its potential. It also provides the reader with a layman's understanding of the nomenclature of the practical and philosophical AI debates so the general reader can join the dialogue. Each chapter includes a summary of key national security policy and law takeaways. Chapter 3, which considers military and intelligence AI applications, addresses hot-button topics like LAWS, swarms, facial recognition, and deep fakes; however, it would be a policy mistake to focus entirely on these issues to the neglect of logistics and manpower where AI is certain to have profound and more immediate effect. Chapter 4 addresses the implications of AI on security. Six risks are addressed: (1) the risk of unintended consequences when humans use or interface with new or complex technologies; (2) the risk that AI will at once cause global instability while enhancing the power and reach of authoritarian regimes; (3) the risks that come from technology "arms races"; (4) that AI will lower the cost of conflict and therefore increase the risk of conflict; (5) that operating at machine speed with AI will compound national security decisionmaking pathologies; and (6) the risks that might, in theory, arise from unfriendly AI and superintelligent artificial intelligence.

In response to the applications and implications of AI on national security practice, in Part II this book turns to its central question: how, if at all, should we, might we, regulate the national security uses of AI? What would a regulatory template look like? Chapter 5 introduces the reader to the three purposes of national security law: (1) the authority to act, and the left and right boundaries of that action; (2) process; and (3) legal values, many of which are also security values. It further explains why each purpose is critical to defining a legal regime.

Much of the Centaur's Dilemma derives from the need to adopt, at the national and tactical levels, decisionmaking processes that can move and respond at machine speed when it is wise to do so, without surrendering the capacity to command and control outcomes. Thus, the chapter contemplates a different kind of Turing test. This is a test of the U.S. government's decisionmaking capacity. Now is the time to get the process right—to design, train, and empower processes that can effectively address the Centaur's Dilemma. Such processes would move from Daniel Webster's imminence to Alan Turing's instantaneousness. Processes that pass this Turing test would be ones that effectively build private-public partnerships, operate day-to-day, and can move at machine speed when needed and human speed when it is wise to do so.

In the absence of clear law, opponents and proponents of the government's actions will litigate the limits of the law. This will place special emphasis on knowing and applying the law and values embedded in the Constitution, upon which both the government and litigants will rely. For this reason, the application of constitutional law to AI is detailed in chapter 6. In the absence of a tailored statutory regime, the government will use existing law to accomplish new missions and employ new AI capabilities. As chapter 7 explains, three laws are central: the International Emergency Economic Powers Act (IEEPA), the Invention Secrecy Act (ISA), and the Defense Production Act (DPA). However, for the reasons articulated, emphasis is given to the DPA, including its common usages as well as its potential AI usages. At present, the DPA is the government's most expansive potential statutory authority to observe and regulate private AI for national security purposes.

In the absence of a comprehensive legal regime, lawyers will also look to adopt, adapt, and apply law by analogy. In this vein, chapter 8 looks at three nonproliferation and arms control regimes addressed to nuclear, biological, and chemical weapons. The chapter considers how these regimes are apt and inapt to AI, including questions involving command and control, safety, and the verification of dual-use technology. Chapter 9, in turn, looks at how the law of armed conflict (LOAC) might offer analogy beyond its obvious application to military operations. In particular, the chapter considers whether and how the doctrine of command responsibility and the requirement to evaluate new weapons for compliance with the law of armed conflict could apply generally to security uses of AI. And, lest advocates of AI controls believe the U.S. government must drive the debate, the chapter also discusses the lessons learned from the grassroots campaign to eliminate anti-personnel mines, which culminated in the Ottawa Treaty. These chapters are not intended for subject matter specialists, but for the generalists who need to quickly understand what analogy the field of arms control or the law of armed conflict might offer to the regulation of AI.

Chapter 10 rounds out the framework by looking at means other than law that might be used to regulate the design and use of AI for security purposes. While law is binding, democratic, and national in its scope, it can also be difficult to pass and often presents lowest common denominator compromises. The chapter considers the strengths and weaknesses of three regulatory mechanisms outside the normal purview of security specialists: (1) ethical codes of professional conduct, including those pertaining to engineers and lawyers; (2) internal review boards, which review the ethics, design, and scope of certain academic experiments and research; and (3) the concept and practice of corporate social responsibility, known as CSR.