Big Data as a National Security Issue

Paul B. Stephan[†]

ABSTRACT

Modern enhancements of data mining have unfolded in a legal near-vacuum. No extant legal system adequately specifies the property rights in the information elements that make up big data. The miners rely on their technical ability to gather and exploit, without waiting to confirm their entitlement to do so. Innovators have tolerated the legal vacuum because, as a practical matter, most potentially valuable information is accessible. Information by its nature implies a sender and a recipient. This sharing relationship necessarily complicates ownership, in particular the power to exclude. At the same time, the capacities that the mining of big data empowers are sufficiently novel to fall outside the scope of traditional regulatory regimes, including those focused on national security.

This Article considers the national security implications of this legal vacuum. It conceives of instances of big data as emergent systems. It argues that the potential benefits and risks of big data demand property rules that optimize the value of data systems, accounting for potential risks as well as benefits, while safeguarding the interests of persons who originate information. The key insight is to distinguish big data as an emergent entity from the countless events that constitute collectable information. The distinction allows us to think separately about the property entitlements and regulatory constraints attributable to the elements and the systems. This analytical step in turn clarifies which legal approaches may advance national security interests consistent with other values and commitments.

I. INTRODUCTION

Over the course of the past quarter-century, big data has gained profound economic and national security significance. It is a necessary predicate for a wide range of essential managerial functions as well as

[†] John C. Jeffries, Jr., Distinguished Professor of Lawand Senior Fellow, Miller Center of Public Affairs, University of Virginia. I am grateful to Stewart A. Baker, Ashley S. Deeks, and Kristen Eichensehr for comments and criticism, and to the editors of *The University of Chicago Legal Forum* for their thoughtful and helpful suggestions. Responsibility for errors, blunders, and misjudgments remains mine alone. My work as Special Counsel to the General Counsel of the U.S. Department of Defense encompassed some of the issues I discuss here, but I do not rely on or otherwise make use of any privileged or classified information that came my way. The views found here are entirely my own and should not be attributed to the U.S. government or the Department of Defense.

predictive analysis. It supports advanced cyber operations by enhancing hacking (cyber penetration of and interference with data sets) as well as cyber defense. It forms the backbone of most states' projects to shape the world to their interests (projecting power) and to defend those interests in a world where adversaries strive to acquire and exploit their own big data resources.

Increasingly ubiquitous artificial intelligence (AI) tools count as the most recent breakthrough in big-data-dependent information processing.[1] AI represents, however, only a jazzier variant of a general long-term practice, the mining of vast stores of information to detect patterns and derive predictions.[2] Data mining, carried out by search-and-classify algorithms that update and reconfigure themselves autonomously, allows people to extract meaning and value from amassed information for myriad benign purposes. At the same time, it offers many opportunities for abuse. It can pollute the information environment with deep fakes and other lies, drive hyper-surveillance that degrades individual autonomy and privacy, and promote risky choices, whether financial or military.[3] It also provides better means to disrupt and destroy the information-based capacities of others.[4]

Modern enhancements of data mining have unfolded in a legal near-vacuum. No extant legal system adequately specifies the property rights in the information elements that make up big data.[5] The data-collectors and miners rely on their technical ability to gather and exploit, without waiting to confirm their entitlement to do so. They have proceeded in the face of the legal vacuum because, as a practical matter, most potentially valuable information is accessible, in the sense that it does not lie behind effective barriers. Information by its nature implies

---

[1] *See* MUSTAFA SULEYMAN & MICHAEL BHASKAR, THE COMING WAVE: TECHNOLOGY, POWER AND THE 21ST CENTURY'S GREATEST DILEMMA 16–19 (2023); *cf.* Dep't of Defense, *Executive Summary: DoD Data Strategy, Unleashing Data to Advance the National Defense Strategy* 4 (2020), https://media.defense.gov/2020/Oct/08/2002514180/-1/-1/0/DOD-DATA-STRATEGY.PDF [https://perma.cc/Q2BW-9QCA] ("Artificial intelligence . . . is long-term data competency grounded in high-quality training-quality datasets . . . that are the pieces of information and associated labels used to build algorithmic models.").

[2] *See* KAI-FU LEE, AI SUPERPOWERS: CHINA, SILICON VALLEY, AND THE NEW WORLD ORDER 14, 104–12 (2018); VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 6–7 (2014).

[3] *See* PAUL B. STEPHAN, THE WORLD CRISIS AND INTERNATIONAL LAW: THE KNOWLEDGE ECONOMY AND THE BATTLE FOR THE FUTURE 276–77 (2023).

[4] *See* Paul B. Stephan, *Big Data and the Future Law of Armed Conflict in Cyberspace*, *in* THE FUTURE LAW OF ARMED CONFLICT 61–62 (Matthew C. Waxman & Thomas W. Oakley eds., 2022).

[5] Admirers of the European Union (EU) would beg to differ, arguing that it has developed within its own legal system rules that will dominate international practice. *See* ANU BRADFORD, DIGITAL EMPIRES: THE GLOBAL BATTLE TO REGULATE TECHNOLOGY 105–45 (2023). Pessimism about the resilience of the EU undercuts that assertion; *see also* STEPHAN, *supra* note 3, at 83, 87–89.

a sender and a recipient.[6] This sharing relationship complicates ownership by requiring stakeholders to consent to any rule of exclusion. Absent agreement, the general default is open access. At the same time, the capacities that the mining of big data empowers often are sufficiently novel to fall outside the scope of traditional regulatory regimes, including those focused on national security.

This Article first gives a brief description of big data and its role in analysis and prediction, including AI. It explains the relevance of the systems-analysis concept of emergence as an analytic tool for understanding the issues. Next, I consider the national security aspects of big data from three perspectives.

First, big data allows a state (this Article focuses on the United States) to extend its capabilities to influence the world. It supports analytical and predictive functions that enable identification of and a timely response to threats. These enhanced capabilities come with downsides, including driving unproductive arms races and encroaching on individual privacy and independence.

Second and symmetrically, I analyze the big-data-assisted capabilities of adversary states as a threat to the United States. Adversaries can probe and disable critical infrastructure that depends on online management. Civil aviation, a wide array of public services, the financial system, oil and gas pipelines, and communications networks become more vulnerable. The United States, as one of the most online states in the world, faces grave risks from adversary exploits fueled by big data.

Finally, I consider big data as a resource that invites targeting and contestation. Both the United States and its adversaries invade each other's big data both for espionage and to damage or destroy big-data-driven capabilities. All states that have big data consider these assets as something to be defended and attacked.

These capabilities and risks raise legal issues. The presence of big data in either the public or private sector depends on a state's fundamental decisions about ownership of and access to data. States must decide whether and how to limit data accumulation and its exploitation for particular instances of data mining. This Article does not offer substantive recommendations as to these choices. Rather, it identifies critical distinctions and basic tradeoffs in the development of applicable law.

---

[6] *See* CLAUDE E. SHANNON & WARREN WEAVER, THE MATHEMATICAL THEORY OF COMMUNICATION 4–6 (1949).

II. BIG DATA AND ITS EXPLOITATION

Farboodi and Veldkamp offer a good working definition of big data, albeit from a business perspective:

> Big data refers to large volumes of data, often from multiple sources, and to the ability to gather, store, and process them to produce new kinds of observations, measurements, and predictions about individual customers.[7]

If we replace "customers" with "actors," the national security issues emerge. Big data consists of organized data sets (databases) that those with access can interrogate with techniques (algorithms) that uncover intelligible patterns and support useful predictions. The practice goes back at least to the seventeenth century and the foundation of actuarial science, used originally to support insurance underwriting.[8] Modern advances in data collection, organization, and interrogation have greatly enhanced the value and importance of big data. The current excitement, although some might say overexcitement, about AI tools provides an exclamation point.

Views differ as to the relative contribution of the three components of modern data mining—collection, organization, and interrogation—to the functionality of big data. Algorithm design is the brainiest and most creative of the components, leading many people whose livelihoods depend on the ability to create and spread knowledge to assign to it paramount importance. Some experts argue, however, that advances in both algorithms and organization techniques, the work of clever coders, tend to disseminate fairly quickly.[9] Data collection is clunkier and relies more on perspiration than inspiration. Yet, these experts claim the size and quality of the data collected, rather than development of better software for organization and interrogation, are the most important variables in determining the usefulness of data mining and the prospects for innovation.[10]

A closer look at big data as the lynchpin of the new analytical and predictive technologies leads to a key insight. Big data functions as an emergent system. Emergence characterizes entities (systems) that have properties attributable to the whole, rather than the entity's

---

[7] Maryam Farboodi & Laura Veldkamp, *Data and Markets*, 15 ANN. REV. ECON. 23, 24 (2023).

[8] *See* James Hickman, *History of Actuarial Profession*, in ENCYCLOPEDIA OF ACTUARIAL SCIENCE 838, 839 (Jozef L. Teugels & Bjørn Sundt eds., 2004); Edmond Halley, *An Estimate of the Degrees of Mortality of Mankind, drawn from curious Tables of the Births and Funerals at the City of Breslaw; with an attempt to ascertain the Price of Annuities upon Lives*, 17 PHIL. TRANS. ROYAL SOC'Y 596 (1693).

[9] *See* STEPHAN, *supra* note 3, at 166–70.

[10] *See* LEE, *supra* note 2, at 14 ("In deep learning, there is no data like more data.").

elements.[11] In physics, observers measure and manipulate the properties of a gas without any need to account for the location and actions of the individual molecules that make it up, much less the state of the particles in the atoms that make up the molecules.[12] In economics, a market produces observable and useful information without requiring isolation or identification of the myriad transactions that it comprises.[13] In medicine, doctors focus on organ systems, not the status of individual cells.[14] So it is with big data. It functions as a whole system, producing useful knowledge without requiring isolation and retrieval of the discrete information elements that it contains.

Recognition of the status of big data as an emergent system allows lawmakers and regulators to separate issues regarding the capacities and risks of data sets from those implicated by appropriation of the elements of data they comprise. Depending on how a data set is designed and protected, extraction of data about individuals may be possible, perhaps even easy. Extraction risk threatens individual privacy and vulnerable persons. All data sets have this problem, to which access controls, with all their limitations, may be the best response.[15] The Biden administration's executive order on transfer of sensitive personal data to problematic countries, ratified in part by the new Protecting Americans' Data from Foreign Adversaries Act, exemplifies a focus on the uses to which collected data are put, rather than on the assembly of data sets.[16]

The dangers as well as benefits posed by a particular system of big data, by contrast, depend on the quality of the database, the capabilities of the algorithms to which it is tied, and the purposes for which it is used. Regulators need to understand the difference, and not to confuse protection of the security of data within a data set with control of the outputs enabled by big data. Each objective is important, but they are

---

[11] *See* Claus Emmeche, Simo Køppe & Frederik Stjernfelt, *Explaining Emergence: Towards an Ontology of Levels*, 28 J. GEN. PHIL. SCI. 83 (1997); PETER CHECKLAND, SYSTEMS THINKING, SYSTEMS PRACTICE 3 (1981) ("The central concept 'system' embodies the idea of a set of elements connected together which form a whole, this showing properties which are properties of the whole, rather than properties of its component parts.").

[12] *See* SEAN CARROLL, THE BIG PICTURE: ON THE ORIGINS OF LIFE, MEANING, AND THE UNIVERSE ITSELF 94–104 (2016).

[13] *See* THOMAS C. SCHELLING, MICROMOTIVES AND MACROBEHAVIOR 47–51 (1978).

[14] *See* Emmeche, Køppe & Stjernfelt, *supra* note 11, at 92.

[15] *See* NAT'L CYBER SEC. CTR., GUIDELINES FOR SECURE AI SYSTEM DEVELOPMENT 14 (2023), https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf [https://perma.cc/9Y2W-7AEA] (agreed principles adopted by 18 state cybersecurity agencies) (calling for "appropriate access controls" regarding the contents of a data set); *see* Exec. Order No. 14,110, 88 Fed. Reg. 75191, 75193 (Oct. 30, 2023); *cf.* DANIELLE KEATS CITRON, THE FIGHT FOR PRIVACY: PROTECTING DIGNITY, IDENTITY, AND LOVE IN THE DIGITAL AGE 148–68 (2022).

[16] *See* Exec. Order No. 15,421, 89 Fed. Reg. 15421 (Feb. 28, 2024); National Security Supplemental Appropriations Act 2024, Pub. L. No. 118-50, Div. I, 138 Stat. 895.

largely independent. Any cost-benefit assessment of the creation of a particular big-data system must account for the capacities, whether for good or harm, of the system, not just the risks associated with unauthorized release of particular data.

III. BIG DATA IN NATIONAL SECURITY

Big data can enhance virtually any governmental function. It can identify incipient pandemics and organize responses to them. Its modeling of protein folding promises to revolutionize the design of vaccines and other pharmaceuticals. It can bolster efforts to grapple with climate change. It can recognize macroeconomic patterns and guide responses. More generally, it can detect preparations by hostile actors, both states and non-state groups, to undertake harmful projects, whether through armed force or subversion. It is an essential component of "smart" weapons that deliver lethal force precisely and efficiently, maximizing the achievement of military objectives while minimizing collateral losses.

Crucially for a discussion of national security, big data underlies advanced forms of hacking as well as measures to defeat cyberattacks. Both kinds of enhancements, viewed from the perspective of a single state, augment national security. Those seeking to invade, compromise, and exploit stored information and on-line operations as well as the defenders of those targets rely on tools that depend on training with big data sets to improve their quality. As Google and Alphabet's chief executive reminds us, artificial intelligence, a capacity that requires big data for its efficacy, has a promising future on both sides of the cyber wars.[17] Throughout we need to balance the greater capacities to achieve good things against the corresponding threats that enhanced abilities create.

In the next two sections, this Article elaborates on the upsides from big data's enhancement of our offensive and defensive national security capacities. Here, I consider the downside of such power, which comprises risk of abuse and greater vulnerability to attacks.

Any enhancement of state capabilities necessarily creates a risk of abuse. First, there is the problem of agency costs. Those in control of state functions, nominally agents of the people, may seek to use their authority for self-preservation and exploitation, rather than for the public good. Liberal democracies rely on institutional arrangements—electoral accountability, separation of powers, dispersion of authorities among levels of government, protection of private capacities such as

[17] *See* Sundar Pichai, *AI Can Strengthen Cyber Defences, Not Just Break Them Down*, FIN. TIMES (Feb. 15, 2024), https://www.ft.com/content/7000ac39-cc0e-467e-96f6-6617f91dc948 [https://perma.cc/F4HZ-WYF6].

free speech and freedom of the press—to check these tendencies. But even in these states, those who exercise power can use big data to frustrate these checking functions.

Consider how big data can enable state intervention in the lives of people. A common concern is the use of big data for surveillance.[18] Revelations by national security workers such as Edward Snowden and Chelsea Manning promote worries that states use data sets as a means of domination and control of the people they are supposed to serve. Privacy advocates voice similar concerns about the use of DNA information in genealogical databases to solve murders.[19] It seems that big data may have the potential to undo the mechanisms that align the interests of states to those over whom they exercise dominion.

Another downside is that big data may lead state actors astray. The enhanced powers bestowed by better analysis and predictive power may encourage risky actions spurred by overconfidence in these capacities. My colleague Professor Ashley Deeks has probed the problems created by artificial-intelligence-enhanced uses of force.[20] In the abstract, a well-designed and hyper-fast analytic and prediction system might do much better than humans, hampered by the fog of battle, in targeting and launching attacks that maximize military benefit while minimizing civilian harm. We reasonably might worry, however, that those responsible for deploying such systems might take their efficacy for granted, rather than testing them and closely monitoring their performance to see how well they actually work.

Deeks's argument generalizes. Data mining extends the ability of actors, including states, to increase the scale and impact of their actions. Greater empowerment can enhance the net benefits of these actions, perhaps by orders of magnitude, but also can serve as a crutch that enables wasteful or dangerous choices. Power always is a cypher normatively. We cannot tell whether greater capacities are desirable until we know how they will be used. An analogy to nuclear arms, a

---

[18] *See* Bruce Schneier, *The Internet Enabled Mass Surveillance. AI Will Enable Mass Spying*, SLATE (Dec. 4, 2023, 11:15 AM), https://slate.com/technology/2023/12/ai-mass-spying-internet-surveillance.html [https://perma.cc/2HEA-XXES]; JOSH CHIN & LIZA LIN, SURVEILLANCE STATE: INSIDE CHINA'S QUEST TO LAUNCH A NEW ERA OF SOCIAL CONTROL 92–113 (2022); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 59–62 (2019).

[19] *See* Teneille R. Brown, *Why We Fear Genetic Informants: Using Genetic Genealogy to Catch Serial Killers*, 21 COLUM. SCI. TECH. L. REV. 114, 115 (2019).

[20] *See* Ashley S. Deeks, *Predicting Enemies*, 104 VA. L. REV. 1529 (2018); Ashley S. Deeks, Noam Lubell & Daragh Murray, *Machine Learning, Artificial Intelligence, and the Use of Force by States*, 10 J. NAT'L SECURITY L. & POL'Y 1 (2019); Ashley S. Deeks, *Coding the Law of Armed Conflict: First Steps*, in THE FUTURE LAW OF ARMED CONFLICT, *supra* note 4, at 41. In the Biden administration, Deeks served as associate White House counsel and deputy legal adviser to the National Security Council when I was special counsel to the General Counsel of the Department of Defense.

capacity that both projects great power and risks terrible destruction, suggests itself.[21] The question thus becomes how to manage these capacities for the best in light of the risks they entail.

For example, the use of data mining to propagate convincing lies helps a state counter its adversaries, if at a substantial potential cost. Misinformation can shift political discourse, distort economic decision-making, and undermine social capital.[22] Data mining can contribute to the creation of more convincing misinformation, known as deep fakes.[23] In the right hands, big-data-assisted deep fakes might turn a conflict in an instrumentally or normatively preferred direction without resort to force. They also may disrupt social cohesion and make the sustenance of a liberal democracy more difficult.

Another issue is the public-private divide.[24] In the United States, what we know about data mining largely involves the private sector. Similar activities undoubtedly take place in the intelligence and national security sectors, but how much and where remains obscure.[25] The government acquires some of these capabilities through contracting out. Reliance on the private sector for important national security functions might be beneficial, to the extent that distributed powers can be both more resilient and susceptible to checking by critics, or problematic, to the extent that misaligned incentives and higher transaction costs make it harder to deploy needed resources.

Data mining in the private sector, unchecked by adequate legal controls, can be bad for reasons independent of state blunders. Private actors may not have the right incentives to pursue the common good. For example, individual privacy and autonomy may suffer while firms may unduly thwart competition in pursuit of monopoly super-profits.

Beyond abuse lies the problem of greater vulnerability. As states pursue the benefits from the exploitation of big data, whether scientific, economic, or cultural, a paradox presents itself: their investments

---

[21] *See* Henry A. Kissinger & Graham Allison, *The Path to AI Arms Control*, FOREIGN AFFS., (Oct. 13, 2023), https://www.foreignaffairs.com/united-states/henry-kissinger-path-artificial-intelligence-arms-control [https://perma.cc/74PU-M4G9].

[22] *See* Stephan, *supra* note 4, at 96, 126–28.

[23] *See* STEPHAN, *supra* note 3, at 201–03; Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753 (2019).

[24] *See* Kristen E. Eichensehr, *Digital Switzerlands*, 167 U. PA. L. REV. 665 (2019); Kristen E. Eichensehr, *Public-Private Cybersecurity*, 95 TEX. L. REV. 467 (2017).

[25] *See* Memorandum from Deputy Sec'y of Def. on Data Advantage to Senior Pentagon Leadership (May 5, 2021), https://media.defense.gov/2021/May/10/2002638551/-1/-1/0/DEPUTY-SECRETARY-OF-DEFENSE-MEMORANDUM.PDF [https://perma.cc/HP7D-QPSX]; *cf.* Gilbert Herrera, *The NSA Warns that US Adversaries Free to Mine Private Data May Have an AI Edge*, WIRED (Mar. 21, 2024), https://www.wired.com/story/fast-forward-nsa-warns-us-adversaries-private-data-ai-edge/ [https://perma.cc/3GQZ-EHT6] (explaining why ban on use of U.S.-source private data limits NSA's ability to develop big data resources).

enhance vulnerability. As states and their subjects become more dependent on big-data driven, online operations, the cost of disruption of these functions grows. The threats posed by other countries' hacking capacities go well beyond espionage and ransom. Adversaries can penetrate and disable targeted networks, stopping the delivery of essential private sector services such as transportation and communication networks, financial systems, utilities, and health care. Computer-dependent governmental services, from military operations to social support, are also vulnerable to these attacks.

In sum, the breakthroughs made possible by recent improvements in data mining pose a general dilemma. They may enable states to deter threats and promote international stability, and perhaps even build peace and prosperity. Their existence, however, threatens harm and unpredictability, perhaps even deeper insecurity. Under plausible conditions, big data and what it engenders might be seen as both essential and terrible.

IV. BIG DATA AS A THREAT TO NATIONAL SECURITY

The concept of national security implies the existence of threats: security has meaning only in reference to those things that disturb it. States do not live in isolation, but rather must account for the effect of their capacities on the decisions of others. Perceptions of the big-data capacities of other states pressure the United States to invest in the technology. This investment can feed either credible deterrence or a wasteful arms race.

Consider as a point of comparison nuclear weapons, for which effective defensive measures seem not to exist. They deter through the threat of retaliation but cannot neutralize another state's capacity directly.[26] In theory, big data might achieve neutralization, not only deterrence. It is conceivable, for example, that analytical and predictive capacities enhanced by data mining might anticipate and counter big-data-based offensive measures, including hostile cyber operations or misinformation campaigns.

The distinction between neutralization and deterrence affects choices about investing in enhanced capacities. The logic of deterrence on its face supports significant investment, but these investments are fundamentally problematic. International relations scholars speak of

[26] *See* THOMAS C. SCHELLING, THE STRATEGY OF CONFLICT 53–80 (1960); THOMAS C. SCHELLING, ARMS AND INFLUENCE 157–68 (1966). The Strategic Defense Initiative of the 1980s aspired to neutralize nuclear missile attacks, but I am aware of no evidence indicating it ever made significant progress toward this objective. Distinguishable are systems that neutralize conventional missile attacks, such as U.S. Patriot Missiles and Israel's Iron Dome. Effective protection against nuclear weapons requires complete success to prevent catastrophic harm, while even partial neutralization of conventional missiles produces substantial benefits.

the security dilemma, a problem that states with adversaries commonly confront.[27] The dilemma posits that persons responsible for a state's security often cannot distinguish improvements in its rival's capacity to deter from enhancement of offensive abilities. With big data, the prospect of neutralization as an alternative to deterrence may at least redirect investments, encouraging states to shy away from technologies with largely offensive potential and to give priority to those that can render attacks ineffective.[28]

Whether through deterrence or neutralization, the present configuration of the international order drives greater investment in big data. In a world containing adversaries, growing use of data mining and its attendant capacities presents a fundamental national security issue. Often the capacities lead to greater reliance on online structures, such as data storage and processing. The move from physical to online storage and operations in turn increase targets for adversaries to hack and disrupt.

It is true that adversaries can exploit big-data dependence through means that do not depend on big data. The relatively recent history of hacking contains many instances where innovative cyber vandals have inflicted great costs without relying on sophisticated, big-data-enhanced methods.[29] Training hacking tools on sophisticated data sets, however, increases the efficacy of both entry and concealment in cyberattacks. The less glamorous and brainy ability to construct data sets matters, no matter how clever the attacker's software or wetware.

How a state addresses the problem depends in part on the nature of the adversary. For the United States, China is considered its main pacing adversary. This status does not preclude cooperation and mutual understanding but frames strategic issues.[30] China's economy is significantly intertwined with those of the United States and Europe, a fact that complicates strategic decisions. Other states, including Russia, Iran, and North Korea, oppose U.S. interests and policy in Europe, the Middle East, and East Asia, respectively. None presents the same economic interdependence issues as does China. Nonstate actors, both foreign and domestic, may exploit the fruits of data mining to propagate

---

[27] *See* Robert Jervis, *Cooperation Under the Security Dilemma*, 30 WORLD POL. 167 (1978).

[28] *See* Kristen Eichensehr & Danielle Keats Citron, *Resilience in a Digital Age*, 2024 U. CHI. L. FORUM 45 (2024). Technological innovation with respect to neutralization, if sufficiently one-sided, may also contribute to the security dilemma. If one side enjoys a breakthrough in neutralization capacity, its adversary may invest in the ability to conduct overwhelming attacks.

[29] *See* SCOTT J. SHAPIRO, FANCY BEAR GOES PHISHING: THE DARK HISTORY OF THE INFORMATION AGE, IN FIVE EXTRAORDINARY HACKS 15 (2023).

[30] *See generally* DMITRI ALPEROVITCH & GARRETT M. GRAFF, WORLD ON THE BRINK: HOW AMERICA CAN BEAT CHINA IN THE RACE FOR THE TWENTY-FIRST CENTURY (2024).

terror and disruption. With them, there are no issues of interdependence.

China, among all states, appears to have the greatest ambitions for unlocking the potential of data mining as a national security resource.[31] It faces few internal constraints, either legal or political, to acquiring the data that its subjects have assembled and using it for state purposes. Many believe that its government has unlimited access to the databases assembled by its large online companies, including ByteDance (TikTok), Tencent (WeChat), and Sina (Weibo), even though those entities are nominally private.[32] It also has shown the ability to access significant databases of its adversaries through clandestine means, most famously the capture of the vast records of the U.S. Office of Personnel Management in 2015.[33] Its ability to integrate state and private collection and organization of data gives it a significant advantage relative to the United States, much less the countries of Europe.

The United States has responded to China's advantages in data-assembly capacity by trying to limit China's ability to mine data sets. The United States restricts exports to China of the advanced semiconductor chips that support AI development.[34] The Biden administration also has secured commitments from Japan and Netherlands, important manufacturers, to similarly limit Chinese access to these chips.[35] The adequacy of this response, however, remains suspect. Export controls are known to leak; China is seeking the knowledge needed to make these chips; and the chipmakers in Taiwan, the technological leaders, face growing pressure to accommodate China.

To date, China has used the capabilities derived from its big data expertise, domestic security aside, mostly to penetrate other state's computer systems and to conduct espionage. Although it probably can shut down any cyber system it can compromise, it has not yet shown any inclination to do so. Intensification of the conflict over Taiwan could

---

[31] *See* Kissinger & Allison, *supra* note 21.

[32] Congress cited government control as a ground for forcing ByteDance to divest from TikTok. See National Security Supplemental Appropriations Act 2024, Pub. L. No. 118-50, Div. H, § 2, 138 Stat. 895.

[33] *See* Dan Efrony & Yuval Shany, *A Rule Book on the Shelf? Tallinn Manual 2.0 on Cyberoperations and Subsequent State Practice*, 112 AM. J. INT'L L. 583, 601–04 (2018) (describing and evaluating episode).

[34] *See* Implementation of Additional Export Controls: Certain Advanced Computing and Semiconductor Manufacturing Items; Supercomputer and Semiconductor End Use; Entity List Modification, 87 Fed. Reg. 62186 (Oct. 13, 2022) (codified at 15 C.F.R. pts. 734, 736, 740, 742, 744, 762, 772, 774).

[35] *See* Gregory C. Allen, Emily Benson & Margot Putnam, *Japan and the Netherlands Announce Plans for New Export Controls on Semiconductor Equipment*, CTR. FOR STRATEGIC AND INT'L STUD. (Apr. 10, 2023), https://www.csis.org/analysis/japan-and-netherlands-announce-plans-new-export-controls-semiconductor-equipment [https://perma.cc/76TV-CAQ6].

change that.[36] Russia, Iran, and especially North Korea, by contrast, have sponsored or directly run many compromising operations, especially but not only through ransomware. These take the form of either the capture of data with the threat of public exposure or the disabling of the target's computer system. When feasible, the attackers may reverse the compromise upon payment of ransom (always in bitcoin).

Some evidence suggests that the U.S. government has learned something from these cyberattacks. The 2021 Colonial Pipeline hack, a ransomware operation attributed to Russian actors that targeted U.S. gasoline infrastructure, resulted in the U.S. government recovering most of the victim's ransom payments by intercepting the attacker's cryptocurrency transfers.[37] The methods used to claw back the money remain secret, but the episode points to the government's development of significant defensive capabilities. There is every reason to believe that collection and interrogation of data relevant to such attacks might strengthen these defensive capabilities even more.

V. BIG DATA AS A TARGET

Large data sets are significant independent of the analytic and predictive functions that they support, making them attractive targets for national security adversaries. The sets represent costly investments in collection and organization. To optimize their value, developers must both test and update them to ensure that they produce useful outcomes. Initial biases—a famous older example is facial-recognition software trained disproportionately on images of people with light skin, while Google's recent Gemini fiasco reminds us that bad curating can destroy the value of large sets—require expensive correction through collection from more sources as well as better data interrogation. The value of data sets depends on both the scale of the data collected and the resources expended to enhance their quality.

Enhancing the value of data sets, however, also increases their vulnerability. The more public or private actors depend on these sets, the more adversaries will find them an inviting target. Adversaries have an interest not only in blocking access to data sets or destroying them outright but also in quietly degrading them through a process known as

---

[36] *See* U.S. DEP'T OF DEF., ANNUAL REPORT: MILITARY AND SECURITY DEVELOPMENTS INVOLVING THE PEOPLE'S REPUBLIC OF CHINA 140 (2023), https://media.defense.gov/2023/Oct/19/2003323409/-1/-1/1/2023-MILITARY-AND-SECURITY-DEVELOPMENTS-INVOLVING-THE-PEOPLES-REPUBLIC-OF-CHINA.PDF [https://perma.cc/Q7XL-FG59].

[37] *See* Press Release, U.S. Department of Justice Office of Public Affairs, Department of Justice Seizes $2.3 Million in Cryptocurrency Paid to the Ransomware Extortionists Darkside (Jun. 7, 2021), https://www.justice.gov/opa/pr/department-justice-seizes-23-million-cryptocurrency-paid-ransomware-extortionists-darkside [https://perma.cc/XL3N-7JJ6].

data poisoning.[38] Ideally, an undetected compromise of a data set can lead those who use it to rely on outputs that are well off the mark, indeed hallucinatory. Once discovered, poisoning compels the proprietor of the data set to rebuild it, often at great time and expense.

Attacks on big data thus present at least two kinds of risk. First, the immediate destruction of a data set may impair the real-world functioning of a critical service. Wiping clean stored financial data, such as bank or stock market records, might shut down a national financial system. Doing the same to geolocational data might cause a transportation system to collapse. Smart weapons might go astray, producing collateral damage rather than targeted interdiction.

Second, sabotage of big data can degrade the outputs of data mining even more dangerously. Immediate destruction at least alerts the data miner to the problem. Poisoned data may support deranged functions for some time, until the user detects the incursion or confronts the bad analysis and predictions resulting from it.

Attacks on big data sets thus present significant national security issues. Sabotage through hidden data poisoning can undermine confidence in existing resources and discourage new investments in the technology. Complete degradation can defeat the capacities that depend on big data, impairing a wide range of military and civilian governmental functions. Either way, the data sets invite adversaries to seek their incapacitation. The states that rely on these data sets in turn must consider both how to defend them and, when defense fails, how to respond.

VI. EXISTING LEGAL RESOURCES AND THEIR DEVELOPMENT

Having described at a high level of generality the kinds of national security problems presented by big data and its applications, this Article now considers the legal questions that follow from these issues. This section frames its points as questions rather than answers. The lack of positive legal sources and well-developed practice precludes definitive statements about the relevant law. Instead, I consider possible analogies and the possibilities for legal development.

A. Who Owns the Data?

A pervasive problem throughout big data is ownership of the elements of data that go into the data sets. Which persons have rights to exclude others from accessing, consuming, and transferring rights in the data? Big data involves the harvesting of observations of events.

---

[38] *See* Gary P. Corn & Eric Talbot Jensen, *"Attacking" Big Data: Strategic Competition, the Race for AI, and the International Law of Cyber Sabotage*, *in* BIG DATA AND ARMED CONFLICT 91, 94 (Laura A. Dickinson & Edward W. Berg eds., 2023).

Data derived from transactions, such as the use of social media or a purchase or sale, begins with the sharing of information. Monitoring of public behavior, whether through security cameras, mobile phone pings, or other technology, implicates a line between behavior that enjoys whatever society considers a reasonable expectation of privacy and that which the curious can observe and collect.

The problem is the paucity of well-developed legal rules. Existing law does not provide clear allocations of ownership, and what rules do exist are legally precarious due to reform projects, both domestic and international. Consider first data derived from transactions.

The giant private firms that dominate cyber transactions, including Meta (Facebook), Alphabet (Google), Amazon, and X (Twitter), rely on term-of-use contracts with customers to give the companies the nonexclusive right to use the data generated by their transactions. Large retailers mostly do the same. The EU and some U.S. states have enacted legislation that seeks to revise these contracts, although they may leave the door open for new arrangements that still permit data sharing.[39] China largely imposes a governmental easement over such data, either formally or in practice. A handful of U.S. lower court cases have so far been unimpressed with the argument that copyright owners have a right that goes further than the traditional first-use rule and that allows them to block harvesting of data derived from covered works. The day is young, though, and courts have not definitely disposed of the claim.[40] EU legislation might yet reach a different, more rigidly anti-collection result.[41] Fundamentally, the legal status of transactionally derived data that goes into these sets remains contested.

Similarly, the question of when data generated in public enters into a property-law commons remains profoundly murky. *Carpenter v. United States* disrupted traditional conceptions of the public sphere by

---

[39] *See, e.g.*, Josephine Wolff, William Lehr & Christopher S. Yoo, *Lessons from GDPR for AI Policymaking*, 27 VA. J. L. & TECH. 1 (2024); Daniel J. Solove, *Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data*, 118 NW. L. REV. 1081 (2024).

[40] *See* Doe v. Github, Inc., 672 F. Supp. 3d 837, 861–62 (N.D. Cal. 2023); *cf.* Andersen v. Stability AI Ltd., No. 23-CV-00201-WHO, 2023 WL 7132064, at 17 (N.D. Cal. Oct. 30, 2023); Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., No. 1:20-CV-613-SB, 2023 WL 6210901, at 14 (D. Del. Sept. 25, 2023). The New York Times on December 27, 2023, filed a copyright suit against OpenAI and Microsoft. Michael M. Grynbaum & Ryan Mac, *The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work*, N.Y. TIMES (Dec. 27, 2023), https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html [https://perma.cc/RS3G-29ZY].

[41] The European Parliament adopted the Act on March 13, 2024, and published a correct text on April 19, 2024. *See* European Parliament Press Release, Artificial Intelligence Act: MEPs Adopt Landmark Law (Mar. 13, 2024), https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law#:~:text=On%20Wednesday%2C%20Parliament%20approved% 20the,fundamental%20rights%2C%20while%20boosting%20innovation [https://perma.cc/XXK4-BSVF].

holding that the owner of a mobile phone enjoys Fourth Amendment protection with respect to locational data collected and retained by private carriers.[42] Even though the carriers gain possession of this information legally, an owner of a phone, the data generator, retains, in the eyes of the U.S. Supreme Court, a power to exclude the government (but perhaps not private actors) from access. In the wake of that decision, at least one court of appeals has barred local government from harvesting publicly observable information that private actors remain free to collect.[43]

As the importance and value of big data grows, one might expect property law (generously conceived to include public regulation) to extend its scope and clarify entitlements in data. At least in the liberal order that has held sway in much of the West, greater specification of interests in property correlates with the increase in social value of an activity, whether it is farming in the medieval period or manufacturing in the nineteenth century.[44] It is reasonable to believe that increases in property law's specificity matter more to the relevant actors than the particular assignment of entitlements.[45]

What might this more developed law of data ownership look like? In some cases, such as deeply personal information, we might expect property law to impose a rule of inalienability so as to maximize the information generator's autonomy and privacy.[46] Most data arising from non-intimate social behavior, however, could be alienable and thus amenable to collection in a data set.[47] All that will be left is haggling over price. As the next subsection argues, the issue of the collectability of data can and should be disaggregated from when and how to extract that information from the set.

B. The Distinction Between Data Elements and Systems of Big Data

This Article does not claim that a fuller appreciation of property in information, especially the distinction between information elements and big data as such, will necessarily lead to greater human security

---

[42] 585 U.S. 296 (2018).

[43] *See* Leaders of a Beautiful Struggle v. Baltimore Police Department, 2 F.4th 330 (4th Cir. 2021) (stating that automobile locational data based on license scanning and overhead aerial tracking is protected by the Fourth Amendment from warrantless government access).

[44] *See* Harold Demsetz, *Toward a Theory of Property Rights*, 57 AM. ECON. REV. 347 (1967); James E. Krier, *Evolutionary Theory and the Origin of Property Rights*, 95 CORNELL L. REV. 139 (2009); Henry E. Smith, *Exclusion Versus Governance: Two Strategies for Delineating Property Rights*, 31 J. LEG. STUD. 453 (2002); Katrina Miriam Wyman, *From Fur to Fish: Reconsidering the Evolution of Private Property*, 80 N.Y.U. L. REV. 117 (2005).

[45] *See* STEPHAN, *supra* note 3, at 159–61.

[46] *See* CITRON, *supra* note 15, at 105–30.

[47] *See* Solove, *supra* note 39, at 1128–35.

and flourishing. The point, rather, is that specifying the property interests is an essential first step in any effort to cope with the contemporary uses of big data. We cannot expect to meet these challenges successfully without getting the basics right.

This Article's central claim is that the risk of harm caused by unauthorized access to the elements of the data set is both conceptually and practically distinct from the risks and benefits associated with data mining. Any collection of data may contain elements that could harm their source if publicized. All data sets have this problem, even if optimal tools to manage the risk may vary among them. Data mining through big data, by contrast, creates specific benefits and risks tied to the analytical and predictive capacities it supports. The risk-benefit analysis for developing any particular data set should focus independently on these capacities, and not on the risk tied to the simple existence of the data in the set.

A story from the headlines illustrates the point. In the wake of overreach by U.S. government agencies in the course of the late, unlamented War on Terror, Congress enacted the Foreign Intelligence Surveillance Act (FISA) Amendments Act of 2008.[48] Section 702, that Act's workhorse provision, allows the government, subject to preclearance from a specialized court, to use electronic surveillance to capture communications involving foreign nationals overseas.[49] If, in spite of the mandate to not target U.S. persons, information turns up involving them, these observations may be preserved in a database, including as metadata that obscures personal information. Specially designated persons in the intelligence community, mostly high-ranking supervisors in the Justice Department or the FBI, may approve limited interrogation of this U.S.-person database to determine whether it contains information relevant to intelligence investigations, including those with law enforcement implications.[50]

Proponents of Section 702 argue that it limits potential harms to U.S. persons by tolerating only accidental and incidental collection of their information. The government disseminates the information only within specific intelligence and law-enforcement circles and only if it is

---

[48] Foreign Intelligence Surveillance Act of 1978 Amendments Act of 2008, Pub. L. No. 110-261, 122 Stat. 2436 (codified at 50 U.S.C. §§ 1801 et seq.).

[49] *See id.* at § 101 (codified at 50 U.S.C. § 1881a). The preclearance applies to collection programs, rather than to surveillance of individual targets. *See id.* at § 1881a(g).

[50] *See* OFF. OF THE DIR. OF NAT'L. INTEL., SECTION 702 OVERVIEW, https://www.dni.gov/files/icotr/Section702-Basics-Infographic.pdf [https://perma.cc/D4S7-RG5N]; *see also* Emily Berman, *Reimagining Surveillance Law*, 23 U. ILL. L. REV. 1235, 1259–65 (2023); Robert S. Litt, *The Fourth Amendment in the Information Age*, 126 YALE L.J.F. 8, 12–15 (2016); Peter Margulies, *Dynamic Surveillance: Evolving Procedures in Metadata and Foreign Content Collection After Snowden*, 66 HASTINGS L.J. 1 (2014). The 2024 amendment makes clear that the interrogation must not be based solely on law enforcement considerations. *See* 50 U.S.C. § 1881a(f)(2) (as amended).

connected to targets of valid investigations. Critics maintain that storage inevitably will lead to abuse, regardless of safeguards. They argue that government investigators need to meet a higher standard than current law applies before getting access to any U.S.-person information in the database and in particular should obtain a search warrant before targeting any specific U.S. persons.

Section 702 is subject to sunset and thus periodic legislative review. In 2023, a coalition of far-left and far-right legislators fought to radically revise it, including by adoption of a blanket judicial warrant requirement for all interrogations of the database. After much debate, Congress rejected the most extreme proposals, although it also shortened the review period to two years.[51]

Whatever Congress produces in the future, what seems striking about Section 702 in its present form is its separation of collection issues from privacy concerns raised by extraction. The law allows those parts of the government with capacity to access international communications (principally the NSA, but also the CIA, the FBI, and other intelligence and law enforcement agencies) to build a database that includes incidentally obtained information about U.S. persons. Government actors cannot access this data at will. Rather, a limited number of officials with a reasonable basis to believe that the base contains information that may advance legitimate investigations may query it to see if their suspicions are well-founded. If they are correct, they may extract relevant material, but only after complying with elaborate internal controls.

Some argue that accountability of this type, without a warrant requirement for database withdrawals focused on individual targets, is insufficient and the risk of abuse too great. Others maintain that the existing restrictions obstruct investigations that might advance national security interests. My point is simply that Section 702 serves as a working example of the conceptual separation that distinguishes building big data, on the one hand, and safeguarding privacy with respect to data elements, on the other.

C. Regulating National Security Capacities

As already observed, the capabilities that depend on big data, however much they extend the reach of government and private actors in valuable way, also have downsides in the form of risk of abuse and heightened vulnerability to adversaries. In the national security context, the principal legal concerns associated with the offensive use of big

[51] *See* Reforming Intelligence and Securing America Act, Pub. L. 118-49, 138 Stat. 862 (codified at 50 U.S.C. § 1881a et seq.).

data are disloyal agents and inadequate data systems. As we see with China and suspect in other authoritarian regimes, big data can support tools to suppress dissent and thwart the correction of policy blunders.[52] Secrecy and demands for national-security deference can produce similar shortcomings in liberal democracies. Moreover, even officials motivated to advance the national interest may fall prey to the shiny-new-toy fallacy, embracing an emerging information technology without adequate testing.

The traditional tools for minimizing the agency costs that arise in the governmental context are transparency and the rule of law. The former entails the release of relevant information into the public domain, while the latter requires collaboration between Congress and the judiciary to hold the executive to account. It has long been recognized that the national security sphere undermines both. Plausible arguments for secrecy frustrate publicity, especially when the courts uphold punishment of leakers.[53] Information asymmetry and claims of expertise also deter legislators and judges from challenging the national security judgments of executive branch officials. Judicial deference manifests itself in procedural doctrines such as standing, evidentiary privilege, and cause of action that have the effect of distancing the judiciary from oversight in this area.[54]

Legal academics with experience within the national security state argue that other mechanisms exist to promote accountability and the rule of law. Insiders can leak evidence of abuse, as Edward Snowden and Chelsea Manning did.[55] An array of actors, within the executive branch and outside, have both opportunity and motive to challenge particular actions and policies, either with or without public disclosure.[56] This constraining function extends to misuses of big data, whether through privacy violations or misplaced confidence in the capacity of particular programs. As a technology premised on disruption of settled expectations ("move fast and break things" applies here), big data is likely to provoke people tied to incumbent projects to push back against new ones by exposing their shortcomings.

---

[52] *See* CHIN & LIN, *supra* note 18.

[53] *See* Haig v. Agee, 453 U.S. 280 (1981) (upholding contract mandating preclearance of publications).

[54] *See, e.g.*, United States v. Zubaydah, 595 U.S. 195 (2022) (state secrets privilege); Fed. Bureau of Investigation v. Fazaga, 595 U.S. 344 (2022) (same); Hernandez v. Mesa, 140 S. Ct. 735 (2020) (cause of action); Clapper v. Amnesty International USA, 568 U.S. 398 (2013) (standing).

[55] In referencing this episode of insider disclosure of communications surveillance practices, I do not mean to imply any view about Snowden's motives. Based on available evidence, I find it impossible to determine whether he was a public-spirited whistle-blower, a narcissist, an agent of a malign foreign power, some of the above, or none of the above.

[56] *See* JACK GOLDSMITH, POWER AND CONSTRAINT: THE ACCOUNTABLE PRESIDENCY AFTER 9/11 205–07 (2012); Ashley Deeks, *Secrecy Surrogates*, 106 VA. L. REV. 1395, 1395–96 (2020).

In short, the use of big data to extend the capacities of the national security state is risky, but not exceptionally so. Any significant technological innovation that extends what a small group of officials can do presents a potential danger to democracy and public welfare. By way of analogy, the capacity to launch nuclear weapons poses a profound threat not just to the country but, potentially, all of humanity. Subjecting the exercise of this power to a popular plebiscite or comparable means of democratic accountability, however, seems implausible in a world of multiple nuclear powers, some of whom are adversaries, and a technological imperative of rapid response. So it is with big data. Ignoring real threats also presents risks. The challenge remains preserving the means for indirect criticism, accountability, and opposition while leaving space for necessary innovation in the face of what people euphemistically call an evolving threat environment.

The United States and China can be seen as conducting a kind of a natural experiment here. The United States uses resources such as inspector generals, whistleblower safeguards, First Amendment protections for journalists and civil society, and legislative oversight to scrutinize the use of data mining for national security resources. There is no guarantee that these institutions identify all the important problems, and they may impede the development of optimal tools for meeting threats. In addition, they leave considerable room for private-sector collection and exploitation of big data, which may both undermine national security imperatives and threaten individual liberty. Yet the U.S. legal checks, combined with tolerance of disbursed research and exploitation, do something to mix innovation with safeguards against extreme concentration of power. China, by contrast, relies much more on top-down supervision and seems to give priority to innovation over safety. Which strategy best meets the needs of the moment will be revealed, if at all, only after much time.[57]

D. Responding to National Security Threats

Threats to national security posed by foreign big data raise the question of permissible responses. As discussed above, the principal tradeoff, given limited resources and attendant budgetary constraints, is between retaliation and neutralization. The latter is more desirable in the short run but requires ongoing vigilance and investment in an array of capacities that respond to possible threats. The former may cost less but is less likely to work effectively except if disproportionately large relative to the injury that triggers retaliation.

---

[57] *See* STEPHAN, *supra* note 3, at 260–64.

Neutralization entails offsetting an adversary's big-data-driven attacks. Defense against attempts to compromise cyber-managed systems rests on enhanced network security. For example, superior detection technology, itself supported by big data, may neutralize harmful deep fakes by exposing the fabrication.

Retaliation, by contrast, need not take the form of cyber operations. In addition to sanctions and retaliatory operations, criminal prosecutions of those responsible for operations or economic sanctions against their state sponsors are plausible responses. On several occasions, the United States has brought criminal charges against state officials involved in offensive cyber operations that big data presumably sustained or enhanced.

For the most part, however, these indictments have been mostly gestures, as the persons indicted remain out of reach. In one case against Russians involved in the 2016 breach of the Democratic National Committee's computer system, an indicted company did submit to U.S. criminal jurisdiction. The Department of Justice then dropped the charges, presumably because the company faced no effective legal accountability due to the absence of assets or personnel in the United States. Justifiably indifferent to the consequences, the company had invoked its due process rights to get access to the Justice Department's files, thus converting the criminal prosecution into an intelligence operation on the part of the nominal defendant.[58]

Economic sanctions rest primarily on the International Emergency Economic Powers Act (IEEPA), a 1977 revision of presidential authorities originally enacted in the 1917 Trading With the Enemy Act.[59] With a few exceptions not relevant here, IEEPA gives the president virtually unlimited power to declare an emergency. The government then can freeze assets belonging to designated foreign actors tied to the emergency and bar designated individuals from traveling to the United States. Sanctionable actors include foreign states and their agencies and instrumentalities as well as foreign firms and individuals.

To my knowledge, the United States has not imposed IEEPA on any state in response to its cyber operations, although presidents have

---

[58] *See* Indictment, United States v. Internet Research Agency LLC, No. 1:18-cr-00032-DLF (2008), https://www.justice.gov/opa/press-release/file/1035562/download [https://perma.cc/M33V-X6WV]; Spencer S. Hsu, *Justice Dept. abandons prosecution of Russian firm indicted in Mueller election interference probe*, WASH. POST (Mar. 16, 2020, 7:29 PM), https://www.washingtonpost.com/local/legal-issues/us-justice-dept-abandons-prosecution-of-russian-firm-indicted-in-mueller-election-interference-probe/2020/03/16/5f7c3fd6-64a9-11ea-912d-d98032ec8e25_story.html [https://perma.cc/ 3EWK-HDKE]. The Department of Justice retained me as an expert witness with respect to issues of Russian law raised by this case, but my views in this Article do not rely on or reflect that work.

[59] Pub. L. No. 95-223, 91 Stat. 1626 (1977) (codified at 50 U.S.C. §§ 1701–06). *See* Paul B. Stephan, *Seizing Russian Assets*, 17 CAP. MKT. L. REV. 276, 278–81 (2022).

used this authority against individuals engaged in malicious (and possibly big-data aided) hacking.[60] Russia, Iran, and North Korea have faced sanctions based on their activities in the material world. A sufficiently grave cyber-attack on the United States might prompt economic sanctions against the responsible state.

The most significant retaliatory economic measure entails confiscation of frozen assets, including sovereign ones. The United States took this step with respect to Iraq at the conclusion of the Gulf War under the authority of a UN Security Council Resolution. The Trump administration relied on the president's recognition power to transfer Venezuelan state assets to a government in exile, and the Biden administration did something similar after the Taliban regained power in Afghanistan. Since World War II, however, no U.S. president has ordered the forfeiture of sanctioned assets of a recognized government without the support of a binding decision of the United Nations.[61] The 2001 Patriot Act amended IEEPA to allow the forfeiture of already frozen foreign-owned property, but only that belonging to states and persons engaged in an armed attack on the United States.[62]

For two years after Russia's invasion of Ukraine, a debate raged within the Biden administration, Congress, and the punditocracy as to whether IEEPA in its present form permitted the government to confiscate Russia's sovereign assets, especially Russian Central Bank deposits made in the New York Federal Reserve.[63] Congress finally broke the impasse by providing the authority needed through the Rebuilding Economic Prosperity and Opportunity for Ukrainians Act.[64] The President now has the power, but not the duty, to confiscate Russia's assets, including the bank deposits, for the use of a fund managed by the Secretary of State. The legislation, in turn, contemplates, but does not require, transfer of the proceeds to an international fund designed to compensate Ukraine for injuries resulting from Russia's internationally wrongful acts. Whether this approach becomes a precedent for using forfeiture more broadly, and particularly in response to actions depending on or compromising big data, remains uncertain.[65]

---

[60] *See, e.g.*, Exec. Order No. 13,694, 80 Fed. Reg. 18,077 (Apr. 1, 2015); Exec. Order No. 13,848, 83 Fed. Reg. 46,843 (Sep. 12, 2018).

[61] *See* Paul B. Stephan, *Response to Philip Zelikow: Confiscating Russian Assets and the Law*, LAWFARE (May 13, 2022, 9:44 AM), https://www.lawfaremedia.org/article/response-philip-zelikow-confiscating-russian-assets-and-law [https://perma.cc/6E35-BBQY].

[62] PATRIOT Act of 2001, Pub. L. No. 107-56, § 106, 115 Stat. 27 (codified at 50 U.S.C. § 1702(a)(1)(C)).

[63] For a review of the arguments, *see* Paul B. Stephan, *How Do We Express Our Outrage at Russia?*, 13 WAKE FOREST J. L. & POL'Y 189 (2023).

[64] *See* National Security Supplemental Appropriations Act 2024, Pub. L. No. 118-50, Div. F, §§ 104-05.

[65] For a consideration of how an international facility might operate, see Oona A. Hathaway,

E. The Law Governing Targeting of Big Data

Big data does more than support offensive and defensive capabilities for purposes of national security. Data sets function as resources, however their owners currently exploit them. They can be repurposed by applying different algorithms or by licensing access to others. They both support current activities, such as administering complex systems, and make new uses possible. They thus have value and strategic significance independent of their current deployments.

This Article stresses the distinction between data elements as discrete events and data sets as emergent systems, as well as the implication of the distinction for legal regulation. This section discusses the application of international law to big data systems. It considers two kinds of issues: how international law addresses incursions on and disturbances of data sets outside of armed conflict, and actions that that implicate the international law governing the right to use armed force (*jus ad bellum*) and the duties and restrictions applicable to the exercise of armed force (*jus in bello*).

1. Big data and international law outside of war

States with the technical capability to do so intrude on foreign data sets all the time. Most often, they seek to collect and record the stored data as part of the time-honored practice of spying. In conventional espionage, the intruder wishes to avoid detection and thus tries not to disturb the data set. A notorious example is the Chinese penetration of the Office of Personnel Management database in 2015, which remained undetected for a considerable period.[66] People in the national security world assume that the United States engages in similar or even more ambitious operations.

The baseline premise of most international lawyers is that classical espionage may violate the domestic law of targeted states but does not transgress any rule of international law. Widespread state practice and the absence of contrary arguments by most if not all states undergird this position. In the wake of the Snowden revelations, academics and civil society actors tried to find a legal check on state spying grounded

---

Maggie M. Mills & Thomas M. Poston, *War Reparations: The Case for Countermeasures*, 76 STAN. L. REV. 971 (2024); Lee C. Buchheit & Paul B. Stephan, *The REPO Act: Confiscating Russian State Assets and Ukrainian Reparations*, LAWFARE (Jul. 7, 2023, 12:30 PM), https://www.lawfaremedia.org/article/the-repo-act-confiscating-russian-state-assets-and-ukrainian-reparations [https://perma.cc/MB9U-S9PQ]; Ashley Deeks, Mitu Gulati & Paul B. Stephan, *What Should the Biden Administration Do With REPO?*, LAWFARE (May 6, 2024, 9:49 AM), https://www.lawfaremedia.org/article/what-should-the-biden-administration-do-with-repo [https://perma.cc/C6N4-QCUZ].

[66] *See* Efrony & Shany, *supra* note 33.

on respect for individual privacy and relevant human rights principles. So far, however, evidence that these aspirations have coalesced into new rules of international law is scant.[67]

States also intrude on foreign databases to wreak havoc, although not necessarily to cause direct harm to people and physical property. The several Russian operations to undermine public confidence in the 2016 U.S. presidential election provide a vivid example. We do not know whether the hackers benefited from the use of big data to enhance their capabilities, but it is at least possible that they did. Data poisoning based on similar intrusive powers, if used to degrade the value of data sets over time and not to immediately impair vital operations that depend on data sets, also falls into the category of non-espionage, non-armed-attack activity.[68]

Most international lawyers believe that some legal limits exist on the activities of one state on the territory of another, even if in cyber form. Determining the conceptual basis for these limits, as well as defining their specific content, is more controversial. All states agree that international law proscribes some kind of interference with important state functions, even if no force is involved. The principle of non-intervention bolsters this point.[69] Some go further and maintain that a general principle of sovereignty forbids uninvited and unfriendly activity within a state's territory, cyberspace included.[70]

---

[67] *See* Ashley S. Deeks, *Confronting and Adapting: Intelligence Agencies and International Law*, 102 VA. L. REV. 599, 601–02 (2016). Deeks argues that the reactions of liberal democratic states to the Snowden revelations suggest the development of new norms to which those states are prepared to commit, but not the creation of international law binding on all states. *See id.* at 669–71.

[68] *See generally* Michael N. Schmitt, *Big Data: International Law Issues Below the Armed Conflict Threshold*, in BIG DATA AND ARMED CONFLICT, *supra* note 38, at 29; *see also* Corn & Jensen, *supra* note 38, at 115–16.

[69] *See* Military and Paramilitary Activities in and against Nicaragua (Nicar. v. United States), Judgment, 1986 I.C.J. Rep. 14, ¶¶ 202, 205 (June 27). The principle forbids coercion with regard to "matters in which each State is permitted, by the principle of State sovereignty, to decide freely." *Id.*; *see also* Caroline Krass, U.S. Dep't. of Def., *DOD General Counsel Remarks at U.S. Cyber Command Legal Conference* (Apr. 18, 2023), https://www.defense.gov/News/Speeches/Speech/Article/3369461/dod-general-counsel-remarks-at-us-cyber-command-legal-confe-ence/ [https://perma.cc/4KPQ-JPEF]. It is widely recognized that prohibited intervention includes two elements: (1) the action must interfere in the matters the targeted State is permitted to decide freely under the principle of sovereignty; and (2) it must be coercive. However, the precise meaning and contours of these elements are not well-defined—and cyberspace magnifies these already existing uncertainties.

[70] *See* Michael N. Schmitt & Liis Vihul, *Respect for Sovereignty in Cyberspace*, 95 TEX. L. REV. 1639 (2017); Schmitt, *supra* note 68, at 42–44. *But see* Paul C. Ney, Jr., *Some Considerations for Conducting Legal Review of U.S. Military Cyber Operations*, 62 HARV. INT'L L.J. ONLINE 22, 39–40 (2020) ("[I]t does not appear that there exists a rule that all infringements on sovereignty in cyberspace necessarily involve violations of international law.") (views of General Counsel of U.S. Department of Defense); Suella Braverman, *International Law in Future Frontiers* (May 19, 2022), https://www.gov.uk/government/speeches/international-law-in-future-frontiers [https://perma.cc/48NL-ANDN] ("The general concept of sovereignty by itself does not provide a sufficient or clear

With respect to non-intervention, consensus breaks down when it comes to distinguishing forbidden coercion from normal international pressure. The Supreme Court of the United Kingdom, for example, recently observed that "the imposition or threat of trade restrictions in order to exert pressure upon other states, and thereby achieve political objectives, has been part of the armoury of the state since classical times."[71] The Court explicitly limited its analysis to the domestic law of duress, and thus did not address the international law principle of non-intervention. It also did not consider cyber measures as such. Still, the statement indicates a view that the use of economic power, even if substantial, to influence another state's policymaking normally is regarded as statecraft, not unlawful meddling in domestic affairs.[72] Presumably this nonapplication of an international rule extends to cyber operations.

The long history of U.S. covert action during the course of the Cold War, beginning with the CIA activities that complemented the Marshall Plan, reinforces the point.[73] Actions that use concealed means to pollute the material-world information space and shift economic outcomes have been an accepted tool for conducting international relations throughout the modern era. Extending this tolerance to interference with data sets, beyond traditional communications media and business relations, does not seem much of a stretch.[74]

Deriving useful legal rules from the principle of sovereignty is even more fraught. All states embrace sovereignty as a fundamental principle in international relations, one embedded in the UN Charter.[75] Yet the derivation of a specific legal rule prohibiting conduct inconsistent with a state's sovereignty is nearly impossible. State interdependence is pervasive, and unbalanced military, economic, and political power the norm. Except where interference involves the use of force, the constraining effect of a clear international legal rule seems invisible.

An example of the confusion that arises where a superficially appealing principle encounters the difficulty of articulating law can be found in the statement of the French Ministry of Defense on cyber

---

basis for extrapolating a specific rule of sovereignty or additional prohibition for cyber conduct going beyond that of non-intervention.") (views of U.K. Attorney General).

[71] The Law Debenture Trust Corp. PLC v. Ukraine, [2023] UKSC 11, [152] (appeal taken from EWCA).

[72] *Cf.* Nicar. v. United States, I.C.J. at 245 (distinguishing assistance to military activities of insurgents from economic sanctions, ruling as to the latter it "unable to regard such action on the economic plane as is here complained of as a breach of the customary-law principle of non-intervention").

[73] *See* JOHN LEWIS GADDIS, GEORGE F. KENNAN: AN AMERICAN LIFE 293–97 (2011).

[74] *See* Krass, *supra* note 69 (distinguishing force to produce a specific outcome versus general deprivation of control).

[75] *See* U.N. Charter art. 2(1) (principle of sovereign equality); art. 2(4) (forbidding threat or use of force against political independence of any state).

operations.[76] Some academics see the statement as a definitive claim about the importance of sovereignty as a source of international legal limitations on cyber activity.[77] A careful reading of the document, however, suggests deft ambiguity, not clarity. It speaks of "international norms and principles that flow from State sovereignty," implying some separation between the norms and the concept.[78] It stakes out only the right to neutralize such interventions, not a clear prerogative to retaliate.[79] At the end of the day, it treats the significant issue as whether a cyber operation crosses the line separating mere misbehavior from use of force, the former subject to domestic responses while the latter bringing in international law.

2. Big data and the right to go to war

Perhaps the most troubling legal question tied to big data is when interference with a data set justifies going to war. Most scholars who have wrestled with the issue resist the idea that the use of cyber capabilities to impair a state's online resources, without more, justifies a kinetic response (death and destruction). U.S. officials, and most experts on the law of war generally, focus on the consequences of the impairment. Destroying the functionality of a database, even if it produces significant economic harm, is different from actions that proximately cause death or personal injury as well as the physical destruction of assets. The mainstream approach uses the material world as the baseline and asks whether cyber events have direct and significant effects in that world. A cyber intervention that crashes the financial system, leading to economic and political chaos but not directly causing violence, might not count as an armed attack, while one that caused a dam to fail with significant material-world loss, akin to airplanes falling from the sky, would qualify.[80]

---

[76] *See International Law Applied to Operations in Cyberspace*, United Nations Off. of Disarmament Affs. (2019), https://documents.unoda.org/wp-content/uploads/2021/12/French-position-on-international-law-applied-to-cyberspace.pdf [https://perma.cc/44WU-FXQ2] (English language report based on the 2019 Ministère des Armées report, as shared by France); *see also* Ministère des Armées, *Droit International Appliqué Aux Opérations Dans Le Cyberespace* (2019), https://web.archive.org/web/20200101220948/https://www.defense.gouv.fr/content/download/567648/9770527/file/international+law+applied+to+operations+in+cyberspace.pdf [https://perma.cc/D7XY-VP2V] (original French language report).

[77] *See* Schmitt, *supra* note 68, at 42.

[78] International Law Applied to Operations in Cyberspace, *supra* note 76, at 6 ; *see also* Krass, *supra* note 69 ("States conducting activities in cyberspace must take into account the sovereignty of other States[.]").

[79] *See* International Law Applied to Operations in Cyberspace, *supra* note 76, at 7.

[80] *See* Stephan, *supra* note 4, at 67–68; *cf.* Corn & Jensen, *supra* note 38, at 120–122 (describing argument that cyber operations without direct material consequences might be regarded as a use of force).

To date, there exists no state practice indicating that data poisoning or other means of compromising valuable and important big data provides a sufficient basis for legalizing the responsive use of armed force, absent direct and considerable material-world consequences. But as big data gains in value and significance around the world, pressure will build to disregard the material-immaterial distinction. If big data is to assume greater importance to our national security, why not protect it to the fullest extent possible?

In prior work, I have worried that states with significant kinetic capacities—the most likely candidates being the United States and Israel, both of which enjoy significant military-resource advantages compared to their peer adversaries—will respond to severe attacks on their data resources with arms, rather than only by cyber retaliation.[81] Some might see the claim as atavistic. The U.S. experience with supposedly-limited warfare in the post-UN Charter era, however, persuades me that the world needs as many plausible arguments as possible to hold back the use of state violence. The distinction between material and immaterial injury may seem increasingly irrelevant to a cyber-dependent world, but I would not surrender it easily.

3. Big data as a lawful target in war

International law addresses two discrete issues: the decision to go to war and what participants in a war may do lawfully. The law of armed conflict (LOAC), also known as international humanitarian law, comprises the latter. It rests largely on a mix of treaties, most prominently the four Geneva Conventions, the two additional protocols to those conventions, and customary international law.

LOAC embraces several principles that it implements with specific rules. The principle of distinction requires states to differentiate legitimate military targets from civilians, the targeting of whom is forbidden. The principle of necessity requires that all uses of armed force have a military objective—that is they must directly contribute to the defeat of an enemy. The principle of proportionality requires a combatant to use a level of force that does not cause "incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."[82] The principle of humanity, an extension of the principle of proportionality, requires a combatant to avoid needless

---

[81] *See* Stephan, *supra* note 4, at 76–77.

[82] Diplomatic Conference on the Reaffirmation and Development of International Humanitarian Law Applicable in Armed Conflicts, *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, U.N. Doc. 34/API, art. 51(5)(b) (Jun. 8, 1977).

cruelty and excessive harm. Some also argue for a precautionary principle, borrowed from EU law, to put a thumb on the scale against the use of force.[83]

Application of these principles to attacks on big data depend on the answer to an antecedent question: are collections of data stored online, and thus separated from the material world, an "object" that can be characterized as "civilian"? If not, this data enjoys no separate protection under LOAC, just regulation of those effects of attacks that spill over into the material world.[84] Some experts dissent and would, in the context of an extant armed conflict, apply LOAC principles to impairment of the functioning of big data, such as data poisoning.[85] If so, degradation of databases that causes significant economic injury to civilians but does not advance military objectives might be illegal.

So far, states have not gone beyond the articulation of general principles. There is a dearth of practice that might define and clarify what LOAC should mean to state cyberwarriors. The best one can do is say that the greater the scale and impairment of an operation against big data, the higher the likelihood that material-world effect, and thus illegality, will result.

VII. CONCLUSION

Assuming continuity in the world's current economic and technological arcs, big data is likely to loom larger not just for the world economy and the global community, but in the national security world as well. States that project power will rely on it, as will states seeking to defend themselves from the actions of other states. Big data is likely to become not only a significant tool to support these projects, but increasingly a target for adversaries.

In a world riven by hot wars as well as broader international instability, the likelihood of enacting formal legal regimes to address these trends seems unlikely, and perhaps unnecessary.[86] Application of

---

[83] Military decisionmakers bear the primary obligation of applying these LOAC principles, with military discipline applicable to violations. Onlookers, both states and civil society actors, including the International Committee of the Red Cross, in turn express views on compliance. In a limited set of circumstances, the International Criminal Court has jurisdiction over grave violations of these rules. *See* Rome Statute of the International Criminal Court, art. 8, 2187 U.N.T.S. 38544. Most states with significant offensive cyber capabilities, including China, Iran, Israel, North Korea, Russia, and the United States, are not parties to the Rome Statute.

[84] *See* Stephan, *supra* note 4, at 68–71; Schmitt, *supra* note 68, at 42, 151, 164.

[85] *See* Corn & Jensen, *supra* note 38, at 118–120 (describing debate).

[86] *Cf.* Compendium of statements in explanation of position on the final report of the Open-ended Working Gp. on Dev. in the Field of Info. and Telecomm. in the Context of Int'l Sec., at 85, U.N. Doc. A/AC.290/2021/INF/2 (2021) ("We remain of the view that [information and communications technologies] are simply not susceptible to traditional arms control arrangements. It would be futile – and a tremendous distraction – to spend a decade or more negotiating a new legally

legacy international law to new problems attributable to the rise of big data will have to rest on analogy and inference, not on new legal instruments.[87] The fundamental question is the adaptability and flexibility of international cooperation outside of formal processes and the availability of nonlegal norm creation.

Most students of international law would admit, if candor were compelled, that formal lawmaking is better for international lawyers than the international community as a whole. By this I mean that international lawyers play a large role in the formal processes and enjoy a disproportionate share of prestige and satisfaction by propounding on their achievements. Yet much of the work of international law involves tacit cooperation, informal understanding, and exercises of restraint or engagement that are not openly attributed to legal compulsion.[88] The regulation of big data in the national security realm is likely to stick to this pattern.

It is not unreasonable, for example, to anticipate that neither the United States nor China will engage in significantly costly cyber operations, such as large-scale data poisoning, except in the context of a material-world dispute. Were armed force to be deployed to alter the legal status of Taiwan, for example, unconstrained attacks on big data might follow. But otherwise, both states would have a strong incentive to maintain an informal cooperative equilibrium. An analogy to the unwritten but nevertheless resilient Cold War norm eschewing assassinations of either side's spies, except when engaged in operations on the target state's territory, suggests itself.[89]

More of a reach, but not completely outside the realm of plausibility, is a tacit agreement by China and the United States to ride herd on respective allies as to efforts to degrade big data assets of others. China may have sufficient economic and political power over North Korea to deter it from these acts, as the United States may have with respect to Israel. Russia's growing military and economic dependence on China might even enable the latter to impose some restraints on the former. By analogy, China apparently has had some success in discouraging Russia from threatening the use of nuclear weapons in its war with Ukraine.

---

binding instrument.") (submission of the United States).

[87] *See* Paul B. Stephan, *The Crisis in International Law and the Path Forward for International Humanitarian Law*, 104 INT'L REV. RED CROSS 2077, 2083–84 (2022).

[88] *See* PAUL B. STEPHAN, APPLYING MUNICIPAL LAW IN INTERNATIONAL DISPUTES 21–28 (2024); ROBERT E. SCOTT & PAUL B. STEPHAN, THE LIMITS OF LEVIATHAN: CONTRACT THEORY AND THE ENFORCEMENT OF INTERNATIONAL LAW 111–27 (2006).

[89] *See* Stephan, *supra* note 4, at 81. Russia's apparent practice in recent years of liquidating its own spies after defection and flight does not represent a break from this pattern.

The broad point is that the growth of big data as a national security factor need not take place in anarchy. States have the means to find the right balance between support and constraint as to their own resources, and international law plus informal understandings can address international threats. The future is not free from danger, but doom and destruction are not inevitable, and perhaps not even likely.