AI, China, Russia, and the Global Order:

Technological, Political, Global, and Creative Perspectives

A Strategic Multilayer Assessment (SMA) Periodic Publication

December 2018

Contributing Authors: Shazeda Ahmed (UC Berkeley), Natasha E. Bajema (NDU), Samuel Bendett (CNA), Benjamin Angel Chang (MIT), Rogier Creemers (Leiden University), Chris C. Demchak (Naval War College), Sarah W. Denton (George Mason University), Jeffrey Ding (Oxford), Samantha Hoffman (MERICS), Regina Joseph (Pytho LLC), Elsa Kania (Harvard), Jaclyn Kerr (LLNL), Lydia Kostopoulos (LKCYBER), James A. Lewis (CSIS), Martin Libicki (USNA), Herbert Lin (Stanford), Kacie Miura (MIT), Roger Morgus (New America), Rachel Esplin Odell (MIT), Eleonore Pauwels (United Nations University), Lora Saalman (EastWest Institute), Jennifer Snow (USSOCOM), Laura Steckman (MITRE), Valentin Weber (Oxford)

Opening Remarks provided by: Brig Gen Alexus Grynkewich (JS J39), Lawrence Freedman (King's College, London)

Editor: Nicholas D. Wright (Intelligent Biology)

Integration Editor: Mariah C. Yager (JS/J39/SMA/NSI)

This white paper represents the views and opinions of the contributing authors.

This white paper does not represent official USG policy or position.

Disclaimers

This white paper represents the views and opinions of the contributing authors. This white paper does not represent official USG policy or position.

Mention of any commercial product in this paper does not imply DoD endorsement or recommendation for or against the use of any such product. No infringement on the rights of the holders of the registered trademarks is intended.

The appearance of external hyperlinks does not constitute endorsement by the United States Department of Defense (DoD) of the linked websites, or the information, products or services contained therein. The DoD does not exercise any editorial, security, or other control over the information you may find at these locations.

OPENING REMARKS: US PERSPECTIVE

Brig Gen Alexus Grynkewich, JS/J39

Given the wide-ranging implications for global competition, domestic political systems and daily life, US policymakers must prepare for the impacts of new artificial intelligence (AI)-related technologies. Anticipating AI's impacts on the global order requires US policymakers' awareness of certain key aspects of the AI-related technologies – and how those technologies will interact with the rapidly changing global system of human societies. One area that has received little in-depth examination to date is how AI-related technologies could affect countries' domestic political systems—whether authoritarian, liberal democratic or a hybrid of the two—and how they might impact global competition between different regimes.

This white paper highlights several key areas where AI-related technologies have clear implications for globally integrated strategic planning and requirements development:

- Since 2012, new AI-related technologies have entered the real world with rapidly accelerating scale and speed. While the character of these technologies currently favors enhanced surveillance, it is currently limited by a need for extensive human involvement and the preparation of big data platforms. This will likely dominate current efforts to incorporate AI into social governance, as we see now in China.
- AI may help enable a plausible competitor to liberal democracy allowing large and industrially sophisticated states to make their citizens rich while maintaining rigid control. China is now in the process of building core components of such a system of digital authoritarianism. Such systems are already being emulated in a global competition with liberal democracy.
- Russia has a different political regime than China. The Russian model is a hybrid that relies on a mix of less overt and often non-technical mechanisms to manipulate online information flows.
- Competition for influence between digital liberal democracy and more authoritarian digital regimes will occur at many levels: international institutions (and norms); nation states; and corporations. The US must adopt a multifaceted approach to influence with allies and crucial swing states. It must also carefully prevent unwanted escalation of this competition – as a number of contributors argue in this white paper, insecurity drives much of Chinese and Russian decision-making.
- China's foreign policy decision-making will not necessarily become more expansionist if its domestic regime becomes more authoritarian. Mapping out AI's effects on foreign policy choices requires mapping them out within the domestic ecosystem and content from which those choices emanate.
- Military dimensions of global competition will change with AI. Hackers become more prominent and new crisis escalation risks emerge. Chinese domestic social governance systems that become ever more reliant on vast digital systems will be tempting targets for adversaries – a fact likely to prompt Chinese regime insecurity that may feed a spiraling security dilemma.

The emerging digital liberal democracy in the US, digital hybrid regime in Russia, and digital authoritarian regime in China will each exert influences far beyond their physical borders. This competition for influence will likely prove a defining feature of the twenty-first century global system. We must not be caught by surprise.

OPENING REMARKS: UK PERSPECTIVE

Sir Lawrence Freedman

In the 1990s, there was talk of a revolution in military affairs (RMA) resulting from the combination of improved sensors, digital communications, and precision guided munitions. In retrospect this was both more and less of a revolution than supposed at the time. It was less of a revolution because the drivers of military conflict were not technological but lay in broader social, economic, and political factors. The new technologies made possible military operations that ran at a faster tempo and used weapons of greater lethality that allowed for greater discrimination. Military power could be directed against vital targets to achieve the optimum effects. It was soon discovered that enemies could limit the advantages these capabilities gave the US and its allies by adopting guerrilla strategies based on ambushes and terrorism. However well suited they might be to fights between regular armies their limitations became evident in struggles over 'hearts and minds.'

Yet it was also more of a revolution than really understood in the 1990s. The RMA was then assumed to represent an advanced stage in a line of technological development that could be traced back to the 1960s when Gordon Moore first observed that the number of components per integrated circuit would double every two years. Yet as we can now see it was really only an interim stage. Over the past two decades we have seen the arrival of smart phones putting data sets, imagery, navigation and forms of communication into the hands of individuals that were once only specialist military tools. Forms of international connectivity have created new opportunities for productive and benign activities but also for mischief and malign influences. The kinetic aspects of conflict have now been joined by non-kinetic forms of struggle including cyber attacks and information campaigns. These have moved the arena of conflict away from the field of battle to the essentials of everyday life and the state of public opinion.

Artificial Intelligence (AI) now points to the next stage. The ability to gather data and interrogate it with scant human engagement now starts to set tests for whole societies: regarding the efficient exploitation of scarce resources on the one hand, and the ability of individuals to live free and fulfilled lives on the other. As this volume makes clear, the government of China is now embarking on a vast experiment in social control that aims to use AI to ensure that individuals are following the party line and rewards or punishes them according to how well they behave. Russia does not have the capacity or the political structures capable of following this example, though it has been a pace-setter in the use of cyber and information operations to undermine its foes (without actually starting a war).

It is worth recalling that the Cold War was decided not by force of arms but because the Soviet system imploded, having failed to deliver for its people and having lost legitimacy as a result of its repressive methods. The military balance of the time, and in particular the fear of nuclear war, maintained a stalemate so that instead of a hot war there was intense ideological competition. Liberal democracy posed a threat to authoritarian systems because it was seen to be better able to meet human needs, including free expression. But during the Cold War, the US and its allies always led the ideological competition and over time demonstrated with relative ease the superiority of their political systems. As before there are formidable reasons for both sides to avoid pushing any contest to open hostilities. This means that there is now a different form of ideological competition. This time it will be tougher because China has invested heavily in the technologies of social control, and in particular in AI, while liberal democracy has lost some of its lustre in unpopular wars and financial crises. The West has yet to work out how to cope with so much personal data being stored and analysed by both private and

state organisations. But liberal democracies must somehow demonstrate that it is possible to take advantage of the new technologies without losing sight of their core values.

Another difference from the Cold War is that China's economy depends on trade with the rest of the world. It has recently started to be viewed as an unreliable partner, for example by getting its technology into the critical systems of Western countries. This issue has acquired more salience because of growing concern over rather old-fashioned geopolitical issues, as China pushes to turn itself into the dominant regional power in the Asia-Pacific region. This takes us back to the question of how much the new technologies have influenced classical forms of military conflict. The answer will depend on how well AI is integrated into command systems, as well as the ability to disrupt enemy systems. In the new era of AI, when humans might be perplexed by what is going on in the machines on which they must depend, the strategies of disruption and disorientation that have been prominently in play in international affairs in recent years, could well move to new levels and become more central than before to the conduct of conflict.

It is unwise to try to predict the future just by following trends, or assuming that the structures of international economics and politics will continue to follow familiar patterns. The US network of alliances, for example, is currently under a lot of pressure. Anticipating the likely path of technological development may therefore be far less difficult than grasping the forms of its interaction with a changing context. The future is unpredictable because it will be shaped by choices between options that are currently barely understood. The great value of this White Paper is that it describes some of the big issues coming our way and urges us to stretch our imaginations when thinking about the challenges that will need to be faced.

EXECUTIVE SUMMARY

Artificial Intelligence (AI) and big data promise to help reshape the global order. For decades, most political observers believed that liberal democracy offered the only plausible future pathways for big, industrially sophisticated countries to make their citizens rich. Now, by allowing governments to monitor, understand, and control their citizens far more effectively than ever before, AI offers a plausible way for big, economically advanced countries to make their citizens rich while maintaining control over them—the first since the end of the Cold War. That may help fuel and shape renewed international competition between types of political regimes that are all becoming more "digital." Just as competition between liberal democratic, fascist, and communist social systems defined much of the twentieth century, how may the struggle between digital liberal democracy and digital authoritarianism define and shape the twenty-first?

The technical nature of AI's new advances particularly well suits all-encompassing surveillance; and as a consequence authoritarianism. New forms of authoritarianism arose with previous waves of global authoritarian expansion: fascism in the 1920s or bureaucratic authoritarianism in the 1960s. China has begun constructing core components of a digital authoritarian state. America's liberal democratic political regime is turning digital, and so too is Russia's hybrid political regime that lies between democracy and authoritarianism.

Swing states from Asia to Africa, Europe and Latin America must manage their own political regimes within the context of this global competition. Several like-minded countries have begun to buy or emulate Chinese systems. Russian techniques are diffusing. To be sure, competing models for domestic regimes must be seen within the broader strategic context—relative military or economic power also matter deeply—but as in the twentieth century it will likely prove a crucial dimension.

This report focuses on the emerging Chinese and Russian models and how they will interact with the global order. We bring together deep expertise on China, Russia, strategy and technology—as well as artists to provide illuminating sidelights.

The key recommendation is that US policymakers must understand the potential for the new AI-related to technologies to affect domestic political regimes (authoritarian, hybrid, and democratic) that will compete for influence in the global order. **We recommend policymakers use the following three-pronged strategy to understand the challenge and develop global policy:**

- US democracy must be kept robust as it adapts to these new technologies. It must respond to both domestic threats (e.g. capture by a tech oligopoly or drift to a surveillance state) and external threats, without becoming governed by a military-industrial complex. US digital democracy, if successful at home, will exert gravitational influence globally.
- The US must exert influence effectively, and manage potential escalation, in the swing states (e.g. in Asia or Europe) and global systems (e.g. norms and institutions) that form the key terrain for competition between the digital regime types. Diplomatic, economic, informational and commercial dimensions will be crucial, with both allies and other states.
- The US should push back on the digital authoritarian and digital hybrid heartlands, but do so in ways that manage the significant risks of spiraling fear and animosity.

Report Overview

We bring together leading experts on China, Russia, strategy, and AI, as well as artists. The report has six sections:

Part I examines the AI-related technologies and their implications for the global order. It provides a framework that describes how the technologies' effects on domestic political regimes may affect the global order. This helps structure the diverse contributions below.

Part II describes specific aspects of the Chinese and Russian regimes in more detail.

Part III examines specific aspects of the export and emulation of the Russian and Chinese models within a global competition for influence.

Part IV explores how AI's potential implications for the Chinese domestic political regime may affect its foreign policy decision-making.

Part V examines specific military dimensions of AI, including in the Chinese and Russian contexts.

Part VI takes a very different approach and provides thought-provoking new viewpoints from artists and perspectives from the humanities.

PART I. INTRODUCTION: AI, DOMESTIC POLITICAL REGIMES AND THE GLOBAL ORDER

In Part I, Nicholas Wright provides an overarching analysis and framework, going all the way from the specific technical characteristics of the new technologies through to the global order.

Chapter 1 examines the AI-related technologies and asks: what specifically is new? By "AI" here we mean a constellation of new technologies: AI itself more narrowly defined (essentially giving computers behaviours that would be thought intelligent in humans), big data, machine learning, and digital things (e.g. the "internet of things"). This constellation is bringing in a new technological epoch. Following a leap in AI research around 2012, we now have: *Automated systems learning directly from data to do tasks that are complicated.* The key leap is that AI's can now do much more complicated tasks (e.g. AI can now do good facial recognition). Crucially, AI has particularly improved for tasks related to "perception"—e.g. perceiving images or speech, or some kinds of patterns in big data—and these are the advances now being rapidly rolled out across diverse real-world uses.

Chapter 2 considers AI's bewildering profusion of implications for the global order, and breaks them down into three more manageable bites. This whitepaper primarily focusses on the first area, which has received by far the least attention.

> (1) The first is how this new technology's potential impacts on *domestic political regimes* (e.g. authoritarian, hybrid, or liberal democratic) may affect competition between them in the *world order*. AI will help enable a plausible competitor to liberal democracy for big industrially sophisticated states to make their citizens rich and maintain rigid control: digital authoritarianism. China is building core components of such a system—which are already being exported and emulated in a global competition with liberal democracy.
>
> (2) An "nth industrial revolution": AI will radically change the means of production across economic and societal sectors, e.g. transport, healthcare or the military.

(3) The "singularity" and the sense of self: In the "singularity", exponentially accelerating technological progress creates an AI that exceeds human intelligence and escapes our control, potentially destroying humanity or disrupting humans' conceptions of themselves.

Chapter 3 examines AI and domestic political regimes in more detail, and introduces three crucial cases: China, Russia, and the US. A domestic political regime is a system of social organization that includes not only government and the institutions of the state, but also the structures and processes by which these interact with broader society. Three broad types dominate globally today: authoritarian (e.g. China), liberal democratic (e.g. the US), and hybrid regimes that fall somewhere in between (e.g. Russia). New variants of these regime types emerge in response to changing times. For instance, historically new forms of authoritarianism emerged in the 1920s (Fascism) and 1960s (bureaucratic authoritarianism). We arguably now see "digital" variants of each regime type emerging: digital authoritarianism (e.g. China), digital hybrid regimes (e.g. Russia) and digital liberal democracies (e.g. the US). However, the character of the new AI-related technologies (i.e. enhanced perception) best suits the augmentation of the surveillance, filtering and prediction in digital authoritarianism, making that perhaps the largest departure of the three.

Chapter 4 discusses global competition, and in particular the export and emulation of these alternative models for influence over swing states—as occurred in the twentieth century between liberal democratic, fascist, and communist regime types. The global competition for influence occurs through active promotion; export of control and surveillance systems; competition between Chinese and US tech titans; as well as battles over global norms and institutions. Swing states across Europe, Africa, Asia etc. are highly heterogenous, and even within states the elites and populations may disagree over the models' relative merits. Of course, the attractiveness or otherwise of the competing models is just one factor in the broader strategic context, as was the case between competing twentieth century regime types. Finally, we also examine two further ways the AI-related technologies may affect global competition: firstly, how AI's potential impacts on domestic political regimes may affect foreign policy decision-making; and second military dimensions.

PART II. DIGITAL AUTHORITARIANISM: EVOLVING CHINESE AND RUSSIAN MODELS

In **Chapter 5, Jeffrey Ding** provides an overview of China's AI strategy. He first places it in the context of past science and technology plans, which helps analyze China's most important current policies and initiatives to further its AI-related industries. Next, he outlines how AI development intersects with multiple areas of China's national interests—and in particular its domestic social governance. He concludes by discussing the main barriers to China realizing its AI dream.

In **Chapter 6, Samantha Hoffman** describes how understanding developments in China's technology-enhanced authoritarianism requires placing them in context of the Chinese Communist Party's political control process known as "social management." The modern "grid management" system, the "Skynet" surveillance project, and "social credit system," are all conceptually linked to long-existing Leninist control processes.

In **Chapter 7, Shazeda Ahmed** describes the Chinese "credit city", in which local governments and tech companies share their data with one another to determine the degree of individuals' and businesses' trustworthiness. The value judgments that come out of assessing these data—in some instances, a numeric score or a verbal rating—becomes a basis for determining the benefits that a person or company can unlock. However, her research on the ground reveals the huge technical and administrative challenges that have yet to be overcome.

In **Chapter 8, Jaclyn Kerr** describes how Russia's innovative and experimental approach to information manipulation and control differs significantly from the more-often discussed Chinese "Great Firewall" system, as well as other approaches that emphasize systemic technical censorship. The Russian model relies on a mix of less overt, and often non-technical, mechanisms to manipulate online information flows, narratives, and framings, to shape public opinion without resort to universal censorship. This model for the domestic control of information not only fits Russia's own political system, but is likely to prove more resonant and easier to emulate in many other countries.

PART III. EXPORT AND EMULATION OF THE MODELS IN GLOBAL COMPETITION

In **Chapter 9, Valentin Weber** provides a more granular view of how the Chinese model is being exported—by the government, state-owned companies, and private companies that make up China's security-industrial complex. This export has been successful in Africa, Asia, the Middle East, and South America. If the US wants to maintain a strategic advantage in regions where it is challenged by China's construction of internet infrastructure and the installation of filtering/surveillance technology, then it requires a global view of the underlying agents that drive exports. This will allow the US to tailor policies that counter the diffusion of information controls.

In **Chapter 10, Laura Steckman** describes China's dual-pronged strategy to become the world's technology leader for AI. Its two primary pathways are: (1) establishing partnerships with nations, organizations, and other entities that demonstrate AI talent; and (2) globally exporting its domestically-developed AI-related technologies. These approaches raise questions for countries with different political and social structures, or that remain wary of using these technologies to shape societies in ways that contradict national values and norms, or more profoundly, to assert control through mechanisms of digital authoritarianism.

In **Chapter 11, Robert Morgus** details the spread of Russia's model. Russian digital authoritarianism is characterized by pervasive communications collection, absent oversight, and government cooption of industry—particularly internet service providers—to do their bidding. Russia's digital authoritarianism is neither as well defined nor as technologically robust or reliant on AI as the Chinese model. The Russian government exports or encourages emulation its model of digital authoritarianism globally and in their near abroad, through diplomatic, informational, and economic means.

In **Chapter 12, James Lewis** takes a skeptical look at ideas of AI and China's unstoppable rise. Judging any Chinese digital authoritarian model's potential attractiveness requires viewing it in strategic context. Not only in the context of a more comprehensive view of what drives influence in the global system, but also in the context of how such influence compares to that of China's major competitor: the US. He outlines five factors that will limit the Chinese model's impact. Although AI ripped from its strategic context can seem powerful or even frightening, given strategic competence the US will remain superior to China.

In **Chapter 13, Chris Demchak** posits that as AI-related technologies rise in criticality for the nations' future economic and political wellbeing, China now has the advantage in three of the four 'horsemen' of AI conflict (scale, foreknowledge, and strategic coherence), leaving only a fourth (speed) to the western democratic societies. To counter China's AI advantages, democratic societies need a new narrative that places their future as minority states in the global order who seek long-term survival – and also novel but practical organizational architecture to implement that vision.

Militaries must also change, preparing to "fight" a constant war in AI-led military operations while collectively embedded in the community of democratic states.

PART IV. AI AND DOMESTIC IMPACTS ON CHINA'S FOREIGN POLICY DECISION-MAKING

In **Chapter 14, Benjamin Chang** asks: How will domestic use of AI affect Chinese foreign policy, particularly with respect to US-China Relations? Drawing on relevant threads of political science, he discusses two possible consequences: (1) significantly worsened US-China relations due to increased ideological friction and opacity, and (2) increased Chinese assertiveness due to increased confidence and a smaller "winning coalition." Finally, he assesses implications for US policy.

In **Chapter 15, Kacie Miura** discusses the implications of increased internal control on China's international behavior. Although a small group of top leaders dictate foreign policy-making in China, several key domestic factors constrain and complicate China's international behavior. These include: regime insecurity, public opinion, factional competition, and bureaucratic discord. AI—if it improves the Chinese leadership's ability to monitor and control societal and elite actors—could presumably reduce the influence of these internal drivers of China's international behavior. This will allow China's leaders to more efficiently advance their aspirations for China's position in the world, regardless of whether they choose to do so through confrontational or cooperative foreign policies.

In **Chapter 16, Rachel Esplin Odell** explores the crucial links between Chinese regime insecurity, its domestic authoritarianism, and its foreign policy. Too often, Western narratives fail to perceive that the Chinese Communist Party's authoritarianism is driven by a deep-seated insecurity about its ability to maintain power while reforming its economy. Moreover, Western observers falsely assume China's domestic authoritarianism infuses its international ambitions, leading China to challenge the existing liberal international order. Instead, if the West recognized China's foreign policy behaviors as largely status quo-supporting efforts to foster economic growth, the West could craft more effective, positive-sum policies in response.

In **Chapter 17, Rogier Creemers** examines the international and foreign policy impact of China's AI and big data strategies. In the past few years, China has embarked upon an ambitious strategy to build up its capabilities in AI and big data. The primary aims for this agenda are domestic: transforming the government's social management and governance abilities, and creating new areas for economic growth. Nonetheless, this agenda also has an international impact, both in terms of foreign governments' responses to China's domestic strategy, and the extent to which Chinese technologies are exported or become part of global cyber processes. This chapter reviews the development of this agenda, and assess its impact for China's foreign policy.

PART V. AI AND MILITARY DIMENSIONS IN INTERNATIONAL COMPETITION

In **Chapter 18, Martin Libicki** argues AI will change the character of warfare by making hacking more important, and by changing hacking. Computer hacking may be understood as the search for vulnerabilities in opposing systems whose exploitation permit leverage: small efforts have great effect. Injecting artificial intelligence into systems systematizes the hackers' search for vulnerabilities. Moreover, AI also multiplies vulnerabilities. Systems can be trained on a corpus of expected environments, but if the other side generates edge cases that the defender failed to imagine; then the receiver's AI may exhibit behavior favorable to the hacker. In sum, as AI becomes more important, searching for such vulnerabilities will likely constitute a growing share of military activity.

In **Chapter 19, Herbert Lin** examines the risks of conflict escalation from AI-enabled military systems. He describes how AI may feed deliberate, inadvertent, accidental, or catalytic escalation. Today's AI—in particular, machine learning—poses particular risks because the internal workings of all but the simplest machine learning systems are for all practical purposes impossible for human beings to understand. It is thus easy for human users to ask such systems to perform outside the envelope of the data with which they were trained, and for the user to receive no notification that the system is indeed being asked to perform in such a manner.

In **Chapter 20, Elsa Kania** examines AI in future Chinese command decision-making. The Chinese People's Liberation Army (PLA) is exploring the use of AI technologies to enhance future command decision-making. In particular, the PLA seeks to overcome admitted deficiencies in its commanders' capabilities and to leverage these technologies to achieve decision superiority in future "intelligentized" (智能化) warfare. Chinese military experts have examined the DARPA program Deep Green, and are inspired by AlphaGo's recent successes. The PLA's apparent expectation that the future increases in the tempo of operations will outpace human cognition could result in a pragmatic decision to take humans "out of the loop" in certain operational environments. In others, the PLA also recognizes the importance of integrating and leveraging synergies among human and machine "hybrid" intelligence.

In **Chapter 21, Lora Saalman** gains insight into Chinese AI research using an illuminating case, Chinese efforts to integrate neural networks into its hypersonic platforms. Based on analysis of over 300 recent Chinese technical journal papers and articles issued by researchers at Chinese university and military institutes, she uncovers three major trends. First, increasingly innovative and prolific research. Second, expanded domestic and international collaboration. Third, is a quantitative and qualitative shift away from defensive countermeasures to offensive platforms, suggesting a trend from China's traditional stance of "active defense" towards a stronger, AI-enabled offense.

In **Chapter 22, Samuel Bendett** examines Russia's expanding AI development. The Russian government's increasing attention to developing AI-assisted and AI-facilitated technologies drives this expansion. Moscow's AI development still lags far behind nearest peer competitors like China and the US. But progress is evident. Specifically, the Russian military is investing heavily in creating the intellectual and physical infrastructure for AI development across its services. The government is also eager to expand debate and cooperation between the country's growing hi-tech private sector and expansive military-academic infrastructure.

PART VI. ARTISTIC PERSPECTIVES AND THE HUMANITIES

Chapter 23 is the short story "Infinite Bio-Intelligence in the World of Sparrows" by **Eleonore Pauwels** and **Sarah Denton.** *Nothing lives or dies without being monitored*. In a future where artificial intelligence, advanced genomics and biotechnologies converge, we will constantly be aware of the biological evidence we unwittingly leave behind as we go about our daily lives. In this fictional, futuristic scenario, the authors attempt to convey the social, political, and ethical implications of deploying such technologies without regard for human rights. What's lost in the fray of the "Internet of Bodies," the ubiquitous bio-surveillance network, are the human stories that emerge from such a system. This is one such story.

Chapter 24 is the visual piece "Two Memos from the Future" by **Lydia Kostopoulos.** In efforts to look backwards into the present, she has chosen futuristic scenarios to help us visualize the future in a way that technical reports do not. Predicting the future in an era of exponential change and rapid

technological convergence is partly making an educated guess based on technological assessments—and partly creative exploration of the status quo and imaginative alternatives. These scenarios are on the horizon in some form or another.

Chapter 25 is the short story "The Parade Cleaners" by **Lt Col Jennifer Snow.** The story explores one of the darker futures of a burgeoning surveillance state. We follow Chad, a security worker responsible for digitally patrolling the prestigious main thoroughfare. The short proposes some challenging questions for our growing global information culture and pushes the reader to consider "what if?" Are these potential technological calamities that could or are becoming real today? Who determines which people benefit and which people do not? The AI programmers? The government? The public? What is the future of free speech, public access, or upward mobility in an increasingly divided global infospace between authoritarian and libertarian ideals?

Chapter 26 is the essay "Beware the Jabberwocky: The AI monsters are coming" by **Natasha Bajema**. Science fiction plays an important role in shaping our understanding of the implications of science and technology and helping us to cope with things to come. This artistic piece describes three AI monsters depicted in science fiction films as one day disrupting the global order and potentially destroying humanity: the automation monster, the supermachine monster, and the data monster. Fears about the implications of the automatic and supermachine monsters distract us from the scariest of them all. Below the surface of our daily lives, the data monster is stealthily assaulting our sense of truth, our right to privacy, and our freedoms.

Chapter 27 is the essay "Is China's AI Future the Snake in the Wine? Or Will Our Future Be FAANGed?" by **Regina Joseph.** China's urgent plan to dominate in AI is characterized in similar world-changing terms to Silicon Valley's. Both portrayals emphasize limitless opportunity, brilliance, and social good. But a different potential lurks beneath. In the US, younger generations seem to slowly recognize the bondage posed by addictive technologies—a fate prophesized by Aldous Huxley's Ultimate Revolution. In China, centralized control and soft coercion stymie public opposition to techno-nationalism, leading to an unchecked zeal for AI expansion that will adversely affect China, the US and beyond.

Table of Contents

Opening Remarks: US Perspective ii
Brig Gen Alexus Grynkewich

Opening Remarks: UK Perspective iii
Sir Lawrence Freedman

Executive Summary v

Acronyms xiv

PART I. AI TECHNOLOGIES, POLITICAL REGIMES AND THE GLOBAL ORDER 1

Chapter 1. The Technologies: What Specifically is New? 1
Nicholas D. Wright

Chapter 2. AI's Three Bundles of Challenges for the Global Order 10
Nicholas D. Wright

Chapter 3. AI and Domestic Political Regimes: Digital Authoritarian, Digital Hybrid and Digital Democracy 16
Nicholas D. Wright

Chapter 4. Global Competition 30
Nicholas D. Wright

Part II: DIGITAL AUTHORITARIANISM: EVOLVING CHINESE AND RUSSIAN MODELS 37

Chapter 5. The Interests Behind China's AI Dream 37
Jeffrey Ding

Chapter 6. Managing the State: Social Credit, Surveillance and the CCP's Plan for China 42
Samantha Hoffman

Chapter 7. Credit Cities and the Limits of the Social Credit System 48
Shazeda Ahmed

Chapter 8. The Russian Model of Digital Control and Its Significance 55
Jaclyn Kerr

PART III. EXPORT & EMULATION OF THE MODELS IN GLOBAL COMPETITION 72

Chapter 9. Understanding the Global Ramifications of China's Information Controls Model 72
Valentin Weber

Chapter 10. Pathways to Lead in Artificial Intelligence 78
Laura Steckman

Chapter 11. The Spread of Russia's Digital Authoritarianism 85
Robert Morgus

Chapter 12. AI and China's Unstoppable Global Rise: A Skeptical Look 94
James A. Lewis

Chapter 13. Four Horsemen of AI Conflict: Scale, Speed, Foreknowledge, and Strategic Coherence 100
Chris C. Demchak

PART IV. AI & DOMESTIC IMPACTS ON CHINA'S FOREIGN POLICY DECISION-MAKING 107

Chapter 14. AI and US-China Relations 107
Benjamin Angel Chang

Chapter 15. The Implications of Increased Internal Control on China's International Behavior 112
Kacie Miura

Chapter 16. Chinese Regime Insecurity, Domestic Authoritarianism, and Foreign Policy 116
Rachel Esplin Odell

Chapter 17. The International and Foreign Policy Impact of China's AI and Big Data Strategies 122
Rogier Creemers

PART V. MILITARY DIMENSIONS 128

Chapter 18. A Hacker Way of Warfare 128
Martin Libicki

Chapter 19. Escalation Risks in an AI-Infused World 133
Herbert Lin

Chapter 20. Artificial Intelligence in Future Chinese Command Decision-Making 141
Elsa Kania

Chapter 21. China's Integration of Neural Networks into Hypersonic Glide Vehicles 153
Lora Saalman

Chapter 22. The Development of Artificial Intelligence in Russia 161
Samuel Bendett

PART VI. ARTISTIC PERSPECTIVES AND THE HUMANITIES 170

Chapter 23. Infinite Bio-Intelligence in the World of Sparrows 170
Eleonore Pauwels & Sarah W. Denton

Chapter 24. Memos from the Future 173
Lydia Kostopoulos

Chapter 25. The Parade Cleaners 176
Jennifer Snow

Chapter 26. Beware the Jabberwocky: The AI Monsters Are Coming 180
Natasha E. Bajema

Chapter 27. Is China's AI Future the Snake in the Wine? Or Will Our Future Be FAANGed? 187
Regina Joseph

Biographies 192

Acronyms

AI	artificial intelligence
AIDP	artificial intelligence development plan
AR	augmented reality
ARF	Advanced Research Foundation
BMI	brain machine interface
BRI	Belt Road Initiative
BRIC	Brazil, Russia, India, China
CCP	Chinese Communist Party
CECIC	China National Electronics Import & Export Corporation
CIA	Central Intelligence Agency
CMC JSD	Central Military Commission Joint Staff Department
CORA	Cyber Operational Resilience Alliance
DARPA	Defense Advanced Research Projects Agency
DOD	Department of Defense
EEG	electroencephalogram
EU	European Union
FAANG	Facebook, Apple, Amazon, Netflix, and Google
GDP	gross domestic product
ICT	information and communication technology
IoT	Internet of Things
ISP	internet service provider
IT	information technology
MIIT	Ministry of Industry and Information Technology
ML	machine learning
MLP	medium- and long-term plan
MOD	Ministry of Defense
NDRC	National Development and Reform Commission
OBOR	One Belt One Road
PLA	People's Liberation Army
PRC	People's Republic of China
R&D	research and development
RMA	revolution in military affairs
SA	situational awareness
SORM	System for Operative Investigative Activities
STEM	science technology engineering math
STES	socio-technical-economic system
UNGA	UN General Assembly
VR	virtual reality

PART I. AI TECHNOLOGIES, POLITICAL REGIMES AND THE GLOBAL ORDER

Nicholas D. Wright
Intelligent Biology; Georgetown University;
University College London; New America
nick@nicholasdwright.com

Chapter 1. The Technologies: What Specifically is New?

Abstract

This chapter discusses the new technologies:

- By "AI" here we mean a constellation of new technologies: AI itself more narrowly defined, big data, machine learning, and digital things (e.g. the "internet of things").
- This constellation of technologies is bringing in a new technological epoch. Following a leap in AI research around 2012, we now have: *Automated systems learning directly from data to do tasks that are complicated.* The key change is that the task is now complicated (e.g. AI can now do good facial recognition).
- Crucially, AI particularly improved for tasks related to "perception"—e.g. perceiving images or speech, or some kinds of patterns in big data—and these are the advances now being rapidly rolled out across diverse real-world uses. AI also improved when choosing actions in tasks that are bounded enough to be very well described by vast amounts of data.
- AI's current technical limitations mean that its current incorporation into social governance must include extensive human involvement; and also, that setting up big data platforms will likely dominate current efforts. This is what we see now in China.
- AI adds a new layer to traditional "cyber."

What are the AI-related technologies?[1]

By the term "AI" here we refer to a constellation of AI-related technologies (AI more narrowly defined, machine learning, big data and digital things) that together provide powerful, wide-ranging and new capabilities (Fig. 1.1).[2] Together they enable a new industrial revolution, taking the vast reams of data now produced by the computers and internet of the preceding revolution – and turning it into useful data (Box 1.1). None of the technologies is entirely new, but there have been big recent improvements (particularly from deep learning, see below) and together the constellation has revolutionary applications. Within the constellation of new technologies, four are crucial[3]:

---

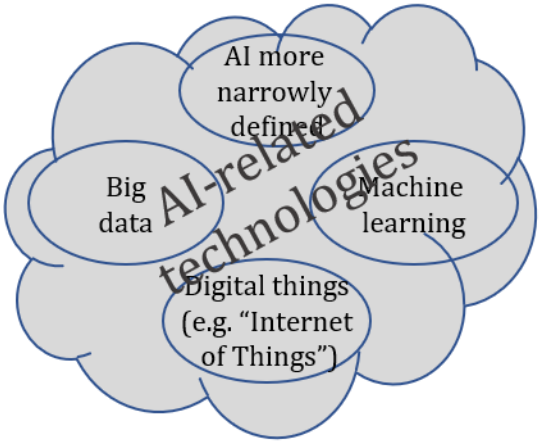
[1] I thank Zeb Kurth-Nelson for insightful discussions on the technical aspects of AI research included here.

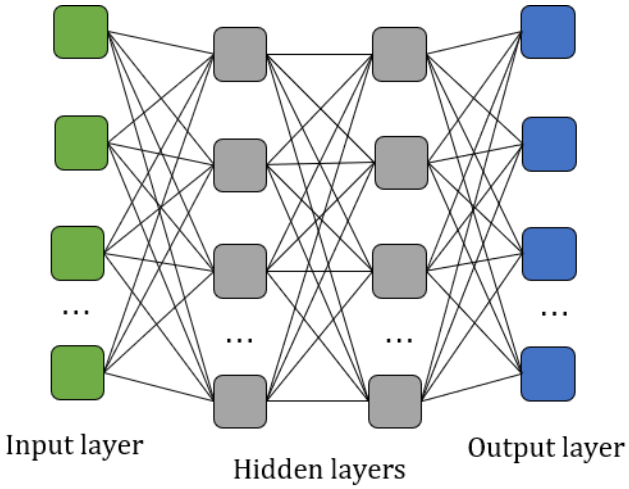
[2] We use a broad characterization here because the term "AI" has come to refer to many significant things that are not captured by narrower definitions. An analogy is the term "rational", which means many things to many people.

[3] This subsection's definitions draw in particular on (ICO, 2017).

- **"AI" more narrowly defined**[4]**:** One can describe AI as the analysis of data to model some aspect of the world, where inferences from these models are then used to predict and anticipate possible future events. Importantly, AI programs don't simply analyze data in the way they were originally programmed. Instead they learn from data in order to respond intelligently to new data and adapt their outputs accordingly. AI is ultimately about "giving computers behaviors which would be thought intelligent in human beings" (ICO, 2017).
- **"Machine learning":** Many of the computational techniques related to AI are actually from a field called machine learning. This can be described as "…the set of techniques and tools that allow computers to 'think' by creating mathematical algorithms based on accumulated data." Arthur Samuel coined the phrase, in 1959, defining it as, "the ability to learn without being explicitly programmed." (McClelland, 2017). **Deep learning** is one method for machine learning – and it is improved deep learning that recently led to big advances in AI (Fig. 1.1 right panel).
- **"Big data":** These are high-volume—as well as often high-velocity and high-variety—information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. The massive recent increase in the amount of big data everywhere is new.
- **Digital things:** Things (e.g. smartphones, "Alexa," toasters, military drones, robots in factories) will increasingly be able to perceive (e.g. facial or speech recognition), decide and act. The things may not be connected to the internet, but may be smart. The "Internet of Things" refers to the growing interconnectedness of things, and getting them onto the Internet.

Together these technologies are more than the sum of the parts. Firstly, consider **"Big data analytics."** One can think of big data as an asset that is hard to exploit, for which AI is a key to unlocking its value, and where machine learning is one technical mechanism for doing AI. When big data analytics is merged with things in the real world, we have **online-to-offline** merging.





[4] This narrower definition of "AI" itself is also highly debated, and is further subdivided in various ways. For instance, one might contrast "general AI" that can apply its intelligence to many tasks, against an AI such as Siri that is programmed to essentially perform a single task (called "narrow AI", although not AI more narrowly defined in the sense we use in this whitepaper that also includes "strong" or "general AI"). This also broadly corresponds to "strong AI" versus "weak AI."

Fig. 1.1 The left panel shows the ***constellation of AI-related technologies****. The right panel illustrates* ***"deep learning."*** *Deep learning is one of many approaches to machine learning. It was inspired by the brain, and in particular the interconnecting of many neurons (Artificial Neural Networks). In deep learning, the key idea is that the neural networks have at least one "hidden layer" in the middle between inputs and outputs, whose "neurons" can take on different weights while learning about the task.*

What was the new big improvement around 2012?

The computer and internet-related revolution made lots of data, and now this AI-related revolution turns that data into usable information. We are early in this new epoch. Around 2012 researchers made a large improvement in the quantity of big data that automated systems can analyze, which was sufficiently large that it essentially provided qualitatively new capabilities. After 2012 we have qualitatively new:

Automated systems learning directly from data to do tasks that are complicated.

We can unpack this. "Automated systems" means AI programs themselves, not relying on humans. "Learning directly from data" means the way the AI doing the job does not depend on hard coding from humans. A task is something like facial recognition. That the task is now complicated is the key change, and how we measure complexity may be via comparisons to human performance, or previous AI performance. For instance, AI can now do good facial recognition. These basic advances were those leveraged to achieve AlphaGo's victory over a top human in 2016.

Two papers signaled and illustrate this change:

(1) (Krizhevsky, Sutskever, & Hinton, 2012): This was a big breakthrough in perception. In a visual object recognition task, they trained a deep convolutional neural network to classify visual images. They trained the neural network on 1.2 million images—all labelled—from a huge and then new dataset called "Imagenet". They roughly halved the error rate of the previous state-of-the-art on the most challenging benchmark to date. Such AIs recently approached human-level performance on some object recognition benchmarks. This paper triggered huge interest in AI research, with some 33,000 Google Scholar citations in under six years.

(2) (Mnih et al., 2015): The AI learned to play a large range of classic "Atari" computer games, with essentially the only inputs being the pixels on the screen and the game score – and it achieved human or superhuman performance on many games. It had to deal with the huge perceptual challenge, and also control actions. They combined ideas from deep learning and reinforcement learning (i.e. learning from the rewards and punishments associated with previous events). Within the tightly bounded environment in each game, the AI could play vast numbers of times to learn from a huge dataset on each game environment.

Such advances were crucial for AlphaGo's famous 2016 victory over a world-class human go player. Go is a lot more difficult than chess. Within the tightly bounded environment of go, before beating world champion Lee Sedol, AlphaGo effectively learned from some 100 million or more games altogether (Lake, Ullman, Tenenbaum, & Gershman, 2017).

What led to this big change? There was no magic bullet. Instead three factors combined:

(1) Raw compute power increased.

(2) Datasets for training became available. For instance, the advance by Krizhevsky et al. (2012) was possible because they had a huge dataset of millions of labelled images on which to train. The imagesets often need to be labelled, so the AI can learn.

(3) Deep learning algorithms were improved. It was not a single innovation, but instead multiple moderate improvements (e.g. "dropout" and "ReLUs" in the 2012 "Imagenet" advance).

Current Strengths and Weaknesses – and Where AI is Going

These advances have been huge but not uniform, and it is important to understand the technical strengths and weaknesses. This helps understand both what we might expect to see in real world applications—for instance in the construction of a surveillance state—and also where the research is likely to go.

Strengths

The new technology has two big new **strengths**.

(1) First, one really huge new improvement in AI capabilities relates primarily to "perception", such as perceiving images or speech, or patterns in some types of big data that humans may not be able to perceive. That is what the 2012 advance in classifying "Imagenet" pictures was all about in the preceding subsection.

Thus, now local devices such as smartphones, digital assistants or cheap cameras in office lobbies can effectively monitor speech or faces – and indeed such technology is already widespread in the West and China. One can see why this is particularly good for surveillance, as discussed later in this whitepaper.

Moreover, being able to learn to perceive well also means that if you reverse those models you can be very good at producing images or audio. In a strategic context that may be useful for fooling others (e.g. "deepfakes").

Databases of data, much of which may have originally been collected for other purposes, can also be examined for patterns – adding value to the "big data" that may just have been sitting there.

(2) Second, AI also improved in choosing actions in tasks that are bounded enough to be very well described by vast amounts of data. Go or the Atari games above are a good example. A well-known real-world example is Google Deepmind training AI on data from Google's datacenters, and so "more accurately predicting when the incoming compute load is likely to land", which reduces power consumption for cooling (Burgess, 2016).

Current limitations

However, there are two major **limitations** in the current AI technology. These help us know what we should expect if these new AI-related technologies were applied in domestic security.

(1) *Huge amounts of data are needed to train the system, and this data often needs to be labelled* (e.g. this is a picture of a cat). The availability of a huge dataset of labelled images—"Imagenet" described above—was a crucial factor enabling the big leap in 2012. The algorithms cannot yet generalize well from learning in one environment to learning in another, and also they cannot

learn things from just a few instances as humans often can.

As discussed in later chapters, this is why it is so important in a surveillance state to add "ground truth" data (e.g. tax returns, criminal records or medical records) that acts like labels for your broader data (e.g. smartphone usage; Fig. 3.3).[5] Often governments are the only parties with such data (e.g. tax returns) or they heavily regulate who can access data (e.g. medical records or genetic data). Without the ground truth data, just having tons of big data by itself will be a lot less useful.

Moreover, this greatly raises the value of having very detailed monitoring specific populations with extensive ground truth data, because you can then use that very detailed data to train your algorithms. For those working on Chinese surveillance, that would be one big advantage of the very heavy physical and online monitoring in Xinjiang province.

Further, this is a good reason why lots of humans will be needed in any AI system of surveillance for the foreseeable future – to do labelling. Indeed, the importance of cheap labor for labeling has even been touted as a key Chinese strength in AI more broadly (Yuan, 2018).

Making the datasets is a huge challenge. Creating the "Imagenet" labelled dataset was a *pre*condition of the leap made in 2012. Similarly, building big datasets that are in right form with the right type of labelling and so on should be the current major effort, if one were building an AI-enabled surveillance state now.

(2) *Context is still very poorly understood by the systems* – that is, they lack common sense (e.g. is this likely to be a picture of a baby holding a toothbrush or a gun?). This is why human-machine teams and semi-automated systems are often the only way to harness the benefits of AI, by adding the human ability to add context.

The challenge of context is another key reason why any plausible surveillance system will only be semi-automated for the foreseeable future – lots of humans would still be needed even if a system built with current cutting-edge AI technology worked perfectly.

Where is AI going in the lab and at scale in the real-world?

Given the state of AI-related research, where might we expect the technologies to go over the next five years or so?

First, we might ask: where will the cutting-edge research go? Efforts to overcome the limitations above are perhaps the two hottest current research areas – and given the huge resources being spent to overcome them, this is where to expect potential research advances. The scientific literature is looking to augment deep-learning methods that learn from experience, for instance by adding more informed models of the world. Just as the human brain does, and as AlphaGo arguably began to do (Lake et al., 2017). To give a flavor of the Defense Advanced Research Project Agency's (DARPA)[6]

[5] One example would be training a program to predict tax payment (or avoidance) based on innumerable aspects of smartphone data. Or training a program to predict criminal acts, including those that may have political dimensions, from smartphone data. An analogy is given by recent reports of the Chinese company Smart Finance, which uses seemingly irrelevant smartphone data, such as the typing speed or battery charge levels, to predict individuals' creditworthiness for loans, with reportedly high accuracy (Lee, 2018). One could also create links between such systems.

[6] Defense Advanced Research Project Agency https://www.darpa.mil/about-us/darpa-perspective-on-ai (retrieved Dec 2018)

goals with AI, they describe a focus on moving beyond what they call "second wave AI" of the type we have now—which is good at perception and learning but not at abstraction and reasoning—and towards "third wave AI." That third wave AI aims at "Contextual Adaptation [in which] Systems construct contextual explanatory models for classes of real-world phenomena."

But while that is the cutting-edge research, crucially we must also allow for time lags: not just from lab to real-world, but also to large-scale in the real-world. The internet, for instance had certainly reached the real-world by the early 1990s—growing up then in London my family had an internet connection—but many of the internet's large-scale real-world impacts took another one decade or two to occur, such as Amazon or Facebook reaching their huge scale.

Thus, second we might ask: where are the AI-related technologies going in the real-world at scale? AI advances in visual or speech perception are now rolling out at huge scale in our smartphones and digital assistants. However, we are still working on how to usefully use all the information this produces, and how to work that into broader commercial or governmental systems. Perceiving patterns in big data that are hard for humans to perceive will almost certainly bring about great advances fields like medicine or predictive policing – but it is important to realize that such real-world applications are only in development ("The Promise and Perils of AI Medical Care," 2018) and fully operational systems are certainly not ready for rollout at huge scales. Driverless vehicles still require a lot more training data, and are now being deployed in very limited circumstances, such as a pilot taxi service in Phoenix or Tesla's partial driving assistance. Overall, for most AI-related technologies the next five years will likely see a lot of piloting to find out what works in the real-world, while building the crucial datasets and also assessing how the technologies can later be rolled out at scale. That is also what we would expect if one were building these powerful new technologies into a surveillance state, and for instance is what we see now in China.

How do the New AI-Related Technologies Relate to Traditional "Cyber"?

AI adds new properties and a new layer of value to what can be done using the Information and Communication Technologies (ICT). It is a bit like an onion. As depicted in Figure 1.2, each new layer adds value to that beneath. The internet increased the value of computers. As so many things have become computers and can communicate electronically this has generated huge amounts of data—big data. The new AI-related technologies help turn this big data into something useful. They add more value.

The prefix "cyber" essentially relates to computers and electronic communication.[7] Much of what happens with information will involve traditional issues in "cyber" rather than AI, so we don't need AI to tell us much about them. Computers still matter – hence the "chip wars" between the US and China ("Chip wars," 2018). Communications still matter – hence the battles over "5G" standards, over social media regulation within and between countries, as well as the global struggle for internet governance. The challenges and opportunities of cyber will still be meaningful, it is just that there will also be other things from the new layer of the onion.

[7] The academic Thomas Rid notes "the increasing use of the word "cyber" as a noun among policy wonks or many a uniformed officer. … I've come to be highly distrustful of "nouners," as they all too often don't seem to appreciate the necessary technical details" p. ix (Rid, 2013). Acknowledging that the term is suboptimal, I use it here as the prefix at least does have meaning amongst scholars and the word amongst practitioners.

Finally, the different applications to which one can put the AI-related and other digital technologies speaks to an important question: *does China have a data advantage over the liberal democracies?* One can make two points here:

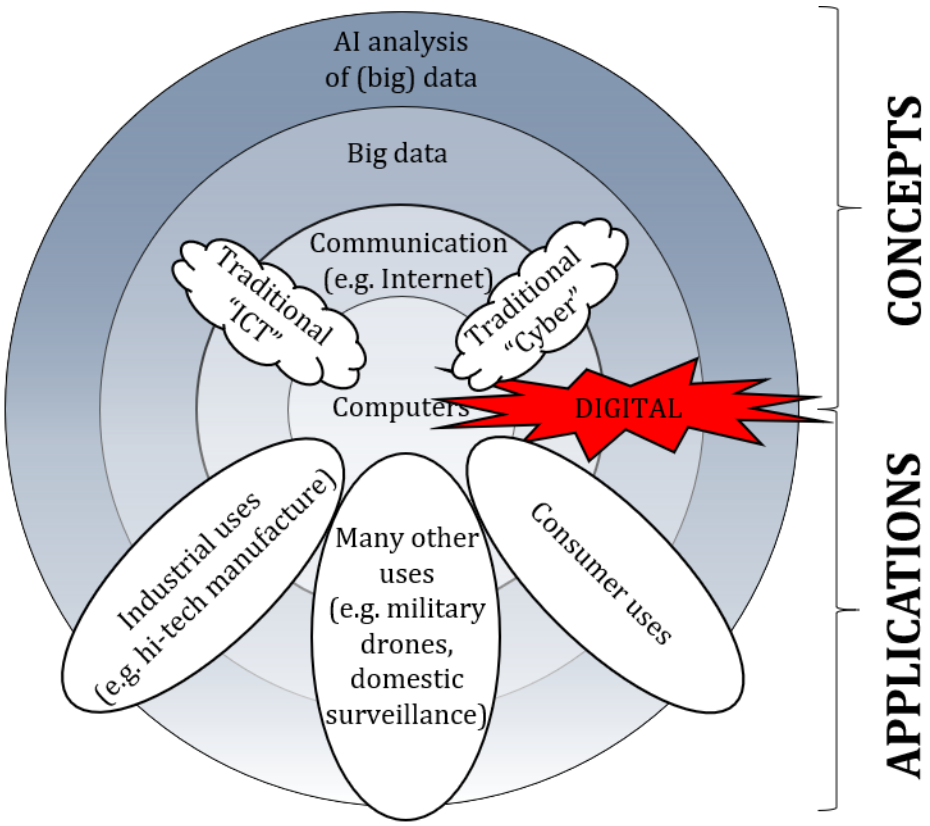


Figure 1.2 The onion of digital technologies

(1) China does not have an advantage in terms of number of users if one includes the global userbases for US tech giants like Facebook or Google. But it does have an advantage in terms of integration of data across platforms[8] and also, most importantly, in terms of combining breadth of data with "ground truth" data (e.g. government data). Training AI depends on both quantity and quality. The liberal democracies *should not* compete.

(2) The above comments relate to human user data, for example for consumer uses or domestic surveillance. However, a lot of important AI training will occur on other types of data such as from sophisticated machines. For example, Germany's AI strategy relies on that alternative aspect of data—which is harnessed from cyber-physical manufacturing processes and the industrial internet of things—that plays to German industrial strengths. Many industrial, military and other applications will rely much more on that type of data than human user or consumer type data.

Conclusions

The advances in the AI-related technologies are very real. They greatly multiply the value derived from the preceding digital technologies, by turning data into useable information. Understanding the current technical strengths and weaknesses helps anticipate and understand its applications in the real world.

A big advance has been in perception; hence the ready application to surveillance and censorship.

[8] Kai-fu Lee's prominent recent book compared Chinese and US approaches to AI. He repeatedly refers to key Chinese apps that bundle together what is done separately in the West by companies like Google, Facebook or Uber (e.g. Chs 1 and 3 in Lee, 2018). An example is the "super app" Wechat, from the tech giant Tencent.

The big limitations are that the AI handles context poorly (hence human-machine teams are key), and that it requires vast amounts of well-labelled data (hence the importance of combining ground truth data—often only from government—with the breadth of data from myriad smart devices and sources). As the big research breakthrough only really began in 2012 we are in the early stages of rolling out many of the new AI capabilities, and so much of what happens now in the real world will essentially be piloting, or the building and preparing of good enough datasets. Indeed, even where AI technologies are being rolled out at massive scale—notably in commercial devices such as smartphones or digital assistants—while their outputs may be dual use for surveillance, usefully harnessing those outputs at societal scale is itself surely a massive additional IT program that requires careful building and piloting.

Finally, in terms of military uses or foreign policy decision-making, these technical characteristics explain why AI's main uses have often been for more perceptual tasks, such as satellite image analysis – as is seen in the Chinese case (e.g. Elsa Kania, Chapter 20 this volume). With current technology, decision support will likely only work well with human machine-teams to provide contextual capabilities – and there are likely considerable risks in allowing many types of decisions to be made with humans "out of the loop."

Box 1.1 Is this new "nth industrial revolution" really distinct from what went before?

This question is closely related to the question: why is AI different to "cyber"? An industrial revolution may be defined as "A general term for the process of the rapid onset of continued economic change and advancement through the application of industrial techniques to traditional forms of manufacture" (Lawrie, 1999). As was recently argued by perhaps the most prominent voice behind the idea that AI reflects a fourth industrial revolution:

"There are three reasons why today's transformations represent not merely a prolongation of the Third Industrial Revolution but rather the arrival of a Fourth and distinct one: velocity, scope, and systems impact. The speed of current breakthroughs has no historical precedent. When compared with previous industrial revolutions, the Fourth is evolving at an exponential rather than a linear pace. Moreover, it is disrupting almost every industry in every country. And the breadth and depth of these changes herald the transformation of entire systems of production, management, and governance." (Schwab, 2015)

I discuss this "nth industrial revolution" in the next chapter.

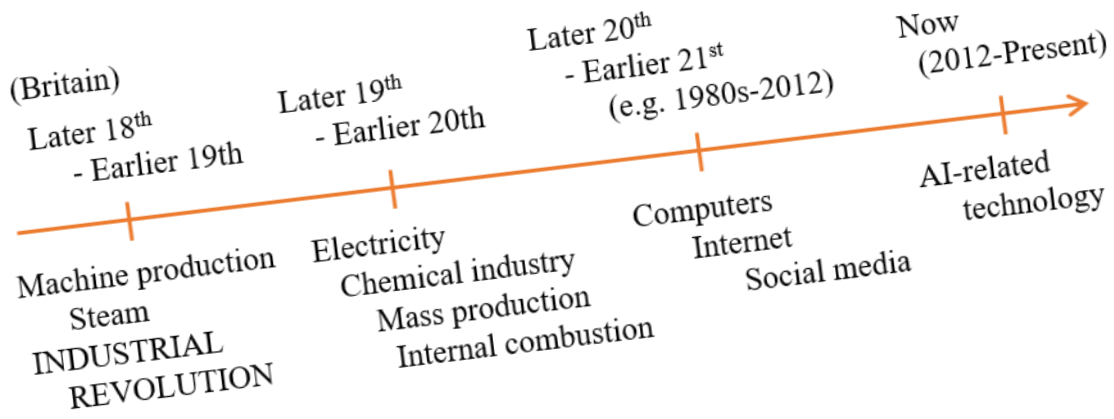


Figure 1.3 Timeline of industrial revolutions.

References

Burgess, M. (2016, July 20). Google's DeepMind trains AI to cut its energy bills by 40%. *Wired UK*. Retrieved from https://www.wired.co.uk/article/google-deepmind-data-centres-efficiency

Chip wars: China, America and silicon supremacy. (2018, December 1). *The Economist*. Retrieved from https://www.economist.com/leaders/2018/12/01/chip-wars-china-america-and-silicon-supremacy

ICO. (2017). *Big data, artificial intelligence, machine learning and data protection v. 2.2*. UK: Information Commissioner's Office. Retrieved from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*.

Lawrie, A. (1999). The New Fontana Dictionary of Modern Thought. HarperCollins.

Lee, K.-F. (2018). AI Superpowers: China, Silicon Valley, and the New World Order. Houghton Mifflin Harcourt.

McClelland, C. (2017, December 4). The Difference Between Artificial Intelligence, Machine Learning, and Deep Learning. Retrieved July 26, 2018, from https://medium.com/iotforall/the-difference-between-artificial-intelligence-machine-learning-and-deep-learning-3aa67bff5991

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529.

Rid, T. (2013). *Cyber War Will Not Take Place* (1 edition). Oxford ; New York: Oxford University Press.

Schwab, K. (2015, December 12). The Fourth Industrial Revolution. *Foreign Affairs*. Retrieved from https://www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution

The Promise and Perils of AI Medical Care. (2018, August 15). Retrieved from https://www.bloomberg.com/news/articles/2018-08-15/the-promise-and-perils-of-ai-medical-care

Yuan, L. (2018, November 27). How Cheap Labor Drives China's A.I. Ambitions. *The New York Times*. Retrieved from https://www.nytimes.com/2018/11/25/business/china-artificial-intelligence-labeling.html

Chapter 2. AI's Three Bundles of Challenges for the Global Order

Abstract

AI raises a bewildering profusion of implications for the global order – which this chapter breaks down into three more manageable bites. This whitepaper primarily focusses on the first area, which has received by far the least attention.

(1) First is how this new technology's may impact on domestic *political* regimes (e.g. authoritarian, hybrid, or liberal democratic) may affect competition between them in the *world order*. AI will help enable a plausible competitor to liberal democracy for big industrially sophisticated states to make their citizens rich and maintain rigid control: digital authoritarianism. China is building core components of such a system – which are being exported and emulated in a global competition with liberal democracy.

(2) An "nth industrial revolution": AI will radically change the means of production across economic and societal sectors, e.g. transport, healthcare or the military.

(3) The "singularity" and the sense of self: In the singularity, exponentially accelerating technological progress creates an AI that exceeds human intelligence and escapes our control, potentially destroying humanity or disrupting humans' conceptions of themselves.

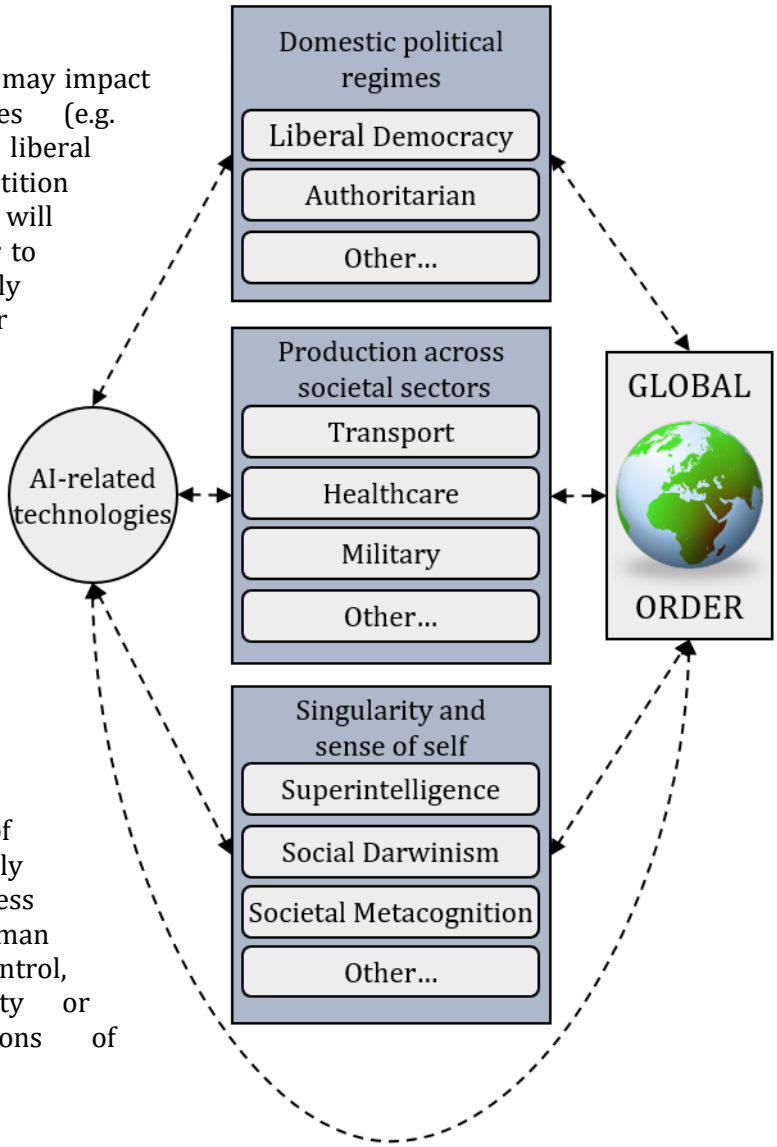


Figure 2.1 AI's impacts on the global order

Introduction[9]

Everybody now seems to agree that AI seems important for everything. From what it means to be human; to the social impacts of laying off Uber drivers once cars drive themselves; to AI propaganda in politics; to the rise of the robots or a superintelligence exterminating humanity. But what does AI's bewildering profusion of implications mean for the global order? Anticipating AI's challenges for the global order requires breaking them down into more manageable bites – because failure in any one

[9] This Chapter draws in part on (Wright, 2018).

of these three distinct bundles of challenges I identify would be catastrophic. As Figure 2.1 shows, they are: (1) Competing types of political regimes in the global order; (2) Change in the means of production across social sectors in an "nth industrial revolution"; and (3) The "singularity" and the sense of self.

This whitepaper focuses primarily on the first area—AI's impacts on competing domestic political regimes in the global order – because it is critical but has been least examined. This chapter also describes the other bundles of challenges for two reasons. Firstly, because global strategy must address all three areas. Each of the three bundles of challenges requires different thinking—and policies—at the level, of nations, of the UN, business and other stakeholders. Second, many who debate the potential importance and/or urgency of AI's impacts on the global order end up discussing different bundles of impacts, and so talk past one another. Here we put them in one space.

(1) Competing Political Regimes in the Global Order

New technologies may affect the form and/or relative attractiveness of different types of *domestic political regimes*—e.g. authoritarian, liberal democratic or hybrids combining features of each—and this may affect competition between such regimes in the *global order*. A domestic political regime is a system of social organization that includes not only government and the institutions of the state, but also the structures and processes by which these interact with broader society. Competition between different types of domestic social system was a crucial feature of twentieth century global politics. While liberal democracy's eventual triumph may now seem inevitable, fascist regimes in the interwar period and communist regimes for much longer were plausible paths forward for big, industrially sophisticated societies to make their citizens rich. Now the new AI-related technologies could crucially help reinvigorate the idea that more authoritarian regimes can make their citizens rich and maintain social control.

Chapter 3 and Part II of this whitepaper examines digital authoritarian regimes in particular in more detail. Chapter 4 and Part III of this whitepaper examines export and emulation of such regimes in global competition.

Before moving on, however, it is important to note multiple reasons why domestic political regimes matter for the global order.

- First, as described above, they may offer competing visions of the future, and these may compete for influence within swing states in global competition. This is only one potential facet of influence between states, but it can be highly significant as it was in the twentieth century. This is particularly the case if large countries develop particular types of domestic regimes, such as Russia in the early twentieth century (i.e. the Soviet-style Communist regime) or potentially China during its rise in the 21st Century.

- Second, aspects of domestic regimes, such as bureaucratic or domestic political audiences, can profoundly affect foreign policy decision-making. Part IV of this whitepaper examines how the development of digital authoritarianism may affect Chinese foreign policy decision-making.

- Third are ideas that some types of regime are inherently less, or more, problematic in the global system. Most prominent is the idea of "democratic peace theory," which identifies a correlation between domestic structure and the absence of war between democracies – a very prominent notion among scholars and indeed practitioners (Russett, Layne, Spiro, & Doyle, 1995).

(2) Change in the Means of Production Across Social Sectors

A second basket of challenges arise because AI and big data will radically change the means of production across many economic and societal sectors. There will be winners, losers and new ways of doing things, which will roil societies across the globe. Consider three sectors. One now classic example is transport: after Uber rolls out self-driving cars, where will all the unemployed drivers work (Edwards, 2017)? Another sector is the military. Drones and AI will likely contribute to a revolution in military affairs, which may be destabilizing (Horowitz, 2018). There may be arms races. A third example is the colossal health sector, accounting for some 18% of U.S. GDP (CMS.gov, 2018), where AI promises to change how medical decisions are made and care delivered ("A revolution in health care is coming," 2018). One can point to essentially any social sector.

But not much so far suggests this will be bigger than other technological impacts, such as those contributing to the industrial revolution itself - or the internet's rise in the 1990s-2000s. Uber drivers being sacked isn't much different to the internet reducing retail jobs with Amazon's rise.[10] The airplane, steamship, machine gun or tank all revolutionized warfare; and so did the internet, with cyber now a military domain alongside land, sea, air and space. Potential change in healthcare is exciting but powerful human, regulatory and institutional factors make the health sector as nimble as a supertanker. One potential caveat is that the rapidity of these changes renders them different, but as Chapter 1 describes making many of the AI-related technologies work in the real-world at scale means overcoming numerous tough practical problems, which downloading a software update won't solve.

We might call this an "*n*th Industrial Revolution", as the popular term fourth industrial revolution has been around since the 1940s (Thornhill, 2018). These changes and their attendant disruptions will require management, just as welfare states were created and adapted to manage the social disruptions from industrialization. It requires sector-by-sector planning. Much will rely on relatively straightforward, although politically challenging, means such as welfare nets and retraining for the swathes of workers whose jobs become obsolete.

Changes in the means of production matter for the global order for multiple reasons:

- First, 20th Century history illustrates how failure to manage domestic social dislocations, such as in interwar Germany, disrupts global order.

- Second, 20th Century history also illustrates how new military technologies can affect the balance of power or strategic stability. Chapter 4 and Part V in this whitepaper examines military aspects.

- Third, how well different countries harness new technologies within their societies can affect the relative balance of power between them. An example is that while Britain dominated economically in the original industrial revolution, instead Germany and the US harnessed twentieth century technologies equally well or even slightly better.

- Fourth, it is possible that the AI-related technologies may alter the inequalities in power between nations. For instance, if robotic manufacturing becomes highly effective, this may remove a key advantage that poor countries have traditionally had when developing – supplies low-skilled workers for labor-intensive manufacturing such as in textiles. The AI-related technologies may

[10] See e.g. (Thompson, 2017) but note e.g. (Manne & Maclean, 2017).

also exacerbate inequalities between the developed economies, as we have seen to some extent with US tech giants totally unmatched in Europe or Japan.

(3) The "Singularity" and the Sense of Self

The singularity is the single biggest concern for many AI scientists. The idea is that exponentially accelerating technological progress will create an AI that exceeds human intelligence and escapes our control ("What is the Singularity?," 2018). This superintelligence may then deliberately or inadvertently destroy humanity, or usher in an era of plenty for its human charges. As Henry Kissinger also describes, the catastrophic consequences may not only be physical but also apply to humans' conceptions of themselves (Kissinger, 2018). For him, the most important question is: "what will become of human consciousness if its own explanatory power is surpassed by AI, and societies are no longer able to interpret the world they inhabit in terms meaningful to them?" Given the rate of progress, the singularity may occur some point this century.

But although clearly momentous, given that nobody knows when, if or how a possible singularity will occur, limits clearly exist on what can sensibly be said or planned for now. Previous existential technologies have emerged: nuclear weapons can obliterate humanity. Indeed, nuclear weapons provide a useful, although imperfect, analogy for global efforts to manage or prevent a singularity. Preventing nuclear war required careful management and luck, which we will need again. Preventing nuclear proliferation is tough, and despite considerable success we couldn't prevent North Korean nuclear weapons. Could one persuade Russian, Chinese or U.S. leaders to stop AI programs viewed as vital for their security? Indeed, this is more concerning than Kissinger's further concerns about human understanding of our own nature. Human egocentrism is remarkably robust – if we can (despite wobbles) deal with Darwin telling us we're just hairless apes, we'll survive this new disclosure.

The bottom line is that, just like nuclear weapons, singularity-related issues will require managing within the international order as best we can, although our best will inevitably be grossly imperfect. The singularity potentially represents a qualitatively new challenge for humanity that we need to think through and discuss internationally. But plenty of other fish also need frying, and a lot sooner.

Conclusions

Global strategy must address all three bundles of challenges that AI presents for the global order. Most attention has been paid to the singularity and a new industrial revolution. Thus, this whitepaper focuses primarily on an equally crucial bundle of challenges for the global order, posed by AI's implications for domestic political regimes.

Box 2.1 What is the global order?[11]

Below I give my working definition and some other examples of definitions, so that the reader can see the concept's broad shape.

My working definition: The global order is a system covering the whole of human society that includes: (1) social institutions around which actors' expectations converge; and (2) the distribution of power amongst key subsystems in the global system, where these subsystems include states (e.g. the US or China), international subsystems (e.g. the global financial system or the UN), and important systems at other levels (e.g. regions below the level of the state, such as Catalonia; or systems above the level of the state, such as during the Cold War there were the liberal international system and the Communist international system).

That is, the global order is a system of systems. It involves material factors, subjective ideas/perceptions, path dependence (i.e. "history matters") and multiple levels.

A textbook definition: "World order is the distribution of power between and amongst states and other key actors, giving rise to a relatively stable pattern of relationships and behaviours." (Heywood, 2013) p. 422.

A prominent academic definition: "International regimes have been defined as social institutions around which actor expectations converge in a given area of international relations. Accordingly, as is true of any social institution, international regimes limit the discretion of their constituent units to decide and act on issues that fall within the regime's domain. And, as is also true of any social institution, ultimate expression in converging expectations and delimited gives international regimes an intersubjective quality." … "The analytical components of international regimes we take to consist of principles, norms, rules, and procedures." (Ruggie, 1982) p. 380.

Henry Kissinger's recent book "World Order" gives the following definition: "World order describes the concept held by a region or civilization about the nature of just arrangements and the distribution of power thought to be applicable to the entire world. An international order is the practical application of these concepts to a substantial part of the globe—large enough to affect the global balance of power. Regional orders involve the same principles applied to a defined geographic area. Any one of these systems of order bases itself on two components: a set of commonly accepted rules that define the limits of permissible action and a balance of power that enforces restraint where rules break down, preventing one political unit from subjugating all others." (Kissinger, 2014) p. 9.

References

A revolution in health care is coming. (2018, February 1). *The Economist*. Retrieved from https://www.economist.com/leaders/2018/02/01/a-revolution-in-health-care-is-coming

[11] Many different terms are used and discussed, such as "World Order", "International Order", "Liberal World Order" or "New World Order". I prefer the term "global system." However, here I avoid including the word "system" to prevent confusion that may arise as this whitepaper also discusses systems at many other levels within the global order. For instance, domestic political regimes may be called systems (see below), while the digital social governance systems used and planned in China are in fact best thought of as systems of systems.

CMS.gov. (2018, December 11). NationalHealthAccountsHistorical. Retrieved December 20, 2018, from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html

Edwards, Ji. (2017, October 1). London's Uber ban shows how driverless cars will cut jobs - Business Insider. Retrieved December 20, 2018, from https://www.businessinsider.com/the-london-uber-ban-and-driverless-cars-2017-9?r=UK&IR=T

Heywood, A. (2013). *Politics*. Macmillan International Higher Education.

Horowitz, M. C. (2018). Artificial Intelligence, International Competition, and the Balance of Power. *Texas National Security Review*. Retrieved from https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/

Kissinger, H. A. (2014). World Order: Reflections on the Character of Nations and the Course of History. London (UK): Penguin.

Kissinger, H. A. (2018, May 15). How the Enlightenment Ends. Retrieved December 20, 2018, from https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/

Manne, G., & Maclean, J. (2017, March 1). Sorry, But Amazon Isn't Actually Annihilating Retail Jobs | WIRED. Retrieved December 20, 2018, from https://www.wired.com/2017/03/sorry-amazon-isnt-actually-annihilating-retail-jobs/

Ruggie, J. G. (1982). International regimes, transactions, and change: embedded liberalism in the postwar economic order. *International Organization*, *36*(2), 379–415. https://doi.org/10.1017/S0020818300018993

Russett, B., Layne, C., Spiro, D. E., & Doyle, M. W. (1995). The Democratic Peace. *International Security*, *19*(4), 164–184. https://doi.org/10.2307/2539124

Thompson, D. (2017, April 10). What in the World Is Causing the Retail Meltdown of 2017? Retrieved December 20, 2018, from https://www.theatlantic.com/business/archive/2017/04/retail-meltdown-of-2017/522384/

Thornhill, J. (2018, January 21). Davos 2018: why automation may be more evolution than revolution. Retrieved December 20, 2018, from https://www.ft.com/content/6136590a-e0ea-11e7-a0d4-0944c5f49e46

What is the Singularity? (2018, May 14). *The Economist*. Retrieved from https://www.economist.com/the-economist-explains/2018/05/14/what-is-the-singularity

Wright, N. D. (2018, December 7). AI & Global Governance: Three Distinct AI Challenges for the UN. Retrieved December 23, 2018, from https://cpr.unu.edu/ai-global-governance-three-distinct-ai-challenges-for-the-un.html

Chapter 3. AI and Domestic Political Regimes: Digital Authoritarian, Digital Hybrid and Digital Democracy

Abstract

This chapter examines AI and domestic political regimes in more detail, and introduces three crucial cases: China, Russia, and the US.

- A domestic political regime is a system of social organization that includes not only government and the institutions of the state, but also the structures and processes by which these interact with broader society. Three broad types of domestic political regimes dominate globally today: authoritarian (e.g. China), liberal democratic (e.g. the US), and hybrid regimes that fall somewhere in between (e.g. Russia).
- New variants of these regime types emerge in response to changing times. For instance, historically new forms of authoritarianism emerged in the 1920s (Fascism) and 1960s (bureaucratic authoritarianism).
- We arguably now see "digital" variants of each regime type emerging: digital authoritarianism (e.g. China), digital hybrid regimes (e.g. Russia) and digital liberal democracies (e.g. the US). However, the character of the new AI-related technologies (notably enhanced perception) best suits the augmentation of the surveillance, filtering and prediction in digital authoritarianism, making that perhaps the largest departure of the three.
- Finally we note that the geopolitical importance of the US, China and Russia means their "really-existing" domestic models will exert disproportionate influence – and thus their particularities matter, just as the Soviet Union's did in the Cold War.

What are Domestic Political Regimes and What Main Types Are There?[12]

A domestic political regime is a system of social organization that includes not only government and the institutions of the state, but also the structures and processes by which these interact with broader society. It is a "system"—or perhaps better a system of systems—in that there are interrelationships within a complex whole, and "political" in that these interrelationships relate to the distribution of power, wealth and resources in society.[13]

A myriad forms of political regimes have existed, but three main types now dominate globally:

(1) **Liberal democratic regimes:** There are many models of democracy, but they consolidate in certain ways. This is a form of democratic rule that balances the principle of limited government against the ideal of popular consent. Its "liberal" features are reflected in a network of internal and external checks on government designed to guarantee liberty and afford citizens protection against the state. Its "democratic" character is based on a system of regular and competitive elections, constructed on a basis of universal suffrage and political equality.

---

[12] I thank Oz Hassan for insightful discussions about his research and the contents of this chapter.

[13] This definition of a domestic political regime, as well as those of liberal democratic and authoritarian regimes, are chosen as typical textbook definitions, in this case from (Heywood, 2013).

(2) **Authoritarian regimes:** A belief in or practice of government "from above", in which authority is exercised regardless of popular consent. Authority rests on legitimacy. Authoritarian regimes emphasize the claims of authority over individual liberty. Authoritarianism and totalitarianism may be distinguished as authoritarianism lacks the more radical goal of obliterating the distinction between the state and civil society. Authoritarian regimes may thus tolerate a significant range of economic, religious and other freedoms.

(3) **Hybrid regimes:** These combine features of democracy and authoritarian systems. Many classifications have been proposed (e.g. Fig. 3.1), which often describe regimes as democracies with a prefix (e.g. illiberal-democracy) or authoritarian with a prefix (e.g. electoral-authoritarian). Here for simplicity we use the term "hybrid." Importantly, such hybrid systems are not just waystations on the way to democracy or full-blown authoritarianism but are often relatively stable regime types (Ottaway, 2013). They do often, however, constitute important "swing-states" in global competition between competing models.

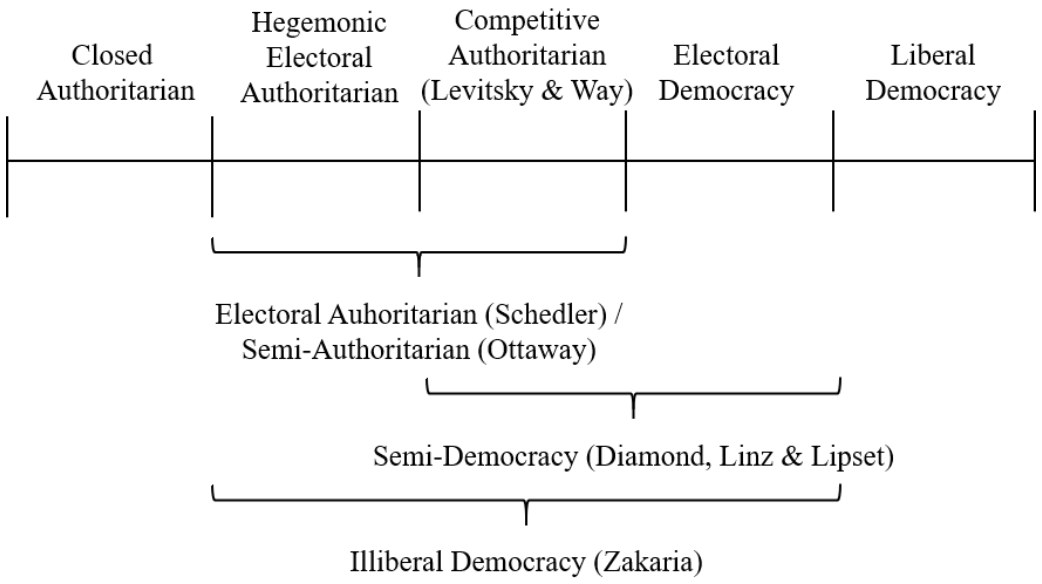


Fig. 3.1 Numerous classifications exist for different types of hybrid regimes. This is taken from (Gilbert & Mohseni, 2011). Here we simplify into authoritarian, hybrid and liberal democracy.

How Do Domestic Political Regimes Change Over Time?

New variants of these regime types emerge in response to changing times. Over the past two centuries, we have seen many changes in the prevalence of different regime types globally. One prominent summary of these changes describes three "waves" of democratization globally, and each wave is then followed by a "reverse wave" of increasing authoritarianism (Table 3.1) (Huntington, 1991). We now live in the third "reverse wave" of increasing authoritarianism (e.g. Fig. 3.2).

Each previous reverse wave of increasing authoritarianism was accompanied by a historically new form of authoritarianism: Fascism emerged in the 1920s; bureaucratic authoritarianism in the 1960s. And now with this reverse wave? Neither authoritarian nor hybrid regimes have been as effective as

liberal democracies at sustainably making the citizens of big, industrialized states rich. AI-related technologies provide a plausible, tangible reason for why things will be different this time. An AI-infused digital authoritarianism enables an approach that may seem appropriate to the needs of the times.

DEMOCRATIC WAVE	REVERSE WAVE
First wave (1820s-1926) Brought some 29 democracies into being.	**First reverse wave (1922-1942)** Began in 1922 with Mussolini gaining power in Italy. By 1942 number of democratic states was 12. Historically new form of authoritarian rule: Fascism was distinguished from earlier forms of authoritarianism by its mass base, ideology, party organization, and efforts to penetrate and control most of society.
Second wave (1945-1960s) World War II allied victory began the second wave. Zenith in 1962 with 36 countries governed democratically.	**Second reverse wave (1960-1975)** Brought number of democracies down to 30. Historically new form of authoritarian rule: Bureaucratic authoritarianism differed from earlier forms of military rule in Latin America with respect to its institutional character, its presumption of indefinite duration, and its economic policies.
Third wave (1974-2006) 1974-1990 at least 30 countries transitioned to democracy. The number of democracies essentially held steady or expanded every year from 1975 until 2007.[14]	**Third reverse wave (THE PRESENT)** The third wave stopped by around 2006. Freedom House argues that the past decade has seen a decline in democracies and a rise in not free countries (Fig. 3.2). New form: **digital authoritarianism.**

Table 3.1 Three waves and three reverse waves. Structure from (Huntington, 1991), with additional data on the third wave from the cited sources. Remarkably, Huntington's 1991 paper predicted that a potential third "reverse wave" may involve an "electronic dictatorship."

[14] E.g. (Diamond, 2015). For an argument that instead democracy held steady see e.g. (Levitsky & Way, 2015).

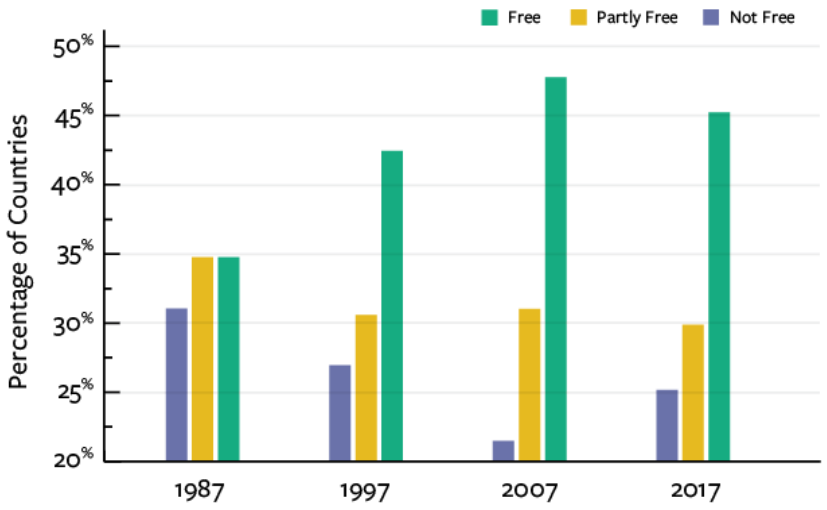


Figure 3.2 The third "reverse wave" of increasing authoritarianism. Figure from https://freedomhouse.org/report/freedom-world/freedom-world-2018 (accessed 20th Dec 2018).

"Digital" Variants of the Regime Types and the Importance of Path Dependence

We arguably now see "digital" variants of each regime type emerging: digital authoritarianism (e.g. China); digital hybrid regimes (e.g. Russia); and digital liberal democracies (e.g. the US). However, the character of the new AI-related technologies (in particular enhanced perception) best suits the augmentation of the surveillance, filtering and prediction in digital authoritarianism (and the authoritarian components of hybrid regimes), making that perhaps the largest departure of the three.

What do I mean by "digital"? I mean that the regime's modes of functioning are critically enabled by the affordances (i.e. possibilities for action) that the digital technologies provide. As Figure 1.2 shows, the digital technologies include computers, communications (e.g. the internet), big data and AI-related processing. Within the past decade, computers, the internet and social media have begun to truly change the sinews of political regimes. For instance, even without big data the communication enabled through social media has changed the ways traditional media and political actors interact. President Trump's election and social media use illustrate such trends.

The AI-related technologies are significant now not so much for what they are already actually doing at scale in the real world now, but because they clearly will bring about revolutionary capabilities at scale. For instance, it provides a clear rationale for building huge, structured databases, because unless the data they store can be analyzed it is not very useful. Indeed, this promise already shapes how the other digital technologies are employed to build databases. Moreover, AI can increasingly filter images and text for sophisticated censorship, so systems can be built now with large human components that can later be gradually reduced or redeployed. It promises to do all of this in a cost-effective way – assuming these societal scale IT projects all go to plan.

Most significantly, the awesome promise of the AI-enhanced digital technologies provides a story—a concrete, tangible narrative—for why this time things will be different for authoritarianism. Previous versions of authoritarianism palpably kept losing to the liberal democracies in the

competition to make the citizens of big, industrially sophisticated societies rich.[15] AI is a value multiplier of the already powerful digital technologies, which provides a crucial element in a story of the future. The importance of a plausible vision of the future cannot be understated: to mass organize a society or lead the core elements of a regime, one needs a story.

Lawrence Freedman's magisterial work on strategy across the military, socio-political and business realms illustrates the centrality of such a narrative element (Freedman, 2013). He defines strategy as the art of creating power, and describes how:

> *"As a practical matter strategy is best understood modestly, as moving to the next stage rather than to a definitive or permanent conclusion. The next stage is one that can be realistically reached from the current stage. … This does not mean it is easy to manage without a view of a desired end state. Without some sense of where the journey should be leading."*

The AI-related technologies driving forward digital systems help provide not only practical next steps—such as the building of colossal new labelled datasets for social governance—but also a vision of where the journey can lead.

Of course, many of the digital technologies are "dual use" so that key parts of digital infrastructure, like ubiquitous smartphones with AI, are being rolled out in authoritarian, hybrid and liberal democratic regimes alike. Thus, key differences between digital domestic political regimes rest in part on how the technologies are embedded and employed in the regimes. Crucial factors include:

- *Regulatory and legal frameworks governing the digital technologies:* Key areas of difference include the degree of privacy protection for individuals under normal circumstances, as well as how far mass surveillance is allowed under normal circumstances (discussed for Russia in Chapters 8 and 11). Regimes may also differ in the degree of integration of different types of data, particularly "ground truth" data such as criminal or medical records (Fig. 3.3).
- *Secret services and police services:* Domestic surveillance by security services for national security will be conducted in all regime types, for instance for counter-terrorism. Regime types may differ in multiple ways, such as: how far such surveillance extends to the broader population; if its use if highly limited to secret services or used by broader state security or police; or whether it is used for domestic political purposes by the leadership or regime.
- *Commercial sector:* Regimes may differ in how far they allow integration between private sector breadth of data, and public sector held or controlled "ground truth" data (Fig. 3.3).
- *Negative control of information (e.g. censorship) and positive control of information (e.g. propaganda):* All regimes likely limit access to some information (e.g. child pornography). However, the amount of censorship may differ, as may the mechanisms (e.g. Chinese technological versus Russian offline or distraction techniques described below). How far regimes use positive means to promote the regime or leadership may also be a crucial difference.
- *Infrastructure:* The physical digital infrastructure is crucial. Large IT projects are hard for any society. Big integrated databases are long-term investments that cannot be built overnight, and such infrastructure will likely differ between regimes. So too will infrastructure used to access

[15] Rich here refers to per capita income. For many decades no other system in any sizeable, industrially sophisticated society has provided a model capable of having rich citizens without an accommodation with liberal democracy. For instance, Singapore is a tiny city state with an extraordinary first leader in Lee Kuan Yew; and whilst China's rise has been remarkable, a potential path forward beyond middle income status hasn't been clear without further opening and liberalization.

digital information at scale, such as the Russian System for Operative Investigative Activities (SORM) system (see e.g. Chapters 8 and 11).

While one might consider regimes types in the abstract, it is also important to examine critical real-world cases. There will be constraints, such as institutional and industrial capabilities, as well as path dependencies in regime structures. In particular the geopolitical importance of the US, China and Russia means their "really-existing" domestic models will exert disproportionate influence – and thus their particularities matter, just as the Soviet Union's did in the Cold War.

Finally, it is important to note that none of these countries will develop in isolation of the others – interactions will likely be critical, although are beyond the scope of this chapter.

Digital Authoritarianism and the Chinese Case

Digital authoritarianism

One can ask: how may the AI-related technologies enable a domestic political regime by which a country can get rich and maintain control?

The new AI-related technologies promise to enable effective control of a society's humans at a bearable economic burden, via several, mutually reinforcing means.

Firstly, AI and big data promise free flow of information for economically creative and productive activities, while simultaneously curbing anti-regime discussions and activities. This is more *selective* censorship of specific topics, and *selective* targeting of specific behaviors. China's "Great Firewall" is an early demonstration of selective censorship (King, Pan, & Roberts, 2013).

Moreover, it enables *predictive* control of *potential* dissenters purely by extrapolating from an individual's data signature: making control more targeted and so cost-effectively reducing the economic burden of an authoritarian apparatus. Think Amazon or Google targeting but immensely turbocharged because its AI can train and draw on data in two crucial ways that should not be allowed in liberal democracies. One is the incredible breadth and volume of data on individuals collected across all the devices or platforms they carry and interact with in their environment. But more importantly such regimes will have no compunction about combining the huge breadth of data with "ground truth" data from tax returns, medical records, criminal records, police records, sexual health clinics, bank statements, genetic data, physical monitoring (e.g. location, biometrics, CCTV face monitoring), family and friends. This matters profoundly as such AI is as good as the data it trains on. Such quantity *and* quality of data on all individuals in society will, sadly, be excellent for training AI (Fig. 3.3).

Even the mere existence of AI's predictive control provides further remarkable advantages to the authoritarian. Self-censorship was perhaps East German Stasi's most important disciplinary mechanism (Wensierski, 2015). Individuals will know that the omnipresent monitoring of their physical and information activities may predict a propensity towards behavior undesired by the regime, even if they are just thinking about it. Computationally, this is no different to AI in healthcare finding patterns in data amongst the seemingly healthy to predict disease in its pre-symptomatic stages.

The two-way nature of humans' constant interaction with their phones and other monitoring devices also builds on a central finding from the cognitive science of influence: making people perform

behaviors can itself change their beliefs or attitudes (Maio & Haddock, 2009). A classic illustration is that making seatbelt use compulsory ended up changing attitudes towards seatbelt use. When omnipresent monitoring of your behaviors—even down to how long your eyes spend looking at different elements on a phone screen—could contribute to a prediction about you, then you cannot avoid *performing* the activities of a "responsible" member of society. Behavior builds belief.

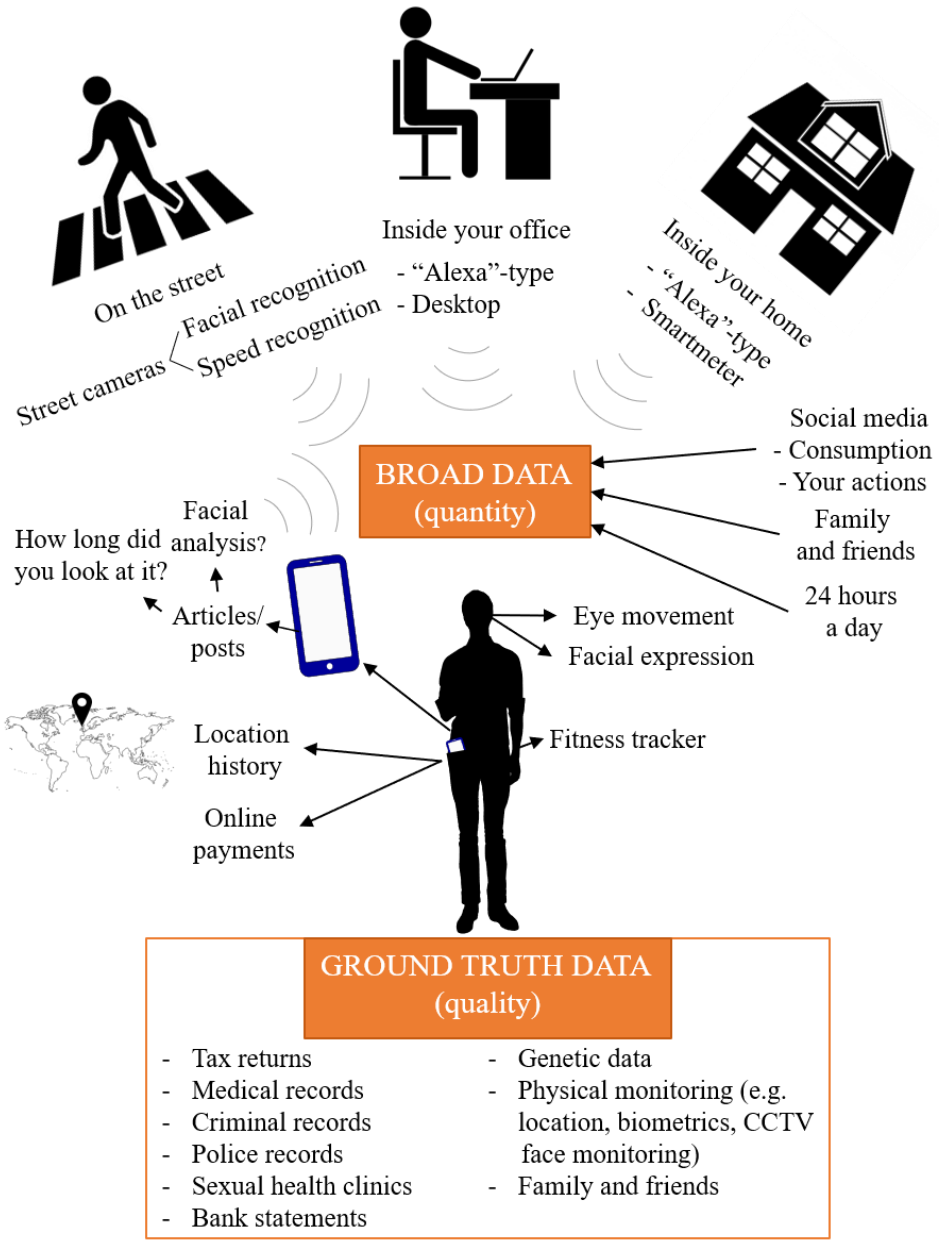


Figure 3.3 Two types of data are important. Firstly, the huge breadth of data from all our interactions with innumerable smart devices, which AI helps collect by, for instance, doing good face and voice recognition. Secondly, the high-quality ground truth data, which is a bit like good labelling of the big data. Together they form a powerful training set for AI.

Thus, AI promises to minimize the costs and enhance the effectiveness of censorship and behavioral control, so unlike in the USSR their costs may not prevent selectively, predictively controlled citizens

becoming rich. But failing central economic direction also hobbled the Soviet economy. Could AI and big data help there too?

Indeed, a further promise of AI is better central planning. As Jack Ma, the founder of Chinese tech titan Alibaba, argues, with enough information central planners can be better at economic central direction, planning and *predicting* market forces (Hornby, 2018). All Western countries marry some degree of central control, for instance in an industrial strategy, with market mechanisms. AI-enabled central planning could shift the balance, and would augment the market signals from the selectively censored information flowing up from the market.[16]

We in the liberal democracies may disagree that a new model of "selective predictive authoritarianism" will work in the long run, but it is a *plausible* model for the People's Republic of China (PRC) and others to aim for. Domestically a regime such as China's needn't aim for an eventual internal accommodation with liberalization, as many recently thought. Communism and Fascism were only defeated when they palpably failed in the real world after having been pretty thoroughly implemented.

Finally, such a system of authoritarian control also provides insurance for the regime. If a country like China eventually ends up stuck in a "middle income trap" anyway, they will by that point possess perhaps the most formidable system of social control ever created to control dissent.

The Chinese case

Next we can ask: could a digital authoritarian be built, for example in China?

Regardless of any potential merits from a regime's perspective, big IT projects are notoriously hard to pull off and this one would be truly mammoth. To consider its feasibility we can look at perhaps the most consequential non-Western country that might build such a system, China, and ask if it has the capability and intention to do so.

China has the capability. It can deliver huge IT projects that span society, such as the Great Firewall of China described by Harvard's Gary King (King, Pan, & Roberts, 2017). It has the funding. Last year China spent at least $196 billion on internal security, a rapidly increasing budget in which big-data platforms likely accounted for a large part of the increase (Chin, 2018). China has good AI expertise (Kania, 2017). Finally, widespread technologies such as smartphones can form the backbone of a personal monitoring system, and Chinese smartphone penetration is similar to Western Europe's (Poushter, 2016). Smartphones in China also provide a lot of data on their users – mobile payments in China are streets ahead of the U.S. or Japan (Wharton, 2018).

Indeed, China is already building core components of such a system. The Great Firewall of China is well established and sophisticated (Clark, 2018), and has been recently tightening ("China's great

[16] One might also take this a step further. An authoritarian state may not only better predict and shape market forces, but in some cases may also marry that with more direct tools of intervention less available to other regimes. These may include subsidies, diverse forms of coercion or espionage. China, for instance, has proven adept at using government actions like cyberespionage to benefit its corporations and maximize competitiveness. I thank Wyatt Hoffman for suggesting this point.

firewall is rising," 2018). Freedom House rates China the world's worst overall abuser of internet freedom ("Freedom on the Net 2017," 2017). China is implementing extensive social surveillance in the physical world (Hersey, 2017). The "citizen credit" scheme announced in 2014 intends to compute various metrics for every citizen's good conduct (Creemers, 2018; Mitchell & Diamond, 2018). The most complete surveillance state is being developed in the restive Xinjinag province with its large Muslim population (Shepherd, 2018), a capability that if desired may be rolled out across China.

But capability is not the same as intending to build a digital authoritarian system. Many of the components of such a system already in existence or under development may partly reflect the continuation of authoritarian practices (Hoffman, Ch 6 this volume). However, consensus opinion is that the China's trajectory is now more towards authoritarianism and away from accommodating greater social liberalization ("How the West got China wrong," 2018). Furthermore, while one is unlikely to ever find a complete blueprint, the Chinese Government clearly sees a big role for AI and big data in enabling this new direction. The 2017 AI Development Plan (Webster, Creemers, Triolo, & Kania, 2017) prominently states that "AI has become a new focus of international competition" and equally prominently goes on to describe how:

> *"AI brings new opportunities for social construction." with "widespread use of AI" in which "AI technologies can accurately sense, forecast, and provide early warning of major situations for infrastructure facilities and social security operations; grasp group cognition and psychological changes in a timely manner; and take the initiative in decision-making and reactions—which will significantly elevate the capability and level of social governance, playing an irreplaceable role in effectively maintaining social stability."*

On one level, whether the Chinese regime inadvertently creates a digital authoritarian regime, or whether its steps reflect an active plan, may not matter: the regime's ever-increasing dependence on its AI and big data systems builds a really-existing digital authoritarian regime.

The Chinese regime is discussed further in Chapters 5, 6 and 7.

Digital Hybrid Regimes and the Case of Russia

Russia does not have the same type of domestic political regime as China. Russia's hybrid model combines features of democracy and authoritarianism.[17] The regime features contested elections combined with numerous restrictions on democratic participation (Zimmerman, 2014). It has also evolved, from a less authoritarian hybrid in the Russia in the mid-1990s to one with more features of a one-party dictatorship under Putin since the turn of the millennium.

Similarly, Russia's approach to information manipulation and control differs significantly from the Chinese system. It emphasizes systemic technical censorship much less. Instead, the Russian model relies on a mix of less overt, and often non-technical, mechanisms to manipulate online information flows, narratives, and framings – so avoiding universal censorship. It also uses positive online means to shape public opinion.

[17] Numerous types of hybrid regime types may be identified, as Fig. 3.1 illustrates. Russia is examined here as it is a highly consequential case – and future work can provide further granularity by examining a range of hybrid regimes.

Finally, it is important to note the extensive domestic surveillance of citizen's online activities carried out by the regime. This centers on the "SORM" equipment installed at key internet locations in Russia, which monitor online activity. However, Russia does not have anything approximating to the Chinese tech giants and therefore its homegrown broader surveillance capabilities using AI-related technologies will inevitably much more constrained unless such systems can be obtained or adapted from abroad.

The Russian regime is discussed in more detail in Chapters 8, 11, and 21 in this volume.

Digital Liberal Democracies and the Case of the United States

How liberal democracies respond to AI's challenges and opportunities depends partly on how they deal with them *internally*, and partly on how they deal with the rise of the selective predictive authoritarian alternative *externally*. On balance, in both cases grounds exist for guarded optimism.

Looking internally, while established democracies like the US or UK require concerted efforts to manage the new technology, the challenges aren't obviously greater than those successfully overcome before. One big reason for guarded optimism is simply path dependence. Countries with strong traditions of individual liberty will likely go in one direction with the tech, while those in a different current condition will likely follow another path. Some tech experts like Jaron Lanier find little difference between the US and China on such tech issues (Kulwin, 2018). But considerable forces in US society push back and limit Government domestic mass surveillance programs, albeit with variable success, as seen with DARPA's efforts in the early 2000s (Harris, 2014) or the domestic programs highlighted by Edward Snowden (Hattem, 2016; Kerr, 2015). Most within liberal democracies acknowledge the need for espionage abroad and surveillance for counter-terrorism domestically, but powerful checks and balances constrain the state's domestic security apparatus. With continued vigilance this will likely continue with new AI technologies.

Second, while oligopolistic tech companies are concentrating power by gobbling up competitors and lobbying Governments, such a challenge has been ameliorated following other technological revolutions. Think of Teddy Roosevelt's trust busting or Microsoft's constraint during the internet's rise.

A third area relates to concerns over the media environment's health, where the tech titans threaten effective media plurality, vital public interest content or a Wild West attitude in political advertising. But such concerns have been tackled with previous radical new technologies (Barnett, 2010). Regulation on who owns "media", who is a "publisher" (Angwin, 2018) and so on will likely catch up with technology. Facebook's Mark Zuckerberg actively resisted labelling political advertising as is required on television, until forced to change last year (Lapowsky, 2017). In 2014 he changed Facebook's motto from "Move fast and break things" to "Move fast with stable infrastructure" (Levy, 2014) – as for his company so for society, where regulation is likely inevitable for systemically significant media companies.

Fourth, given the checks and balances outlined above, liberal democracies are unlikely to allow the commonplace, unfettered integration of their domestic populations' data to include two crucial sources: their "ground truth" data (e.g. from medical, tax or police records); or the breadth of data from across the multiple platforms individuals carry and interact with in their environment. Limiting the quality and quantity of data limits AI's power.

Moreover, more broadly neither most people nor other key stakeholders in established liberal

democracies are yet on the lookout for a new system of social organization. The opinion that democracy is "essential" may be declining in established liberal democracies (Breene, 2017). But this is a far cry from genuinely fragile democracies such as Brazil, where polls report the share who think Brazil needs "a strong leader who will break the rules" rose from 48% in November 2016 to 89% in March this year ("How a strike by lorry drivers will shape Brazil's elections," 2018).

How will liberal democracies respond to the external challenge of a new selective predictive authoritarian competitor – perhaps paradoxically it may strengthen liberal democracy. The human tendency to frame competition in "them and us" terms may lead the liberal democracies to, at least in part, define their attitudes to censorship and surveillance in opposition to this new competitor. Witnessing the selective predictive authoritarian state emerge may sharpen the pressure to prevent that happening "here", and highlight the paths to avoid. The nitty gritty of data policies is pretty boring to most people, and some hypothetical future harm seems pretty abstract arising from commonplace integration of "ground truth" data or data across multiple diverse platforms. But when this clearly underpins a dystopian selective predictive authoritarian regime in the real world it is neither boring nor abstract. Governments and the tech oligopoly in liberal democracies will have to explain how they—we—are different.

Of course, the response to the external threat may not work out that way. In future competition with adversaries' AI-fueled offensive attempts to spread disinformation, will we ever more rely on upon AI capabilities concentrated in very few powerful public or private "stewards" for our security (Brundage et al., 2018)? Perhaps. But as President Eisenhower forcefully argued, mitigating the dangers of a garrison state governed by a "military-industrial complex" was a key Cold War aim (Westad, 2017), and one in which the U.S. and its allies saw considerable success.

Conclusions

The AI-related technologies promise to be a massive force multiplier of the existing digital technologies – and this will shape the domestic political regimes of key countries across the globe. This provides a crucial part of a new authoritarian story about why this time it will be different; why this time authoritarianism can successfully make the citizens of big, industrially sophisticated countries rich. Observers in the liberal democracies may disagree on how likely that is to work out, but unfortunately it is at least plausible.

Such stories reach multiple audiences. While stories about domestic political regimes are typically mainly for domestic purposes, those regimes also compete in a global system. A country's domestic political regime, particularly when the country is as large as China, may serve as a model for others – and exert influence over the development of the others' own domestic political regimes. Much as the Soviet Union did. We next turn to such competition.

References

Angwin, J. (2018, April 5). Four Strategies for Fixing Facebook. Retrieved December 21, 2018, from https://www.theatlantic.com/technology/archive/2018/04/four-ways-to-fix-facebook/557255/

Barnett, S. (2010). What's wrong with media monopolies? A lesson from history and a new approach to media ownership policy. Retrieved December 21, 2018, from http://www.lse.ac.uk/media@lse/research/mediaWorkingPapers/pdf/EWP18.pdf

Breene, K. (2017, June 8). Millennials are rapidly losing interest in democracy. Retrieved December 21, 2018, from https://www.weforum.org/agenda/2017/06/millennials-are-rapidly-losing-interest-in-democracy/

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., … Filar, B. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *ArXiv Preprint ArXiv:1802.07228*.

Chin, J. (2018, March 7). China Spends More on Domestic Security as Xi's Powers Grow. *Wall Street Journal*. Retrieved from https://www.wsj.com/articles/china-spends-more-on-domestic-security-as-xis-powers-grow-1520358522

China's great firewall is rising. (2018, January 4). *The Economist*. Retrieved from https://www.economist.com/china/2018/01/04/chinas-great-firewall-is-rising

Clark, G. (2018, November 6). The Great Firewall of China. Retrieved from https://www.bloomberg.com/quicktake/great-firewall-of-china

Creemers, R. (2018). *China's Social Credit System: An Evolving Practice of Control* (SSRN Scholarly Paper No. ID 3175792). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=3175792

Diamond, L. (2015). Facing up to the democratic recession. *Journal of Democracy*, *26*(1), 141–155.

Freedman, S. L. (2013). *Strategy: A History*. Oxford ; New York: OUP USA.

Freedom on the Net 2017: Manipulating Social Media to Undermine Democracy. (2017, October 27). Retrieved December 21, 2018, from https://freedomhouse.org/report/freedom-net/freedom-net-2017

Gilbert, L., & Mohseni, P. (2011). Beyond Authoritarianism: The Conceptualization of Hybrid Regimes. *Studies in Comparative International Development*, *46*(3), 270. https://doi.org/10.1007/s12116-011-9088-x

Harris, S. (2014, July 29). The Social Laboratory. Retrieved December 21, 2018, from https://foreignpolicy.com/2014/07/29/the-social-laboratory/

Hattem, J. (2016, December 25). Spying after Snowden: What's changed and what hasn't [Text]. Retrieved December 21, 2018, from https://thehill.com/policy/technology/310457-spying-after-snowden-whats-changed-and-what-hasnt

Hersey, F. (2017, November 22). China to have 626 million surveillance cameras within 3 years · TechNode. Retrieved December 21, 2018, from https://technode.com/2017/11/22/china-to-have-626-million-surveillance-cameras-within-3-years/

Heywood, A. (2013). *Politics*. Macmillan International Higher Education.

Hornby, L. (2018, May 6). Living Marxism: the Chinese Communist party reasserts control. Retrieved December 20, 2018, from https://www.ft.com/content/766d2a42-419d-11e8-803a-295c97e6fd0b

How a strike by lorry drivers will shape Brazil's elections. (2018, June 9). *The Economist*. Retrieved from https://www.economist.com/the-americas/2018/06/09/how-a-strike-by-lorry-drivers-will-shape-brazils-elections

How the West got China wrong. (2018, March 1). *The Economist*. Retrieved from https://www.economist.com/leaders/2018/03/01/how-the-west-got-china-wrong

Huntington, S. P. (1991). Democracy's third wave. *Journal of Democracy*, *2*(2), 12–34.

Kania, E. B. (2017, December 5). Artificial Intelligence and Chinese Power. *Foreign Affairs*. Retrieved from https://www.foreignaffairs.com/articles/china/2017-12-05/artificial-intelligence-and-chinese-power

Kerr, O. (2015, April 9). Edward Snowden's impact. Retrieved December 21, 2018, from https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/04/09/edward-snowdens-impact/

King, G., Pan, J., & Roberts, M. E. (2013). How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review*, *107*(2), 326–343. https://doi.org/10.1017/S0003055413000014

King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *American Political Science Review*, *111*(3), 484–501. https://doi.org/10.1017/S0003055417000144

Kulwin, N. (2018, April 17). Jaron Lanier Q&A: 'We Won, and We Turned Into Assholes.' Retrieved December 21, 2018, from http://nymag.com/intelligencer/2018/04/jaron-lanier-interview-on-what-went-wrong-with-the-internet.html

Lapowsky, I. (2017, September 22). Why It Won't Be Easy for Facebook to Regulate Political Ads. *Wired*. Retrieved from https://www.wired.com/story/why-facebook-will-struggle-to-regulate-political-ads/

Levitsky, S., & Way, L. (2015). The myth of democratic recession. *Journal of Democracy*, *26*(1), 45–58.

Levy, S. (2014, April 30). Mark Zuckerberg on Facebook's Future, From Virtual Reality to Anonymity. *Wired*. Retrieved from https://www.wired.com/2014/04/zuckerberg-f8-interview/

Maio, G. R., & Haddock, G. (2009). *The Psychology of Attitudes and Attitude Change*. Los Angeles ; London: Sage Publications Ltd.

Mitchell, A., & Diamond, L. (2018, February 2). China's Surveillance State Should Scare Everyone. Retrieved from https://www.theatlantic.com/international/archive/2018/02/china-surveillance/552203/

Ottaway, M. (2013). Democracy Challenged: The Rise of Semi-Authoritarianism. Carnegie Endowment.

Poushter, J. (2016, February 22). 2. Smartphone ownership rates skyrocket in many emerging economies, but digital divide remains. Retrieved October 10, 2016, from http://www.pewglobal.org/2016/02/22/smartphone-ownership-rates-skyrocket-in-many-emerging-economies-but-digital-divide-remains/

Shepherd, C. (2018, February 27). "Big data" predictions spur detentions in China's Xinjiang: Human... *Reuters*. Retrieved from https://www.reuters.com/article/us-china-rights-xinjiang-idUSKCN1GB0D9

Webster, G., Creemers, R., Triolo, P., & Kania, E. (2017). Full Translation: China's "New Generation Artificial Intelligence Development Plan" (2017). Retrieved December 21, 2018, from https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

Wensierski, P. (2015, July 10). Web of Surveillance: East German Snitching Went Far Beyond the Stasi. *Spiegel Online*. Retrieved from http://www.spiegel.de/international/germany/east-german-domestic-surveillance-went-far-beyond-the-stasi-a-1042883.html

Westad, O. A. (2017). *The Cold War: A World History* (1 edition). New York: Basic Books.

Wharton. (2018, January). The Mobile Payments Race: Why China Is Leading the Pack - for Now. Retrieved December 20, 2018, from http://knowledge.wharton.upenn.edu/article/how-will-chinas-overseas-mobile-payment-systems-fare/

Zimmerman, W. (2014). *Ruling Russia: Authoritarianism from the Revolution to Putin*. Princeton: Princeton University Press.

Chapter 4. Global Competition

Abstract

This chapter discusses three ways in which the AI-related technologies may affect global competition.

- First is the export and emulation of these alternative models for influence over swing states—as occurred in the twentieth century between liberal democratic, fascist, and communist regime types. The global competition for influence occurs through active promotion; export of control and surveillance systems; competition between Chinese and U.S. tech titans; as well as battles over global norms and institutions. Swing states across Europe, Africa, Asia and so on are highly heterogenous, and even within states the elites and populations may disagree over the models' relative merits. Of course, the attractiveness or otherwise of the competing models is just one factor in the broader strategic context, as was the case between competing regime types in the twentieth century.
- Second, are AI's potential impacts on domestic political regimes may affect foreign policy decision-making
- Third, are impacts on the military dimensions of global competition.

(1) Global Competition: Export and Emulation of the Models

Global competition between alternative domestic political regime types means that their proponents compete for influence. Liberal democracy has been actively exported by the U.S. and others for decades—albeit patchily (Lagon, 2011)—and its soft power drove emulation from South Korea to South America. The liberal democracies have also promoted their views on individuals' digital freedoms. Now we will likely also see competition from export and emulation of the digital authoritarian and hybrid regimes. We discuss three aspects of this global competition.

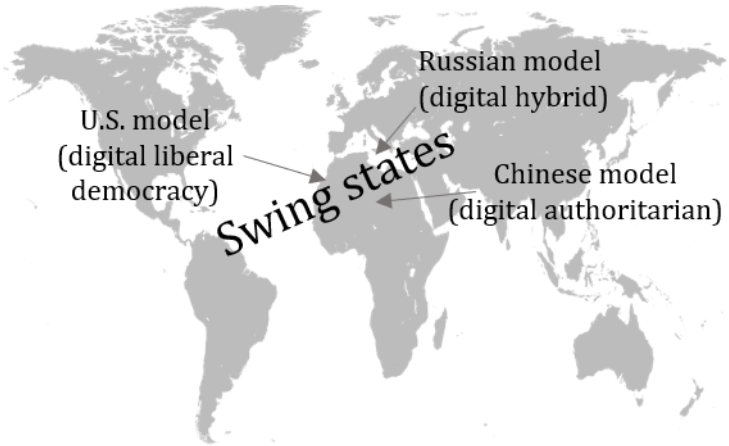


Figure 4.1 Heartlands and competition over the non-heartland swing states. *In one deliberately simplified conception, the US, China and Russia can be thought of as great powers competing for influence in the world system. The domestic political regime in each is different, and each provides a distinct model for other powers to look towards. Those other states can be described in many overlapping ways, for instance: as middle powers or small powers; they may be allies of the great powers; and they may be "balancers" or "bandwagoners." Particularly important are potential "swing states", which comprise much of the developing world, and which might plausibly tend towards or lend support to the different models.*

Supply and demand related to the digital authoritarian and hybrid models

We already see export of new surveillance and control systems. There is **supply**, for instance from China and Russia (Weber, 2017). China's Great Firewall approach has diffused to Vietnam and

Thailand. Chinese experts have reportedly provided relevant support in Sri Lanka and equipment in Iran (Stecklow, 2012), Zambia, Zimbabwe and Ethiopia – and even Russia. This year, Chinese AI firm Yitu reportedly supplied "wearable cameras with artificial intelligence powered facial-recognition technology" to Malaysian law enforcement; and prepared to bid for a Singapore government surveillance project that includes facial recognition in public spaces. This volume's Parts II, III, and V discuss and compare the Chinese and smaller Russian spread.

Crucially for the shape of this future competition in supply, only the US and China truly have tech giants. The US has the "FAANGs" (Facebook, Apple, Amazon, Netflix and Alphabet's Google) China has the tech titans Alibaba or Tencent, each worth many hundreds of billions of dollars, and many other key companies. Russia has no such tech giants. A country such as the UK may be home to Deepmind that built AlphaGo or ARM that leads the world in chip design – but neither is now UK-owned. The US and Chinese tech giants are now vying for influence across emerging markets and are increasingly going head to head in these swing states ("Chinese and US tech giants go at it in emerging markets," 2018).

Box 4.1 Domestic political regimes, models and ideologies

The *domestic political regime* in a state such as the US or China can provide a *model* that may influence the domestic political regimes in other states. There may be competition between such models. This may not be best described as a competition between *ideologies*—defined below—although it shares some features of such competition.

Consider the example of China. As scholar Thomas Christensen noted, the Chinese Communist Party (CCP) "has, by way of market reforms, all but obliterated the second of the two adjectives in its name … [so] nationalism is the sole ideological glue that holds the People's Republic together and keeps the CCP government in power" (Christensen, 1996). Instead, for example a "China model" did start to gain international attention, particularly following the 2008 international financial crisis, which was much more a model of statist development (Breslin, 2011). As China develops its digital authoritarian state as a domestic political regime, this may act as a model for others.

Definition of ideology for comparison: "A more or less coherent set of ideas that provides a basis for organized political action, whether that is intended to preserve, modify or overthrow the existing system of power relationships. All ideologies therefore (1) offer an account of the existing order, usually in the form of a "world-view", (2) provide a model of a desired future, a vision of the Good Society, and (3) outline how political change can and should be brought about." (Heywood, 2013)

And there is **demand**, from regimes that may want their countries to develop while maintaining control or who may just want effective mechanisms of control. Of course, within states the elites and populations may disagree over the competing models' relative merits. But importantly even population groups that may not want to import aspects of digital authoritarian systems, may not object to importing much of the "dual use" apparatus—such as smartphone or digital assistants—on which such systems will come to rely (Fig. 3.3).

This "dual use" may also affect demand compared to past authoritarian systems, as there will be much lower cost barriers to adoption of the new AI-related authoritarian control systems. Now the vast majority of the world's states are already witnessing huge uptake of digital technologies such as smartphones, which will also form crucial components of digital authoritarian monitoring (Poushter, 2016). This markedly differs from many previous versions of surveillance states. The twentieth century's Stasi or KGB systems required very large, sophisticated and expensive technical machinery

(Soldatov & Borogan, 2015). North Korea's surveillance state during the Cold War employed a vast and hugely expensive network of human eyes and ears (Lankov, 2013). Now, key systems—although to be sure not all—will already be in place.

Another factor that will shape demand for different models is the considerable heterogeneity between the swing states across Europe, Africa, Asia, and so on. Such heterogeneity may include factors that favor some models. Because of path dependence, many countries won't have the institutions of control or capabilities that China has – but for instance former Soviet republics may have a successor to the KGB that would be able to relatively easily adopt a version of the Russian digital hybrid model.

Global institutions and norms

Global institutions and norms also form a significant arena for competition. More broadly, China and Russia have pushed back against a, perhaps idealistic, conception of a free, borderless global internet. China uses its market power to influence technical standards, 'normalize' domestic control and shape norms of behavior through international organizations. Such states may conceive of these as strategically defensive measures necessary to ensure domestic control, but to observers they may seem offensive.

Context

The US, Chinese and Russian models' potential attractiveness will only be one factor in these states' global competition for influence – albeit potentially an important one if the twentieth century competition between regimes offering plausible, competing versions of the future is a guide. Other critical factors will include relative power, economic self-interest and historical grievances (e.g. Sino-Japanese antagonism) – as well as a good dose of luck.

However, while one must view such competition between alternative types of digital domestic political regimes against the broader backdrop of a rising China, resurgent Russia and enhanced "Grey Zone" competition between states – in many ways that context renders a plausible alternative to liberal democracy even more significant. Scholars of Chinese global influence such as David Shambaugh have long noted a gap in its social system's appeal as a competitor to liberal democracy (Shambaugh, 2013), which these new technologies may help fill.

Part III of this whitepaper examines the export and emulation of the models in global competition.

(2) Domestic Political Regimes and Foreign Policy Decision-Making

A second way that domestic political regimes affect global competition is by affecting states' foreign policy decision-making. States' decision-making is crucially affected by both internal *and* external factors, with arguments for the primacy of one or the other overstating the case (for discussions see e.g. Waltz, 2001; Zakaria, 1992).

Aspects of domestic regimes, such as bureaucratic or domestic political audiences, can profoundly affect foreign policy decision-making. This is seen in the case of democracies, for example where the specter of repeating a quagmire like "Vietnam" constrained various US administrations. Historical analysis of multiple episodes show the importance of public opinion in various different ways (Snyder & Borghard, 2011; Trachtenberg, 2012). The bureaucratic level also matters, shown for instance in classic studies of decision-making during the Cuban Missile crisis (Allison & Zelikow,

1999). Various interest groups also matter in authoritarian states, and may matter differently in different types of authoritarian states (Weeks, 2008). Russia under Vladimir Putin is not the same as Russia under the incredibly powerful mid-nineteenth century Tsar Nicholas II. Arguably, Chinese leader Xi Jinping has consolidated far more power than his immediate predecessor Hu Jintao.

Part IV of this whitepaper examines how the development of digital authoritarianism may affect Chinese foreign policy decision-making. For example, if digital authoritarian controls mean Chinese leaders have less to fear from their popular disquiet, they may be able to take more risks and back down (or ramp up) tensions in crises. Future work must extend this to examine Russia and the US.

(3) Global Military Dimensions

The AI-related technologies also affect the military dimensions of global competition. This may act on the longer term, such as through fears on all sides of spiraling AI arms races. It may act on escalation during crises (e.g. Herb Lin, Chapter 19). It may increase the importance of "hacking" more broadly within warfighting (e.g. Martin Libicki, Chapter 18). AI in information operations may play a key role in the "Gray Zone" conflict that has become such a feature of global competition since 2012 (Wright, 2017). The domestic security implications of AI discussed above may also directly feed into thinking about the use of force or other means externally. Domestic security thinking informing external operations were arguably seen with Russia's recent external use of information operations (Soldatov & Borogan, 2015); and are arguably seen historically in the links between the People's Republic of China's thinking on domestic security and its external use of military force (Scobell, 2003).

One critical challenge that may arise from the development of sophisticated digital authoritarian states relates to the profound asymmetry in vulnerability that will create between the US and China. Such asymmetries can be a cause of profound instability, as is seen now by the US asymmetric dependence on space making that a dangerously tempting target in Sino-US escalation scenarios (Wright, 2018). Recent events show the US political system's potential vulnerability to foreign digital interference – but if the Chinese regime builds the indefensibly vast digital systems of social governance that it plans, consider how vulnerable they may feel in 5-10 years' time if that were threatened with disruption. Regime security is often held to be the Chinese leadership's primary motivation, and attacks on that system may be perceived as threats to the regime. What would happen in a crisis 10 years hence if the then crucial social governance systems in a major city such as Chongqing were essentially turned off? Chinese domestic social governance systems that become ever more reliant on vast digital systems will be tempting targets for adversaries – a fact likely to prompt Chinese regime insecurity that may feed a spiraling security dilemma.

How these many such challenges are understood in China, Russia, and the US may also be a cause for misperception if they are understood differently. Thus, they must be thought through and discussed. The urgency of such discussions is illustrated by recent experience with more traditional "cyber" technologies: even basic concepts from the key technologies associated with cyberspace are still understood differently in these three key actors (Giles & Hagestad, 2013).

Part V of this whitepaper examines these military dimensions, including examinations of aspects of Chinese and Russian thinking.

Conclusions

The AI-related technologies may profoundly affect global competition via a number of mechanisms,

as discussed in Chapter 2. This whitepaper focuses on the key areas illustrated in Figure 4.2.

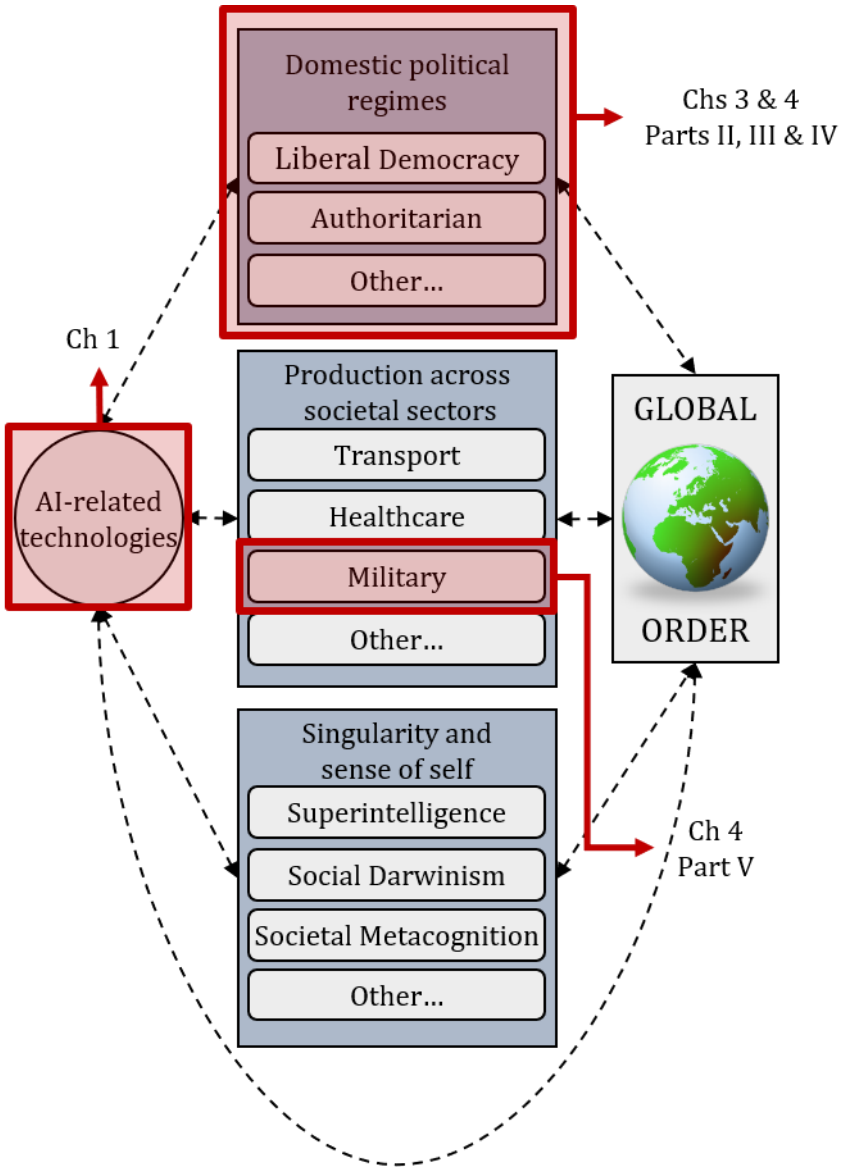


Figure 4.2 *AI's impacts on the global order. This report focusses on those areas highlighted in red.*

References

Allison, G., & Zelikow, P. (1999). Essence of Decision: Explaining the Cuban Missile Crisis (2nd ed.). Pearson.

Breslin, S. (2011). The 'China model'and the global crisis: from Friedrich List to a Chinese mode of governance? International Affairs, 87(6), 1323–1343.

Chinese and US tech giants go at it in emerging markets. (2018, July 7). The Economist. Retrieved from https://www.economist.com/business/2018/07/07/chinese-and-us-tech-giants-go-at-it-in-emerging-markets

Christensen, T. J. (1996). Chinese Realpolitik. Foreign Affairs, 75(5), 37–52. https://doi.org/10.2307/20047742

Giles, K., & Hagestad, W. (2013). Divided by a common language: Cyber definitions in Chinese, Russian and English. In Cyber Conflict (CyCon), 2013 5th International Conference on (pp. 1–17). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6568390

Heywood, A. (2013). Politics. Macmillan International Higher Education.

Lagon, M. (2011, January 11). The Whys and Hows of Promoting Democracy. Retrieved December 21, 2018, from https://www.cfr.org/expert-brief/whys-and-hows-promoting-democracy

Lankov, A. (2013). The Real North Korea: Life and Politics in the Failed Stalinist Utopia (1st ed. edition). Oxford: OUP USA.

Poushter, J. (2016, February 22). 2. Smartphone ownership rates skyrocket in many emerging economies, but digital divide remains. Retrieved October 10, 2016, from http://www.pewglobal.org/2016/02/22/smartphone-ownership-rates-skyrocket-in-many-emerging-economies-but-digital-divide-remains/

Scobell, A. (2003). China's Use of Military Force: Beyond the Great Wall and the Long March. Cambridge, UK: Cambridge University Press.

Shambaugh, D. (2013). China goes global: The partial power. Oxford University Press.

Snyder, J., & Borghard, E. D. (2011). The Cost of Empty Threats: A Penny, Not a Pound. American Political Science Review, 105(3), 437–456. https://doi.org/10.1017/S000305541100027X

Soldatov, A., & Borogan, I. (2015). The Red Web: The Struggle Between Russia's Digital Dictators and the New Online Revolutionaries. New York: PublicAffairs.

Stecklow, S. (2012, March 22). Special Report: Chinese firm helps Iran spy on citizens. Reuters. Retrieved from https://www.reuters.com/article/us-iran-telecoms-idUSBRE82L0B820120322

Trachtenberg, M. (2012). Audience Costs: An Historical Analysis. Security Studies, 21(1), 3–42. https://doi.org/10.1080/09636412.2012.650590

Waltz, K. N. (2001). Man, the State, and War: A Theoretical Analysis. Columbia University Press.

Weber, V. (2017, December 12). Why China's Internet Censorship Model Will Prevail Over Russia's. Retrieved December 21, 2018, from https://www.cfr.org/blog/why-chinas-internet-censorship-model-will-prevail-over-russias

Weeks, J. L. (2008). Autocratic Audience Costs: Regime Type and Signaling Resolve. International Organization, 62(1), 35–64. https://doi.org/10.1017/S0020818308080028

Wright, N. D. (2017). From control to influence: Cognition in the Grey Zone (p. 159). Birmingham, UK: University of Birmingham, UK. Retrieved from www.nicholasdwright.com/publications

Wright, N. D. (2018). Mindspace: Cognition in Space Operations. London (UK): Intelligent Biology.

Zakaria, F. (1992). Realism and Domestic Politics: A Review Essay. International Security, 17(1), 177–198. https://doi.org/10.2307/2539162

Part II: DIGITAL AUTHORITARIANISM: EVOLVING CHINESE AND RUSSIAN MODELS

Chapter 5. The Interests Behind China's AI Dream

Jeffrey Ding
University of Oxford
Jeffrey.ding@magd.ox.ac.uk

Abstract

Marked by the State Council's release of a national artificial intelligence development plan (AIDP) in July 2017, China's aggressive pursuit of AI has been regarded as both a wake-up call for China's increasing technological prowess as well as a precursor for concerning applications of AI in surveillance and military domains. Deciphering China's AI dream requires understanding *how* China is cultivating AI as a starting point to unlocking the broader implications of China's AI development. In this high-level overview of China's AI dream, I first place China's AI strategy in the context of its past science and technology plans, providing an analysis of the most important policies and initiatives China is currently engaging in to further its AI-related industries. Then, I outline how AI development intersects with multiple areas of China's national interests – including social governance. I conclude with a discussion of the main barriers to China realizing its AI dream.

Understanding the Context Behind China's AI Push

The key, guiding document of China's AI strategy in both the domestic and international realms is the State Council's July 2017 AI Development Plan (AIDP). The plan laid out key benchmarks for China's AI industry, sent a clear signal that AI was a national-strategic level priority, and emphasized priority areas where government policy and action could cultivate a favorable environment for sustainable, technical advances. The plan outlines a three-stage progression toward China's ambition of leading the world in AI:

1. By 2020, China's AI industry will be "**in line"** with the most advanced countries, with a core AI industry gross output exceeding RMB 150 billion (USD 22.5 billion) and AI-related industry gross output exceeding RMB 1 trillion (USD 150.8 billion).

2. By 2025, China aims to reach a "**world-leading"** level in some AI fields, with a core AI industry gross output exceeding RMB 400 billion (USD 60.3 billion) and AI-related industry gross output exceeding RMB 5 trillion (USD 754.0 billion).

3. By 2030, China seeks to become the world's "**primary"** AI innovation center, with a core AI industry gross output exceeding RMB 1 trillion (USD 150.8 billion) and AI-related gross output exceeding RMB 10 trillion (USD 1.5 trillion).

In a broad sense, these benchmarks map neatly onto three strategic phases of AI development: (1) catching up to the most advanced AI powers, (2) becoming one of the world leaders in AI, and (3) achieving primacy in AI innovation (Ding, 2018a).

In addition to these benchmarks the plan had at least three further aims. Firstly, the plan had a

massive signaling effect, prompting many local governments to publish their own AI plans and set up AI funds. Second, the plan prioritized key policy levers, especially the construction of technical standards — the Chinese word for standards (标准) appears 24 times in the AIDP, compared to the Chinese word for policy (政策) which appears 26 times — that could enable Chinese companies to become the world's leading AI backbone enterprises (Ding, Triolo, & Sacks, 2018). Lastly, the plan called for international cooperation and the establishment of more comprehensive AI regulations and ethical norms, though it did not present any concrete proposals in this area.

China's AIDP did not appear from thin air. Thus, it is important to consider the broader context behind China's policy support for AI development. In fact, Chinese government support for AI development, emphasis on indigenous innovation, and prioritization of frontier technologies traces back to February 2006, when the State Council issued a "National Medium- and Long-Term Plan (MLP) for the Development of Science and Technology (2006-2020)." The designation of "Artificial Intelligence 2.0" as a megaproject follows the framework set by the MLP, which provided significant research funds for frontier technologies. Alongside the MLP, the "Made in China 2025" initiative, issued in 2015, set explicit targets to strengthen indigenous innovation and advance China up the value chain in high-end manufacturing, both of which inform the background for AI policy. Other government policies, including the "'Internet Plus' and AI Three-Year Implementation Plan" and the Ministry of Industry and Information Technology (MIIT)'s own three-year action plan to implement the AIDP, reflect the key takeaway: China's AIDP does not exist in a vacuum but it interacts with other initiatives for strategic technologies and is also linked to a historical imperative to support those strategic technologies.

China's National Interests in AI

Appraising China's interests in AI requires an understanding of two key assumptions: (1) success in AI must be judged across various interests: economic competitiveness, military strength, and social stability; and (2) China is not a monolithic actor and different stakeholders (e.g. bureaucratic departments, military, tech giants) are pursuing their own notions of success in this field. I consider China's national interests in AI in three areas below: economic, military and social governance.

In the economic realm, China's potential gains from AI development are enormous. Per research by PwC in 2017, China had the most to gain from AI technologies, garnering a forecasted 26% boost in gross domestic product (GDP) from benefits attributable to AI advances (PwC, 2017). A McKinsey Global Institute report supports this view, estimating that 51% of work activities in China can be automated, which means that China has the largest labor force associated with such activities out of any country in the world (McKinsey Global Institute, 2017). The stakes for global economic preeminence are stark: industries under the AI umbrella have the potential to become the new "commanding heights" of the world economy, as reflected by "winner-take-all" and "first-mover" dynamics in Internet-based industries in the social network and e-commerce space. Finally, as China's population ages and it loses its demographic dividend, the integration of AI systems could improve overall productivity levels, enabling China to sustain economic growth and meet GDP targets.

In the military arena, some Chinese thinkers view AI as a revolutionary technology that could affect the balance of power. Lieutenant General Liu Guozhi, director of the Central Military Commission's Science and Technology Commission, stated in reference to military applications of AI, that the world is "on the eve of a new scientific and technological revolution," and "whoever doesn't disrupt will be disrupted!" (Kania, 2017). Still, since military applications of AI require a great deal of testing, there is as yet no consensus in Chinese strategic thinking about the degree to which AI will disrupt military

affairs. Thus, success for China in this area could range from on the one hand bolstering its abilities in order to asymmetrically counter adversaries in key zones such as the South China Sea; to on the other hand upending the current military balance of power by developing AI capabilities that function as a "trump card" in military competition.

China's National Interests in AI: Social Governance

Lastly, Chinese government officials view AI as a double-edged sword, with significant implications for social governance. On the one hand, Chinese officials are concerned that AI could accelerate the "digital divide" by placing a premium on high-skilled workers, thereby exacerbating existing divisions in Chinese society, including income and gender inequality, the urban rural divide, and the coastal/inland opportunity gap. In a wide-ranging special lecture to China's National People's Congress, Tan Tieniu, a Professor of Computer Vision and Pattern Recognition and Deputy Secretary-General of the Chinese Academy of Sciences, called for systematic study of the social impacts of AI, warning that "Water can keep the boat afloat but can also sink it (水能载舟，亦能覆舟)" – a phrase often used in the context of regime stability (i.e. the people can support a political regime but can also overturn it (Hickert & Ding, 2018).

On the other hand, China also seeks to employ AI technology to maintain social stability. In relatively clear terms, the State Council's AIDP states that AI will play an "irreplaceable" (不可替代) role in maintaining social stability. China aims to integrate AI across a broad range of public services, which includes judicial reviews, medical care, and public security. China's expansion of its public security apparatus has been most noticeable in Xinjiang, a situation which a UN human rights pane has described as "mass surveillance disproportionately targeting ethnic Uighurs" (Nebahay, 2018). Translations of readouts from the annual China-Eurasia Security Expo in Xinjiang show that some of China's leading AI startups, including facial recognition upstarts Sensetime and Megvii (Face++), are partnering closely with local companies and public security bureaus to boost security and surveillance (Ding, 2018b).

However, it is also important to note that the expansion of surveillance in Xinjiang is part of a broader, nationwide effort to build "safe" and "smart" cities. With this broader lens in mind, other AI applications besides facial recognition play an essential role, including AI-enabled censorship through better identification of patterns and predictive policing measures.

Obstacles to China's AI Dream

As China chases down its AI dream, different interest groups and departments are also staking out their own claims to that dream. Different bureaucratic departments, in particular the Ministry of Science and Technology and the MIIT, are fighting over claims to guiding AI development as they hope to get the political credit for advancing this strategic technology (Council on Foreign Relations, 2017). Divergent interests and multiple stakeholders have resulted in the formation of "data islands," which negatively affects the degree to which data can be integrated, a key bottleneck to AI development.

Among Chinese thinkers and international analysts alike, there seems to be a flawed assumption that some form of "industrial policy" targeted toward AI is better than doing nothing at all — the historical record, however, is not all that clear. When it comes to strategic technology, China's over-hyped government plans usually under-deliver. For instance, the Chinese government has only invested $12 billion of an announced $150 billion semiconductor fund since the fund's establishment in 2014

(Zwetsloot, Toner, & Ding, 2018). For context, Samsung spent nearly $27 billion in capital expenditures for its semiconductor group in the last year alone. Finally, even if money does get spent, it does not always have beneficial effects; historically, Chinese S&T megaprojects have often diverted funding from high-quality labs toward more politically-connected entities.

Thus, for China to realize its AI dream, balancing these competing interests from different levels of government, industry, and academia, while avoiding the pitfalls of industrial policy will be an essential endeavor.

References

Council on Foreign Relations (2017, August). Beijing's AI Strategy: Old-School Central Planning with a Futuristic Twist. Retrieved from https://www.cfr.org/blog/beijings-ai-strategy-old-school-central-planning-futuristic-twist

Ding, J. (2018, March). Deciphering China's AI Dream. *Future of Humanity Institute Technical Report*. Retrieved from https://www.fhi.ox.ac.uk/deciphering-chinas-ai-dream/

Ding, J. (2018, September). "ChinAI Newsletter #29: Complicit - China's AI Unicorns and the Securitization of Xinjiang." ChinAI Newsletter. https://chinai.substack.com/p/chinai-newsletter-29-complicit-chinas-ai-unicorns-and-the-securitization-of-xinjiang

Ding, J., Triolo, P., & Sacks, S (2018, June 20). Chinese Interests Take a Big Seat at the Ai Governance Table. *New America*. Retrieved from https://www.newamerica.org/cybersecurity-initiative/digichina/blog/chinese-interests-take-big-seat-ai-governance-table/

Hickert, C. & Ding, J. (2018, November 29). The Innovative Development and Social Impact of Artificial Intelligence. *New America*. Retrieved from https://www.newamerica.org/cybersecurity-initiative/digichina/blog/read-what-top-chinese-officials-are-hearing-about-ai-competition-and-policy/

Kania, E (2017, June 8). 数字化 – 网络化 – 智能化: China's Quest for an AI Revolution in Warfare. *The Strategy Bridge.* Retrieved from https://thestrategybridge.org/the-bridge/2017/6/8/-chinas-quest-for-an-ai-revolution-in-warfare#_edn23

McKinsey Global Institute (June, 2017). Artificial Intelligence, The Next Digital Frontier? Retrieved from https://www.mckinsey.com/mgi/overview/2017-in-review/whats-next-in-digital-and-ai/artificial-intelligence-the-next-digital-frontier

Nebahay, S. (2018, August 30). U.N. calls on China to free Uighurs from alleged re-education camps. *Reuters.* Retrieved from https://www.reuters.com/article/us-china-rights-un/u-n-calls-on-china-to-free-uighurs-from-alleged-re-education-camps-idUSKCN1LF1D6

PwC (2017). Sizing the Prize. Retrieved from PwC. "Sizing the Prize." 2017. https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-theprize-report.pdf

State Council (2015, May 19). Zhongguo Zhizao 2025 [Made in China 2025]. Retrieved from http://www.gov.cn/zhengce/content/2015-05/19/content_9784.htm

State Council (2006, February 9). Guojia Zhongchangqi Kexue he Jishu Fazhan Guihua [National Medium- and Long-Term Plan for the Development of Science and Technology]. Retrieved from http://www.gov.cn/jrzg/2006-02/09/content_183787.htm

Zwetsloot, R., Toner, H., & Ding, J (2018, November 16). Beyond the AI Arms Race: America, China, and the Dangers of Zero-Sum Thinking. *Foreign Affairs.* Retrieved from https://www.foreignaffairs.com/reviews/review-essay/2018-11-16/beyond-ai-arms-race

Chapter 6. Managing the State: Social Credit, Surveillance and the CCP's Plan for China

Samantha Hoffman
Visiting Academic Fellow, Mercator Institute for China Studies (MERICS)
contact@samanthahoffman.net

On July 20th 2017, the Chinese government released its *Next Generation Artificial Intelligence Development Plan* (see Webster, Triolo, Kania,& Creemers (2017) for a translation of the plan).[18] The plan has gained significant media attention in part because it links AI with another topic that has drawn a considerable amount of attention, China's "social credit system" (社会信用体系). Social credit uses big-data collection and analysis, to monitor, shape and rate individual's behavior. While advances in artificial intelligence (AI), and the growth of the surveillance state are all noteworthy on their own, China's social credit program explicitly links them as parts of a broader political control process known as "social management" (社会管理).

The phrases "social management", and the more recent version "social governance", may seem like pseudo-scientific jargon, but in fact, are given clear importance by China's top leaders.[19] In 2016, General Party Secretary Xi Jinping highlighted the concept, noting: "people working in political and legal affairs and comprehensive social governance have focused on dealing with outstanding problems and innovating *social governance methods* in recent years, achieving greater results," (Xinhua, October 12, 2016). Elsewhere, the Party has clearly explained that it sees operationalizing social management as its blueprint for maintaining power. Far from being a narrow, isolated political concept, "social management" gives cohesion to an array of concepts ranging from Hu Jintao's signature "Scientific Development" to Xi's push for military-civil integration, as part of this power maintenance process (People.com.cn, April 17).

Social Management

Social managements' roots are in the core ideology of the Chinese Communist Party (CCP). The CCP defines itself as the "vanguard of the people"—the Leninist idea that a small group of scientifically guided and educated cadres can lead the people in the direction of social equality and prosperity. Mao Zedong's organizational guide, the "mass line" describes the same concept. The CCP leadership is explicitly at the top of this hierarchical mass line system. It takes the "scattered and unsystematic ideas of the masses" and forms them into "concentrated and systematic ideas" before taking them back to the masses to "propagate and explain these ideas until the masses embrace them as their own"—Meaning management along scientific principles (Heath, 2013, p. 3-6; Mao, 1943).

Social management describes a "scientific" Leninist machinery for shaping, managing, and responding. It is best summarized as a complex systems management process through which the

[18] This chapter was first published in China Brief Volume: 17 Issue: 11, https://jamestown.org/program/managing-the-state-social-credit-surveillance-and-the-ccps-plan-for-china/

[19] Social management (社会管理) and social governance (社会治理) are two phrases that, in practice, have the same definition and are implementing exactly the same process, but the shift from "social management" to "social governance" under Xi Jinping has more to do with political power and ensuring the effectiveness of the social management process, than actual conceptual change (Hoffman, 2012, p 5-8; Hoffman & Mattis, 2013).

Party leadership attempts to manage the Party itself, and through which Party leadership attempts to manage the Party's interactions with society as a whole. Social management is aimed at ensuring China's "holistic" or "comprehensive" state security (国家安全). This holistic state security concept is not fundamentally new under Xi Jinping. It includes the western "national security" concept, but, more significantly, is focused on two internal security dimensions. First, managing the Party itself, and second is managing social order (Xinhua, April 15, 2014; *Qiushi*, April 15, 2017; *PLA Daily* [archive], December 13, 2000).

Social management itself is not a new concept and dates to the PRC's founding in 1949, when it was first integrated into the CCP's discussion of law and social order. The concept became increasingly prominent in the Party leadership's rhetoric between the late 1990s and early 2000s. When the 12th Five-Year Plan for National and Social Development was released in March 2011, social management was enshrined as a key objective (China.com.cn, March 16, 2011). In the plan, the Party set targets for speeding up the construction of a social management system that combined governance measures to address problems at their source, dynamic management, and emergency response—while adhering to the core leadership of the Party.

In its ideal form, the social management process optimises interactions vertically (within the Party), horizontally (between agencies), and holistically, between the Party and society. At every linkage, the goal is to improve governance capacity to shape, manage, and respond to social demands. Social management must efficiently solve problems to succeed. Such problems include: allocation of public resources, preventing and controlling risks associated with man-made and natural disasters, stopping dissent, and pre-empting and managing social conflict. The process involves both coercive and co-optative tactics, constantly acting together, to force individuals and to incentivize individuals to participate in social management.

For the social management process to succeed, particularly when in a crisis response mode, an automation of the interactions between the state and society, as well as the interactions within the Party itself, is required. The modern-day "grid management" (网格化管理) and the "social credit system" (社会信用体系) are unique compared to previous versions of similar social control mechanisms because they employ modern technology. They represent the attempted automation of social management.

The concept of automating social management is not new under Xi Jinping (2012-Present) or his predecessor Hu Jintao (2002-2012). In fact, the concept emerged in the late 1970s when "social management" was directly connected to complex systems theories the Party-State's theorists were drawing from to design a *Leninist* governance system to recover power after Mao and the Cultural Revolution. The basic ideas originated around 1957 when Qian Xuesen ("father of Chinese rocketry") called on the Chinese Academy of Social Sciences to take the concept seriously as a way of solving social problems [People's Daily [archive], 28 May 1957]. By the 1970s and throughout the 1980s complex systems thinking [largely via Qian Xuesen's promotion of engineering cybernetics] was clearly tied to "social management". For example, one report from 1984, "On the New Technological Revolution" (新技术革命) said:

> Leaps and bounds in science and technology [since the 1940s have] influenced or given rise to transformations in the way social management agencies work. The theory and practice, perspective and method of systems engineering were born and developed from these changes." It elaborated that it is impossible to manage effectively through individuals or a small number of people, and, "only if we fully

> grasp [the concepts of] information, data, systems analysis, and decision modeling, can we truly possess 'foresight and sagacity', and generate the courage and a bold vision consistent with the flow of history.[20]

The report further laid out what steps were needed to implement systems engineering in the "social domain". It included, among other things, defining what targets systems management should reach, establishing facilities to ensure information flow, and planning and developing methods and procedures for systems analysis. This explains why systems thinking is key to understanding not only how social credit fits into social management, but overall how the social management system is being designed.

System Construction

Rather than being relatively new conceptions, modern surveillance techniques and social credit are merely the newest developments in realizing the automated social management objective. Advances in AI and big data management further improve their function, from a technical perspective. These advances describe what the Party refers to as social management "innovation".

The first major step in the technological development of social management's automation was the implementation of grid management (网格化管理). Structurally, it advanced what has been described as a multilateral "vertical and horizontal integration" (纵横结合) of resources, people and agencies involved in social management. The political-legal and public security apparatus, including neighborhood and street-level committees, largely responsible for the technical side of its day-to-day implementation. Grid management enabled the organization of data to generate better situational awareness and predictive capacity, as well as enhanced tracking and monitoring of individuals (*People's Daily*, October 15, 2006).

The first modern grid(-ized) (网格化) policing was implemented between 2001 and 2002 in cities like Shanghai (*People's Daily* [archive], August 3, 2001; *China File*, August 10, 2016). It organizes information gathering by dividing an urban space into grids, each of these grid spaces is assigned grid managers who help to collect data and pre-empt and solve problems within their grid. The modern informatized grid enables faster emergency response and improved prevention and control. The photographs and videos police take at the scene of almost every protest are one example of the kind of data fed into the grid system.

Grid management's application to social management was significantly expanded between 2002–2012 under the direction of Zhou Yongkang, first as minister of Public Security and later as head of the Central Political-Legal Affairs Committee. Advances in integrated e-government resources in the internal security apparatus, namely the Golden Shield Project, greatly enabled grid management. The Golden Shield Project is not an internet monitoring project updating the Great Firewall. Rather it is an e-government project creating an organizational network connecting the Ministry of Public Security with its local-level bureaus, which was already being employed at provincial and city levels by 2002 (*China Brief*, June 3, 2011; *People's Daily* [Archive], April 26, 2002). The "Shield" was part of an expanded series of systems engineering projects, originally initiated in 1993 and later expanded

[20] The People's Daily [archive], September 13, 1984

as "Golden Projects". Each of the Golden Projects were e-government projects designed to build and streamline information systems, and connect agencies to improve their operational capacity.

This eventually included the multi-phase Golden Shield project, which was being implemented under the guidance of the State Informatization Leading Small Group by the late 1990s and early 2000s (Zhou, Hongyuan, Yuxian, Changsheng, & Xinhong, 2003). For public security bureaus, it improved both efficiency and surveillance. Software applications were developed to integrate data by requiring "real name" registration for travel booking, telecom services, and other services, information from hotel check-in and at customs clearance could be linked to law enforcement databases. The major contribution of the Golden Shield Project to the overall social management program was that it created a capacity to automate information sharing. Ostensibly, the Golden Projects were the technological starting point for building the social credit system, and perhaps social credit was an end goal much earlier in the process. E-government in China has always been designed to improve governance capacity and operate as a feedback loop with social management functions. The timing of social credit implementation probably is explained more by improved technical capacity than by changing policy objectives.

Automated Social Management?

The social credit system relies on the technology enabling and the organizational capacity created through the grid management system. Effectively "social credit" is the technological marriage of individual "responsibility" mechanisms and social control methodologies. Responsibility is a concept underlying the social management process, and it implies that the entire Party and all of society are responsible for upholding the Communist Party's leadership. This is also why individual responsibility is a key theme of all major state security-relevant legislation passed under Xi Jinping (IISS Voices, May 26; *The National Interest*, May 17, 2016). Enabled through the same resources and technology found in grid management, social credit creates a simultaneous co-optative and coercive responsibility systems function, and when fully implemented comprehensively covers all of society. Society is co-opted to participate because the same technology is directly linked to conveniences that improve everyday life, for instance electronic payment. Society is coerced to participate, for instance by self-censoring online, because increasingly technology systems are improving the government's capacity to enforce "responsibility" to the party-state. Not participating could have consequences not only for the individual but also their personal networks. These functions will only become further advanced through plans such as "Internet Plus", as the same technology applications used to provide social and commercial services feed directly into government information gathering and sharing processes (Gov.cn, February 1st 2017).

In the construction of the social credit system, current research and development is largely focused on areas such as big data analysis and integration to support the collection of information and ensure its effective use for intelligence. This is one of many areas where advances in AI would help streamline social management processes and, perhaps ideally, even automate them. Two major problems, however, confront this automated version of social management.

The first is the struggle for power within the Party. The Party members in charge of day-to-day implementation of social management are also responsible to the Party. As the systems were being enabled in the early 2000s, these agencies had a large amount of relatively unregulated power. The age-old problem of an authoritarian system is that security services require substantial power in order to secure the leadership's authority. The same resources enabling management of the Party-society relationship can be abused by Party members and used against others within the Party (*War*

on the Rocks, July 18, 2016). This appears to be the case with Zhou Yongkang, Bo Xilai, and others ahead of the 18th Party Congress. The problem will not disappear in a Leninist system, which not subject to external checks and balances. And it is why ensuring loyalty is a major part of the management of the party side of "state security."

The second probably is a symptom of the first: disaggregated security agencies. In an ideal form, agencies tasked with different aspects of social management can cooperate to address state security problems that have "integrated" characteristics. Usually such threats involve the 'three evil forces' of splittism, terrorism and extremism, and often specifically are related to Tibet, Xinjiang and Falun Gong. Because these are described as threats that have domestic and international connectivity, cooperation between domestic departments, intelligence, and foreign affairs is required for operational success (UIR Center for International Strategy and Security Studies, 2014, 133). It is particularly applicable in massive multi-agency operations such as "Operation Skynet", tracking down fugitives from the Party-state (Huang, 2015).

Both problems are explanations for structural changes that put Xi Jinping in charge of leading groups on State Security, Cyber Security, and so on. Using the example of the Central State Security Commission, there are now local government-level iterations in the form of state security work small leading groups in nearly every province, as well as the counties and cities within them. All are led by the relevant party secretary of the locality, and, where data is available, their membership appears to include (but is not limited to) the heads of Political-Legal Affairs Committees, Ministry of State Security bureaus, Armed Police, and Propaganda departments. Similar committees have been set up to mirror other new central leading groups. The membership overlaps significantly. Such leading groups are not new, but the evidence points to the system being utilized not only to re-center power away from the Central Political-Legal Affairs Committee and local versions, but also to develop a more effective system for mobilizing the social management process. For as much as the changes may be geared toward re-centering internal security power, the changes probably serve a dual purpose of creating a capacity for departments to function like a holistic "system of systems". It would address problems by issue—rather than as separate systems addressing overlapping problems.

Conclusion

Chinese information technology research and development, including the priorities outlined in the 2017 AI development plan, are interesting on their own because they mark advances in important research areas. But, as the language of the AI development plan indicates, these advances cannot be separated from Beijing's social management and state security policy. Applied to the social management process, they are aimed at improving governance capacity—automating the "carrot" and "stick" processes that ensure the Party-state's power. Senior CCP leadership hopes that through automation the Party will be able to more effectively anticipate and react to emerging problems, pre-empting crises before they become serious threats to stability.

References

Heath, T. (2013). Xi's mass line campaign: realigning party politics to new realities. *China Brief, 13*(16).

Hoffman, S. (2012). Portents of change in China's social management. *China Brief,* 12(15), pp. 5-8;

Hoffman, S., & Mattis, P. (2013, November 21).China's proposed "State Security Council": Social governance under Xi Jinping. *China Policy Institute: Analysis.*

Huang, K. L. (2015, March 26). China ramps up global manhunt for corrupt officials with opertion 'Skynet." *South China Morning Post.* Retrieved from https://www.scmp.com/news/article/1748113/china-ramps-global-manhunt-fugitive-corrupt-officials-skynet

Mao, Z. (1943). Some questions concerning methods of leadership. Marxists.org, Retrieved from https://www.marxists.org/reference/archive/mao/selected-works/volume-3/mswv3_13.htm

UIR Center for International Strategy and Security Studies (中国关系学院国际战略与安全研究中心) (2014). *Annual Report on China's National Security Studies 2014* (中国国家安全研究报告 2014), Blue Book of National Security (Beijing: Social Sciences Academic Press (China).

Webster, G., Triolo, P., Kania, E., & Creemers, R. (2017). A next generation artificial intelligence development plan. China Copyright and Media blog. Retrieved from https://chinacopyrightandmedia.wordpress.com/2017/07/20/a-next-generation-artificial-intelligence-development-plan/

Zhou, H., Hongyuan X., Yuxian Z., Changsheng W., & Xinhong Z.. (2003). China E-Government Development Report No. 1 (中国电子政务发展报告) *Blue Book of Electronic Development* (电子政务蓝皮书)

Chapter 7. Credit Cities and the Limits of the Social Credit System

Shazeda Ahmed
University of California, Berkley
shazeda@ischool.berkeley.edu

Abstract

The "credit city" is one in which local governments and tech companies share their data with one another to determine the degree of individuals' and businesses' trustworthiness. Exploring the "credit cities" concept, which tech companies and the government are both using to pilot aspects of the "social credit system", enables this chapter to shed light on early efforts at state-firm collaborations to construct the social credit system's technological infrastructure. I examine two examples of credit cities—Suzhou, which has partnered with Ant Finanacial, and Fuzhou, which works with JD Finance—along with the notable central-government led effort the "Xinyi+" project. A review of the Mandarin literature suggests that China's major tech companies are collaborating with the state on more bounded, localized projects that to date make little to no use of AI. Policymakers and academics have also identified critical challenges in this process, which indicate concerns about low data quality and siloed databases that must be improved before the system can progress.

Introduction

The opening of almost any news article written in English about China's social credit system presents a world in which digital sensors and cameras are everywhere, recording and judging people's every action—a scene which is far from the truth, while still raising the question of what the Chinese government realistically hopes to achieve with the host of data they collect from their citizens. In *Divining a Digital Future*, Genevieve Bell and Paul Dourish (2011) capture the now decades-old sense of mythmaking that initially drove future visions of "ubiquitous computing," in which computers and sensors would be embedded in all imaginable settings of everyday life. Their efforts to examine the "ubiquitous computing of the present" serves as a reminder that even though many of the initial ideals of ubiquitous computing have more or less been attained, the idea continues to be repackaged anew to keep pace with major technological and cultural developments. The perpetually just-out-of-reach ideal of ubiquitous computing in smart cities, for example, is being refashioned in China in ways that differ from such projects in the United States. A recent spin on the smart city concept, that of the "credit city" (信用城市), has arisen from information technology companies' efforts to aid the Chinese government in building a social credit system.

The Social Credit System

The social credit system is a nationwide effort to give pre-existing Chinese laws teeth through a mix of blacklists, intragovernmental and public-private data sharing, and rewards for so-called trustworthy (守信) behavior (Daum 2017) – which is a far cry from what Vice President Pence referred to as "an Orwellian system premised on controlling virtually every facet of human life" (Pence 2018). Scholars of Chinese law have debunked some foreign media coverage that misreported on the social credit system's current state and reliance on technology (Horsley 2018). Core beliefs many observers outside of China hold about the system—that it feeds into a single numerial score, and that facial recognition-enabled cameras and other digital sensors are constantly updating a central government database that calculates these scores—are mistaken. A more balanced view is

required, informed by evidence from on the ground.

The design and implementation of "credit cities" provides one crucial window into the reality on the ground in China, on which I focus in my field research. Understanding "credit cities" provides an entry point into understanding how China's government (at the national and municipal levels) works with tech companies to assess the trustworthiness of individuals and companies, as well as to publicize these judgments. Credit cities form one experimental facet of the much broader goal of establishing a social credit system, reveal an understanding of what the social credit system is anticipated to achieve, and help identify what the roadblocks are to fulfilling these expectations.

What Exactly is a "Credit City," and What is its Purpose?

The credit city is one in which local governments and tech companies share their data with one another to determine the degree of individuals' and businesses' trustworthiness. Such a characerisation arises from the speeches and reports from the two annual Credit Cities Construction Summit (中国城市信用建设高峰论坛) meetings that have been held since 2017, hosted by the National Development and Reform Commission (NDRC) in partnership with city governments and tech companies. Roughly four dozen cities were represented at the most recent Summit. The value judgments that come out of assessing a mix of public and private sector data—in some instances, a numeric score or a verbal rating—becomes a basis for determining the benefits that a person or company can unlock in a credit city. Benefits for individuals include deposit-free rentals of hotel rooms, apartments, offices, and bicycles, for instance. These same judgments can be used to restrict individuals and enterprises from taking certain actions, although thus far there have been few examples where additional punishments are meted out to those who already find themselves on blacklists.

NDRC deputy director Lian Weiliang, a notable leader at the Credit Cities Summit, has argued that "In the construction of the social credit system, cities are without a doubt in an important position to be first movers in experimentation, and local governments are without a doubt the best practitioners to lead urban credit construction" (Lian 2017).

Many government representatives at these and similar conferences lament an "information islands" (信息孤岛) phenomenon in which government departments have failed to share "public credit information" (公共信用信息) with one another, and furthermore lack access to sufficient "market credit information" (市场信用信息). Examples of the former include whether or not someone is on a blacklist for behavior such as tax evasion, whereas the latter could include data about online shopping activity through e-commerce platforms such as Chinese tech giant Alibaba's Taobao. Credit cities comprise platforms that link up public and market credit data, providing publicly accessible online lookups of blacklists, redlists, and in some cases ratings, along with a growing range of additional applications that different cities have begun to develop.

In these laboratories for influencing behavior, two kinds of projects stand out:

- national government collaborations with tech companies to build new online platforms for domain-specific credit data monitoring;
- municipal government contracting of tech companies to create local rating systems for residents of their cities.

The "Xinyi+" (信易+) Project

In one notable central-government led effort, the NDRC has enlisted tech companies that dominate the markets for their respective services— including financial technology (digital payments, micro-lending, investment) firm Ant Financial, the "Uber of China" Didi Chuxing, and travel booking website CTrip—to create new information-sharing platforms under the "Xinyi+" (信易+) project. The project is marketed as bringing users reward incentives in exchange for model behavior, yet it may also provide the central government the veneer of greater control over tech companies by ensuring that the latter are using government redlist data on their platforms. Roughly translating to "credit convenience," "Xinyi+" is broken into five separate yet interconnected systems. Examples include:

- The "Xinyi transportation" (信易行) offshoot, for which Didi Chuxing signed a memorandum with NDRC declaring that "when individuals who are on the 'redlist' [for] trustworthiness use Didi Chuxing's software, they will be prioritized in calling cabs, [receive] discounts on rides, and can rent bicycles at a discount and without paying a deposit" (China News Network 2018).
- "Xinyi rental," (信易租) which has used data from Ant Financial's Sesame Credit product to make decisions about renting homes and office space to potential tenants (Credit China 2018).

It remains unclear if these companies are relying on redlists alone or are developing evaluatory models that incorporate their own proprietary user data with state-supplied data. This question becomes more complicated when examining credit cities that are considered exemplary for their technological achievements.

City Governments

Thus far, state-lauded examples of credit cities in China tend to involve collaborations between a city government and a single major tech firm. Two cities that have earned the government's praise are Suzhou, with a population of 10.6 million people and located some 100 km from Shanghai, as well as Fuzhou, a city of 7.6 million in 2017.

Suzhou's municipal government consulted Alibaba spin-off Ant Financial in 2015 in order to produce the local Osmanthus Points (桂花分) scoring system, which won the city an innovation award at this year's Credit Cities Summit (Xinhua 2018). Residents of Suzhou receive scores that start from a base of 100 points and can reach a maximum of 200. According to a feature in *Modern Suzhou* magazine, Osmanthus Points' "foundational data come from public security, civil affairs, family planning, social security, and other government bureaus, as well as business units" (*Modern Suzhou* 2016). If an individual's name appears on any state blacklists, this publicly available information would lower the person's Osmanthus Points. Scores increase for donating blood, volunteer work, and winning awards or special honors. To date, it would appear that the "punishment" resulting from a low score is losing out on the benefits that come with having a high score. Although the city has yet to roll out its full range of benefits for high scores, the list of potential institutions and departments that may offer rewards includes libraries, public transportation, and the education, medical, job recruitment, and public service sectors.

Similarly, the municipal government of the southwestern coastal city of Fuzhou is working with Ant Financial's competitor JD Finance on a series of local credit city initiatives. As the fintech branch of

major ecommerce company JD.com, JD Finance is one of the few tech firms to openly advertise its use of artificial intelligence in building what it refers to as a "smart city credit platform" that Fuzhou and other cities have adopted (*Securities Times* 2018). Yet the company is unclear about whether this deployment of AI features in the models that assess citizens' trustworthiness, in cameras that use facial recognition to confirm users' identities, or in other components of their credit city services. The "Three Lanes and Seven Alleys" (三坊七巷) historic neighborhood of Fuzhou is offering benefits for those who are highly rated within the platform JD Finance has created for the city, "Jasmine Points" (茉莉分). These rewards range from deposit-free umbrella rentals to discounts at JD's unmanned supermarket, again suggesting that the stakes are low for those with poor scores.

The Suzhou and Fuzhou examples are instructive because they are playing out in smaller cities whose perceived successes would be harder to replicate at the scale of Beijing or Shanghai. Moreover, their localized point systems seem to be largely ignored as they can neither make nor break people's social statuses in their current state.

Public Awareness and Attitudes

It is difficult to gauge how many people are aware of or interested in the credit city component of the social credit system, although at least one survey Tsinghua University and *Xiaokang* magazine jointly conducted sheds light on some views about the initiative (Liu 2016). A little over half (55.5%) of respondents believe that rewards for trustworthy behavior and joint punishments for untrustworthy behavior should be undertaken in constructing credit cities, but only a third (33.8%) of those surveyed think that blacklists and redlists should be publicized. Furthermore, under a third (29.2%) approve of "credit information sharing" practices writ large, and a similar proportion of respondents (26.5%) support "using honesty networks and similar platforms to expose untrustworthy behavior," with the networks in question here referring in part to websites and apps that name and shame blacklisted entities. This survey's results raise questions about considerations of privacy and the social consequences of blacklisting, redlisting, and joint rewards and punishments. For example, are certain social groups more frequently blacklisted or redlisted than others? Notably, officials and academics who consult the government on building credit cities have their sights trained on a different set of issues.

Chinese Officials' Appraisals of their own System

At the most recent Credit Cities Summit, NDRC deputy director Lian Weiliang identified shortcomings (Lian 2016) in the credit cities initiative. These included inadequate execution of the "double publicity" (双公示) requirement (government bureaus' publication of both punishments and penalties levied against individuals and companies), a lack of mechanisms for timely updating of credit information, and the prevalence of low-quality data. Economist and head of Peking University's China Credit Research Center Zhang Zheng has consulted the NDRC on developing the social credit system, and cautions that "there are still shortcomings in public provisions on the collection, processing, use, and sharing of data on urban subjects' credit information" (Zhang 2017). These opinions from leading voices in the social credit system's development are further compounded by the complexities of ensuring regular information sharing between government bureaus. Since the passage of the national cybersecurity law, many of these departments may be even more reluctant to share data with one another for fear of punishment were a data breach to occur (Dai 2018). Moreover, the problems Chinese experts identify as pressing reveal the gaps between hyperbolized foreign portrayals of the social credit system's technological sophistication and its current growing

pains.

Much of the exaggerated overseas media reportage on social credit conflates the system with highly publicized projects such as Alibaba's City Brain—a smart city offshoot primarily known for its monitoring and guiding of urban traffic—and similar in-house developments from China's other tech giants. These efforts more closely resemble the work of Sidewalk Labs and other US counterparts that partner with local law enforcement, yet there is no current evidence that they are related to the social credit system. Compared to much of the hype surrounding China's biggest tech firms' smart city developments, the credit city concept is not geared toward generating new data so much as it repurposes data that are already routinely collected for other purposes. Nor is there any indication in current policy documents that credit cities are going to integrate Internet of Things (IoT) technologies to gather more granular data on individuals and companies. Even though the fundamentals of credit cities are far less complex than corporate plans for smart cities, a survey of mayors at the most recent Credit Cities Summit revealed that 81% of them believed "full realization of credit cities will probably require another ten years approximately" (Computerization of Finance 2017).

There are still unaddressed risks in how so-called "market credit information"—such as e-commerce and ride-sharing data—is collected and used in credit cities, and the full extent of tech companies' cooperation with the state remains unclear. For now, the stakes for those who obtain poor ratings in credit cities that actually provide scoring systems are still low. These ratings are not used in socioeconomically meaningful scenarios such as loan applications or job screenings, and to date it appears uncertain whether they eventually will be a factor in these selection processes. The vision of ubiquitous computing in credit cities is underwhelming compared to misreported accounts of social credit in China that assume all security cameras are constantly recording individuals' behavior in order to dock points from a centralized score. Despite the prioritization of investment in and research on AI in China, much more ink has been spilled on how to overcome less thrilling bureaucratic hurdles in the social credit system's development than on how to apply AI in credit cities.

Conclusions

The experimental nature of credit cities may produce results policymakers will seek to replicate across additional cities in China, or conversely they may prove ineffective and ultimately be replaced with other solutions. Even if the integration of public and market credit data is considered achieved within the next few years, the nature of ubiquitous computing ideals suggests that by that time goals will have likely shifted once again. The shelf life of catchy project names like "credit cities" can be short in China; the integration of public and private data that makes a "credit city" unique today may make it indistinguishable from any other city in a few years as the systems of reward and punishment underlying this model become a form of infrastructure taken for granted. Yet before this term of art disappears, it merits following because it indicates that at this point in time, there is a division between the data collection and analysis capabilities of the state, and those of the tech sector. As critical as it is to understand what is and isn't true about the social credit system, it's even more important to distinguish what the state and the tech giants need from one another, and how their cooperation subtly changes the fabric of everyday life in China.

References

Bell, G. and Dourish, P. (2011) Divining a Digital Future: Mess and Mythology in Ubiquitous Computing. Cambridge, MA: MIT Press. pp. .

China News Network. (2018) 国家发改委与滴滴签署信用信息共享协议 [National Development and Reform Commission Sign Credit Information Sharing Agreement with Didi Chuxing]. Accessed at: http://www.chinanews.com/business/2018/03-07/8461965.shtml

金融电子化［Computerization of Finance]. (2017). 热议信用城市，助推城市信用建设提速［Credit Cities a Hot Topic in Promoting the Construction of Urban Credit].

Dai, X. (2018) Toward a Reputation State: The Social Credit System Project of China. Accessed at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3193577.

Daum, J. (2017) China through a glass, darkly. China Law Translate. Accessed at: http://www.chinaalawtranslate.com/seeing-chinese-social-credit-through-a-glass-darkly/?lang=en.

Horsley, J. (2018) China's Orwellian Social Credit Score Isn't Real. Foreign Policy. Accessed at: https://foreignpolicy.com/2018/11/16/chinas-orwellian-social-credit-score-isnt-real/.

Jiangsu Economic and Information Technology Commission. (2015) 苏州市政府和阿里巴巴签署战略合作协议 协力推进社会信用体系建设 [Suzhou Municipal Government and Alibaba Sign Strategic Cooperation Agreement, Unite in Effort to Promote Construction of Social Credit System]. Accessed at: http://www.jseic.gov.cn/xwzx/ztzl/dqzt/sxyb/hzyjl/201507/t20150721_203790.html

连维良［Lian Weiliang]. (2017) 让更多信用城市撑起更高水平的信用中国［Allow More Credit Cities to Support a Higher Level of 'Credit China'].《中国经贸导刊》[China Economic and Trade Herald] (22), pp. 4-6..

连维良［Lian Weiliang]. (2017) 信用城市到底长什么样？[What does a credit city look like?].《杭州(党政刊A)》[Hangzhou Party and Government Journal A].

刘源隆 [Liu Yuanlong] 创建信用城市，谁是最大"赢家"? [Who is the Biggest "Winner" in Establishing Credit Cities?].《小康》[Xiaokang] (22), pp. 74-77.

《现代苏州》[Modern Suzhou]. (2016) 想知道苏州人"人品"高低，刷刷"桂花分" [If You Want to Know Whether a Suzhou Resident Has High "Moral Character," Swipe "Osmanthus Points"].

Pence, M. (2018) Remarks by Vice President Pence on the Administration's Policy Toward China. Accessed at: https://www.whitehouse.gov/briefings-statements/remarks-vice-president-pence-administrationspolicy-toward-china/

《证券时报》[Securities Times]. (2018) 京东金融打造智能城市信用平台 用AI和大数据服务城市信用［JD Finance Establishes Smart City Credit Platform, Uses AI and Big Data to Serve Credit City]. Accessed at: http://baijiahao.baidu.com/s?id=1603207392402141026.

新华网［Xinhua]. (2018) 苏州市夺得全国信用信息平台建设城市组冠军 [Suzhou City Wins First Place for National Credit Information Platform Construction]. Accessed at: http://www.js.xinhuanet.com/2018-09/13/c_1123426630.htm.

信用中国 [Credit China]. (2018) 来自2018中国城市信用建设高峰论坛的信之声（系列之二）[Speeches from the 2018 China Credit Cities Construction Summit (Series II)]. Accessed at: http://www.sohu.com/a/235804006_589061.

章政 [Zhang Zheng]. (2017）让信用城市成为中国特色市场经济的领跑者［Allow Credit Cities to Become Frontrunners of a Market Economy With Chinese Characteristics]. 信用中国 [Credit China]. Accessed at: https://www.creditchina.gov.cn/xinyongkanwu/zazhi/201712/P020171213697992309785.pdf.

Chapter 8. The Russian Model of Digital Control and Its Significance

Jaclyn Kerr
LLNL[21]
jackiekerr@gmail.com

Abstract

Russia has emerged as an exemplar of an innovative and experimental alternative approach to information manipulation and control. This differs significantly from the more-often discussed Chinese "Great Firewall" system and other approaches with an emphasis on systemic technical censorship. The Russian model relies on a mix of less overt, more plausibly deniable, legalistic, and often non-technical mechanisms to manipulate online information flows, narratives, and framings, to affect and shape public opinion without resort to universal censorship. The government uses surveillance, a panoply of vague laws, the prosecution or censorship of exemplars, proxy actors and hard to track extra-legal pressures, hacking and leaks, and a heavy emphasis on content production and manipulation to influence narratives and shape public opinion. This model for the domestic control of information not only fits with Russia's own political system. It is also likely to prove more resonant and easier to emulate across many other countries in which a systematic-censorship approach is not technologically or politically feasible. The learning and experimentation involved in this type of domestic information manipulation also has direct applicability to the use of information operations in international political and military competition. The future of this model will likely depend on continuing innovation, not least on the leveraging of advances in AI and big data analysis. If successful, however, this might look very different from the future of information control in China – and have significantly different repercussions for democracies and the international system.

Introduction

The Internet and new information and communication technologies (ICTs) were once hailed as "liberation technologies" – tools to enable the free flow of information, allowing individual freedoms of expression and organization, and breaking down the last vestiges of authoritarianism. Through the course of the 2000s and early 2010s, while democratic states largely converged on a norm of non-censorship, the most closed authoritarian regimes tended to be early adopters of high-censorship overtly-restrictive approaches to Internet control, adopting such approaches as domestic Internet use levels grew and the required technological solutions became affordable on global or regional markets (Gallagher, 2012; Kerr, 2016, 2018; Marquis-Boire et al., 2013; Wagner, 2012; York, 2015). But observers questioned the long-term survivability of such adaptations, looking to events such as Iran's Green Movement and the Arab Spring as proof of the vulnerability of non-democratic systems to the new global flows of information and the transformational uses of the digital technologies (Diamond, 2010; Earl & Kimport, 2011; Farrell, 2012; Garrett, 2006; Howard, 2010, 2011; Meier, 2011; Shirky, 2011; Zayani, 2011).

During this period, "hybrid regimes" – non-democratic regimes that still based their domestic and international legitimacy in part on democratic institutions and rights protections – seemed particularly vulnerable to the sorts of critical discourse and mass protest mobilizations enabled by

[21] This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC. LLNL-TR-764577.

the new technologies (Balzer, 2003; Diamond, 2002; Heydemann, 2007; Karl, 1995; Levitsky & Way, 2010; Schedler, 2002, 2010; Zakaria, 1997). Fraudulent elections, illegal government actions, corruption, and the inadequate protection of constitutional rights all appeared as potential flashpoints – possibly critical to regime survival but also the potential sources of mass protest around official hypocrisy. These regimes were less quick to adopt systemic censorship or other approaches to Internet control that would overtly violate democratic norms. Such overt actions could further undermine their legitimacy at home and abroad. But they faced increasing pressures to do something – to find alternative approaches to manage the stability risks caused by the increasing use of digital technologies in their societies (Kerr, 2016, 2018).

As threats to regime survival in hybrid regime type countries became clear (often following major domestic mass protest mobilizations) these countries began to experiment with alternative mechanisms to rein in the destabilizing influences of the new technologies (Deibert & Rohozinski, 2010; Deibert, Palfrey, Rohozinski, & Zittrain, 2011; Deibert, Palfrey, Rohozinski, Zittrain, & Haraszti, 2010; Deibert, Palfrey, Rohozinski, Zittrain, & Stein, 2008). These approaches were distinctive from the earlier high-censorship models first adopted in more closed authoritarian regimes, and ultimately more akin to the "low-intensity coercion" approaches these regimes often followed in other areas of domestic political control (Levitsky & Way, 2010). This included efforts to utilize democratic legal mechanisms and institutions in combination with pro-regime content production and plausibly deniable forms of disruption to alter online discourse and narratives without recourse to pervasive censorship. Russia emerged as an exemplar of this alternative approach to information control (Deibert & Rohozinski, 2010; Deibert, et al., 2008, 2010, 2011; Kerr, 2016, 2018).

The Russian model has implications both for ongoing global authoritarian learning concerning domestic information control, and for emerging new forms of information warfare and their potential global proliferation. As this model continues to develop, its future will depend on continuing innovation. AI and big data analysis are likely to play a critical role.

A Russian Model of Information Control?

As early as the late 1990s and early 2000s, the Russian government was concerned by the potential destabilizing impact and national security repercussions of information flows within society. In 1995 Russia adopted the "Law on Operational Investigations," giving the Federal Security Service (FSB) authority "to monitor all private communications" of citizens, including electronic communications, and the first "System for Operative Investigative Activities" (or SORM) infrastructure was built – extended in 1998 (SORM-2) to allow monitoring of Internet traffic (Kerr, 2016, 2018; Soldatov & Borogan, 2012, 2013, 2015). Beginning in 1998, Russia submitted nearly-annual resolutions to the United Nations General Assembly concerning "Developments in the field of information and telecommunications in the context of international security," and a 1999 submission to the UN Secretary-General contained a proposed set of "principles in international information security"[22] (Kerr, Loss, & Genzoli, 2018; Korzak, 2017; McKune, 2015). In these submissions it was clear that the concern related as much to international flows of information *content* as to the growing field of cybersecurity. On September 9th 2000, following the immensely negative media coverage of the Kursk submarine tragedy the previous month, Vladimir Putin (then in his first year of office) signed the new

[22] The promotion of international standards for information non-aggression became a consistent theme, with Russia also leading blocks of states in efforts. In 2011 and 2015 it collaborated with other countries from the Shanghai Cooperation Organization to submit joint proposals to the UN General Assembly for an "International Code of Conduct for Information Security," for example (Anderson, 2011; Carr, 2011; Grisby, 2015; McKune, 2015; Rõigas, 2015; Permanent Representatives, 2011, 2015).

"Information Security Doctrine of the Russian Federation" that had been developed by his Security Council. The document declared formal support for freedom of speech and the media, but it also indicated supposed threats to national security related to the flow of information (Kerr, 2016, 2018; Russian Federation, 2000; The Jamestown Foundation, 2000).

Importantly, despite these moves, during this period the Russian government also took steps towards fuller participation in the global digital economy and to assure their burgeoning domestic Internet industry of this commitment (Kerr, 2016).[23] The Russian Internet (colloquially called "RuNet") developed into a vibrant new space of public discourse, with little or no censorship throughout the 2000s, even as restrictions over mainstream media and civil society tightened (Alexanyan, 2013; Alexanyan, et al., 2012; Breininger, 2013; Etling, et al., 2010; Kerr, 2016, 2018). But this is not to say that no effort was made to control the new technology's impact on political stability. This effort increased precipitously following the experience of social-media-fueled mass protest at home.

By the early 2010s, and especially following the 2011-2012 White Ribbon Protest Movement and Vladimir Putin's return to the presidency, Russia emerged as an exemplar of an innovative and experimental – though not always completely consistent or successful – alternative approach to information manipulation and control that differed significantly from the more-often discussed Chinese "Great Firewall" system and other approaches with an emphasis on systemic technical censorship. It has pioneered a distinct model that uses a variety of less overt, more plausibly deniable, legalistic, and often non-technical mechanisms to manipulate online information flows, narratives, and framings, to affect and shape public opinion. It so far does not utilize the level of pervasive censorship observed in China and other settings (Allnut, 2011; Deibert & Rohozinski, 2010; Elder, 2012a; Fedor & Fredheim, 2017; Kerr, 2016, 2018; Ragan, 2012; Subbotovska, 2015). This model for the domestic control of information not only fits with Russia's own domestic political system, but is likely to prove more resonant and easier to emulate across many other countries – including but not limited to other hybrid regimes – in which a systematic-censorship approach is not technologically or politically feasible (Kerr, 2018).

Since 2012, Russia has had a blacklist of legally censored websites. This was a stark change after years in which the Internet was essentially uncensored (Elder, 2012b; Kerr, 2016). But it uses this list parsimoniously, providing legal justifications for each category of restricted content and usually applying these to exemplars rather than systematically. To be clear, pressures on the producers and hosts of controversial online content have increased significantly in Russia in the post-2012 period. But these pressures often take the form of new laws and quasi-democratic processes, financial dealings between companies, or behind-the-scenes (and plausibly deniable requests). A laundry list of new laws have created legal bases for the blocking of a wide variety of content during this period, while also increasing the systematic collection of user data and placing a heavy burden of liability on content intermediaries (HRW, 2017; ICNL, 2016; Kerr, 2016).

---

[23] This included, notably, a widely-recalled December 1999 meeting between then-Prime Minister Putin and members of the Russian Internet community in which, under pressure from the assembled bloggers and ISP-directors, Putin rejected a considered plan for more centralized government control over the Internet and promised that they would be consulted before further policy decisions (Kerr, 2016; Soldatov & Borogan, 2015; Zasoursky, 2003). But some dynamics of consultation continued throughout the 2000s and beyond. During Dmitriy Medvedev's presidency, 2008-2012, Internet entrepreneurship was also avidly promoted as part of his economic modernization program. Medvedev toured Silicon Valley, met with young ICT entrepreneurs, and himself utilizing social media (Hodge, 2009; Kerr, 2016; Siegler, 2010).

These new laws include, for example:

- The 2013 "Anti-Piracy Law" – This law, meant to prevent the online spread of copyrighted materials, put extreme burdens of liability on Internet intermediaries. It was passed despite a broadly coordinated Internet user and platform protest campaign modeled on the successful protests that led to the rejection of the similar Stop Online Piracy (SOPA) legislation in the United States (Kerr, 2016; Rothrock, 2013; Omidi, 2013).
- The 2014 "Anti-LGBT Propaganda Law" – This extended the original 2012 Blacklist for the protection of children from child pornography and content related to illegal drugs and suicide, also requiring the blocking of content that could be seen as "propaganda" for alternative sexual orientations directed at children. The sites of LGBT-youth support groups have been among the first targeted under the law (Elder, 2012b, 2013; Gribova, 2015; Kerr, 2016; Luhn, 2015).
- The 2014 "Law on Pre-Trial Blocking of Websites" – Also called the "Lugovoi Law" after the lawmaker who proposed it, this law permitted the immediate blocking of sites on court order that are deemed to contain "incitement to extremism or riots." It was used to abruptly block several leading oppositional news outlets and blogs at the height of the Crimea Annexation crisis (HRW, 2017; Kerr, 2016).
- The 2014 "Blogger's Law" – Passed as part of an "Anti-Terrorist" package of laws in summer 2014, this law required that all bloggers with a daily audience of more than 3000 register on a national list and follow media regulations for fact-checking their posts (Davidoff, 2014; HRW, 2017; Kerr, 2014, 2016; MacFarquhar, 2014).
- The 2014 "User Data Storage" Law and 2016 Amendments – This law, a version of which was originally passed as part of the same 2014 legal package, began going into force in 2015 and was further updated by the 2016 "Yarovaya Amendments." The later version required that all telecommunications, Internet Service Provider (ISP) and Internet platform companies collecting data from Russian users must store content data for six months and that the telecoms and ISPs further store metadata for three years. Data must be stored on servers located within Russia, providing for government access (Kerr, 2016; HRW, 2017; Shackelford, Richards, Raymond, Kerr, & Kuehn, 2016, 2017; Soldatov, 2015; Whittaker, 2014).
- The 2016 "Anti-Encryption Law" – Also included in the 2016 package were provisions requiring that all encrypted services provide the Federal Security Service (FSB) with encryption keys or other means of decoding transmitted data. An administrative statute adopted at the same time further prohibited the use of uncertified encryption services (HRW, 2017; Shackelford, Richards, Raymond, Kerr, & Kuehn, 2016, 2017).

Such laws, though almost never systematically enforced, create significant chilling effects both for content producers and intermediaries as well as providing legal grounds for subsequent blockings or prosecutions.

Online media outlets and social media platforms face the threat of potential financial takeovers and pressures to swap editors, CEOs, or other key personnel, if they fail to bow to content restriction pressures. Pavel Durov, the founder of Russia's most popular social network, VKontakte, left the country in April 2014 after being fired as CEO and forced to sell his shares in the company leaving it majority owned by oligarchs close to the Kremlin. Durov publicly stated that the conflict had resulted from his unwillingness to disclose user information or block pages relating to Alexei Navalny's anti-corruption campaign and the conflict in Ukraine (Hakim, 2014; Kerr, 2016; Rothrock, 2014; Walker, 2014). Durov subsequently founded the popular encrypted messaging app, Telegram, which the Russian communications regulator Roskomnadzor ordered blocked in April 2018 for refusing to turn

over encrytion keys. The blocking prompted protests and had limited success, with attempts causing temporary blockage to countless other popular sites while the Telegram application itself remained accessible (Burgess, 2018; Deahl, 2018; MacFarquhar, 2018).

Changes in surveillance laws and capabilities have been an important area of increased government control in the post-2012 period – though it is not always clear to exactly what extent and ends the collected data is being utilized. Internet service providers (ISPs) and social media platforms alike have faced pressure to quickly implement new requirements such as the purchase and installation of surveillance equipment on their networks or the storage of and government access to all user metadata and communications. Russia's mass surveillance system, SORM, is grounded on a legal framework allowing for the "lawful interception" of communications by a number of KGB-successor security organs and other government bodies. It also involves particular technological systems and infrastructures used to implement the data storage and access. Both the SORM regulation and technology have received recent enhancements. Whereas earlier SORM-2 systems had only operated at the ISP level, an August 2014 decree required all social media platforms operating in Russia to install SORM monitoring equipment. The new SORM-3 system, announced also in 2014, was to permit the storage of all communications and tracking of data streams by particular users and IP addresses (Franceschi-Bicchierai, 2014; Kerr, 2016, 2018; Kozlovsky, 2014; Paganini, 2014; Soldatov & Borogan, 2012, 2013, 2015; Soldatov, Borogan, & Walker, 2013).

In the Russian approach to information control, in addition to surveillance and legal and extra-legal pressures, new forms of pro-regime content mass-production and narrative manipulation as well as the limited use of plausibly deniable cyberattacks and hacking play critical roles in efforts to undermine and marginalize the voices of opposition movements and leaders, while also shaping broader public opinion without a sense of dramatic restriction. The leveraging of youth organizations (such as Nashi), third-party botnets, independent hackers, contracted video-producers, and pro-regime bloggers in coordinated actions provides a further degree of deniability of government involvement. Bots, trolls, leaks of compromising or manipulated content, distributed denial of service (DDoS) attacks causing temporary "technical failures," and other difficult-to-attribute techniques are combined with occasional legal prosecutions or site-blockages for exemplary offenders under vague laws and mass digital surveillance, creating an overall online environment which still appears relatively unrestricted – with the ability to produce and access wide varieties of content, including content critical of the government – but in which the government exerts significantly more control over the overall development of content and narratives (Deibert & Rohozinski, 2010; Fedor & Fredheim, 2017; Kerr, 2016, 2018; Ragan, 2012; Subbotovska, 2015).

In the realm of content production, Russia has shown significant experimentation in its effort to gain greater control over domestic opinion and dampen sources of political instability. While originally seeking to sway public opinion primarily through television content, the approach has been updated in recent years to adjust for the growing domestic political significance of Internet content consumption. The new 2016 version of the Russian "Doctrine of Information Security" explicitly discussed the roles of the Internet and social media as well as other mediums for information production and consumption (Russian Federation, 2016). While some forms of propaganda and tools of narrative manipulation are repeated across all platforms in coordinated efforts, other techniques appear to have been developed explicitly to take advantage of the capabilities and vulnerabilities created by the digital media ecosystem.

Russian content production and manipulation efforts often pay careful attention to framing and agenda-setting. This plays off existing biases, identities, societally-resonant symbols, and the

manipulation of emotion.[24] In some cases, efforts aim to promote particular narratives. In others they plant numerous alternatives to existing narratives sowing confusion and uncertainty (e.g. "who downed MH17?" "who was behind chemical weapons attacks in Syria?"). They also interrupt and distract politically critical conversations, dilute potentially critical discourse contexts with fun apolitical content[25] (e.g. the discussion surrounding politically salient hashtags), or seed different content into different echo chambers to further exacerbate existing tensions (Fedor & Fredheim, 2017; Woolley & Howard, 2017; Kerr, 2018; Lin & Kerr, 2017; Pomeranzev, 2014a,b; Stewart, Arif, & Starbird, 2018; Sanovich, 2017). These techniques can aim to drown out criticism or break potential protest coalitions, preventing critical discourse from leading to political mobilization without the need for frequent censorship.

Relationship to International Information Conflict

The Russian approach to domestic control of information within society has direct applicability to the leveraging of information operations in international political and military competition. It also is closely tied conceptually. Throughout the 2000s, as concern about the existential threat to regime survival posed by mass protest events grew, the leadership increasingly came to worry about the roles of transnational information flows as part of military and strategic competition and as potential sources of domestic political instability. While the U.S. and democratic allies promoted "Internet freedom" as a distinct issue from growing attention to national cybersecurity and the military cyber domain, the Russian understanding of "information security" and international information aggression subsumed both the transnational networked flows of media and information and the networked computer systems and data that were generally the focus of cybersecurity analysis (Clinton, 2010; Kerr, 2016, 2018; Kerr, Loss, & Genzoli, 2018).

This consideration is clear in an often-quoted article by Russia's then chief of the general staff, General Valery Gerasimov, that focused on Arab Spring type events as part of an analysis of the current military-technological and geopolitical threat landscape (Gerasimov, 2013). He suggested that "broad use of political, economic, informational, humanitarian, and other non-military measures – applied in coordination with the protest potential of the population" were playing increasingly significant roles in contemporary forms of strategic international competition. Speaking of threats posed to Russia, he stressed Russia's need to also utilize such combined efforts, engaging in "cognitive-psychological" and "digital-technological" forms of influence. Suggesting that strategic goals could be achieved with little resort to armed conflict[26] through influencing perceptions and decision-making processes, he stressed the importance of "information spaces" and the possibility of

---

[24] In some cases, framing and agenda setting appear to be given particularly systematic attention, even utilizing the broader global information flows to the regime's benefit. In the period following the White Ribbon movement's mass mobilization of a diverse coalition to protest regime corruption and electoral fraud, the imprisonment of members of the feminist punk girl band Pussy Riot (for staging a protest inside a cathedral), and the ratcheting up of pressure on LGBT groups both seemed calculated to draw attention to the less traditional values expressed by small subsets of the protest movement. This attention – reflected back and magnified through Western civil society and governmental attention and outrage – helped to reframe the protest movement as one concerned primarily with these progressive issues, weakening the cross-coalition bonds between different protest participant groups and reducing the resonance for the majority of participants who had mobilized around economic and political rights. At the same time, moderate protest mobilizations in Moscow concerned with peace with Ukraine and media freedom received little such coordinated media attention.

[25] This sometimes includes content which could pass as either satire or propaganda and thus is spread by supporters and opponents. Such content sometimes plays to pop-cultural tropes, memes, and content-production patterns, blending with other popular content and even inspiring copycat content production.

[26] He suggested a 4-to-1 ratio of non-military to military operations.

exploiting asymmetric vulnerabilities, even against more militarily-powerful adversaries (Adamsky, 2015; Gerasimov, 2013; Lin & Kerr, 2017).

Evidence today suggests that Russia utilizes information operations abroad, both in regional and international theaters, at levels targeting individuals, groups or entire populations. These are applied to undermine credibility or intimidate, plant particular narratives and distract from others, sow confusion and uncertainty, exacerbate divisions, galvanize protest, and slow or influence decision-making processes. Goals appear to include efforts to influence elections, undermine support for political parties and candidates, support extremism and polarization, and undermine the legitimacy of institutions not aligned with Russian foreign policy. Techniques sometimes involve both technical (hacking, malware) and informational (content) components, including actions such as leaks of compromising material, DDoS attacks, and website defacements. They often take advantage of plausible (or even implausible) deniability, and can occur during peacetime, grey zone (sub-threshold) conflicts, or wartime, in combination with special operations, direct military action, or diplomatic interaction. In addition to state-organ-led efforts, Russia appears to also sometimes leverage hacktivists, youth organizations, criminal networks, and paid troll farms (e.g. the Saint Petersburg-based Internet Research Agency) as state proxies to conduct operations, aiming to obfuscate direct governmental links (Giles, 2016; Kerr, 2016, 2018; Lin & Kerr, 2017; MacFarquhar, 2016; Parlapiano & Lee, 2018; Polyakova, 2018; Timberg, 2017).

As elements of international geopolitical competition, these techniques draw on a long tradition within Russian and Soviet military strategy. Soviet "active measures" and the concepts of "maskirovka" and "reflexive control" in Russian military theory each involve the use of information and deception, ambiguity and illusion, and deniable and indirect activities, for the purposes of psychological manipulation and asymmetric influence (Dailey & Parker, 1987; Schultz & Godson, 1984; Thomas, 2004). The more recent cyber-enabled information operations are differentiated, however, by the ubiquity and capabilities of the digital technologies being utilized. Using the new technologies, significant influence effects can be achieved remotely, quickly, on scale, and at relatively low cost – at least in theory. Some of these cyber-enabled information and influence operations are undoubtedly more effective than others. As in the domestic sphere, there is evidence of experimentation to develop more effective uses of the current tools for information manipulation (Giles, 2016; Lin & Kerr, 2017).[27]

The international applicability of aspects of the Russian domestic model for information control suggests that ongoing learning and experimentation within authoritarian regimes will have continuing relevance to international information contestation. This also brings into question the Cold War era assumption that democracies, having less to fear from public discourse and free expression, are always more resilient to international flows of information than are non-democratic regimes. Democratic countries may in fact have some important vulnerabilities that are different and greater in the face of the new information operation techniques.

---

[27] While early examples occurred in the 2000s, including operations during diplomatic and military conflicts with Estonia (2007) and Georgia (2008) respectively, more recent events, particularly since the 2014 beginning of the conflict with Ukraine and surrounding elections in the United States and Europe, show ongoing experimentation and learning. The more recent campaigns have used the hacking and leaking of confidential information to manipulate media discussion (e.g. DNC hack), leveraged major social media platforms through bots and other fake accounts to disseminate content, and used micro-targeting and advertising technologies and the manipulation of existing fringe or activist echo chambers and group sites to introduce false information or alternative narratives, exacerbate social divisions, influence public discourse, and catalyze real-world protest events.

So What? And What Next?

The Russian model of Internet control should not be reified. It has emerged out of ongoing experimentation, and sometimes seems as much shaped by opposing internal inclinations or by a failure to adequately implement more robust censorship models as by an intentional effort to maintain some semblance of democratic legitimacy. Why then, if at all, is the distinctiveness of the Russian approach worth noting? There are at least two significant reasons.

The first is the applicability of some subset of this model's features to regional and international theaters. This means that experimentation and learning around information control at home can drive advances in "political" or "information" warfare capabilities in international competition. The second is the potential broader diffusion of this model – both the domestic and international elements – to countries for which a sophisticated censorship approach might, for various reasons, not be within grasp.[28] The continued success and diffusion of the model's domestic approach promises a potential path forward for hybrid regimes in the digital age. The demonstration of its utility in regional and international conflict is likely also to serve as inspiration for many copycats.

But this leaves several unresolved questions. Can this model continue? Is it possible, long-term, to retain as much (or sufficient) control over public opinion through content production, surveillance, and limited censorship as through ubiquitous censorship? The continuing success of this approach will require ongoing innovation. So the answer might depend on the next steps. How is this likely to develop further in the near future? Should we expect an eventual convergence with or continued distinction from the Chinese model?

Of particular significance for the future of digital authoritarian models and global information-power competition will be the interrelated roles of AI and big data. Advances in machine learning are now driving breakthroughs in a variety of technologies relevant to online discourse and its monitoring, censorship, or emulation. As demonstrated by Cambridge Analytica, AI and big data permit ever-more-precise forms of micro-targeting – whether for advertising or propaganda. Algorithms also now permit the production of increasingly inexpensive and realistic deep fakes – fabricated lifelike audio and video files that can make it appear that someone said or did something that they did not. Improvements in sentiment analysis and natural language processing allow better analysis of emotion – useful for targeting and engaging individuals and populations. Meanwhile, chatbots are finally passing the "Turing Test," with some experiments showing subjects unable to differentiate between interactions with real people and computer agents in certain settings (Barnes & Chin, 2018; Horowitz, Allen, Kania, & Scharre, 2018; Polyakova, 2018; Powers & Kounalakis, 2017; Wooley & Howard, 2017; Wright, 2018). In states which have struggled to implement systematic Internet and

[28] The domestic model is likely of particular ease to emulate across the former Soviet region, as these states share legal and institutional legacies, participate in common regional organizations, and also often share overlapping media markets and Internet resources. There is already evidence of significant diffusion of aspects of this model across the region, including, for example, the SORM surveillance infrastructure and accompanying "lawful intercept" legal frameworks permitting access for KGB-successor organs, national security frameworks focused on the role of information, and many similar hacking and content production and manipulation tactics. The model is also of clear merit to other hybrid regimes, which wrestle with the same conflicting pressures, however. And some aspects of this approach are likely to prove valuable to states of various non-democratic regime types that for technical, financial, human capital, or organizational reasons have more adequate capacities to implement an approach that relies less on technical systematic censorship and more on the prosecution or censorship of exemplars, use of broad legal rules, and content production (Bourgelais, 2013; Deibert and Rohozinski, 2010; Hall and Ambrosio, 2017; Horowitz, 2010; Kerr, 2016, 2018; Kucera, 2010; Michel, 2015; Omanovic, 2014; Soldatov and Borogan, 2012).

information controls, shouldn't these tools permit more ubiquitous censorship and more perfect law enforcement?

In a September 2017 speech, Vladimir Putin noted the importance of AI. "Artificial intelligence is the future," he told the nation's students, "not only for Russia, but for all humankind" (RT, 2017). A great deal of attention has focused on the Chinese government's access to large quantities of data – critical to the training and effective use of AI algorithms. But with all the input from the SORM surveillance systems and recent data storage requirements, the Russian government is likely also to have significant data with which to experiment.

How might the role of AI and big data in information control look different in a Russian context? One could imagine this as the solution to all of Russia's censorship and enforcement woes. But if the content production and limited censorship approach continues to prove effective, it seems more likely that Russia would use AI in ways consistent with that model: more precise micro-targeting, more emotional manipulation, more believable and impactful propagandistic content – and, importantly, the use of these same tools at home and abroad. This is something worth anticipating and preparing for…

Reference

Adamsky, D. (2015). *Cross-Domain Coercion: The Current Russian Art of Strategy*. Paris, France: Institut Français des Relations Internationales. Retrieved from http://www.ifri.org/sites/default/files/atoms/files/pp54adamsky.pdf

Alexanyan, K. (2013). *The Map and The Territory: Russian Social Media Networks and Society*. (PhD dissertation). New York, NY. Retrieved from https://academiccommons.columbia.edu/doi/10.7916/D8XP7C49

Alexanyan, K., Barash, V., Etling, B., Faris, R., Gasser, U., Kelly, J. Palfrey, J. G., & Roberts, H. (2012). Exploring Russian Cyberspace: Digitally-Mediated Collective Action and the Networked Public Sphere. Cambridge, MA: Harvard University, Berkman Center for Internet and Society. Retrieved from http://papers.ssrn.com/abstract=2014998

Allnutt, L. (2011, March 23). Russia's 30-Ruble Army. *RadioFreeEurope/RadioLiberty: Tangled Web*. Retrieved from http://www.rferl.org/content/russias_30_ruble_army/2347318.html

Anderson, N. (2011, September 20). Russia, China, Tajikistan propose UN "code of conduct" for the 'net. *Ars Technica*. Retrieved from http://arstechnica.com/tech-policy/news/2011/09/russia-china- tajikistan-propose-un-code-of-conduct-for-the-net/

Balzer, H. (2003). Managed pluralism: Vladimir Putin's emerging regime. *Post-Soviet Affairs*, *19*(3), 189–227.

Barnes, J. E. & Chin, J. (2018, March 2). The New Arms Race in AI. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/the-new-arms-race-in-ai-1520009261.

Bourgelais, P. (2013, June). Commonwealth of surveillance states. *Access Now*. Retrieved from https://s3.amazonaws.com/access.3cdn.net/279b95d57718f05046_8sm6ivg69.pdf

Breininger, O. (2013, March 28). Banderlogs and Network Hamsters: The Language of Political Protest in Russia. *openDemocracy*. Retrieved from http://www.opendemocracy.net/od-russia/olga-breininger/banderlogs-and-network-hamsters-language-of-political-protest-in-russia

Burgess, M. (2018, April 28). This is why Russia's attempts to block Telegram have failed. *Wired*. Retrieved from https://www.wired.co.uk/article/telegram-in-russia-blocked-web-app-ban-facebook-twitter-google

Carr, J. (2011, September). 4 problems with China and Russia's international code of conduct for information security. *Digital Dao*. Retrieved from http://jeffreycarr.blogspot.com/2011/09/4- problems-with-china-and-russias.html

Permanent Representatives of China, Kazakhstan, Kyrgyzstan, the Russian Federation, Tajikistan and Uzbekistan. (2015, January). *Developments in the field of information and telecommunications in the context of international security: Letter dated January 9, 2015, to the United Nations addressed to the secretary-general* (UN A/69/723). Retrieved from https://ccdcoe.org/sites/default/files/documents/UN-110912-CodeOfConduct_0.pdf

Permanent Representatives of China, the Russian Federation, Tajikistan and Uzbekistan. (2011, September). Developments in the field of information and telecommunications in the context of international security: Letter dated September 12, 2011, to the United Nations addressed to the secretary-general (UN A/66/359). Retrieved from https://ccdcoe.org/sites/default/files/ documents/UN-110912-CodeOfConduct_0.pdf

Clinton, H. (2010, Jan 21). Remarks on Internet Freedom. Remarks of Secretary of State Hilary Rodham Clinton. Washington, D.C.: The Newseum. Retrieved from https://2009-2017.state.gov/secretary/20092013clinton/rm/2010/01/135519.htm

Dailey, B. D., & Parker, P. J., eds. (1987). *Soviet Strategic Deception*. Hoover Institution Press: Lexington, MA.

Davidoff, V. (2014, April 27). An Internet Censorship Law Right Out of '1984.' *The Moscow Times*. https://themoscowtimes.com/articles/an-internet-censorship-law-right-out-of-1984-34692

Deahl, D. (2018, March 20). Russia orders Telegram to hand over users' encryption keys. *The Verge*. Retrieved from https://www.theverge.com/2018/3/20/17142482/russia-orders-telegram-hand-over-user-encryption-keys

Deibert, R., & Rohozinski, R. (2010). Chapter 2: Control and Subversion in Russian Cyberspace. *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. The MIT Press: 15-34. Retrieved from http://www.access-controlled.net/wp-content/PDFs/chapter-2.pdf.

Deibert, R., Palfrey, J., Rohozinski, R., Zittrain, J. & Stein, J. G., eds. (2008). *Access Denied: The Practice and Policy of Global Internet Filtering*. Cambridge, Mass: The MIT Press.

Deibert, R., Palfrey, J., Rohozinski, R., & Zittrain, J., eds. (2011). *Access Contested: Security, Identity, and Resistance in Asian Cyberspace*. Cambridge, MA: The MIT Press.

Deibert, R., Palfrey, J., Rohozinski, R., Zittrain, J., & Haraszti, M. (Eds.). (2010). *Access controlled: The shaping of power, rights, and rule in cyberspace*. Cambridge, MA: MIT Press.

Diamond, L. (2002, April). Elections Without Democracy: Thinking About Hybrid Regimes. *Journal of Democracy* 13, no. 2: 21–35.

Diamond, L. (2010, July). Liberation Technology. *Journal of Democracy* 21, no. 3: 69–83.

Earl, J., & Kimport, K. (2011). *Digitally Enabled Social Change: Activism in the Internet Age*. Cambridge, Mass: The MIT Press.

Elder, M. (2012a, Feb 7). Polishing Putin: Hacked emails suggest dirty tricks by Russian youth group. *The Guardian*. Retrieved from https://www.theguardian.com/world/2012/feb/07/putin-hacked- emails-russian-nashi

Elder, M. (2012b, Nov 12). Censorship Row over Russian Internet Blacklist. *The Guardian*. Retrieved from https://www.theguardian.com/world/2012/nov/12/censorship-row-russian-internet-blacklist

Elder, M. (2013, June 11). Russia Passes Law Banning Gay 'Propaganda.' *The Guardian*. Retrieved from https://www.theguardian.com/world/2013/jun/11/russia-law-banning-gay-propaganda

Etling, B., Alexanyan, K., Kelly, J., Faris, R., Palfrey, J. G. and Gasser, U. (2010). Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization. Cambridge, MA: Harvard University, Berkman Center for Internet and Society. Retrieved from https://cyber.law.harvard.edu/node/95073

Farrell, H. (2012). The Consequences of the Internet for Politics. *Annual Review of Political Science* 15, no. 1: 35–52. doi:10.1146/annurev-polisci-030810-110815.

Fedor, J., & Fredheim, R. (2017). We need more clips about Putin, and lots of them: Russia's state-commissioned online visual culture. *Nationalities Papers*, *45*(2), 161–181.

Franceschi-Bicchierai, L. (2014, Jan 28). With Olympics Looming, Russia Ramps Up Online Surveillance. *Mashable.com*. Retrieved from http://mashable.com/2014/01/28/russia-internet-olympics/

Gallagher, S. (2012, Sept 27). Big Brother on a Budget: How Internet Surveillance Got so Cheap - Deep Packet Inspection, Petabyte-Scale Analytics Create a 'CCTV for Networks.' *Ars Technica*. Retrieved from http://arstechnica.com/information-technology/2012/09/big-brother-meets-big-data-the-next-wave-in-net-surveillance-tech/

Garrett, R. K. (2006). Protest in an Information Society: A Review of Literature on Social Movements and New ICTs. *Information, Communication and Society* 9, no. 2: 202–24. doi:10.1080/13691180600630773.

Gerasimov, G.V. (2013, Feb 27). The Value of Science in Prediction. *Military-Industrial Kurier*. Article by then Chief of the General Staff of the Russian Federation. Original article retrieved from

http://vpk- news.ru/sites/default/files/pdf/VPK_08_476.pdf A non-authoritative English translation is available by Robert Coalson at https://www.facebook.com/notes/robert-coalson/russian-military-doctrine-article-by-general-valery-gerasimov/10152184862563597/

Giles, K. (2016). *Handbook of Russian Information Warfare*. Rome: NATO DEFENSE COLLEGE. Retrieved from http://www.ndc.nato.int/news/news.php?icode=995

Gribova, D. (2015, Sept 22). It Gets Worse for Russia's Most-Prominent LGBT Youth Support Group. *Global Voices*. Retrieved from https://globalvoices.org/2015/09/22/it-gets-worse-for-russias-most-prominent-lgbt-youth-support-group/

Grisby, A. (2015, Jan 28). Net politics: Will China and Russia's updated code of conduct get more traction in a post-Snowden era? *Council on Foreign Relations—Net Politics*. Retrieved from http://blogs.cfr.org/cyber/2015/01/28/will-china-and-russias-updated-code-of-conduct-get- more-traction-in-a-post-snowden-era/

Hakim, D. (2014, Dec 2). Once Celebrated in Russia, the Programmer Pavel Durov Chooses Exile. *The New York Times*. http://www.nytimes.com/2014/12/03/technology/once-celebrated-in-russia-programmer-pavel-durov-chooses-exile.html.

Hall, S., & Ambrosio, T. (2017). Authoritarian learning: A conceptual overview. *East European Politics*, *33*(2), 143–161.

Heydemann, S. (2007, Oct). Upgrading authoritarianism in the Arab world. Washington, D.C.: The Brookings Institution. Retrieved from http://www.brookings.edu/research/papers/2007/10/arabworld

Hodge, N. (2009, April 28). Kremlin 2.0: Russian Prez Discovers Social Media. *WIRED*. Retrieved from https://www.wired.com/2009/04/kremlin-20-russian-prez-discovers-social-media/

Horowitz, M. (2010). The diffusion of military power: Causes and consequences for international politics. Princeton, NJ: Princeton University Press.

Horowitz, M. C., Allen, G. C., Kania, E. B., & Scharre, P. (2018). Strategic Competition in an Era of Artificial Intelligence. Washington, DC: Center for a New American Security. Retrieved from http://files.cnas.org.s3.amazonaws.com/documents/CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf

Howard, P. N. (2010). The Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam. Oxford: Oxford University Press.

Howard, P. N. (2011, Jan 29). How Digital Media Enabled the Protests in Tunisia and Egypt. *The Great Debate*. Retrieved from http://blogs.reuters.com/great-debate/2011/01/28/how-digital-media-enabled-the-protests-in-tunisia-and-egypt/

Human Rights Watch (HRW). (2017, July 18). Online and On All Fronts: Russia's Assault on Freedom of Expression. Retrieved from https://www.hrw.org/report/2017/07/18/online-and-all-fronts/russias-assault-freedom-expression#

International Center for Not-for-Profit Law (ICNL). (2016, July 21). Overview of the Package of Changes into a Number of Laws of the Russian Federation Designed to Provide for Additional Measures to Counteract Terrorism. Washington, DC: The International Center for Not-for-Profit Law. Retrieved from http://www.icnl.org/research/library/files/Russia/Yarovaya.pdf

Karl, T. L. (1995). The hybrid regimes of Central America. *Journal of Democracy*, *6*(3), 72–86. doi:10.1353/jod.1995.0049.

Kerr, J. (2014, May 13). Chill of Victory: Russia Targets Bloggers Amid Celebrations. *Independent Journalism 101 [Previously: Journalism for Change]*. Retrieved from https://www.scoop.it/t/russia-by-ricehaus/p/4021297381/2014/05/14/chill-of-victory-russia-targets-bloggers-amid-celebrations

Kerr, J. (2016). Authoritarian management of (cyber-) society: Internet regulation and the new political protest movements. (Doctoral dissertation). Retrieved from: http://hdl.handle.net/10822/1042836

Kerr, J. (2018). Information, Security, and Authoritarian Stability: Internet Policy Diffusion and Coordination in the Former Soviet Region. *International Journal of Communication* 12, 3814–3834.

Kerr, J., Loss, R., & Genzoli, R. (2018, April). Cyberspace, Information Strategy, and International Security: Workshop Summary. Livermore, CA: Center for Global Security Research.

Korzak, E. (2017, July 31). UN GGE on Cybersecurity: The End of an Era? The Diplomat. Retrieved from https://thediplomat.com/2017/07/un-gge-on-cybersecurity-have-china-and-russia-just-made-cyberspace-less-safe/

Kozlovsky, S. (2014, Aug 15). Russia Just Doubled Its Internet Surveillance Program. *Global Voices Advocacy*. Retrieved from https://advox.globalvoices.org/2014/08/15/russia-just-doubled-its-internet-surveillance-program/

Kucera, J. (2010, Dec). CSTO fires salvo in information war. *EurasiaNet.org*. Retrieved from http://www.eurasianet.org/node/62639

Levitsky, S., & Way, L. (2010). *Competitive authoritarianism: Hybrid regimes after the Cold War*. New York, NY: Cambridge University Press.

Lin, H. & Kerr, J. (2017). On Cyber-Enabled Information/Influence Warfare and Manipulation. Forthcoming in *Oxford Handbook of Cyber Security*. Ed. Paul Cornish. Oxford: Oxford University Press. Retrieved from https://cisac.fsi.stanford.edu/sites/default/files/cyber-enabled_influence_warfare-ssrn-v1.pdf

Luhn, A. (2015, July 29). LGBT Website Founder Fined under Russia's Gay Propaganda Laws. *The Guardian*. Retrieved from https://www.theguardian.com/world/2015/jul/29/lgbt-yelena-klimova-fined-russia-gay-propaganda-laws

Macfarquhar, N. (2014, May 6). Russia Quietly Tightens Reins on Web With 'Bloggers Law.' *The New York Times*. Retrieved from http://www.nytimes.com/2014/05/07/world/europe/russia-quietly-tightens-reins-on-web-with-bloggers-law.html

MacFarquhar, N. (2016, Aug 28). A Powerful Russian Weapon: The Spread of False Stories. *New York Times*. Retrieved from https://www.nytimes.com/2016/08/29/world/europe/russia-sweden- disinformation.html

MacFarquhar, N. (2018, April 30). 'They Want to Block Our Future': Thousands Protest Russia's Internet Censorship. The New York Times. Retrieved from https://www.nytimes.com/2018/04/30/world/europe/russia-telegram-digital-resistance.html

Marquis-Boire, M., Anderson, C., Dalek, J., McKune, S., & Railton, J. S. (2013). Some Devices Wander by Mistake: Planet Blue Coat Redux. Toronto, Canada: University of Toronto, The Citizen Lab. Retrieved from https://citizenlab.org/2013/07/planet-blue-coat-redux/

McKune, S. (2015, Sept). An analysis of international code of conduct for information security: Will the SCO States' Efforts to Address 'Territorial Disputes' in Cyberspace Determine the Future of International Human Rights Law? Toronto, Canada: University of Toronto, The Citizen Lab. Retrieved from https://citizenlab.org/2015/09/international-code-of-conduct/

Meier, P.P. (2011) *Do 'Liberation Technologies' Change the Balance of Power Between Repressive States and Civil Society?* (Doctoral dissertation). Retrieved from https://irevolution.files.wordpress.com/2011/11/meier-dissertation-final.pdf

Michel, C. (2015, Sept 27). The Eurasian Economic Union's 'single information field.' *The Diplomat*. Retrieved from https://thediplomat.com/2015/09/the-eurasian-economic-unions-single-information-field/

Omanovic, E. (2014). Private interests: Monitoring Central Asia. Privacy International. Retrieved from https://privacyinternational.org/report/837/private-interests-monitoring-central-asia

Omidi, M. (2013, Aug 1). Blackout: Why Russian Internet Sites Are Going Dark over Anti-Piracy Laws. *The Calvert Journal*. Retrieved from http://calvertjournal.com/articles/show/1273/blackout-why-russian-internet-sites-are-going-dark-over-anti-piracy-laws

Paganini, P. (2014, Aug 18). New Powers for the Russian Surveillance System SORM-2: The Russian Prime Minister Dmitry Medvedev Has Signed a Decree That Will Extend the Use of SORM-2 to Social Network Surveillance. *Security Affairs*. Retrieved from http://securityaffairs.co/wordpress/27611/digital-id/new-powers-sorm-2.html

Parlapiano, A., & Lee, J. C. (2018, Feb 16). The Propaganda Tools Used by Russians to Influence the 2016 Election. *The New York Times*. Retrieved from https://www.nytimes.com/interactive/2018/02/16/us/politics/russia-propaganda-election-2016.html

Polyakova, A. (2018, Nov 15). Weapons of the weak: Russia and AI-driven asymmetric warfare. Washington, DC: The Brookings Institution. Retrieved from https://www.brookings.edu/research/weapons-of-the-weak-russia-and-ai-driven-asymmetric-warfare/

Pomerantsev, P. (2014a). Nothing Is True and Everything Is Possible: The Surreal Heart of the New Russia. New York: PublicAffairs.

Pomerantsev, P. (2014b, Sept 9). Russia and the Menace of Unreality. *The Atlantic*. Retrieved from http://www.theatlantic.com/international/archive/2014/09/russia-putin-revolutionizing-information-warfare/379880/.

Powers, S., & Kounalakis, M., eds. (2017, May). Can Public Diplomacy Survive the Internet? Bots, Echo Chambers, and Disinformation. Advisory Commission on Public Diplomacy, U.S. Department of State. Retrieved from https://www.state.gov/documents/organization/271028.pdf

Ragan, S. (2012, Feb 20). Political activism gives way to hacktivism in Russia. *SecurityWeek.Com*. Retrieved from http://www.securityweek.com/political-activism-gives-way-hacktivism-russia

Rõigas, H. (2015). An updated draft of the code of conduct distributed in the United Nations—What's new? *NATO CCDCOE International Cyber Developments Review*. Retrieved from https://ccdcoe.org/updated-draft-code-conduct-distributed-united-nations-whats-new.html

Rothrock, K. (2013, June 15). Orphaned in US, SOPA Finds Home in Russia. *A Good Treaty*. http://www.agoodtreaty.com/2013/06/14/orphaned-in-us-sopa-finds-home-in-russia/

Rothrock, K. (2014, April 21). So Long, Mr. Durov, and Thanks for All the Fish. *Global Voices*. https://globalvoices.org/2014/04/21/so-long-mr-durov-and-thanks-for-all-the-fish/

RT. (2017, Sept. 1). 'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day. Retrieved from https://www.rt.com/news/401731-ai-rule-world-putin/

Russian Federation. (2000, Sept 9). INFORMATION SECURITY DOCTRINE OF THE RUSSIAN FEDERATION. United Nations International Telecommunications Union (ITU) Archive. Retrieved from https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/Russia_2000.pdf

Russian Federation. (2016, Dec 5). Doctrine of Information Security of the Russian Federation. Ministry of Foreign Affairs: Approved by Decree of the President of the Russian Federation, No. 646. Retrieved online from http://www.mid.ru/en/foreign_policy/official_documents/-/asset_publisher/CptICkB6BZ29/content/id/2563163?p_p_id=101_INSTANCE_CptICkB6BZ29&_101_INSTANCE_CptICkB6BZ29_languageId=en_GB

Sanovich, S. (2017). Computational Propaganda in Russia: The Origins of Digital Misinformation. Report for the Project on Computational Propaganda. Oxford, England: Oxford Internet

Institute. Retrieved from http://comprop.oii.ox.ac.uk/publishing/working-papers/computational-propaganda-in-russia-the-origins-of-digital-misinformation/

Schedler, A. (2002). The menu of manipulation. *Journal of Democracy*, *13*(2), 36–50.

Schedler, A. (2010). Authoritarianism's last line of defense. *Journal of Democracy*, *21*(1), 69–80.

Shultz, R. H., & Godson, R. (1984). *Dezinformatsia: Active Measures in Soviet Strategy*. Pergamon Press: Washington, D.C.

Shackelford, S., Richards, E., Raymond, A., Kerr, J., & Kuehn, A. (2016, October 20). Decrypting the Global Encryption Debate. *Huffpost*. Retrieved from https://www.huffingtonpost.com/entry/decrypting-the-global-encryption-debate_us_5808d3f9e4b00483d3b5d0bf

Shackelford, S., Richards, E., Raymond, A., Kerr, J., & Kuehn, A. (2017). iGovernance: The Future of Multi-Stakeholder Internet Governance in the Wake of the Apple Encryption Saga. *North Carolina Journal of International Law and Commercial Regulation*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2851283

Shirky, C. (2011). The political power of social media: Technology, the public sphere, and political change. *Foreign Affairs*, *90*(1), 28–41.

Siegler, M. G. (2010, June 23). Russian President Medvedev Sends His First Tweet At Twitter. *TechCrunch*. Retrieved from http://social.techcrunch.com/2010/06/23/medvedev-twitte/

Soldatov, A., & Borogan, I. (2012, Dec 21). In ex-Soviet states, Russian spy tech still watches you. *WIRED*. Retrieved from https://www.wired.com/2012/12/russias-hand/

Soldatov, A., & Borogan, I. (2013, Fall). Russia's surveillance state. *World Policy Journal*, *30*(3). Retrieved from http://www.worldpolicy.org/journal/fall2013/Russia-surveillance

Soldatov, A., & Borogan, I. (2015). The red Web: The struggle between Russia's digital dictators and the new online revolutionaries. New York, NY: Public Affairs.

Soldatov, A. (2015, Dec 16). Dear Facebook, Please Don't Hand Our Data to The Kremlin - Transitions Online. *Transitions Online: Regional Intelligence*. Retrieved from http://www.tol.org/client/article/25373-facebook-please-dont-hand-our-data-to-putin.html

Soldatov, A., Borogan, I., & Walker, S. (2013, Oct 6). As Sochi Olympic Venues Are Built, so Are Kremlin's Surveillance Networks. *The Guardian*. Retrieved from https://www.theguardian.com/world/2013/oct/06/sochi-olympic-venues-kremlin-surveillance

Stewart, L. G., Arif, A., & Starbird, K. (2018, Feb). Examining Trolls and Polarization with a Retweet Network. MIS2: Misinformation and Misbehavior Mining on the Web Workshop Proceedings. Marina Del Rey, CA. Retrieved online from https://faculty.washington.edu/kstarbi/examining-trolls-polarization.pdf

Subbotovska, I. (2015, May). Russia's online trolling campaign is now in overdrive. *Business Insider*. Retrieved from http://www.businessinsider.com/russias-online-trolling-campaign-is-now-in- overdrive-2015-5

The Jamestown Foundation. (2000, Sept 15). Putin Signs Information Security Doctrine. *Eurasia Daily Monitor* 6, no. 171. http://www.jamestown.org/single/?tx_ttnews[tt_news]=22353&tx_ttnews[backPid]=214&no_cache=1.

Thomas, T. (2004). Russia's Reflexive Control Theory and the Military. *Journal of Slavic Military Studies* 17: 237–256.

Wagner, B. (2012). Exporting Censorship and Surveillance Technology. The Hague, Netherlands: Humanist Institute for Co-operation with Developing Countries (Hivos). Retrieved from https://hivos.org/sites/default/files/exporting_censorship_and_surveillance_technology_by_ben_wagner.pdf

Walker, S. (2014, April 2). Founder of Vkontakte Leaves after Dispute with Kremlin-Linked Owners. *The Guardian*. Retrieved from https://www.theguardian.com/media/2014/apr/02/founder-pavel-durov-leaves-russian-social-network-site-vkontakte

Whittaker, Z. (2014, April 24). Facebook, Gmail, Skype Face Russia Ban under 'Anti-Terror' Data Snooping Plan. *ZDNet*. Retrieved from http://www.zdnet.com/article/facebook-gmail-skype-face-russia-ban-under-anti-terror-data-snooping-plan/

Woolley, S. C., & Howard, P. N. (2017). Computational propaganda worldwide: Executive summary. Computational Propaganda Project Working Paper No. 2017.11. Oxford, England: Oxford Internet Institute. Retrieved from http://275rzy1ul4252pt1hv2dqyuf.wpengine.netdna-cdn.com/wp-content/uploads/2017/07/Casestudies-ExecutiveSummary-1.pdf

Wright, N. (2018, July 10). How Artificial Intelligence Will Reshape the Global Order. *Foreign Affairs*. Retrieved from https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order

York, J. C. (2015, April 16). The Surveillance Marketplace. *openDemocracy*. Retrieved from http://www.opendemocracy.net/opensecurity/jillian-c-york/surveillance-marketplace.

Zakaria, F. (1997, Nov–Dec). The rise of illiberal democracy. *Foreign Affairs,* 22–43. Retrieved from https://www.foreignaffairs.com/articles/1997-11-01/rise-illiberal-democracy

Zasoursky, I. (2003). *Media and Power in Post-Soviet Russia*. Armonk, N.Y.: Routledge.

Zayani, M. (2011). Social Media and the Reconfiguration of Political Action in Revolutionary Tunisia. *Democracy & Society* 8, no. 2: 2–4.

PART III. EXPORT & EMULATION OF THE MODELS IN GLOBAL COMPETITION

Chapter 9. Understanding the Global Ramifications of China's Information Controls Model

Valentin Weber
Centre for Technology and Global Affairs, University of Oxford
valentin.weber@cybersecurity.ox.ac.uk

Abstract

China's system of online information controls has its roots in the early 1990s. Then it consisted of rudimentary restrictions on domestic internet users. It has since matured into a high-tech model that has diffused to countries far abroad. The model is being exported by the government, state-owned companies, and private companies that make up China's security-industrial complex. It has been successful in Africa, Asia, the Middle East, and South America. If the US wants to maintain a strategic advantage in regions where it is challenged by China's construction of internet infrastructure and the installation of filtering/surveillance technology, then it requires a global view of the underlying agents that drive exports. This will allow the US to tailor policies that counter the diffusion of information controls.

What is China's Information Controls Model?

The Chinese system distinguishes itself from other models, such as the Russian system, through its dependence on highly sophisticated filtering and surveillance technology (Weber, 2017).[29] A combination of human and automated filtering prevents information deemed harmful to regime security from spreading. Private companies play a crucial part in censorship. Bytedance, a content provider, employs 6,000 censors (Chin, 2018). At Weibo, China's Twitter, a largely automated system of censoring deletes 30% of contentious posts within 5-30 minutes and 90% within 24 hours of posts being submitted (Zhu, Phipps, Pridgen, Crandall, & Wallach, 2013).

A sophisticated restrictive legal and regulatory framework underpins this pervasive technical filtering and blocking. Real name registration has been incrementally implemented starting in 2012. Most recently, a new regulation has been announced that will extend the police's authority to inspect internet service providers (Gan, 2018).

In addition to filtering and regulation the government relies on the steering of discourse in online forums. 2 million Chinese citizens (also known as the 50 Cent Party) are tasked with posting positive comments about the Communist Party (King, Pan, & Roberts, 2017).

China's control state is based on physical surveillance infrastructure. It encompasses surveillance equipment at key internet chokepoints, 200 million surveillance cameras (Grenoble, 2017), and also

[29] "In essence, China filters the information as it is posted whereas Russia tries to scare people from posting offending material in the first place, as well as overwhelming any information that evades the chilling effect." (Weber, 2017)

hundreds of millions of citizens' mobile phones that the state has access to through private companies and law enforcement. Mobile phones are perhaps the most effective tool for surveillance purposes, as they are with the owner at all times and include a whole ecosystem of apps that add everyday actions of citizens into a social credit system (Karsten & Darrell, 2018).

How is it Being Exported?

The Chinese model of managing information is being exported via three entities: state agencies, state-owned companies, and private companies.

State agencies: A plethora of government agencies and individuals play a role in disseminating Chinese practices. Beijing sent the PLA's intelligence division to train Sri Lankan officials on how to filter websites (Sirimanna, 2010). China's Ministry of Public Security aided the Cambodian National Police to install surveillance cameras in Phnom Penh (Xinhua, 2015). China also sent high-level individuals to Russia, such as Fang Binxing, who is known as the father of the Great Firewall of China, to share practices (Soldatov & Borogan, 2016). In another development, Beijing has been active in socializing journalists from abroad into Chinese information controls techniques. The China Public Diplomacy Association, which formalizes media fellowships and cooperations, states that the goal is to train "500 Latin American and Caribbean journalists over five years, and 1,000 African journalists a year by 2020" (Lim & Bergin, 2018).

State-owned companies: These play another important part in the export of technology. The China National Electronics Import & Export Corporation (CEIEC) for instance handles national security projects abroad. Its areas of activity include border, public, and cybersecurity thereby effectively doing "real-time monitoring, comprehensive analysis and emergency response to borders, cities and Internet space (CEIEC, n.d.). A project that CEIEC managed was the Integrated Monitoring and Assistance System in Venezuela, which encompassed 30,000 new surveillance cameras for the country. The deal was valued at US $1.2 billion (Mallett-Outtrim, 2013). CEIEC was also involved in the implementation of the ECU911 Integrated Security Service in Ecuador, where it provided hardware for facial recognition technology (Mai, 2018).

Private companies: The most crucial element of China's security export model are private companies. Those include Huawei, ZTE, and Tencent. Even though those companies are designated as private it is necessary to mention that most large technology companies have Communist party committees on their decision-making levels (Feng, 2017). The private sector's exports have focused mostly on filtering and surveillance technology, such as projects in Iran, Zambia, and Zimbabwe (Freedom House, 2013; Reporters Without Borders, 2007; Stecklow, Fassihi, & Chao, 2011), and Kazakhstan and Uzbekistan, where China has been updating ageing Russian SORM equipment (Privacy International, 2014).[30]

More recently, the security portfolio has expanded to include surveillance cameras with facial recognition capacities to enable the "safe city" project in more than 100 cities worldwide (Muthethya, 2016), equipment to combat cybercrime (Xinhua, 2015), and a national identity card for Venezuela (Berwick, 2018). An important take away of this increasing variety of products being exported is that what is seen in China today in security terms will be seen globally very soon thereafter.

Backdoors are to be expected in Chinese built internet infrastructure. The African Union's

[30] SORM is a technical surveillance system developed by Russia.

Headquarters for instance have been the target of a Chinese espionage campaign. For years, information was shared in an unauthorized manner from Addis Ababa to the Chinese mainland. Once the campaign had been revealed, the African Union bought its own computer servers and implemented encryption (Statt, 2018). Similar concerns remain relevant as China is poised to build the Economic Community of West African States' headquarters in Nigeria (Campbell, 2018).

In sum, the Chinese model is driven by an ecosystem of state and non-state actors, which I label the *security-industrial complex* (Weber, 2018). The different actors have varying incentives: the Chinese state and its state-owned companies wish to promote the Chinese vision of handling information, while private companies are driven by profit maximization and are hence the most fervent actors creating new export markets.

Why is this Model Successfully Exported?

The Chinese information controls model has proliferated to Africa, Asia, the Middle East, and South America. It is successful for several reasons.

Firstly, China has demonstrated that online information controls and high economic growth rates are not mutually exclusive. Countries across the world have taken notice of it and have started emulating the Chinese model. Thailand, for instance, laid out plans that would create its own Great Firewall in the image of China's (Bernard, 2015).

Secondly, China has the technology to provide and maintain security equipment abroad. China has a large domestic demand for surveillance and security gear. This includes the Ministry of Public Security, which is always avid for new law enforcement technology acquisitions (Cadell & Li, 2018). The domestic demand allows companies to mature at home and become competitive for the global market. A large part of the export is not centrally led. It is rather driven by the private sector and the goal of profit maximization, which makes resource allocation more efficient. This sophisticated technology based on an innovative private sector gives China also a competitive edge over Russia's model in global markets (Weber, 2017).

Thirdly, and most crucially, China caters to regimes' aspirations for security. Information has become an essential resource to the extent that its free flow cannot remain unchecked (Powers & Jablonski, 2015). Internet security concerns of autocratic regimes were reinforced by the 2009 Iranian election protests as well as the Arab Spring, both of which were largely organised via social media (Hempel, 2016; Landler & Stelter, 2009).

What to Do About It

If the US aims to retain a strategic advantage in countries in which it vies for influence with China, it must:

- Have a good understanding of who the public and private entities driving China's export model are;
- Engage China on its multipronged export strategy – at the state and private sector level;
- Devise a comprehensive strategy to slow down the diffusion of information controls-related hardware, software, and ideas. This may include sanctions against Chinese companies whose information controls hardware or software have been tied to human rights abuses. On the

ideas level, the US should establish media fellowship programs that foster critical reporting on events and expose journalists to a democratic media landscape;

- Encourage non-Chinese built infrastructure by developing alternative funding for new internet infrastructure projects abroad.

References

Bernard, D. (2015, September 28). Thailand Set to Build China-like Internet Firewall. Retrieved 11 August 2017, from https://www.voanews.com/a/thailand-set-to-build-china-like-internet-firewall/2982650.html

Berwick, A. (2018, November 14). A New Venezuelan ID, Created with China's ZTE, Tracks Citizen Behavior. Retrieved 17 November 2018, from https://www.reuters.com/investigates/special-report/venezuela-zte/

Cadell, C., & Li, P. (2018, May 30). At Beijing Security Fair, an Arms Race for Surveillance Tech. *Reuters*. Retrieved from https://www.reuters.com/article/us-china-monitoring-tech-insight/at-beijing-security-fair-an-arms-race-for-surveillance-tech-idUSKCN1IV0OY

Campbell, J. (2018, April 11). China to Build New ECOWAS Headquarters in Abuja. Retrieved 17 November 2018, from https://www.cfr.org/blog/china-build-new-ecowas-headquarters-abuja

CEIEC. (n.d.). Public Safety. Retrieved 18 October 2018, from http://www.ceiec.com/solution/publicsafety

Chin, J. (2018, April 11). New Target for China's Censors: Content Driven by Artificial Intelligence. *Wall Street Journal*. Retrieved from https://www.wsj.com/articles/new-target-for-chinas-censors-content-driven-by-artificial-intelligence-1523446234

Feng, E. (2017, October 10). Chinese Tech Groups Display Closer Ties with Communist Party. Retrieved 28 December 2017, from https://www.ft.com/content/6bc839c0-ace6-11e7-aab9-abaa44b1e130

Freedom House. (2013, March 7). China Media Bulletin: Issue No. 82. Retrieved 9 August 2017, from https://freedomhouse.org/china-media/china-media-bulletin-issue-no-82#5

Gan, N. (2018, October 5). Chinese Police Get Power to Inspect Internet Service Providers. *South China Morning Post*. Retrieved from https://www.scmp.com/news/china/politics/article/2167240/chinese-police-get-power-inspect-internet-service-providers

Grenoble, R. (2017, December 12). Welcome to the Surveillance State: China's Ai Cameras See All. *HuffPost*. Retrieved from http://www.huffingtonpost.com/entry/china-surveillance-camera-big-brother_us_5a2ff4dfe4b01598ac484acc

Hempel, J. (2016, January 26). Social Media Made the Arab Spring, But Couldn't Save It. *WIRED*. Retrieved from https://www.wired.com/2016/01/social-media-made-the-arab-spring-but-couldnt-save-it/

Karsten, J., & Darrell, W. M. (2018, June 18). China's Social Credit System Spreads to More Daily Transactions. Retrieved 18 October 2018, from https://www.brookings.edu/blog/techtank/2018/06/18/chinas-social-credit-system-spreads-to-more-daily-transactions/

King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *American Political Science Review*, *111*(03), 484–501. https://doi.org/10.1017/S0003055417000144

Landler, M., & Stelter, B. (2009, June 16). Washington Taps Into a Potent New Force in Diplomacy. *The New York Times*. Retrieved from https://www.nytimes.com/2009/06/17/world/middleeast/17media.html

Lim, L., & Bergin, J. (2018, December 7). Inside China's Audacious Plan for Global Media Dominance. *The Guardian*. Retrieved from https://www.theguardian.com/news/2018/dec/07/china-plan-for-global-media-dominance-propaganda-xi-jinping

Mai, J. (2018, January 22). Ecuador Is Fighting Crime Using Chinese Surveillance Technology. *South China Morning Post*. Retrieved from https://www.scmp.com/news/china/diplomacy-defence/article/2129912/ecuador-fighting-crime-using-chinese-surveillance

Mallett-Outtrim, R. (2013, November 27). 30,000 More Security Cameras and 17,000 Less Guns on Venezuelan Streets. *Venezuelanalysis.Com*. Retrieved from https://venezuelanalysis.com/news/10198

Muthethya, E. (2016, November 4). New Vision for Big Data: Safe Cities. *China Daily Europe*. Retrieved from http://europe.chinadaily.com.cn/epaper/2016-11/04/content_27269942.htm

Powers, S. M., & Jablonski, M. (2015). *The Real Cyber War: The Political Economy of Internet Freedom*. Urbana, Chicago, and Springfield: University of Illinois Press.

Privacy International. (2014). Private Interests: Monitoring Central Asia.

Reporters Without Borders. (2007, August 6). All Communications Can Now Be Intercepted under New Law Signed by Mugabe. Retrieved 9 August 2017, from https://rsf.org/en/news/all-communications-can-now-be-intercepted-under-new-law-signed-mugabe

Sirimanna, B. (2010, February 14). Chinese Here for Cyber Censorship. *The Sunday Times*. Retrieved from https://web.archive.org/web/20100215081800/www.sundaytimes.lk/100214/News/nws_02.html

Soldatov, A., & Borogan, I. (2016, November 29). Putin Brings China's Great Firewall to Russia in Cybersecurity Pact. *The Guardian*. Retrieved from

https://www.theguardian.com/world/2016/nov/29/putin-china-internet-great-firewall-russia-cybersecurity-pact

Statt, N. (2018, January 29). China Denies Claims It Built Backdoors into African Union's Headquarters for Spying. *The Verge*. Retrieved from https://www.theverge.com/2018/1/29/16946802/china-african-union-spying-hq-cybersecurity-computers-backdoors-espionage

Stecklow, S., Fassihi, F., & Chao, L. (2011, October 27). Chinese Tech Giant Aids Iran. *Wall Street Journal, Eastern Edition; New York, N.Y.* Retrieved from https://www.wsj.com/articles/SB10001424052970204644504576651503577823210

Weber, V. (2017, December 12). Why China's Internet Censorship Model will Prevail over Russia's. Retrieved from https://www.cfr.org/blog/why-chinas-internet-censorship-model-will-prevail-over-russias

Weber, V. (2018, July 17). The Rise of China's Security-Industrial Complex. Retrieved 8 August 2018, from https://www.cfr.org/blog/rise-chinas-security-industrial-complex

Xinhua. (2015, December 22). China Donates Traffic Cameras, Anti-Cybercrime Equipment to Cambodia - People's Daily Online. *Xinhua*. Retrieved from http://en.people.cn/n/2015/1222/c90000-8994022.html

Zhu, T., Phipps, D., Pridgen, A., Crandall, J. R., & Wallach, D. S. (2013). The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions. Presented at the USENIX Security Symposium, Washington, D.C. Retrieved from https://www.usenix.org/conference/usenixsecurity13/technical-sessions/paper/zhu

Chapter 10. Pathways to Lead in Artificial Intelligence

Laura Steckman
The MITRE Corporation
lsteckman@mitre.org

Abstract

China, in establishing multiple national policies and plans to become the world's technology leader for AI and other emerging technologies, has developed a dual-pronged approach to cementing that status. Its two primary pathways for establishing leadership include establishing partnerships with nations, organizations, and other entities that demonstrate AI talent and on globally exporting its domestically-developed AI technologies. While these approaches further China's goals, they raise questions for countries that have different political and social structures. Many countries remain wary about China's potential to use technology to pursue its own social goals, such as shaping societal impacts in ways that contradict soverieign national values and norms, or more profoundly, asserting control through mechanisms of technological authoritarianism.

Introduction

China intends to become the world leader in AI by 2030. Its vision places the country at the pinnacle of the discipline and as a primary driver of AI development internationally. It also avails the reaping of economic benefits from the technology, including providing an opportunity to change the country's reputation as a mass exporter of cheap goods to one that focuses on the development and sale of high-quality technical products. This goal to lead the world in technological research and development is part of President Xi Jinping's Chinese dream, which includes China's evolution to a world-class technology superpower (Chan & Lee, 2018). The strategy takes a long-term approach to achieve real and perceived technological superiority. Numerous documents set out government plans to support indigenous development of AI.[31] China has also invested heavily in AI and sought external investors, making it the recipient of more than half of the world's investments in AI between 2013-2018 (Burrows, 2018). As the publication of these policy documents demonstrates, China aims to become a strong backer behind AI-technologies for its domestic sector and seeks recognition as a technical global leader. They also demonstrate that China is already several years into its plan to dominate AI and other technologies, which may have contributed to its successful ability to align with talent and attract AI-related investments.

Along these lines and in accordance with its national plans, China has identified two pathways to promote itself as an international AI leader.

- One is to place itself at the forefront of developing an international AI community of practice, working in cooperation with other nations, universities, and AI talent.
- The other path is to become a primary exporter of AI technologies around the globe, increasing China's reputation for technical development.

[31] 2015 Made in China 2025; 2016 Three-Year Guidance for Internet Plus Artificial Intelligence; 2017 New Generation Artificial Intelligence Development Plan; 2018 Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry; and the 2018 White Paper on Artificial Intelligence Standardization.

These paths are not mutually exclusive. Whether together or separately, they allow China to use technology for the dual purposes of influence and economic gain. Their potential ramifications include social and cultural impacts, data and intellectual privacy concerns, and questions surrounding how much Chinese ideology will transfer between it and its customers. AI is an emerging technology with the potential to enable the direct and indirect spread of soft power, reshaping the world through the lenses of the technology's developer.

Both pathways form part of an engagement model designed to further China's One Belt One Road (OBOR) intiaitve. OBOR seeks to construct a modern-day Silk Road trade network connecting Asia, Africa, and Europe. This economic model offers China the opportunity to transform its reputation as a manufacturing-based society that exports cheap goods abroad into one of cutting-edge technical leadership based on technological innovation. The country's drive for partners and its desire to export AI technologies could increase the reach of OBOR physically and virtually. The pathways also demonstrate China's seriousness in becoming the world leader in AI, in that its focus is not solely on domestic research and development but also on outreach, partnerships, and export opportunities that advance the country's objectives and reposition it in terms of global technical leadership.

The First Pathway: Develop AI together—Outreach and Partnerships

China's vision to become a world AI leader requires it to become central to everything AI. China espouses that technological development, to include advances in AI, will be more rapid when performed in partnership with the possibility to "advance together." To this end, China developed its first pathway approach to increase its visibility as an international AI leader: it actively seeks partners and consortia to cooperate and collaborate on AI research and development. Its partnerships include other governments, academics, companies, and most recently, with bodies that collectively develop international standards for AI development, such as the US-led Partnership on AI that has most recently accepted its first Chinese firm as a member.

Vice-Premier Liu He announced China's intention to seek international partnerships to develop AI as a field at the 2018 World Artificial Intelligence Conference in Shanghai. This approach allows Chinese companies access to a large volume of data, while partnerships in themselves tend to advance capabilities faster than one organization acting alone by pooling talent and resources. It also paves the way for establishing and strengthening ties with other entities, such as those invested in or on the periphery of OBOR.

During that same 2018 conference, SenseTime, a Chinese firm dubbed the world's most valuable AI start-up, launched a 15-university consortium to connect talent from primarily Chinese schools, but also from American and Singaporean ones (Ping, 2018). The timing of the consortium's launch underscored China's announcement, synchronizing Vice-Premier Liu He's speech with an action that further demonstrated the country's commitment to partnering on AI development.

China also seeks to partner with other national governments. It has already established bi-lateral national and corporate partnerships worldwide and continues to pursue new opportunities. In March 2018, China's CloudWalk Technology entered into a strategic partnership with the Zimbabwean government. CloudWalk develops facial recognition technology for multi-purpose use, and the company actively seeks business opportunities across Africa. According to Chutel (2018), CloudWalk is only one example of a Chinese AI start-up that could gain access to data cheaply in ways that circumvent the ethical and legal concerns that other nations impose on AI projects. The country has, of course, not limited its partnerships to Africa and the other regions of interest for OBOR. It also has partnerships in Latin America where it has already exported AI-enabled technologies.

China's AI outreach and partnerships, as shown by the country's current efforts, support its global ambitions for becoming the center of AI development.

The Second Pathway: A Reputation Shift – Exporting AI

In conjunction with developing partnerships, China has a second approach to bolster its credibility as a world AI leader that intertwines with its building partnerships model. The second approach, or pathway, concentrates on the country becoming the central hub for developing and exporting cutting-edge AI-related technologies. To achieve this goal, China will continue to invest heavily in AI development and market its technologies abroad. In doing so, it initiates a paradigm shift: China will change its reputation as a mass exporter of cheap goods and technologies to one that offers high-quality, high-tech products and services. The transformation started with the vision to make China the global AI leader by 2030; its progress is attached to how quickly it can develop and export viable AI technologies that improve society and advance AI-related research.

China has started exporting its AI technologies through its partnerships and businesses. It continues to seek new opportunities to develop and export its new technologies. The full impact of its AI exports is not yet known; as with any new technology, the global impact could be positive and/or negative. On one hand, the Artificial Intelligence and Life in 2030 report provides an optimistic view of the benefits of AI globally (Stone et al., 2016).[32] China desires to be the purveyor of these benefits and societal transformations.

Taking another perspective, researchers focused on international technological trends and development have provided a China-specific view on AI. For example, Wright (2018) asserts that Chinese exports may contribute to the spread of digital authoritarianism, while Benaim & Gilman (2018) provide a more micro, AI-centric perspective that they believe will result in the spread of algorithmic authoritarianism. The discussion of algorithmic authoritarianism, for instance, involves Chinese export of algorithms that enable nations to oppress domestic and, potentially, foreign populations. We must examine China's ambitions fully.[33] Such ambitions may go beyond coding, algorithms, and the digital realm with a longer-term projection to include additional nascent, newly-emerging technologies and platforms that will incorporate AI, many of which could be used to exert influence, access data without privacy restrictions, and spread an ideology that permits governments to have complete control over populations and patterns-of-life—restricting movement, expression, thinking, and decision making. The consequences of the country's push to dominate AI and other advanced technologies are already under way; China's government and corporations already export AI technologies, and their strategies focus on increasing total AI exports.

Examples of Chinese corporations focused on exporting AI are numerous. Yiwue launched in 2014 and released Squirrel AI, a technology intended to transform education, in 2017. The program identifies weaknesses in a student's knowledge and provides enhancements in nano chunks to improve learning. Squirrel AI operates across China in more than 700 schools in 100 cities (Yixue, 2015). The company has labs in the United States and now actively seeks international partnerships

---

[32] Note that the report mentions China only once to indicate that it only entered the world economy in the past fifteen years, a claim that is unclear in terms of AI or employment, the report section containing the reference.

[33] China's "Made in China 2025" strategy indicates that it wants to dominate many technologies beyond Artificial Intelligence. This strategy asserts China wants to lead in multiple technologies, such as alternative energy-fueled vehicles, emerging IT and telecommunications , advanced robotics, aerospace engineering, biomedicine, new synthetic materials, and other advanced equipment and infrastructure. Through the development and sale of these technologies, particularly when they incorporate digital elements, the Chinese partnership and export model may precipitate the spread of its ideology, worldview, and values.

for expansion and export. Technologies exported globally potentially allow Chinese companies a high degree of control over their users' experience. In the case of education, questions about who sets and approves the curricula are key, otherwise a company could have sole discretion over what information a student learns, with implications that could alter or curtail course materials and change how people perceive the world.

One of China's major AI successes is SenseTime, China's largest AI start-up, valued at more than $1.5 billion (Bloomberg, 2018). It has offices throughout China, Hong Kong, and Japan. Recently, it entered into agreements with companies in Singapore to increase its access to the Asian and Southeast Asian markets. It also recently signed an agreement with Qualcomm, headquartered in the United States, to provide such AI-enabled solutions as facial recognition technology across the world (Dai, 2018). The export of facial recognition technology worldwide is an area in which China has had success; not only has it exported its technology, but it has also donated the technology to governments across the globe.

China's emphasis on exporting AI is a natural outgrowth of its investment in the technology and its need to increase demand and keep profits high for continued reinvestment. Exporting the technologies is also necessary to demonstrate, gain, and maintain a position of technical leadership, which is one of China's professed objectives. This approach raises questions about the effects the exported technologies could have on other people and societies, particularly since few populations view the world similarly to China.

What is at Stake: Societal Freedom versus Control through AI Implementation

Who or what controls AI from a research and development perspective is not a trivial matter. China's ambition to take the lead on AI internationally poses some challenges for non-authoritarian societies. Specifically, it leads to tensions between societal freedom and social control. China's current pathways, the outreach/partnership and export modes, position China as an active player in the AI field. That position affords China the opportunity to shape the solutions used by other societies, which means the decisions it makes during the AI development process can affect these societies. There are four areas where social freedom versus control stand out as primary concerns.

First, one area on which China has set its sights is developing the standards for AI. Presently, there are no accepted, standardized international guidelines for AI-related ethical, legal, and privacy questions. As they are developed, the key players will have the ability to shape the standards in ways that may align more with the shapers' values and goals than the recipients' needs. Reflecting back to ideas such as digital authoritarianism, technology and its subcomponents may be used to influence, exploit data, and privilege a foreign ideology to exert control over one or more populations. China may use opportunities to create standards and norms to insert its ideology and values into the technology in ways that affect unwitting consumers.

Second is restricting the free flow of information. Bytedance, a Chinese news aggregation platform that integrates AI, announced in April 2018 that it has plans to export its technology to market around the world. Zhang Timing, the founder and CEO, explained that this strategy will keep the company current with globalization (TMTPost, 2018). Exporting AI technology attached to news raises the question of whether the algorithms will contain biases in favor of the Chinese or the same censorship standards used in China. Around the world, AI researchers currently acknowledge that unfairly trained AI or imbalanced data sets can create bias in technologies using AI; the issue of data restriction is not limited to China, and it also extends to tech companies. Cultivated news sources can use conscious and unconscious biases to overemphasize one worldview over another and restrict

counterviewpoints that promote critical thinking and healthy public debate. The same challenges may be apparent in AI-enhanced education, which may privilege some perspectives over others, especially in the social sciences and humanities.

Third, there is an inherent risk for any country that adopts a technology developed on another's values. In the case of China, its national values, which will likely be embedded into algorithm development and potentially in the selected training data, may not reflect those of a nation that imports and implements its technology. To mitigate potential unforeseen effects, indigenous AI experts would have a role in developing China's technical solutions or, at minimum, would perform critical reviews of the imported technology. Without such a review, the potential exists that the technology will not meet the nation's needs or will introduce errors that contradict its cultural norms. The country that exports such technologies could, therefore, have the ability to shape the recipients of its technical solution at the societal level, depending on how that recipient adopts and applies the solution. This issue is, of course, not limited to China; any country or company that develops algorithms has the potential to insert values into a technology that could influence the end user.

Fourth, data privacy, particularly corporate data that constitutes intellectual property, is another potential area of concern. China's dealings with foreign companies, particularly those that operate within its borders, often require that the companies' inventions and products undergo government review. Reviewers can gain insights into the companies' intellectual property through extensive review of the data. The AI partnerships likely include similar requirements, which will give China access to significant amounts of personal data through its partnerships and exports. Some agreements may allow the Chinese access to private data without the affected populations' consent.

Overall, the implementation of AI will have unknown effects and consequences in multiple societal sectors. While this is true for AI regardless of who develops the technology, China's approach may also create more socioeconomic issues because it may be released too early. Lee Kai-fu, chairman of Sinovation Ventures and former head of Google China, stated that China has a "techno-utilitarian" approach in which "[t]he government is willing to let technology launch, to see how it goes, and then rein it back if needed" (Dai & Shen, 2018). A strategy of launch-and-see could turn populations into experimental test beds for new technologies without their knowledge or consent.

Conclusion

The AI-enabled future is still a vision for most of the world, though the AI evolution has already begun and will continue to gain momentum. In fact, many scholars, politicians, and developers expect that AI will transform the globe on multiple fronts to advance humanity. The extent to which these changes promote positive transformations is likely contigent on who—or what—leads the field. China's ambitions include global leadership of AI and other related technologies that could revolutionize human society. Its two pathways to leadership, the outreach/partnership and export models, provide it a dual-pronged approach to gain access to the world's cutting-edge researchers to develop AI faster, and the ability to export its internally-developed technologies, whether developed entirely domestically or in collaboration with partners, to (re)shape the world through AI. In the process, China may influence educational curricula, set international standards for AI, selectively highlight or impede the spread of news and other information, gain access to extensive personal data, and use the technologies to disseminate its ideological perspective. The potential outcome of China's role and influence on AI leads to questions about the benefits and the consequences. While these questions apply to other nations and technology companies experimenting with AI, China's momentum to attract investiments, its focus on indigenous AI development, and its desire to be the forerunner for AI and related technologies demonstrate its commitment to be an international AI

player. As China advances its China dream, it invests and markets itself heavily as a world-class AI and technology leader. Whether it achieves its ultimate objective, China's current dual-pronged approach to immerse the world in Chinese-made, high-tech AI solutions will undoubtedly leave some long-lasting marks and lingering questions.

References

Benaim, D., & Gilman, H. R. (2018, August 9). China's Aggressive Surveillance Technology Will Spread Beyond Its Borders. Retrieved from https://slate.com/technology/2018/08/chinas-export-of-cutting-edge-surveillance-and-facial-recognition-technology-will-empower-authoritarians-worldwide.html

Bloomberg. (2018, April 8). China Now Has the Most Valuable AI Startup in the World. Retrieved from https://www.bloomberg.com/news/articles/2018-04-09/sensetime-snags-alibaba-funding-at-a-record-3-billion-valuation

Burrows, I. (2018, October 5). Made in China 2025: XiJinping's plan to turn China into the Al world leader. Retrieved from https://www.abc.net.au/news/2018-10-06/china-plans-to-become-ai-world-leader/10332614

Chan, E., & Lee, A. (2018, September 25). 'Made in China 2025': is Beijing's plan for hi-tech dominance as big a threat as the West thinks it is? Retrieved from https://www.scmp.com/business/china-business/article/2163601/made-china-2025-beijings-plan-hi-tech-dominance-big-threat

Chutel, L. (2018, May 25). China is exporting facial recognition software to Africa, expanding its vast database. Retrieved from https://qz.com/africa/1287675/china-is-exporting-facial-recognition-to-africa-ensuring-ai-dominance-through-diversity/

Dai, S. (2018, July 3). Tech start-ups push to make China's facial recognition systems part of daily life across Asia. Retrieved from https://www.scmp.com/tech/start-ups/article/2153471/tech-start-ups-push-make-chinas-facial-recognition-systems-part-daily+&cd=4&hl=en&ct=clnk&gl=us

Dai, S., & Shen, A. (2018, October 2). 'Made in China 2025': China has a competitive AI game plan but success will need cooperation. Retrieved from https://www.scmp.com/tech/article/2166177/made-china-2025-china-has-competitive-ai-game-plan-success-will-need

Ping, C. K. (2018, September 18). China wants to work with other countries to develop AI. Retrieved from https://www.straitstimes.com/asia/china-wants-to-work-with-other-countries-to-develop-ai

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., … Teller, A. (2016). Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence: Report of the 2015 Study Panel, 52. https://doi.org/https://ai100.stanford.edu

TMTPost. (2018, April 25). Zhang Viming's globalization strategy: export Al technology & make localized operation. Retrieved from https://medium.com/@TMTpost/zhang-yimings-globalization-strategy-export-ai-technology-make-localized-operation-517 d7 4dcf8f4

Wright, N. (2018). How Artificial Intelligence Will Reshape the Global Order. Foreign Affairs. Retrieved from https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order

Yixue. (2015). Yixue Squirrel AI's current scale. Retrieved from http://www.squirrelai.us/school.html

Chapter 11. The Spread of Russia's Digital Authoritarianism

Robert Morgus
New America
morgus@newamerica.org

Abstract

Russian digital authoritarianism is characterized by pervasive communications collection, absent oversight, and government cooption of industry—particularly internet service providers (ISPs)—to do their bidding. Although Russia's digital authoritarianism is neither as well defined nor as technologically robust or reliant on artificial intelligence as the Chinese model, the Russian government nonetheless takes actions that promote digital authoritarianism globally and diffuse their model and technology in their near abroad. This chapter explains how Russia is exporting or encouraging emulation of models of digital authoritarianism around the world, through diplomatic, informational, and economic means.

Key points:

- Russia's digital authoritarianism is an appealing, relatively low-tech alternative to the connected and "smart" Chinese model.
- Russian companies have had limited success exporting the technology underpinning Russian digital authoritarianism outside of its near abroad in Belarus and Central Asia, but the Russian state does take action to promote digital authoritarianism around the world more generally.

Introduction

Weber (2017) argues that the Chinese model for digital authoritarianism will "prevail" over Russia's. Nonetheless, the Russian system for internet control offers a plausible, lower-tech alternative to the tech-heavy Chinese approach. In the context of the diffusion of the Russian model for digital authoritarianism, it is important to understand three observations about the system.

First, rather than architecting internet infrastructure to filter massive amounts of content, the Russian system is more reliant on a combination of self-censorship and intimidation underpinned by complex, but ultimately highly restrictive speech and expression laws and pervasive and overt telecommunications surveillance. As with the Chinese system for internet control, the Russian state leans heavily on the private sector—particularly the ISPs—in the country to implement their filtering and surveillance policies through the SORM system, which is explained in greater detail below. In essence, where China's digital authoritarianism increasingly relies on code to enforce its authoritarian approach, the Russia model relies on law.

Second, the Russian state produces information in high volume and velocity. Both China and Russia spend resources on internal propaganda. However, the Russian system, absent pervasive censorship, relies heavily on information manipulation. In other words, rather than tightly limiting the supply of information on the internet, the Russian state seeks to inundate the information market with pro-government propaganda, drowning out negative press.

Third, Russia lacks the high-tech industrial robustness characteristic of Chinese industry. In practical terms, this hampers Russian industry's ability to engage in foreign markets outside of its near abroad,

though Russian surveillance companies have been able to make limited inroads in authoritarian-leaning parts of the world, as explained in greater detail below. Nonetheless, Russia's ability to spread authoritarianism via trade is minute in comparison to China's.

What these three observations mean in practical terms is that Russia possesses limited capacity to "export" its model for digital authoritarianism outside of its near abroad in the traditional economic sense. However, it is adept at utilizing information and diplomacy to further its goals. Here, I provide a brief refresher on the Russian model for digital authoritarianism, explain the mechanisms Russia uses to export its model, briefly discuss the cooperation and competition between Russian and Chinese models, and offer concluding thoughts for U.S. policymakers.

The Russian Model

Russian president Vladimir Putin has long viewed the global internet as an American Central Intelligence Agency (CIA) project (MacAskill, 2014). For Putin, everything from the technical architecture to the governance that underpins it has been carfully constructed in favor of American values and interests. For Putin, it follows logically that it is therefore in Russia's interest to subvert existing architecture and governance structures. In response to these paranoias, Russia has constructed a "red web" in its authoritarian image.

Russian digital authoritarianism displays four notable characteristics:

1. Less technical filtering of content than a comparable Chinese system, but greater reliance on intimidation and social norms around self-censorship, underpinned by robust technical surveillance system on all internet traffic, known colloquially as SORM;
2. Complex, but ultimately highly restrictive speech and expression laws;
3. Corporate capture, particularly state capture of ISPs; and
4. Heavy-handed state manipulation of the market for information domestically.

Because the Russian approach is so reliant on surveillance, its system necessarily involves strict laws and technology to enable law enforcement.

Legal structure: As Human Rights Watch (2017) notes, since 2012 the Russian state has gradually updated the legal system to outlaw extreme speech online. In Russia, extreme speech has been selectively cast to include relatively benign criticism of the government (Human Rights Watch, 2017). Rather than filtering or blocking content, the Russian state relies on ISPs and other providers to comply with a series of laws mandating law enforcement access to their data and servers via SORM-compliant technologies, which allow law enforcement to collect or monitor traffic without the knowledge of the service provider. A set of regulations issued by the Federal Council of Ministers and the Ministry of Communications and Information Technologies codify legal requirements for ISPs to use SORM-compliant technology and install SORM "black boxes" on their networks (Global Legal Monitor, 2016). The SORM network intercepts and stores all internet traffic in Russia. Due to arcane, Soviet Era-legal orders, once law enforcement has court permission to access SORM data, the scope of lawful access under the order is largely unrestricted (Soldatov & Borogan, 2015, p. 78). In addition, as Marecal (2016, p. 33) notes:

> *Surveillance can begin before the warrant is granted (or even requested), the warrant need not be shown to anyone (whether the surveillance target or the telecom operator), and it is only*

> *required for the retrieval of collected communications content, and not for the metadata that is often just as revealing as content, if not more so.*

Technology: To enforce laws, law enforcement requires communcations content and data. In order to scoop up this information, the FSB relies on SORM black boxes which mirror online traffic, sending the original on to its intended destination and a copy of all traffic to FSB owned and operated servers. The FSB then employs technical solutions from a myriad of Russian and non-Russian companies to conduct deep packet inspection, decrypting communications and gaining critical access as needed. A second major technology leveraged by the Russian security services is the Semantic Archive Platform, provided by Analytical Business Solutions, a Russian software developer. The Semantic Archive Platform provides a means for security services to aggregate open-source online data (media, social networks, forums, etc.), process this data, and analyze it (Analytical Business Solutions, Date Unknown). The Semantic Archive Platform utilizes algorithms to both identify and extract key data, as well as to process the data for easier use by operators (Analytical Business Solutions, Date Unknown).

The Russian SORM-3 system allows most information and data to flow through the internet—the exception being data and content from applications and platforms that refuse to provide data access to security services via SORM devices. Because of this relatively free flow, the Russian state therefore engages in widespread pro-state propaganda to flood the market for information and manipulate online narratives.

Exporting Russian Digital Authoritarianism

As Jacklyn Kerr (2018) notes, authoritarian adoption of digital solutions that shape their local information environment is likely driven by:

> *The internationally available Internet control solutions of which a regime is aware, a regime's financial and organizational capacity to implement these or to access assistance in their implementation, and the policies selected by other states in the regime's reference group or endorsed by regional and international organizations with which a state is closely engaged.*

The adoption of digital authoritarian practices is, as such, largely driven by the availability of models and products and the ease with which those products and models fit with existing capacities and legal frameworks. Russian approaches to digital authoritarianism are alluring to countries that have existing legal frameworks with similarities to Russia. The majority of Russian export of authoritarian enabling technology occurs in Russia's own near abroad. As explained below, however, Russia is adept at promoting its model globally via means beyond trade, such as through diplomacy and the strategic use of information. It is important to note that, while some of this activity is clearly a concerted effort on the part of the Russian state, some activities that spread or promote digital authoritarianism likely do so unintentionally.

Diplomacy

Russian president Vladimir Putin has long viewed the global internet as an American Central Intelligence Agency project. For Russia, the global rules governing the internet have been carefully crafted by western powers. It is therefore a primary objective of the Russian state to not only assert Russia's sovereignty over the network within its borders, but to also "make other countries,

especially the United States, accept" this right (Soldatov and Borogan, 2015, p. 223). In pursuit of this goal, the Russian Ministry of Foreign Affairs is one of the leading proponents of sweeping international "cybersecurity" treaties. Often, support for these treaties is driven by desire to allow state entities to reassert "sovereignty" over the information space, legitimizing authoritarian approaches to internet censorship and surveillance. Using events like the rampant misinformation around liberal-democratic elections in Europe and the United States, as well as the Snowden revelations, Russia seeks to blur the line between cybersecurity (computer network defense) and information control in international forums. The linkage of the two issues—and the promotion of this view globally—is a direct threat to the flow of information on the internet and ease the task of authoritarians who seek to manipulate narratives to fulfill their objectives. By linking information control to cybersecurity in global forums, Russia, China, and their Shanghai Cooperation Organization (SCO) partners have long sought to obscure their true intentions through semantics.

Russia uses a myriad of multilateral bodies to advocate for sovereignty, including at the United Nations and with Brazil, Russia, India, & China (the BRICs). At the United Nations, Russia is a consistent participant in UN General Assembly (UNGA) discussions regarding cyber and information security. Since 2011, Russia, along with allies in the SCO have consistently submitted a letter to the UNGA leadership proposing a code of conduct for information security, which seeks to undermine existing human rights and international law by legitimizing authoritarian "sovereignty" over domestic information space (McKune, 2015). In the most recent UNGA (November 2018), Russia proposed a resolution, which was passed, to mandate further discussion about the adoption of a convention to codify this code of conduct (Grigsby, 2018). In addition, the Russian government has sought to bring the internet under the jurisdiction of the UN's International Telecommunications Union, a move that would consolidate governance of the space in the hands of states, a move that many view as impractical and antithetical to liberal values (Soldatov & Borogan, 2015, p. 237).

Russia also seeks to build cooperation in other international forums, like the BRICs, who have the "common strategic intention to reform the global cyberspace governance system," (Wanglai, 2018) and the Commonwealth Security Treaty Organization (Kerr, 2018). In particular, Russia has pushed for a BRICS (including South Africa) undersea cable to link the networks in BRICs countries without routing traffic through the United States (Ozores, 2015).

Information

Russia uses informational means to spread interest in its model for internet and content control in global political circles in several ways. I highlight two examples.

Second, Russia has successfully shown the fragility of the global market for information through rampant online mis- and disinformation operations in the western world. The effects of these operations are not limited to the direct objectives of the operations, like undermining confidence in elections. Indeed, the 2016 interference in the US election and the Brexit vote heightened western policymakers' interest in an increased role for governments in controlling information.[34] As such, Russia gradually legitimizes its own approach to information controls and increases global interest in greater sovereign control over the internet and information space (Morgus, 2016). While sovereignty is not necessarily at odds with liberal and democratic ideals, in parts of the world with

[34] This is best evidenced by the increased public discourse around misinformation and disinformation. However, on more than one occasion this author has experienced western policymakers suggest that we could learn something from the way Russia handles information domestically.

less robust privacy and human rights protections, sovereign control over the internet quickly becomes sovereign control over information. It is unclear whether this was a considered objective on the part of the Russian state or whether it was simply a positive externality (from the Russian perspective) of ongoing information operations.

Trade

According to Kerr (2018), as of early 2018, "at least nine states (in addition to Russia) of the former Soviet Union have emulated aspects of the Russian legal, technical, and institutional approach to electronic surveillance." In many cases, the legal and institutional emulation may be a byproduct of legacy policies adapted for a connected world. However, the technical emulation is undeniably underpinned by SORM compliant technology from both Russian telecommunications and surveillance companies (Bourgelais, 2013).

In the early- to mid-2010s, two companies founded by Russian nationals—Protei and Peter-Service—became the subject of a great deal of attention for the technology they develop and had been providing broadly. Protei, which "produced all kinds of technology from SORM-I to SORM-3" and has openly written about the deployment of an internet filtering systems in Kyrgyzstan, Belarus, Uzbekistan, and others, states that it serves more than 300 customers in 26 countries in "Russia, Eastern Europe, Central Asia, Latin America, the Middle East & North Africa" (Protei, 2018a). Protei's Middle East and North Africa customers include telecommunications companies in Bahrain, Iraq, Jordan, Palestine, Qatar, Sudan, Tunisia, and Yemen (Protei, 2018b). Based on publically available news releases, Protei's Latin American customers are most likely Cuba, Mexico, and Venezuela (Protei, 2017, Protei, 2016a, & Protei, 2016b). Peter-Service, a Russian based company that offers similar technology currently has customers in Belarus, the Abkhazia region of Russia, Georgia, and Ukraine (Peter-Service, 2018). The majority of the technology offered by Protei and Peter-Service enables SORM-like surveillance on local ISPs.

The export of equipment and services does not stop with SORM equipment. Analytical Business Solutions, the Russian company that developed the Semantic Archive Platform, has installed its system in Belarus, Ukraine, and Kazakhstan, in addition to Russia (Analytical Business Solutions, Date Unknown).

Russia and China: Coopetition?

It is unclear whether the diffusion of authoritarian models for internet control are driven by suppliers of technology and governance models, like Russia and China, or demand from third countries. Russian and Chinese industrial and governmental players are therefore influential in the spread of digital authoritarianism. However, there is little beyond normative efforts in multilateral forums to suggest that the two countries work in coordination to push digital authoritarianism.

While the broad goal of legitimizing their own authoritarian approaches and shaping the future internet may be shared, the ways the two countries go about shaping the internet differ both at home and abroad. According to reporting from Soldatov and Borogan (2016), the Russian state brought Chinese experts in an effort to build and configure its own "Great Firewall." Nonetheless, increaslingly Russian companies like Analytic Business Solutions, Peter-Service, and Protei are alluring to authoritarians. However, lawful application of some of these technologies—particularly the SORM technologies—requires tailored regulation similar to the SORM laws in Russia, but countries with Soviet legal legacies are likely to posses the at least the legal foundation to craft such regulations. It follows therefore, that companies from the respective countries that supply the technology to enable

digital authoritarianism are likely in competition with one another, especially in markets that both Russia and China see as strategically important, like the Central Asian Republics.

Conclusions

Today, the global struggle between authoritarianism and liberalism is mirrored in the digital space. Digital authoritarianism offers a viable route to reaping the benefits of a digital society for dictators and despots who were unnerved by the Arab Spring. In the struggle between digital authoritarianism and the alternative—which is currently ill-defined and rife with contradictions (Morgus and Sherman, 2018)—two battlegrounds exist: (1) a group of as-of-yet undecided countries (which I call the *Digital Deciders*) and (2) international legal bodies (Morgus et al., 2018). Today, the global struggle to bring order to the digital space will materially impact the future of the broader global order. In pursuit of American interests, the U.S. Department of Defense (U.S. DOD) should take the following steps.

1. *Work with partner countries to build cybersecurity capacity*. Anecdotal data suggests that the adoption of digital authoritarian practices is driven in part by perceptions of cyber, information, and state insecurity. The U.S. DOD should redouble its efforts to build cybersecurity capacity via military to military engagement. However, the United States' engagement on this front cannot stop with military to military engagement. As such, the U.S. DOD should advocate for increased funding for the U.S. Department of State and U.S. Agency for International Development to invest n building cybersecurity capacity in *Digital Decider* countries (see: Morgus, 2018).

2. *Improve our and our allies' strategic messaging about alternatives to digital authoritarianism*. Today, the foreign policy promoting a free, open, interoperable, secure, and resilient internet is losing (Morgus and Sherman, 2018 and Hohmann and Benner, 2018). In order to provide the *Digital Deciders* with a viable alternative to digital authoritarianism, the United States must lead on developing and messaging around a model that is both viable and compelling in developing markets.

3. *Advocate to limit the export of digital surveillance tools made in the United States and allied countries to authoritarians*. Many American and western companies manufacture digital surveillance tools. In open societies, these tools play a crucial role in maintaining security and, when kept in check by requisite legal and regulatory oversight, provide immense good to society. However, in the wrong hands, these same tools can provide the foundation for digital authoritarianism. The U.S. DOD should advocate for the adoption and strong enforcement of export controls like the "IP network communications surveillance systems" control proposed at the Wassenaar Arrangement in 2013 (Kehl and Morgus, 2014).

Finally, several open, empirical questions need to be answered, either by the national security community or by researchers in the private sector: (1) How much of the export of digital authoritarianism is a concerted effort by Russia and China? (2) How much coordination exists between Russia and China on this issue? (3) To what extent are Russia and China competing in this space and does this competition represent an opportunity for the United States and allies? (4) What is fueling the adoption of digital authoritarianism around the world?

References

Analytic Business Solutions. (Date Unknown). Semantic archive information analysis platform. *RusTrade*. Retrieved from http://www.rustrade.hu/07_Kommercheskie_predlojenia_i_zaprosi/07_01_Predlojenia_ross_export/07_01_01_Tovar/07_01_01_16_Tovari/Semantic%20Archive%20presentation.compressed.pdf

Bourgelais, P. (2013). Commonwealth of surveillance states. *Access Now*. Retrieved from https://www.accessnow.org/cms/assets/uploads/archive/docs/Commonwealth_of_Surveillance_States_ENG_1.pdf

Global Legal Monitor. (2016, March 2). ECHR, Russian Federation: Breaches of human rights in surveillance legislation. *Library of Congress*. Retrieved from http://www.loc.gov/law/foreign-news/article/echr-russian-federation-breaches-of-human-rights-in-surveillance-legislation/

Grigsby, A. (2018, November 15). The United nations doubles its workload on cyber norms, and not everyone is pleased. *Council on Foreign Relations*. Retrieved from https://www.cfr.org/blog/united-nations-doubles-its-workload-cyber-norms-and-not-everyone-pleased

Hohmann, M & Benner, T. (2018, June 28). How European internet foreign policy can compete in a fragmented world. *The Global Public Policy Institute*. Retrieved from https://www.gppi.net/2018/06/28/how-european-internet-foreign-policy-can-compete-in-a-fragmented-world

Human Rights Watch. (2017, July 18). Online and on all fronts: Russia's assault on freedom of expression. *Human Rights Watch*. Retrieved from https://www.hrw.org/report/2017/07/18/online-and-all-fronts/russias-assault-freedom-expression

Kehl, D & Morgus, R. (2014, March 31). The Dictator's little helper. *Slate – Future Tense*. Retrieved from https://slate.com/technology/2014/03/export-controls-how-to-stop-western-companies-from-sending-surveillance-tech-to-dictators.html

Kerr, J. A. (2018). Information, security, and authoritarian stability: Internet policy diffusion and coordination in the former Soviet region. *International Journal of Communication* (12). 3814-3834.

MacAskill, E. (2014, April 24). Putin calls the internet a 'CIA project' renewing fears of web breakup. *The Guardian*. Retrieved from https://www.theguardian.com/world/2014/apr/24/vladimir-putin-web-breakup-internet-cia

Marecel, N. (2016, October 30). Networked authoritarianism and the geopolitics of information: Understanding Russian internet policy. *Media and Communication* (5). 29-41.

McKune, S. (2015). An Analysis of the International Code of Conduct for Information Security. *Citizen Lab*. https://citizenlab.ca/2015/09/international-code-of-conduct/

Morgus, R. (2018, April 23). Securing digital dividends. *New America*. Retrieved from https://www.newamerica.org/cybersecurity-initiative/reports/securing-digital-dividends/

Morgus, R. (2016, December 5). Russia gains an upper hand in the cyber norms debate. *Council on Foreign Relations*. Retrieved from https://www.cfr.org/blog/russia-gains-upper-hand-cyber-norms-debate

Morgus, R. & Sherman, J. (2018, July 26). The Idealized Internet vs. Internet Realities (version 1.0). New America. Retrieved from https://www.newamerica.org/cybersecurity-initiative/reports/idealized-internet-vs-internet-realities/

Morgus, R., Woolbright, J., & Sherman, J. (2018, October 23). The digital deciders. New America. https://www.newamerica.org/cybersecurity-initiative/reports/digital-deciders/

Ozores, P. (2015, October 27). Russia pushes for BRICS undersea cable. *BNamericas*. Retrieved from http://www.bnamericas.com/en/news/privatization/russia-pushes-for-brics-underseas-cable

Peter-Service. (2018). Customers. Peter-Service. Retrieved from https://www.billing.ru/en/customers

Protei. (2018a). PROTEI Company Profile. Protei. Retrieved from http://www.protei.com/company/

Protei. (2018b). MENA Customers. Protei. Retrieved from http://protei.me/Protei-MENA-Customers

Protei. (2017). PROTEI in Havana. Protei Past Events. Retrieved from http://www.protei.com/events/past/

Protei. (2016a). Protei participates in Redknee project for DIGITEL BSS network transformation. Protei News. Retrieved from http://www.protei.com/news/

Protei. (2016a). Mexican MVNE deployed PROTEI HLR/HSS. Protei News. Retrieved from http://www.protei.com/news/

Soldatov, D., & Borogan, I. (2015). The Red Web.

Soldatov, D., & Borogan, I. (2016, November 29). Putin brings China's Great Firewall to Russia in cybersecurity pact. *The Guardian*. Retrieved from https://www.theguardian.com/world/2016/nov/29/putin-china-internet-great-firewall-russia-cybersecurity-pact

Wanglai, G. (2018, January 18). BRICS cybersecurity cooperation: Achievements and deepening paths. *China International Studies No. 68 January/February 2018.* Retrieved from http://www.ciis.org.cn/gyzz/2018-01/18/content_40192957.htm

Weber, V. (2017, December 12). Why China's internet censorship model will prevail over Russia's. *Council on Foreign Relations*. Retrieved from https://www.cfr.org/blog/why-chinas-internet-censorship-model-will-prevail-over-russias

Chapter 12. AI and China's Unstoppable Global Rise: A Skeptical Look

James A. Lewis
Center for Strategic and International Studies
jalewis@csis.org

Abstract

Judging any Chinese digital authoritarian model's potential attractiveness requires viewing it in strategic context – and not only in the context of a more comprehensive view of what drives influence in the global system, but also in the context of how such influence compares to that of its major competitor: the US. From this perspective I outline five reasons that will limit the model's impact. (1) China is not attractive as a governance model, not only because key elements such as Xi Jinping thought possess very limited soft power, but also because recent Chinese coersive diplomacy leads to antagonistic reactions in many countries. (2) The model's exportability will be limited by the CCP's declining domestic legitimacy. (3) China's innovative surge over the past decade, which has so impressed observers, will likely slow relative to the US as China turns to a more statist model. This will not be helped by a non-existent Chinese data advantage. (4) Relative military power affects states' relative influence, and while AI will change how states engage in warfare it is very unclear if China will make better use of this than the US. (5) A pervasive surveillance state may be attractive to a few governments but will not be to their citizens – leading to turbulence in countries lacking China's poweful and unexportable institutions of social control. For these reasons, although AI ripped from its strategic context can seem powerful or even frightening, given strategic competence the US will remain superior to China.

Introduction

Will AI and big data reshape the global order by allowing authoritarian regimes, led by China, to offer a new model of economic growth while using pervasive surveillance (guided by AI) to ensure political control? Any such prediction must be examined in broader strategic context – and within that context there are many reasons to be skeptical that China will obtain an edge over the United States. Long experience shows that new technologies do not by themselves increase national power or provide competitive advantage unless they are embedded in effective doctrine and policy for their use.

It is easy to overestimate China, in part because the Chinese government spends lavishly to encourage this overestimation, and in part because of a crisis of confidence in Western nations. We can usefully examine some of the weaknesses that will limit export and emulation of the Chinese model for innovation and growth. I discuss five below.

(1) Limited "Soft Power" in the Context of Regional Coercion

The most important limitation is that China is not that attractive as a governance model. Its market is attractive, as is its money, but the idea of soft power based on Xi Jinping thought is a contradiction in terms. China has managed to alienate many of its neighbors—Malaysian Prime Minister Mohammed Mahathir refers to Chinese policy as the "New Colonialism" (Hornby, 2018)—and in other regions. China's most effective tools for influence are coercion and bribery. China has used market access as a lever for forty years. These tools have proved most useful in defending against the reputational damage that arises from its domestic policies, but have also generated an antagonistic reaction in many countries.

(2) Limited "Soft Power" in the Context of Declining Regime Legitimacy

Second, the backdrop for the exportability of the China model is what the history of Chinese reform predicts for the model's future. A key question is whether the Party has reached its "sell-by" date. The Chinese Communist Party (CCP) is part of a longer line of reform in China that dates back to the mid-19th Century, but this history did not end in 1949. Leninism says that once the Party has seized power, it is irreversible, but the increasingly draconian efforts of the CCP to retain control belie this. The CCPs twin, the Kuomintang, abandoned single party rule in Taiwan some time ago without social collapse, a return of warlords, foreign dismemberment - any of the outcomes the party predicts if it loosens its grip. Some of the millions of Chinese who visit Taiwan realize this. It remains unclear if pervasive surveillance, emotional appeals to strident nationalism, and nostalgia for Mao is enough to prevent paralysis in the face of a declining legitimacy as each generation of Chinese leaders is more distant from the legitimizing Revolution.

(3) Limited "Soft Power" in the Context of Innovation Declining Relative to the US

Third, these trends will affect innovation in China – and as China under Xi returns to a more statist model, the past decade's innovative surge that has so impressed others will likely slow. China's success in technology must be assessed carefully, given its uneven nature. China has made immense strides in income since 1949, and the programs behind "two bombs, one satellite" (see "News of the Communist Party of China", 2009; Wangshu, 2015) remain a justifiable source of pride (and is now an annual award given to leading scientists), but it is still dependent on the West for most advanced technologies. It has attempted for decades to remedy this by the acquisition, licit and illicit, of western technology and by significant investment in China's research base.

Despite these investments in innovation, even the Chinese do not expect to reach technological parity with the US before 2030. Sino-US comparisons require considering both Chinese and US factors. The rate of progress is conditioned by Chinese domestic politics, since more restrictive policies create an outflow of money and talent, and also by the status of American spending on science, since static federal support for research, education, and immigration have a decelerating effect (see Henry, 2016; McCarthy, 2017). The real issue is less that China is speeding up and more that the U.S. is slowing down.

Several factors allow us assess competition with China on AI and estimate China's rate of progress. American companies substantially outspend China on AI R&D. A recent survey showed that the country with the most technology professionals working in AI is the United States, while China ranked seventh (Huang, 2017). Whether Chinese government funding for AI research will compensate for this is unclear. Research and development (R&D) for AI relies more on open collaboration between researchers using widely shared software, suggesting that open environments have an advantage over closed for AI research (generally true for innovation) and making it important that U.S. does not cut itself off from the flows of investigation and talent.

It is worth noting, however, that Chinese money still flows to the U.S. in search of advanced technology and skills. China relies on western universities for advanced scientific and technological training. Resurgent demands in China for party loyalty may increase the outflow of Chinese talent (as has been the case in Russia). On balance, Chinese attendance at American universities should not be a problem, but it could become one because of the inability to retain Chinese graduates in the United States, combined with a lack of incentives for American students to enter STEM fields.

Two further factors related to innovation are worth discussing in more detail: data; and central planning.

Innovation: Data from human users and privacy

China does not have a data advantage. This is a fundamental misunderstanding that is surprisingly common in the West. Yes, Alibaba and other Chinese companies have access to the data of hundreds of millions of Chinese users, but they are limited to China by a distrust of their services in foreign markets (an outgrowth of widespread negative perceptions of China's pervasive surveillance and the degree of control it exercises over its companies).[35] In contrast, Facebook, Google and others service a global market and have access to twice as much data as Chinese companies. Facebook has 3.4 billion users, more than twice the population of China. Different kinds of AI require different kinds and quantities of data, so that comparing user data numbers is simplistic; what is interesting is the willingness of observers to accept this kind of shallow analysis about China's innate advantage.

Where China may have an advantage is in the scope of privacy regulations. These could hamper the access of western companies to their larger data pool. Chinese privacy regulations are likely to be less restrictive. Restrictive or badly implemented privacy regulation in the West, along with efforts at data localization in countries like India, could give China an advantage in the development of some kinds of AI. The Chinese do have privacy regulations that are loosely modeled on Europe's GDPR, but their effect is less limiting on data use. U.S. privacy regulation is still in a formative period and how much of an advantage China could gain will depend on the outcomes for new U.S. privacy rules.

The best example of the advantage conferred by weak Chinese regulation is in biotechnology, where China has made rapid progress in developing its own biotech industry. In China, companies enjoy streamlined and accelerated clinical trials, which lower costs (see China's biotech revolution, 2018).

Innovation: Central economic planning

China is attempting to layer a centrally directed economic policy to create indigenous innovation atop a market-driven, global innovation system and supply chain. It is in effect, choosing the less productive path, gambling that its combination of investment, espionage, and the continued allure of the China market for western companies will make this an effective strategy. At the same time, it is extending the Party's influence in tech company decision-making. None of these decisions are likely to advance China's innovation capabilities, but domestic political necessities drive unsound economic policies. In most cases, countries ruled by a "President-for-Life" have not seen happy outcomes.

Indeed, China still depends on the West for advanced technology. China still lags in the production of advanced semiconductors, something that worries Beijing, but which despite massive investment, it has been unable to rectify. Some Chinese companies are able to design specialized chips for AI purposes and then have them manufactured elsewhere using "fabless" processes, but so far China still relies on foreign suppliers for the most advanced chips (Kubota, 2018). This is obscured by intense propaganda about China's progress in AI, accompanied by intensified efforts to illicitly acquire chip technology from the West. The state of the Chinese semiconductor industry supports a general conclusion that Chinese technology investments since 1979 have had mixed results, and that

[35] Expanded use of encryption and clashes with the U.S. serves to insulate American companies from similar suspicions.

China has made faster progress when it relied on market to direct investment rather than central planning.

The current vogue for "Civil-Military Integration" (Laskai, 2018; Lei, 2018) is an improvement over China's traditionally closed, state-owned approach to military procurement, but integrating tech companies with a "Central Commission for Integrated Military and Civilian Development" led by the head of state sounds like an effort to graft modernized central planning onto China's free-wheeling tech industry. China needs Civil-Military Integration if it is to develop the rules and mechanisms for private companies to compete as defense contractors, something that China did not have or need in the past. Government media sources also say that the intent is "to injecting new momentum into the country's private sector" and speed the transfer of technology from State-owned defense companies to private companies. This is the opposite of how technology flows usually work in other countries. A decision to move away from the Soviet style defense-industrial complex to a more modern contracting approach makes sense for China, but it may not prove to be a font of innovation.

China is not a market economy. A version of the Soviet central planning organ, Gosplan, reinforced by AI and leavened with an irregular reliance on market tools, is unlikely to translate into economic or technological advantage. AI, like other areas of tech competition such as 5G, is a competition between economic models, between State-centric and market-led approaches to investment and research. In most (but not all), the market-led model is both more efficient and more productive. Key variables are whether there is market demand for an innovation, how much investment in non-commercial research is needed, and how far in time the innovation is from being marketable. Using these criteria, a market approach is likely to be more effective in developing AI. In the U.S., this is complicated for now by uneven relations between some Silicon Valley firms and DOD, but this should not affect the overall pace of AI innovation (making the DOD challenge adopting commercial solutions to military problems).

It is also easy to overestimate the validity of Vladimir Putin's statement that the nation that leads in AI will rule the world. AI will not save Russia's economy, crippled by corruption and crime. If what Putin meant was that the nation best able to capture the benefits of the next several generations of digital innovation will be wealthier and perhaps more powerful than others, the statement is accurate. AI will improve economic performance in combination with next generation networking technologies and improved data analytics, but this will be incremental, (albeit at a faster pace than other technological changes) and economic performance by itself does not confer advantage or power.

(4) AI, Warfare, and Information

Relative military power affects states' relative influence, and while AI will change how states engage in warfare it is very unclear if China will make better use of this than the US. AI will change how countries engage in warfare, but the scope and pace of change will depend not only on the acquisition of new technologies but, more importantly, on the development of doctrine, tactics and operations strategies to take advantage of the new technology. Greater automation in weapons and sensors will improve performance, but the advantage this confers depends on if and how more advanced weaponry is used. This an area where China has lagged (although it is making good progress in developing doctrine). We can speculate on using later generations of AI to accelerate the process of creating doctrine, and AI could contribute to better decision-making, although strategy and policymaking at high levels remains an intensely human function.

Fears that countries will exploit AI for information operations tend to rely on abstract assumptions about effect. These operations are most effective when they exploit existing fissures in western societies: the Russians did not invent racism or income inequality. There has to be a degree of cultural awareness. The Russians have been studying American society or a century and are themselves a "western" country.' Other nations lack this touch – the recent Iranian efforts on social media were crude and AI will not improve the Chinese Communist Party's ideology to the point where it becomes attractive or persuasive. Finally, people are far more suspicious of social media and social media companies are making efforts of varying degrees of feebleness to defeat hostile operations, AI can increase the speed and scope of attack, but the effects will be marginal.

The use of AI will create a new target set – what some call algorithmic warfare. An essential goal for cyber operations is to interfere with the opponent's decision making, to expand the fog of war and make opponents uncertain, slow and confused. Manipulating opponent algorithms to produce these results or to better predict opponent decisions will be a source of military advantage. The U.S. should assume that its cyber peer opponents – Russia and China – will be at least as good as we are. Russia may be better, given its long focus in military doctrine on cognitive effect, with the chief of the Russian Armed Forces' General Staff, saying "the 'rules of war' themselves have changed significantly, nonmilitary options have come to play a greater role in achieving political and strategic goals and, in some situations, are greatly superior to the power of weapons" (McKew, 2017).

(5) AI, Surveillance, and Espionage

AI, combined with improved sensors and network technologies, and using mass data analytics, creates the ability for a country to improve pervasive surveillance and conduct it at lower cost on its citizens and its opponents. This will be attractive to a few governments but not their citizens in countries more turbulent than China. AI enhanced surveillance works in China because it is married to powerful institutions of social control, something that almost all other countries lack and which is not exportable.

The effect on both domestic and foreign intelligence operations will be pronounced as the space for secret (or private) activity continues to shrink, but benefits to economic growth, innovation, and political stability may be overstated. If Chinese governance was innately stable, the Party would not need these massive expenditures. While political unrest is unlikely, there is an increase in discontent. More importantly, the combination of an unattractive surveillance state buttressed by increasing nationalism (and more coercive foreign policies) will reduce China's international influence. The U.S. may be unable to take advantage of this trend but conversely, declining U.S. influence does not automatically translate into increased influence for China.

Conclusions

Countries that lead in science and technology do better economically and could do better at exercising power and influence, but this is not a surprising conclusion. That said, leadership in science and technology by itself does not guarantee power or influence (although it may guarantee wealth). Ripped from the larger strategic context, AI can seem powerful and perhaps frightening, but the issue is not whether you have the technology but how you use it. Technological advantage combined with inadequate strategy and doctrine is no recipe for victory, and it is not only technological innovation that is needed but innovation in their application. Large, wealthy countries with cutting edge military and technological assets, and strategic competence, can turn leadership in technological innovation into power. In all but the latter category, the U.S. remains superior to China.

References

China's biotech revolution. (2018, August 02). *UBS*. Retrieved from https://www.ubs.com/global/en/wealth-management/chief-investment-office/our-research/discover-more/2017/china-biotech.html

Henry, M. (2016, November 8). US R&D spending at all-time high, federal share research record low. *American Institute of Physics*. Retrieved from https://www.aip.org/fyi/2016/us-rd-spending-all-time-high-federal-share-reaches-record-low

Hornby, L. (2018, August 20) Mahathir Mohamad warns against 'new colonialism' during China visit. *Financial Times*. Retrieved from https://www.ft.com/content/7566599e-a443-11e8-8ecf-a7ae1beff35b

Huang, E. (2017, August 25). Half of the top 10 employers of AI talent in China are American. *Quartz.* Retrieved from https://qz.com/1062035/half-of-the-top-10-employers-of-ai-talent-in-china-are-american/

Kubota, Y. (2018, May 6). China plans $47 billion fund to boost its semiconductor industry. *The Wall Street* Journal. Retrieved from https://www.wsj.com/articles/china-plans-47-billion-fund-to-boost-its-semiconductor-industry-1525434907

Laskai, L. (2018, January 29). Civil-military fusion: The missing link between China's technological and military rise. *Council of Foreign Relations.* Retrieved from https://www.cfr.org/blog/civil-military-fusion-missing-link-between-chinas-technological-and-military-rise

Lei, Z. (2018, March 3). Civil-military integration will deepen. *China Daily.* Retrieved from http://www.chinadaily.com.cn/a/201803/03/WS5a99d67ca3106e7dcc13f437.html

McCarthy, N. (2017, February 2). The countries with the most STEM graduates [infographic]. Forbes. Retrieved from https://www.forbes.com/sites/niallmccarthy/2017/02/02/the-countries-with-the-most-stem-graduates-infographic/#1b239597268a

McKew, M. K., (2017, Sept/Oct). The Gerasimov doctrine. Politico. Retrieved from https://www.politico.com/magazine/story/2017/09/05/gerasimov-doctrine-russia-foreign-policy-215538

News of the Communist Party of China. (2009, August 25). Retrieved from http://english.cpc.people.com.cn/66113/6738142.html

Wangshu, L. (2015, January 10). H-bomb work news scientist top award. *China Daily*. Retrieved from http://www.chinadaily.com.cn/china/2015-01/10/content_19286624.htm

Chapter 13. Four Horsemen of AI Conflict: Scale, Speed, Foreknowledge, and Strategic Coherence

Chris C. Demchak[36]
Cyber and Innovation Policy Institute
chris.demchak@usnwc.edu

Abstract

With the shoddy creation of the global cyberspace two decades ago, a new form of intergroup struggle—'cybered' conflict'—emerged to massively enrich bad actors across the world through five novel offense advantages. The resulting massive transfers of wealth to nonwestern nations has enabled China to amass resources to ensure increasingly dominant economic, demographic, and technological power. As AI-related technologies rise in criticality for the future economic and political wellbeing of nations, China now has the advantage in three of the four 'horsemen' of AI conflict (scale, foreknowledge, and strategic coherence), leaving only a fourth (speed) to the western democratic civil societies. To counter systemically the four AI advantages accruing to China, democratic civil societies need both a new narrative placing their future globally as minority states in terms that ensure long term survival, and a novel but practical organizational architecture with which to implement that vision. Militaries will have to change as well, preparing to"fight" a constant war in AI-led military operations while collectively embedded in the community of democratic states.

Introduction

AI emerges upon a world already roiled by the wild west infancy of global cyberspace. When optimistic actors in the 1990s widely connected otherwise unprotected critical digitized functions to the rest of globe, the cyberspace substrate became a global playing field for millions of malicious actors across states able to reach across thousands of miles and extract massive amounts national resources from democracies with impunity. By 2014, senior cyber scholars and business analysts estimated the annual losses through cyber insecurity to consolidated democratic civil society to be on the order of 1-2 percent of their individual GDP (Price Waterhouse Cooper, 2014).

As the backdrop to the AI challenge today, the cyberspace vulnerabilities enabled the unprecedented and rapid rise of an otherwise impoverished authoritarian majority of states – especially China, and accelerated a new form of intergroup struggle – 'cybered' conflict' – emerging systemically across nations. From peace to war, the then new form of cyber conflict offered five offense advantages historically only available to emperors or near neighbors (Demchak & Dombrowski, 2011).

- First, nearly any actor, group, or state could organize for little cost vast dispersed armies of humans or compromised computers (a 'botnet') at a *scale of organization* only emperors or neighbors could afford throughout history.
- Second, for little but time and net access, any of these could reach a victim at any *proximity* from five to five thousand kilometers away – to gain critical intelligence and even strike digitally with possible physical harm.
- Third, cybered conflict relieved the search for *precision*, driven throughout history by a need

[36] The ideas in this piece are solely those of the author, and do not reflect the position of the US government, Department of Defense, the Department of the Navy, or the US Naval War College.

to reduce unnecessary or unaffordable costs and time spent in overproducing armies and navies for the chosen targets. Given the massive global underground cybercrime market, large and small aggressors can choose from a wide variety of malicious applications, of targets, and of operational tempos or simultaneity.

- Fourth, cybered conflict made *deception in tools* critical, lest years of effort can be wasted if an attacker's cyber tool is correctly understood before or during its use, encouraging preemption in use.
- Fifth, *opaqueness in originators* throughout the attack and beyond became an imperative, deeply complicating deterrence. Both aggressors and victims must operate through large-scale socio-technical-economic systems (STES) and if the aggressor is known, retaliation can come through the same mechanisms (Schmitt, 2013).

AI-related technologies now add new and seemingly existential challenges to this list – at least for western civil societies. This exceptionally shoddy cyber substrate exposed the wealthy states' financial and other infrastructural wellsprings to, in particular, China's aggressive economic ambitions and acquisition tsunami - and positioned China well to leapfrog into the new technological era. Today the advantage in three of the four 'horsemen' of AI conflict – scale, foreknowledge, and strategic coherence – leans towards China, leaving only a fourth – speed in employing AI – to the western democratic civil societies. The system set up after WWII is receding rapidly, and will be displaced over time by the system created as the other and more authoritarian 90% of the world's population rises. The new world is currently well set to be led by China's models and preferences – economically, politically, and digitally. Unless actions by the formerly dominant westernized states alter current trends, the continued global rise of China's dominance in AI-led technology carries with it the commensurate decline in futures of democratic civil societies – and the liberal economic international governance system they built (Chang, 2014).

This memo addresses each of the four horsemen of AI conflict and the imperatives for rethinking global governance and the role of militaries in beleaguered democracies. I conclude with recommendations to counter the Chinese advantages in each of the four areas.

Four Advantages in AI Conflict– and Whether China or the West Currently Hold Them

The global outcome of the US-China conflict involving AI depends heavily on who holds or develops strong advantages in four areas: scale, speed, foreknowledge, and strategic coherence. AI is not magical but its competent deployment strongly enhances a state's chances to prevail in any conflict, especially a whole-of-society cybered conflict (Dickson, 2018a). Military history suggests consistent success comes if one has foreknowledge of adversaries' actions at faster speeds and the larger scale ability to act disruptively, especially if married to an organizationally coherent state able to create and pursue a coordinated, overarching, and longer term strategy. With its aggressive national program to acquire and dominate in artificial intelligence in the next twenty years (Segal, 2018), China's central leadership is determined to dominate AI and related technologies as the global cyber power (Buckley, 2013; Gow, 2017). At present, the US and its allies are holding their own only in one area – commercial speed of adoption. Without serious and near term whole-of-society efforts, the future of the US and its westernized allies as prosperous, democratic, and open civil societies is bleak.

Scale

Scale in demographic size multiplies the AI advantage when the large state's resources are able to employ it strategically. As a global system tool, however, scale was underappreciated by modern democratic defense and commercial leaders informed by the seventy years of the Cold War global

dominance by the otherwise relatively small western population. During this era, the major largescale adversaries – Russia and China – self-isolated economically, allowing the delusion of permanent control to infuse throughout western governance and defense thinking. Due to China's population alone, its preferences in economics and, inevitably, political terms would begin to dominate global system choices even without direct coercion and the usage of technological and overt state power (Helleiner & Kirshner, 2014). However, China has already demonstrated strategic success in using its demographic scale to fan out globally in every niche. Its corporate, university, and government agencies are, acquiring the enormous volumes of data needed for AI conflict from a massive variety of legal and illegal economic, political, and technological means.

With generous – if publicly denied – state subsidies and protected from failure by China's assertive (and punishing) economic statecraft, China's technology state champions are rising to the top of global corporations in fields critical for the digital future and, eventually AI and following technologies.

China now leads the world in numbers of internet users, computer science, and science, technology, engineering, and math (STEM) college graduates, home owners, billionaires, and technology investors, as well as basic science investments. Its scale allowed it to surpass the global economic giant – the US – within a mere 18 years of joining the World Trade Organization (Wang, 2017). Under Xi Jinping, the government is also encouraging a rising nationalist sense of superiority, technological optimism, and entitlement to a globally dominant position. This socialization then motivates individual expectations and then complementary actions across science and industry at massive scale, which advance strategic interests without direct governmental guidance (Yang, 2017).

The demographic and economic scale advantage enhancing AI dominance thus currently accrues to China. It cannot be contained; rather, its preferences will have to be accommodated up to the points where the democracies are undermined. No westernized civil society alone has the scale to exploit AI or any technology sufficiently enough to balance this advantage.

Speed

Speed of analysis and action far beyond the currently developed cyberspace is the second advantage that data, tools, talent, techniques, and algorithms of the AI-related technologies confer on states – if they are implemented, protected, and updated as rapidly as required. With the real revolution in AI found in the emerging applications of "deep neural learning" requiring massive computational resources (Leetaru, 2018), the rise of quantum computations will massively increase the speed at which AI- related technologies produce results of trustable analysis. With this advantage, national actors can compute likely outcomes across societal-scale problems and threats, and then coordinate unprecedentedly rapid actions to enhance, dampen, disrupt, or destroy the essential elements of targeted processes in opponent states. This AI horseman dramatically increases the offense advantages within the deceptive and opaque conditions of cybered conflict at distance, with precision, and materially inexpensively (Dickson, 2018b).

For the moment, the advantage in speed of innovation currently rests with the information technology (IT) capital goods industries of the democracies and that in speed of analysis with some militaries, notably DOD (Hymas, 2018; Wong, 2018). However, China is determinedly seeking AI talent, tools, techniques, and dominance – including exceptional efforts aimed at commanding heights in quantum computing. Without more collective efforts across defending democracies to enhance the defense of their own and allied IT capital goods industries (Blenkinsop, 2018) and infrastructure from the tireless Chinese economic and other predations (Wong, 2018), today's

success is at best labeled staying afloat (Bennett & Bender, 2018).

Foreknowledge

Foreknowledge is the sine qua non of strategic power – knowing what the adversary knows and can do – and AI's emergent analytical promises will confer critical weight to adversaries in any conflict. The widespread embedded use of AI technologies by state-sponsored actors particularly enhances likely acquisition of more systemic and longitudinal trends and nearer real time comprehensive situational awareness (SA) critical to national interests. Any state able to gather the near and longer term, highly accurate foreknowledge of perceptions and action options over a wide range of adversaries at the scale of a China and speed of an US National Security Agency has a massive source of influence in any exchange.

While neither the Chinese nor the defending democracies currently have the level of foreknowledge either would like, the elements leaning the advantage to the Chinese are already emerging. The ability to foresee what to do next for success can be built from a variety of sources from massive data heists such as the Office of Personnel Management (OPM) heist of 2016 or the elemental penetration of critical infrastructures across nations (Smyth, 2018). The intelligence feedback in SA from widespread Chinese presence allows for the effective identification of individual, corporate, and political leaders vulnerable to what is here called the 'Four B's' of aggressive economic and political behaviors common in developing and nonwestern societies: (*hostile) buy, bribe, bully, and blackmail* (Reilly, 2014; Norris, 2016). What capturing the Enigma machine did for the Allies in WWII, the rising Chinese dominance across embedded AI technologies will do for China's leadership in providing foreknowledge to undermine democratic state and corporate resilience in resisting China's new world preferences.

Strategic coherence

Strategic coherence is the ability to declare and implement systemic programs across a large-scale socio-technical-economic system (STES) without losing its benefits or the purposes due to internal political-economic battles. In the pursuit of AI dominance, a strategically coherent nation is more likely to be able to announce strategic goals in investments, R&D, and education, and streamline actions to achieve those advances across sectors. rtificial intelligence has a multiplier effect, confering increasing advantage in strategic coherence when the other three advantages (scale, speed, foreknowledge) are also present. With robust AI technologies distribued across the nation, senior leaders will have the tools to create coherent strategic objectives and then – in principle – comprehensively monitor national and allied lines of effort towards those objectives. Beyond that, authoritarian nations have the advantage in forcing national actors to act coherently as a group in coordinated pursuit of AI objectives, despite existing economic or political battles.

This advantage now increasingly rests with China under Xi Jinping, whose leadership has developed this strategic coherence with respect to using cyber means to ensure the nation's rise to center stage globally. The Chinese government seeks AI to monitor and control its own domestic and overseas actors with singular strategic coherence. The westernized democracies continue to struggle with any strategic common view across their normally fractious political and economic internal and intergroup interactions.

Implications for Global Governance and Defense of Democracies

China under Xi Jinping is moving outward aggressively at large scale. Democracies lack a response to China's current command of three of the four AI horsemen. They are on trend to lose their remaining speed advantage. To change these trends, the small group of civil societies need a collective approach to mitigate the imbalance in scale, in the foreknowledge potential flowing to China, and in the overmatch in strategic coherence. Especially missing are a new narrative and a new model of defense, survival, and prosperity able to guide the outnumbered consolidated democracies in defending their wellbeing in the future.

It is difficult for the once dominant westernized states to accept that their future is as a minority of states with divergent culture and institutions from a largely postwestern and likely hostile, more authoritarian world. As a rising power seeking new international system rules for its benefit, China has become an "alt-authoritarian anchor state" with its presence and abundant state resources (Goldman 2010). China actively promotes how nonwestern states may avoid democracy and prosper by adopting the Chinese model of economic and digitized state control (Dahir, 2018). With Xi Jinping's leadership, the AI-related technologies make that model increasingly desirable, comprehensive, and reliable for control by central authorities in nondemocratic nations.

Imperative – Counter China's Four Horsemen of AI, Both Operationally and Collectively

To counter the four AI advantages accruing to China systemically, democratic civil societies need both a new narrative placing their future globally as minority states in terms that ensure long term survival, and a novel but practical organizational architecture with which to implement that vision.

A new narrative is needed to outline a robust and sustainable a modus vivendi with the larger authoritarian behemoth and its fellow states. It must put paid to postCold War myths that China can be contained or that the the westernized international liberal economic order can be saved as it stands. The new narrative must include support for collective democratic mechanisms (trust, tolerance, and transparency) while sustaining acceptable national wellbeing for the coming AI-driven era. The story must energize public spirited independent actions to collectively enable longer term survival of these nations even if out-numbered and out-financed (eventually) by the other ninety percent of the globe's population living under authoritarian rule. It must in short inspire and explain the need for continuous defense across the whole of these societies.

The narrative must have a practical implementation path: Given the current conditions and the rapidity of the decline of westernized global dominance, the goal is to buy time for the laggardly democracies to adapt the defense of their future wellbeing for the long term. Only a collective and operational response in the near term that matches and undermines China's longer term dominance of the four AI advantages can stall the current processes from producing overwhelming Chinese systemic dominance in the coming technological era.

With the markets and IT talent of roughly 900 million educated citizens, the democratic civil societies have the scale needed to create a cross-democracy Cyber Operational Resilience Alliance (CORA) organized to defend in the near term and lay the foundation for the longer term. As the AI-era advances, the CORA has the best – perhaps the only – chance to alter the current trends placing AI advantages firmly in China's hands. Democracies as a united group forming a peer power are more likely to be able to deter - and even prevent - China from its aggressive forms of cybered conflict. The democracies can pool their talent and funds at the scale of a China .

US military thinking, planning, and operating especially needs to update its strategic understanding and shared common operating pictures. Military and civilian national security forces will play new, more integrated national roles as well, especially in their interactions with private sector IT industries. Allies are not supplements to the US military capabilities.

China rose not in the expected fifty years, but in twenty to challenge the US and its allies economically, politically, and technologically. It is changing the world system and democratic militaries will either adapt to the collective digitized defense across nations and sectors, or fail profoundly to defend their nations.

References

Bennett, C., & Bender, B. (2018 May 22). How China acquires 'the crown jewels' of U.S. technology. *Politico.* Retrieved from https://www.politico.com/story/2018/05/22/china-us-tech-companies-cfius-572413

Blenkinsop, P. (2018, November 20). With eyes on China, EU agree investment screening rules. *Reuters*. Retrieved from https://uk.reuters.com/article/us-eu-china-investment/with-eyes-on-china-eu-agrees-investment-screening-rules-idUKKCN1NP1IJ

Buckley, C. (2013, August 23). China takes aim at Western ideas. *The New York Times*. Retrieved from https://www.nytimes.com/2013/08/20/world/asia/chinas-new-leadership-takes-hard-line-in-secret-memo.html

Chang, A. (2014). Warring state: China's cybersecurity strategy. *Center for New America Security.* Retrieved from http://www.cnas.org/chinas-cybersecurity-strategy#.VeHZlM5REs

Dahir, A. L. (2018, November 20). China is exporting its cyber surveillance to African countries. *Defense One*. Retrieved from https://www.defenseone.com/threats/2018/11/china-exporting-its-cyber-surveillance-african-countries/152537/?oref=d-topictop

Demchak, C. C., & Dombrowski, P. J. (2011). Rise of a cybered westphalian age. *Strategic Studies Quarterly, 5* (1):31-62.

Dickson, B. (2018a, November 21). The difference between AI and machine learning, explained. *The Next Web*. Retrieved from https://thenextweb.com/syndication/2018/11/21/the-difference-between-ai-and-machine-learning-explained/

Dickson, B. (2018b, November 27). The malware of the future will have AI superpowers. *Gizmodo*. Retrieved from https://gizmodo.com/the-malware-of-the-future-will-have-ai-superpowers-1830678865

Gow, M. (2017). The core socialist values of the Chinese dream: Towards a Chinese integral state. *Critical Asian Studies, 49*(1):92-116.

Helleiner, E. & Kirshner, K. (2014). The politics of China's international monetary relations. In Eric Helleiner & Jonathan Kirshner (Eds), *The Great Wall of money: Power and politics in China's international monetary relations*. Ithaca, NY: Cornell University Press.

Hymas, C. (2018, October 9). China is ahead of Russia as 'biggest state sponsor of cyber-attacks on the West.' *The Telegraph*. Retrieved from https://www.telegraph.co.uk/technology/2018/10/09/china-ahead-russia-biggest-state-sponsor-cyber-attacks-west/

Leetaru, K. (2018, November 14). Today's deep learning "AI" is machine learning not magic. *Forbes*. Retrieved from https://www.forbes.com/sites/kalevleetaru/2018/11/14/todays-deep-learning-ai-is-machine-learning-not-magic/#4b47d92a6875

Norris, W. J. (2016). Chinese economic statecraft: Commercial actors, grand strategy, and state control. Ithaca, NY:Cornell University Press.

Price Waterhouse Cooper. (2014). Global State of Information Security® Survey 2015. In Annual State of Information Security Survey.

Reilly, J. (2013, November). China's economic statecraft: Turning wealth into power. *Lowy Institute for International Policy*. Retrieved from https://www.lowyinstitute.org/publications/chinas-economic-statecraft-turning-wealth-power

Saarinen, J. (2018, October 26). China systematically hijacks internet traffic: researchers. *Itnews*. Retrieved from https://www.itnews.com.au/news/china-systematically-hijacks-internet-traffic-researchers-514537

Schmitt, M. N., (2013). Tallinn manual on the international law applicable to cyber warfare. Cambridge University Press.

Segal, A. (2018, January 8). Year in review: Chinese cyber sovereignty in action. *Council on Foreign Relations*. Retrieved from https://www.cfr.org/blog/year-review-chinese-cyber-sovereignty-action

Smyth, J. (2018, August 22). Australia bans China Huawei's 5G rollout over security fears. *Financial Times*. Retrieved from https://www.ft.com/content/c5f24650-a66b-11e8-8ecf-a7ae1beff35b

Wang, Z. (2017). The economic rise of China: Rule-taker, rule-maker, or rule-breaker? *Asian Survey, 57*(4):595-617.

Wong, J. (2018, November 21). Short sellers vs. Chinese companies—a war, not a battle. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/short-sellers-vs-chinese-companiesa-war-not-a-battle-1542798883

Yang, Y. (2017, December 29). China's Communist party raises army of nationalist trolls. *Forbes*. Retrieved from https://www.ft.com/content/9ef9f592-e2bd-11e7-97e2-916d4fbac0da

PART IV. AI & DOMESTIC IMPACTS ON CHINA'S FOREIGN POLICY DECISION-MAKING

Chapter 14. AI and US-China Relations

Benjamin Angel Chang[37]
Massachusetts Institute of Technology
bachang@mit.edu

Abstract

How will domestic use of Artifical Intelligence (AI) affect Chinese foreign policy? Drawing on relevant threads of political science, I discuss two possible consequences: (1) significantly worsened US-China relations due to increased ideological friction and opacity, and (2) increased Chinese assertiveness due to increased confidence and a smaller "winning coalition." Finally, I briefly assess implications for US policy.

US-China Relations

The key effect of AI on Chinese authoritarianism is likely to be scalability. Currently, China deploys over two million Internet censorship workers (Xu and Albert, 2017), and up to 1,000 censors per individual site (King et al., 2013). As facial recognition technology is in its infancy, processing data from China's 200 million physical cameras still depends on masses of flesh-and-blood humans poring over photos and files (Mozur, 2018).

With progress in AI, however, the human labor required per monitored citizen may become minimal. While initial investments are sure to be costly, improvements in efficiency and inducement of a culture of self-censorship mean AI is likely to make authoritarianism significantly more sustainable, over the medium term. AI could independently process millions of hours of footage, carry out intelligently automated censorship on untold volumes of social media posts, and generate predictions as to which people might hold undesired views, organize protests, or attempt to flee the country (Wright, 2018). Just as automation obviated many factory jobs, so too might AI put much of the lower levels of China's sprawling security apparatus out of business, given the dramatically increased efficiency of those remaining.

Such developments are likely to significantly worsen US-China relations for two reasons.

First, barring dark futures in which the United States itself loses its democratic character, intensified PRC authoritarianism would necessarily widen the perceived and actual ideological gap between both societies. As crystallized in Vice-President Pence's October 4th speech at the Hudson Institute (Pence, 2018), US-China relations have become increasingly conflictual in recent years. There is an emerging bipartisan consensus that China, contrary to decades of American hopes, will not liberalize as a result of sustained economic growth and contact with the US-led international order. In other words, if life in China increasingly reminds Western audiences of Orwellian scenes from dystopian fictions (in fact, in a strange twist of humor, China has named its video surveillance system Skynet

[37] Many thanks to Torin Rudeen for his comments on an earlier draft of this article.

[天网]), this is likely to directly drive more antagonistic, ideologically framed views of US-China strategic competition.

Indeed, some political science research finds that ideological distance tends to predict conflict. For Haas (2005), the history of European great power relations reveals a common theme: leaders seek to legitimate their own forms of government, and tend to identify their sense of self with the fortunes of other ideologically similar states. Napoleonic France and the Soviet Union, for example, were assessed by many as threatening in significant part because of how different their revolutionary ideologies were from other states.

Second, an AI-empowered PRC is likely to be increasingly opaque to external observers. In recent years, pro-China sentiment among Western businesses has progressively dampened, due to intellecual property theft, forced tech transfers, and an otherwise increasingly hostile business environment – in fact, this has been a significant factor in freeing US Congresspersons to favor a more competitive posture toward China (Dickinson, 2018). As the PRC becomes increasingly authoritarian and its tools of control become increasingly intrusive, the slow-down in Western investment may deepen and become an exodus. A similar story can be told about student exchanges and academic collaborations. Cornell, for example, recently suspended two exchange programs with Renmin University over academic freedom (Weiss, 2018). Separately, even as the digital era provides various new collection opportunities, AI-empowered surveillance tools may further complicate US HUMINT efforts due to the increased difficulty of maintaining cover in-country (Jackson et al., 2017). Recently, the CIA suffered a significant setback with the loss of its network of agents in China from 2010-2012, due to a breach of the agency's communications network (Dorfman, 2018).

Overall, the resultant opacity is likely to be harmful to US-China crisis stability. Back-and-forth movements of people represent valuable flows of information for both governments, increasing their mutual cultural, linguistic, and diplomatic intelligibility. The ability of states to accurately and precisely understand each other's intentions is limited even at the best of times, to say nothing of the fog which pervades crisis situations (Rosato, 2014/15). Especially from the US side of the Pacific, while US political debates are broadcast daily on television for all to see, Chinese elites seldom discuss grand strategic matters in easy view of American eyes. All this may blunt our ability to signal and communicate diplomatically during crises. Moreover, while reports of US relative decline are likely somewhat overstated, it is worth mentioning that according to one historical study, autocratic opacity tends to induce uncertainty in other states and thereby magnify the volatility of power transitions (Kliman, 2014). In short, such conditions increase the risk of war.

Chinese Behavior

Independent of effects on the US-China relationship, intensified PRC use of AI for domestic security may also encourage greater Chinese assertiveness. Again, I highlight two potential reasons.

First, a pacified domestic sphere might free up attention for expanded external aims. China's relative ability to weather the 2008 financial crisis significantly motivated its recently more assertive turn (Chen & Wang, 2011). Whereas many Chinese intellectuals had previously sought to emulate Western economic development, viewing the American stage of development as if a higher rung on a universally climbable ladder, the crisis incubated the view that, instead, the Chinese model might be a fine endpoint in and of itself. Similarly, if the CCP were to feel AI had successfully and permanently allowed it to address the full panoply of possible sources of broad public unrest, ranging from unbalanced growth to Xinjiang to income inequality, it would likely see this as one of the Party's

crowning achievements in its leadership of the Chinese people. Chinese spending on domestic security has exceeded spending on external defense since 2010, with the gap increasing each year. In 2017, according to the best open-source estimate available, the former exceeded the latter by 18.6 percent (Zenz, 2018). Were the domestic sphere to be "solved," some of this attention might then be turned outward.

Second, concentrations of power generally tend to lead to more belligerence on the international stage. In particular, by substituting technology for manpower in carrying out the state's policing functions, an AI-empowered PRC may enable ever-smaller groups of elites to retain equivalent amounts of power. For de Mesquita et al. (1999, 2003), as the size of the coalition required for political survival (the "winning coalition") shrinks, corruption and war may become more likely, as leaders no longer fear being punished by other domestic actors for selfish arrangements or military defeats.[38]

Implications for US policy

Technology is not the only input into how authoritarian the PRC will be. Xi's choices to repress human rights lawyers, abolish term limits, and detain perhaps a million Uighurs in "re-education camps" in Xinjiang are, in fact, choices, and Xi or any successors could also reverse these trends. Indeed, we have seen inflection points presaging such a reversal before, such as with Deng Xiaoping's "Reform and Opening-Up" strategy starting in 1978. Nonetheless, given the frustrated hopes of engagement advocates over the past several decades, this seems fairly unlikely. Instead, more autocratic ability will likely translate smoothly into more autocratic action, which the above discussion gives reasons to believe will both worsen US-China relations and encourage greater Chinese assertiveness.

As such, care must be taken to manage the risk of conflict as US-China competition progresses, especially as AI itself may generate separate and novel challenges to global governance just as the Sino-American relationship begins to badly fray. For example, US-China cooperation on AI may be needed to regulate the spread to non-state actors of AI-empowered long-range precision strike capability, in the form of cheap UAVs, as well as to address the increasing black-market availability of cyber activities that currently require high-skill labor (Allen & Chan, 2017). Generally, as AI enables humans to relinquish control over military technology, the risk of accidents may increase (Danzig, 2018). In the nuclear realm, computing advances may separately herald a "new era of counterforce," with arsenals increasingly vulnerable to high-accuracy, low-fratricide munitions (Lieber and Press, 2017). Given previous Chinese crisis behavior during the 1969 Sino-Soviet border conflict, some scholars have recently assessed nuclear war with China to be a genuine tail risk in the relationship (Talmadge, 2017). If Chinese and American views of AI become framed primarily in dueling ideological terms, these and other issues may go unaddressed.

References

Allen, G., & Chan, T. (2017). Artificial Intelligence and National Security. Harvard Kennedy School. Retrieved from www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf

[38] See also Kacie Miura's chapter in this volume, however, on the possibly stabilizing benefits of an internally better coordinated PRC.

Chen, D., and Wang, J. (2011). Lying Low No More? China's New Thinking on the Tao Guang Yang Hui Strategy. China: An International Journal, 9, 195-216.

Danzig, R. (2018). Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority. Center for a New American Security. Retrieved from https://s3.amazonaws.com/files.cnas.org/documents/CNASReport-Technology-Roulette-DoSproof2v2.pdf?mtime=20180628072101

de Mesquita, B., Morrow, J., Siverson, R., and Smith, A. (1999). An Institutional Explanation for the Democratic Peace. American Political Science Review, 93, 791–807. doi:10.2307/2586113

de Mesquita, B., Smith, A., Siverson, R., and Morrow, J. (2003). The Logic of Political Survival. Cambridge: MIT Press.

Dickinson, S. (2018). The New Normal in US-China Relations and What to do About that. China Law Blog. Retrieved from https://www.chinalawblog.com/2018/11/the-new-normal-in-us-china-relations-and-what-to-do-about-that.html

Dorfman, Z. (2018). Botched CIA Communications System Helped Blow Cover of Chinese Agents. Foreign Policy. Retrieved from https://foreignpolicy.com/2018/08/15/botched-cia-communications-system-helped-blow-cover-chinese-agents-intelligence/

Haas, M. (2005). The Ideological Origins of Great Power Politics, 1789-1989. Ithaca: Cornell University Press.

Jackson, R., Pattar, S., and O'Connor, S. (2017). Fast forward: analysing changes to the intelligence landscape in the 2020s. Jane's Intelligence Review. Retrieved from https://www.janes.com/images/assets/461/76461/Fast_forward_analysing_changes_to_the_intelligence_landscape_in_the_2020s.pdf

King, G., Pan, J., and Roberts, M. (2013). How Censorship in China Allows Government Criticism but Silences Collective Expression. American Political Science Review, 107, 1-18. doi:10.1017/S0003055413000014

Lieber, K., and Press, D. (2017). The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence. International Security, 41, 9-49. doi:10.1162/ISEC_a_00273

Mozur, P. (2018). Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras. The New York Times. Retrieved from https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html

Pence, M. (2018). Vice President Mike Pence's Remarks on the Administration's Policy Towards China. Hudson Institute. Retrieved from https://www.hudson.org/events/1610-vice-president-mike-pence-s-remarks-on-the-administration-s-policy-towards-china102018

Rosato, S. (2014/15). The Inscrutable Intentions of Great Powers. International Security, 39, 48-88. doi:10.1162/ISEC_a_00190

Talmadge, C. (2017). Would China Go Nuclear? Assessing the Risk of Chinese Nuclear Escalation in a Conventional War with the United States. International Security, 41, 50-92. doi:10.1162/ISEC_a_00274

Weiss, J. (2018). Cornell University suspended two exchange programs with China's Renmin University. Here's why. The Washington Post. Retrieved from https://www.washingtonpost.com/news/monkey-cage/wp/2018/11/01/cornell-university-suspended-two-exchange-programs-with-chinas-renmin-university-heres-why/?

Wright, N. (2018). How Artificial Intelligence Will Reshape the Global Order. Foreign Affairs. Retrieved from https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order

Xu, B., & Albert, E. (2017). Media Censorship in China. Council on Foreign Relations. Retrieved from https://www.cfr.org/backgrounder/media-censorship-china

Zenz, A. (2018). China's Domestic Security Spending: An Analysis of Available Data. The Jamestown Foundation. Retrieved from https://jamestown.org/program/chinas-domestic-security-spending-analysis-available-data/

Chapter 15. The Implications of Increased Internal Control on China's International Behavior

Kacie Miura
Massachusetts Institute of Technology
kkmiura@mit.edu

Abstract

Although a small group of top leaders dictate foreign policy-making in China, several key domestic factors constrain and complicate China's international behavior. These include: regime insecurity, public opinion, factional competition, and bureaucratic discord. Artificial intelligence, if it improves the Chinese leadership's ability to monitor and control societal and elite actors, could presumably reduce the influence of these internal drivers of China's international behavior. Whether increased internal control will lead China to adopt a more or less confrontational foreign policy, however, is unclear. On the one hand, China's leaders may no longer need to prioritize external cooperation to balance against internal threats to the regime. On the other hand, China's leaders may also be less likely to escalate international disputes to appease nationalist publics or factions that support more hardline policies. Greater control over domestic actors, particularly elites and bureaucrats, will lead to a more tightly coordinated foreign policy. This will allow China's leaders to more efficiently advance their aspirations for China's position in the world, regardless of whether they choose to do so through confrontational or cooperative foreign policies.

Introduction

China's top leaders have long conducted foreign policy with an eye towards managing and preempting domestic threats to the regime. AI, by increasing the capacity of the CCP to monitor and control domestic actors, can have profound consequences for China's international behavior. This chapter explores the potential foreign policy implications of increased control over state and society. In particular, it addresses the following question: what does increased internal control over both societal and elite actors, aided by advancements in AI, mean for China's foreign policy?

Increased Societal Control

Since being appointed Secretary General of the CCP, Xi Jinping has worked overtime to consolidate his authority. However, Xi and the Party remain vulnerable to a number of domestic challenges, including internal unrest and a slowing economy, that have potentially dangerous implications for the security of the regime. While AI could help China's leaders minimize internal threats and sources of pressure, it is not immediately clear whether a more secure CCP will pursue a more cooperative or confrontational foreign policy. AI, as I explain below, could reduce the leadership's incentives to pursue both compromise and conflict in its foreign relations.

In the past, regime insecurity has occasionally prompted the CCP to engage in international cooperation, particularly with respect to its territorial disputes (Fravel, 2005). In order to divert more resources to addressing internal threats, the CCP has sought to reduce tensions in its external affairs. For example, in the early 1990s, in the aftermath of the Tiananmen crackdown and amidst acute ethnic unrest in Xinjiang, China's embattled leaders compromised in territorial disputes with Central Asian neighbors in exchange for public agreements to refrain from assisting separatists (Fravel, 2005).

By enhancing the CCP's ability to monitor, police, and repress restive populations, AI could be a potentially powerful remedy to the age-old problem of regime insecurity. For the CCP, the ability to preempt popular uprisings not only reduces motivations to address the root causes of unrest – such as systemic ethnic discrimination, socioeconomic dislocation, and poor governance – but also reduces the need to minimize external tensions, eliminating a potentially important driver of international cooperation.

On the other hand, regime insecurity, particularly stemming from the CCP's concerns about its legitimacy, has led the Party to stoke popular nationalism and anti-foreign sentiment. China's leaders have sought to boost the regime's popularity through nationalist appeals, urging the masses to "never forget national humiliation" and reminding them of the CCP's role in national rejuvenation (He, 2007; Wang, 2012). Having primed its population, some scholars caution that the CCP has painted itself into a corner by driving up the political costs of pursuing compromise in international disputes and de-escalating foreign policy crises (Christensen, 2011; Shirk, 2007).

AI, by strengthening the CCP's censorship capacity and ability to shape popular opinion, could free China's foreign policy decision-makers from the grip of popular pressure. The CCP, by repressing and shaping social media commentary, is unlikely to be motivated by public opinion to adopt a hardline foreign policy position.[39] China's leaders may therefore find it easier to resolve or shelve international disputes. The CCP might also have more flexibility to practice strategic restraint with respect to China's sovereignty claims over Taiwan and in the East and South China Seas.

In sum, a CCP unfettered by regime insecurity or nationalist public pressure will be freer to pursue both a more confrontational and cooperative foreign policy. Being more fully in the driver's seat of foreign policy, China's leaders will be better able to steer the country in whichever direction they prefer.

Increased State and Party Control

Despite Xi Jinping's efforts to weed out rampant official corruption and to overhaul the Chinese bureaucracy (Hancock, Hornby, & Wildau, 2018), intra-Party threats remain a critical source of insecurity for China's top leaders. The CCP's recent removal of constitutional term limits for the presidency and sidelining of seniority norms have disrupted cadres' expectations of their career prospects, paving the way for intensified factional infighting. The current climate of heighted uncertainty is likely to exacerbate instability in not only the top leadership, but also in civil and military bureaucracies and local governments.

Elite divisions, as experts have observed, tend to create incentives for factions and sub-state actors to take more hardline foreign policy positions in order to avoid accusations by rivals for failing to protect national interests (Jakobson, 2014; Reilly, 2013). However, just as AI can be used to increase control over society, it can also be deployed to more closely monitor and dictate the behavior of the many actors within China's sprawling Party apparatus. If top leaders can better identify and eliminate potential rivals, foreign policy will no longer be susceptible to factional competition.

Furthermore, AI could facilitate the CCP's efforts to constrain the autonomy of bureaucratic agencies and local governments. Increased control over sub-state actors will lead to a more coordinated foreign policy, minimizing the chances of *faits accompli* by disobedient sub-state actors. Greater

[39] See Weiss (forthcoming) for a counterargument about the impact of public opinion on China's foreign policy.

foreign policy coordination will reduce the likelihood of inadvertent crises and will allow China to send clearer signals to adversaries, decreasing the chances of miscalculation and misperception.

At the same time, however, increased control over actors within the Party and state system risks silencing elite debates on foreign policy issues. Foreign policy advisers and other elites may engage in self-censorship and are likely to become even more fearful of offering critical assessments of the CCP's foreign policies. Leaders operating within such repressive environments are less likely to learn from foreign policy failures, and will be more prone to making avoidable mistakes.[40] Therefore, even if AI is able to limit bias in decision-making, if self-censorship means that the inputs that inform decision-making are biased or fabricated, then foreign policy decisions may be more rather than less prone to human error.

In sum, if AI enables China's leaders to harness control over other elites, foreign policy will be less susceptible to factional competition and narrow bureaucratic interests. However, tightened intra-Party control could stifle internal dissention and policy evaluation. Under such circumstances, even a well-oiled foreign policy machine will be prone to making avoidable mistakes.

Conclusions

AI, by strengthening the CCP's ability to control state and societal actors, could be a potential game-changer for China's foreign policy. Because internal threats to regime security have long shaped China's external behavior, it is not immediately clear what China's foreign policy would look like if *not* dictated by domestic concerns. A more internally secure CCP regime, as discussed above, might for different reasons be less prone to pursue either international cooperation or conflict.

What is clear, however, is that if AI silences internal debate, China's leaders will be less likely to critically examine policy decisions and make course corrections in their foreign pursuits. Even if AI facilitates greater foreign policy coordination, it will lead to a more personalized foreign policy. Therefore, whether China engages in more or less confrontational international behavior will depend upon the top leaders' foreign policy aspirations and their vision for China's position in the world.

References

Christensen, T. J. (2011, March/April). The advantages of an assertive China: Responding to Beijing's abrasive diplomacy. *Foreign Affairs.* Retrieved from https://www.foreignaffairs.com/articles/east-asia/2011-02-21/advantages-assertive-china

Fravel, M. T. (2005). Regime insecurity and international cooperation: Explaining China's compromises in territorial disputes. *International Security, 30*(2).

Hancock, T., Hornby, L., & Wildau, G. (2018, March 13). China revamps bureaucracies as Xi tightens grip. *Financial Times*. Retrieved from https://www.ft.com/content/3c0d4596-2666-11e8-b27e-cc62a39d57a0

[40] See Woods, Lacey, & Murray (2003) and Van Evera (2002) for examples of the role of self-censorship and misinformation on foreign policy decision-making.

He, Y. (2007). "Remembering and forgetting the war: Elite mythmaking, mass reaction, and Sino-Japanese relations, 1950-2006." *History & Memory*, *19*(2), 43–74.

Jakobson, L. (2014). China's unpredictable maritime security actors. Lowy Institute. Retrieved from https://www.lowyinstitute.org/sites/default/files/chinas-unpredictable-maritime-security-actors_3.pdf

Reilly, J. (2013). Strong society, smart state: The rise of public opinion in China's Japan policy. Columbia University Press.

Shirk, S. (2007). *China: Fragile superpower*. Oxford: Oxford University Press.

Van Evera, S. (2002). "Why states believe foolish ideas: Non-self-evaluation by states and societies." Manuscript. Retrieved from https://core.ac.uk/download/pdf/4382722.pdf

Wang, Z. (2012). Never forget national humiliation: Historical memory in Chinese politics and foreign relations. Columbia University Press.

Weiss, J. C. (forthcoming). "How hawkish is the Chinese public? Another look at "rising nationalism" and Chinese foreign policy."

Woods, K., Lacey, J., & Murray, W. (2006). Saddam's delusions: The view from the inside. *Foreign Affairs*. Retrieved from https://www.foreignaffairs.com/articles/iraq/2006-05-01/saddams-delusions

Chapter 16. Chinese Regime Insecurity, Domestic Authoritarianism, and Foreign Policy

Rachel Esplin Odell
Massachusetts Institute of Technology
rodell@mit.edu

Abstract

The Chinese Communist Party's increased dependence on AI-related technologies to monitor and control its population has been described in the West as a prime manifestation of China's drift toward centralized and monolithic authoritarianism over the past five to ten years. Too often, however, Western narratives about China fail to perceive the source of the CCP's authoritarianism: a deep-seated insecurity about its ability to effectively maintain and exercise power as it seeks to reform its economy in order to ensure long-term growth. This misperception directly contributes to a false conflation of two key themes in Western understandings of China: the character of China's grand strategy and its domestic authoritarianism. China's illiberalism at home is assumed to infuse its international ambitions, leading it to challenge the existing liberal international order in areas as diverse as trade, international development, and maritime security. If the West were to instead recognize China's foreign policy behaviors in these areas as largely status quo-supporting efforts to grapple with its need for ongoing economic growth to satisfy an ambitious and restive populace, it could craft more effective, positive-sum policies in response.

Western Narratives on China Fail to Grasp the Core Chinese Motivation

Western narratives about China in the past five to ten years, especially during the tenure of Chinese president Xi Jinping, have often stressed the authoritarian entrenchment of the Chinese Communist Party (CCP). Initiatives such as the anti-corruption campaign and the reform of the People's Liberation Army are seen as moves by Xi to consolidate his power, while China's suppression of internet and press freedom and persecution of ethnic and religious minorities are identified as examples of authoritarian entrenchment (Schell, 2018; The New York Times Editorial Board, 2018b). The CCP's increased dependence on AI-related technologies to monitor and control its population has been heralded as a particularly ominous manifestation of this increased tendency toward centralized and monolithic authoritarianism (Campbell & Ratner, 2018; Wright, 2018).

Although accurately describing many of the repressive consequences of Chinese political developments, these narratives frequently misinterpret such repression as evidence of Chinese strength rather than weakness. Such misperception arises from a failure to grasp the core motivation of the Chinese regime. In fact, the CCP's efforts to shore up its control at home reveal a deep-seated insecurity about the ability of the party to effectively maintain and exercise power as it seeks to reform its economy and ensure sustainable long-term growth. In order to escape the so-called middle-income trap, wherein developing countries sucessfully rise out of extreme poverty only to plateau at per capita income levels below $12,000, China will need to undertake difficult economic reforms that shift economic growth strategies away from reliance on low capital and labor costs and heavy government investment and toward an emphasis on higher productivity, efficiency, and innovation. Such a shift requires painful and disruptive reform that alienates established sectoral interests and state-owned enterprises (Overholt, 2018).

Despite the challenges associated with such reforms, many current Chinese leaders nonetheless

perceive them to be necessary. This perception is to a large degree driven by the fact that CCP legitimacy since its post-1978 opening and reform has largely depended on its ability to promote strong economic growth and improve the quality of life for the Chinese people, as opposed to its adherence to the rule of law or any particular civic creed (Shirk, 2007). Without reforms, the CCP's ability to continue satisfying that demand for high performance could be crippled in the medium to long term.[41] At the same time, the CCP's reliance on performance-based legitimacy also imbues it with a paranoid fear that if it somehow fails to perform, it will be faced with massive unrest that could topple the regime. It is this paranoia that has led Beijing to grasp at technological levers of power such as AI to manipulate the population into quiescence.

However, even though artificial intelligence may facilitate CCP management and control of the mass public, it cannot necessarily be used to stifle elite-level dissatisfaction and dissent to nearly the same degree (see also Miura, Chapter 15 this volume). And it is precisely such internal elite power struggles that CCP leaders have long perceived as among the greatest threats to regime survival, especially if dissenting elites or elite factions were to assume the mantle of leadership over mass movements whose membership is cross-cutting, i.e. drawn from different demographic segments of society (Shirk, 2007). In fact, the far-reaching anti-corruption campaign that has toppled many titans within the CCP, state-owned enterprises, and the People's Liberation Army, is inspired in part by this fear. The great irony and paradox faced by the CCP is that such efforts to consolidate power also have the potential to exacerbate problems of elite dissension and damage intraparty solidarity, further weakening the party's hold on power.

Western Narratives on China Misinterpret Chinese Grand Strategy

The West's failure to fully recognize the insecurity of the CCP has led it to misperceive Chinese authoritarianism—epitomized by the CCP's ambitious program to use AI to monitor and control its populace—as evidence of Chinese strength rather than weakness. This misinterpretation has in turn influenced Western and U.S. perceptions of Chinese foreign policy and grand strategy.

Western narratives about Chinese foreign policy have moved from stressing the "assertiveness" of China in the post-2008 financial recession period (Johnston, 2013; Chen, Pu, & Johnston, 2013; Swaine, 2011) to fretting about the implications of the Chinese Dream concept promulgated during the Xi Jinping era (Callahan, 2016; Thayer & Friend, 2018). Concerns over the past 15 years about a growing "Beijing Consensus" (Ramo, 2004; Halper, 2010)[42] have now been augmented with concerns that China is seeking to supplant the "rules-based liberal international order" with an alternative and mercantilist web of institutions and economic ties (Carter, 2015; Trump, 2017; The New York Times Editorial Board, 2018a).

Concerns about Chinese power in the United States in particular have intensified as Western observers have come to interpret China's foreign policy through the lens of its domestic repression (Allison, 2017; Campbell & Ratner, 2018; The New York Times Editorial Board, 2018b; Diamond & Schell, 2018). Xi astride the CCP has become a metaphor of sorts for the monolithic and illiberal power that China exerts in the world. Xi's efforts to strengthen party control over China through the

---

[41] For a theoretical account of the political instability and unrest that can result when societal welfare is improving, only to be followed by a rapid reversal, see Davies (1962).

[42] This term was coined by Ramo (2004) to describe Beijing's model of economic development as an alternative to the traditional "Washington consensus" on economic growth. It is generally seen as referring to pragmatic, technocratic, state-led, rapid economic growth undergirded by authoritarian politics.

accretion of titles to himself and the dismantling of competing power networks within the country are seen as reflective of China's efforts to expand its influence in other countries and regions of the world in a zero-sum competition with other regional and global powers.

But there is a danger when Western governments treat China as a powerful behemoth whose policies are strategically calculated and deliberately executed components of a grand strategy of supplanting the West through the imposition and promulgation of an authoritarian Chinese model of national and global governance. The misplaced fear and zero-sum thinking embedded in such perceptions leads Western governments to adopt policies toward China that exacerbate the security dilemma between China and the West. This dynamic unfolds in part by weakening moderates within China who favor constructive positive-sum engagement with the West, while strengthening hardliners who use U.S. containment strategies to support their own more revisionist policy preferences. In contrast, if the West recognizes China as a deeply insecure power with an uncertain future that can be shaped in part by the strategies other states pursue toward Beijing, then Western governments' policies toward China are more likely to be sensitive to the costs of confrontation with Beijing and the benefits of collaborative global governance and non-zero-sum competition.

Three Examples of the Narratives Driving Western Policy

This mistaken Western interpretation of Chinese motivations can be illustrated by three areas of Chinese policy:

- *International trade:* China's initiatives to expand trade with other states have often been decried by U.S. strategists as counterfeits of bona fide free trade agreements, mercantilist outgrowths of Beijing's statist approach to internal economic governance.
- *International development*: China's investments in countries throughout the developing world, including projects undertaken as part of the Belt and Road Initiative, are frequently seen as deliberate efforts to undermine or replace existing development institutions with institutions that reflect its illiberal ideology.
- *Maritime security:* China's expanded presence in the South China Sea and beyond the first island chain, is seen as an effort to strengthen its military presence in order to expand its strategic periphery and underwrite a strategy of regional hegemony.

Guided by such zero-sum assessments, the West has responded to Chinese initiatives in these three areas with accusations that Beijing is seeking to undermine the "rules-based international order." Washington has married these rhetorical accusations with efforts to bolster its own military presence in the Indo-Asia-Pacific region and enhance traditional security relationships with countries in the region in an effort to build up a counterbalancing coalition to contain Beijing. Meanwhile, the United States has pursued trade agreements that were exclusive of China (under the Obama administration) before pulling out of those negotiations only to impose broad-spectrum tariffs on Beijing (under the Trump administration). It has refused to participate in and even lobbied against Chinese-led development institutions such as the Asian Infrastructure Investment Bank. It has stepped up its military surveillance and operations in the South China Sea and implicitly taken sides in China's maritime jurisdictional and territorial disputes.

Alternative Approaches

Instead, if the West were to recognize these efforts as part and parcel of China's effort to grapple with its need for ongoing economic growth to satisfy an ambitious and restive populace, it would be able

to make less alarmist interpretations of Chinese policy in all three of these areas and craft more effective, positive-sum policies in response. Such policies would more effectively enable the United States to coordinate responses to global challenges, resolve collection action dilemmas in the international arena, and reduce the likelihood of crisis escalation and conflict outbreak in East Asia.

International trade: The United States could craft a trade strategy that recognized the reality of Chinese economic strength and its advantages for U.S. and global interests, even while applying leverage to secure fairer terms for U.S.-Chinese economic exchange. Such a strategy would prioritize the negotiation of regional trade agreements that are inclusive of China and promote America's integration in an Asian regional economy that is inescapably intertwined with the Chinese economy. At the same time, it would insist on the negotiation of bilateral agreements that provide enhanced market access and strengthened intellectual property protections for American companies in China.

International development: Washington could view Chinese infrastructure and development projects in the developing world as constructive supplements to existing development institutions, even while augmenting its own efforts in this regard as a form of non-zero-sum competition. To this end, U.S. policymakers could frame the $60 billion aid and investment package that Congress recently passed in the form of the BUILD Act not as a tool for containing Chinese influence, but rather as a complementary and competitive alternative option for developing states that would press Beijing to improve the standards of its own investments.

Maritime security: Finally, in the maritime realm, America could negotiate new initiatives of cooperative maritime security with the Chinese, building upon successful inclusive crisis management and prevention mechanisms such as the Code for Unplanned Encounters at Sea (Western Pacific Naval Symposium, 2014). Washington could also seek to establish clearer mutual understandings of international maritime law with Beijing that reflect both countries' strong interests in freedom of navigation for commercial and military vessels alike.

Conclusions

Accurate assessment of other states' capabilities and intentions is the first step in successful strategy formation. Accordingly, Western governments' strategies toward China must be informed by a realistic understanding of the motivations underlying both China's high-tech domestic authoritarianism and its changing role in the world.

The United States in particular must resist the simplistic assumption that an illiberal and repressive Chinese state has no interest in a liberal international order. On the contrary, China's domestic insecurity has led it to embark on a process of reforming its economy toward consumption-driven growth bolstered by strong international commerce, while also applying its newfound power and influence to exercise constructive leadership in the world.

These developments present a positive-sum opportunity for the West to work with Beijing as a partner in crafting effective global governance institutions. If the West seizes this opportunity, it can help ensure that China's global engagement supports and complements rather than supplants and undermines the values and priorities at the heart of the postwar international order, such as trade, economic development, and freedom of navigation.

References

Allison, G. (2017). China vs. America: Managing the Next Clash of Civilizations. *Foreign Affairs*, 96, 80–89.

Callahan, W. A. (2016). China 2035: from the China Dream to the World Dream. *Global Affairs*, 2(3), 247–258. https://doi.org/10.1080/23340460.2016.1210240

Campbell, K. M., & Ratner, E. (2018). The China Reckoning: How Beijing Defied American Expectations. *Foreign Affairs*, 97(2), 60–70.

Carter, A. (2015, November). Remarks on "Strategic and Operational Innovation at a Time of Transition and Turbulence." Speech by U.S. Secretary of Defense presented at the Reagan National Defense Forum, Simi Valley, California. Retrieved from https://www.defense.gov/News/Speeches/Speech-View/Article/628146/remarks-on-strategic-and-operational-innovation-at-a-time-of-transition-and-tur/

Chen, D., Pu, X., & Johnston, A. I. (2013). Debating China's Assertiveness. *International Security*, 38(3), 176–183

Davies, J. C. (1962). Toward a Theory of Revolution. *American Sociological Review*, 27(1), 5–19. https://doi.org/10.2307/2089714

Diamond, L., & Schell, O., co-chairs. (2018). *Chinese Influence & American Interests: Promoting Constructive Vigilance*. Working Group on Chinese Influence Activities in the United States. Stanford, CA: Hoover Institution Press.

Halper, S. (2010). Beijing Consensus: How China's Authoritarian Model Will Dominate the 21st Century. New York: Basic Books.

Johnston, A. I. (2013). How New and Assertive Is China's New Assertiveness? *International Security*, 37(4), 7–48

The New York Times Editorial Board. (2018, February 5). China, Elbows Out, Charges Ahead. *The New York Times*. Retrieved from https://www.nytimes.com/2018/02/05/opinion/china-elbows-out-charges-ahead.html.

The New York Times Editorial Board. (2018, February 27). Xi Jinping Dreams of World Power for Himself and China. *The New York Times*. Retrieved from https://www.nytimes.com/2018/02/27/opinion/xi-jinping-power-china.html

Overholt, W. H. (2018). *China's Crisis of Success*. Cambridge: Cambridge University Press.

Ramo, J. C. (2004). *The Beijing Consensus*. Foreign Policy Centre.

Schell, O. (2016). Crackdown in China: Worse and Worse. *New York Review Of Books*, 63(7).

Shirk, Susan L. (2007). *China: Fragile Superpower*. Oxford; New York: Oxford University Press.

Swaine, M. D. (2011). China's Assertive Behavior—Part One: On "Core Interests." *China Leadership Monitor*, (34). Retrieved from www.hoover.org/research/chinas-assertive-behavior-part-one-core-interests

Thayer, B. A., & Friend, J. M. (2018, October 3). The World According to China. *The Diplomat*. Retrieved from https://thediplomat.com/2018/10/the-world-according-to-china/

Trump, D. (2017, December). National Security Strategy of the United States of America. Retrieved from https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf

Western Pacific Naval Symposium. (2014). Code for Unplanned Encounters at Sea (CUES) (Version 1.0). Qingdao, China. Retrieved from https://news.usni.org/2014/06/17/document-conduct-unplanned-encounters-sea

Wright, N. (2018, July 10). How Artificial Intelligence Will Reshape the Global Order. *Foreign Affairs*. Retrieved from https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order

Chapter 17. The International and Foreign Policy Impact of China's AI and Big Data Strategies

Rogier Creemers
Leiden University
r.j.e.h.creemers@law.leidenuniv.nl

Abstract

In the past few years, China has embarked upon an ambitious strategy to build up its capabilities in AI and big data. The primary aims for this agenda are domestic: transforming the government's social management and governance abilities, and creating new areas for economic growth. Nonetheless, this agenda also has an international impact, both in terms of foreign governments' responses to China's domestic strategy, and the extent to which Chinese technologies are exported or become part of global cyber processes. This chapter will review the development of this agenda, and assess its impact for China's foreign policy.

Introduction

Since the 1990s, China has embarked on an ambitious agenda of "informatization" (*xinxihua*), which seeks to upgrade social, economic and political processes through the use of information technologies (ITs). This agenda originally mainly covered the digitization of existing government information, as well as increasing interagency openness, interoperability and information sharing. In recent years, however, these ambitions have grown to encompass new forms of systems, which are driven by big data technologies and AI. The goals for this agenda include automating decisionmaking and social management, attaining a global leadership role in these strategic technologies, and fostering new forms of economic growth.

In the light of the greater push for domestic state control, and China's growing assertiveness on the international stage, which have together characterized the Xi Jinping era, the question is: what will be the international impact of these plans? Information technology plays an increasingly central role in the growing tensions in the relationship between China and the US,[43] as well as – although somewhat less acrimoniously – European countries,[44] This paper will therefore explore how the Chinese data and AI agenda have become significant elements of competition and concern in these relationships, as well as how China's plans will be influenced by international responses.

How is China's AI and Data Agenda Developing?

Since the early days of the Xi Jinping administration in 2012, the Chinese leadership has started to pay considerably greater attention to information technology than it did in the past. Although this focus started out in the realm of social media, it rapidly spread to encompass areas as diverse as smart manufacturing, predictive policing and surveillance, social management and financial reform.

---

[43] Zhong, R. and Mozur, P. (2018, 23 March). For the U.S. and China, a Technology Cold War That's Freezing Over. *New York Times*, retrieved from https://www.nytimes.com/2018/03/23/technology/trump-china-tariffs-tech-cold-war.html

[44] Cerulus, L. (2018, 9 May), Europe Turns Cool on Chinese Tech. *Politico*, retrieved from https://www.politico.eu/article/europe-reaches-moment-of-reckoning-over-5g-security-huawei-us-donald-trump-cybersecurity/

On the one hand, China has come to see information technologies as the driver of a "fourth industrial revolution," in which it can rapidly attain global leadership and for which its domestic environment is well suited. Yet on the other hand, China remains lagging behind in required elements ranging from operating systems and core chipsets to security software, market power for its companies and discursive influence at the global governance level.

Institutionally, this growing importance is reflected in the creation of new leadership bodies. Most notable amongst these are the Central Commission for Cybersecurity and Informatization,[45] established in 2014, and the Cyberspace Administration of China, which is in charge of daily policy coordination and has some direct regulatory powers. It has also resulted in the publication of a series of high-level policy document, including a dedicated Five-Year Plan,[46] the "Internet Plus" plan[47] to modernize traditional economic sectors, a national big data strategy[48] and a national AI strategy.[49] Private businesses have also rapidly expanded their capabilities, evidenced by the construction of large, platforms combining functions ranging from e-commerce and online payment to ride sharing and online dating, by businesses such as Alibaba and Tencent, and the establishment of specialized AI labs at home and abroad, most notably on the American West Coast, by Tencent and Baidu.

These efforts have already led to preliminary achievements. For instance, Tencent has launched a healthcare programme, Miying, which assists medical professionals in making diagnoses and better integrating patient data. At the governmental level, policing and surveillance increasingly rely on facial recognition software, with businesses even working on tools enabling identification of an individual's gait. Under the various development plans, considerable funding has been made available to expand data and computing curricula in higher education, support R&D in businesses, and – through government procurement – create a market for government-oriented applications. Alibaba is building the second iteration of its "City Brain" technology, to assist with traffic management and emergency response, in Hangzhou.

Nonetheless, many of the goals in this plan, thus far, remain exactly that: future goals. The governmental social credit system, a project intended to amplify currently underperforming law enforcement mechanisms, for instance, is often touted as the posterchild example of a big-data driven model of constant autocratic surveillance and control. However, it currently does not seem to contain any automated decision-making, machine learning or big data analysis processes at the moment.[50] Similarly, barring specific military or intelligence applications for AI in areas such as image recognition, these systems are completely absent from assistance in foreign policy-related decision-

---

[45] The Central Commission is a Party body, chaired by Xi Jinping personally, which includes all senior officials whose portfolio concerns technology, including the military. It coordinates the technology agenda across bureaucratic boundaries, and sets overall policy directions.

[46] State Council. (2016, 15 December). 13th Five-Year Plan for National Informatization, retrieved from http://www.gov.cn/zhengce/content/2016-12/27/content_5153411.htm

[47] State Council, (2015, 1 July). Guiding Opinions concerning Vigorously Moving Forward the "Internet Plus" Plan. Translation retrieved from https://chinacopyrightandmedia.wordpress.com/2015/07/01/state-council-guiding-opinions-concerning-vigorously-moving-forward-the-internet-plus-plan/

[48] State Council, (2015, 31 August). Outline of Operations to Stimulate the Development of Big Data. Translation retrieved from https://chinacopyrightandmedia.wordpress.com/2015/08/31/outline-of-operations-to-stimulate-the-development-of-big-data/

[49] State Council, (2017, 20 July). New Generation Artificial Intelligence Development Plan. Translation retrieved from https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

[50] Creemers, R. (2018). China's Social Credit System: An Evolving Practice of Control. *Working Paper*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3175792

making. Consequently, the extent to which data and AI technologies shape China's foreign policy often largely depend on perceptions, rather than realities.

Rightly or wrongly, both China and the United States have come to see AI as one small set of disruptive, transformational zero-sum capabilities. For China, it is an area in which the leadership believes it can rapidly acquire competitive status with the US. For its part, the US believes AI to be essential in maintaining its military and economic advantage over China. With regard to data, China and Europe have both come to the conclusion that the considerable economic and strategic value of data require greater national control and regulation.[51]

To what Extent are Chinese AI and Data Technologies Internationalizing?

If the implementation of Chinese AI and sophisticated big data analysis capabilities remains largely incipient domestically, they are also largely absent – so far – from the regional or global stage. This contrasts with the rapid expansion of China's traditional telecommunications exports: Huawei in particular has become the world's largest supplier of telecommunications infrastructure components and handsets. To a significant degree, this is due to the fact that they remain under development or only operate at the trial stage, or are geared specifically to Chinese local needs and require further adaptation for international applicability. Natural language-based AI tools developed for Mandarin Chinese, for instance, are of little import outside Greater China. It is also unclear what the priority for exporting these technologies is in the wider background of China's overall foreign policy. That said, various non-democratic governments have expressed interest in China's technological social management and control capabilities, and may be interested in acquiring them in the future.

From a governmental policymaking perspective, perhaps the most important future potential for these tools lies in the Belt-Road Initiative (BRI). While BRI started out mainly focusing on infrastructure and logistics, ITs have gained increasing prominence in recent years, under the guise of the Digital Silk Road. The idea behind this is to add a layer of digital connectivity and smart applications to the more traditional infrastructure-oriented BRI. At the 2017 Wuzhen World Internet Conference, China's showcase event for its digital sector, several governments from the region jointly concluded a "Proposal for International Cooperation on the "One Belt, One Road" Digital Economy," promising greater collaboration on matters such as connectivity, smart cities, and telecommunications.

Nevertheless, as with the BRI more broadly, the Digital Silk Road sometimes comes across as an overly broad and abstract list of intentions, without clear prioritization and sometimes questionable support from target governments. The Digital Silk Road is deemed to play many roles, including providing new markets for China's online giants and hardware providers, providing territory for the expansion of China's homegrown Beidou navigation system, and regulatory integration in support of the developing e-economy. In the 2017 Wuzhen document noted above, AI is only mentioned as one potential area for collaboration between smart cities, while big data is not mentioned at all.

[51] In Europe, this is best reflected in the General Data Protection Regulation , which came into force in 2018. China included certain aspects of data protection in the 2017 Cybersecurity Law. Subsequently, the CAC's subordinate standardization committee has issued a raft of technical data protection standards, and specializzed legislation is reportedly in an advanced stage of drafting.

For the sake of analytical clarity, it is probably most useful to gauge the future export potential of data and AI technologies in two streams: commercial and political.

- *Commercial:* It is very likely that China's online giants, such as Tencent or Alibaba, will continue to expand in the region. Their development trajectory in China has prepared them well to deal with environments characterized by less wealthy customers, suboptimal infrastructure, and underdeveloped market circumstances. They also provide products that may be attractive to governmental users. Kuala Lumpur is in the process of introducing Alibaba's "City Brain" for traffic management.[52]
- *Political:* However, this is different from China pushing particular social management approaches, and their concomitant technological underpinnings, onto other governments. Not only would this stretch China's expressed commitment to non-interference in foreign governments, there is also a risk that it might backfire by galvanizing anti-China opposition in these countries. Nonetheless, one objective of the Digital Silk Road project is regulatory harmonization for the digital economy, as well as infrastructure construction and interoperability. With these building blocks in place, BRI governments themselves might be attracted to aspects of China's social management capabilities, and seek to obtain them.

Outside of the Digital Silk Road, Chinese businesses, often with government support, do seem to have become more engaged in investments in foreign data and AI start-ups in the past five years. One example is the US. Baidu, Alibaba, and Tencent, as well as Chinese Venture Capital funds are increasingly active in investing in budding Silicon Valley businesses. Some of these have obtained US government contracts. In some cases, the ownership structure of Chinese investors is opaque. Sometimes, investments were blocked by an increasingly watchful Committee on Foreign Investment in the United States. It is not always clear what the risk of potential Chinese investments is, often because it is far from certain a particular start-up will succeed, or the future potential ramifications of technology are unknown, but that does not mean the risk you should be discounted. Moreover, Chinese funding often provides opportunities for US tech businesses that might otherwise not be forthcoming. It is thus difficult to assess the balance between risks and benefits arising from such investments, but it is also worth considering which actions the US could actively take to reduce the influence of Chinese investment where deemed harmful, while stimulating the continued development of US tech capabilities.

How do these Developments Affect China's Broader Relations with the US and EU?

The increasing heavy-handedness of the Xi administration is undoubtedly one of the major factors in the souring of the US-China relationship, and technology has become the core arena where this tension manifests itself.[53] Key in this process has been a shift of perception of technology: until recently, it was largely seen as an economic matter, with interests largely confined to the field of international trade. Yet both Beijing and Washington have recently come to see technology as crucial

[52] Alibaba (2018, 29 January). Alibaba Cloud Launches Malaysia City Brain to Enhance City Management. Retrieved from https://www.alibabacloud.com/press-room/alibaba-cloud-launches-malaysia-city-brain-to-enhance-city-management

[53] See, for instance, Pence, M. (2018, 4 October) Remarks by Vice President Pence on the Administration's Policy Toward China. Retrieved from https://www.whitehouse.gov/briefings-statements/remarks-vice-president-pence-administrations-policy-toward-china/

to national security interests as well. Both countries see the competition in AI and data technology as a zero-sum game in their foreign policy interactions.[54]

Therefore, both the US and China will likely seek to reduce their future reliance on each other, potentially disentangling the highly globalized tech industry. Instead, China will seek to ever more rapidly achieve parity with the United States on technological capabilities, while the US will likely become ever warier about the Chinese role in technology supply and financing chains. It is also likely there will be greater scrutiny on investments in both directions, but particularly in the United States. Insofar as these technologies are used in some of the Chinese state's more repressive operations, such as in Xinjiang, they may also serve to galvanize US public opinion to be ever more critical of China. China in return will come to see the US as ever more of an existential adversary, bent on preventing its rightful rejuvenation.

Data sovereignty is another area in which such tension will manifest. In this area, it is the European Union that has, thus far, made the most impactful policy moves, most notably the General Data Protection Regulation. The EU is not only motivated by concerns about the economic value of data, but also the privacy rights of its citizens. While it acts on the basis that there are nearly no significant European online businesses, it is likely that this concern about privacy will also influence the EU's relationships with China, which promises to play an increasingly important role in the global data economy. For reasons different to Europe, and encompassing political and ideological security, China holds similar concerns, but has – so far – gone less far in regulating the export of data. Under the Cybersecurity Law, only "important data" held by critical infrastructure operators is required to be stored on Chinese servers. So far, these provisions have remained vague. Nonetheless, new legislation and new industry standards to provide further clarity are in the drafting stage.

Conclusions

Recently, digital technology has shifted from being largely an economic and trade matter, to one also crucial to national security among all major players. AI and big data are core elements of that process. This will likely have a severe impact on the development of the tech industry. As major players will seek to develop indigenous capabilities and reduce mutual interdependence, the highly integrated and globalized value and supply chains that exist today will likely come under threat.

One option is that the Internet will become feudalized: that increasingly countries, businesses and consumers will have to choose between Chinese and Western infrastructure, gadgets and technologies. While these technologies may remain interoperable, rather like an Apple user rarely switches to Android or vice versa, a certain amount of path dependency will emerge. This, in turn, will influence the extent to which countries are able to maintain decision-making power about how major businesses will use and process citizens', data, but also how they govern their societies.

Moreover, it must not be simply assumed that the "like minded" axis (US-Europe) will hold on its own. There are considerable differences between the EU and the US on privacy and data security: the GDPR primarily targets conduct of American businesses in the EU marketplace, while the European parliament has long since evinced a skeptical attitude vis-à-vis the American approach. If this axis is

[54] See, for instance, Trump trade advisor Peter Navarro: 'Zero-sum game' between China and the rest of the world' (2018, 19 July). Retrieved from https://www.cnbc.com/2018/07/19/peter-navarro-zero-sum-game-between-china-and-the-rest-of-the-world.html. Similar sentiments have been expressed by Xi Jinping in speeches on digital development.

to hold, and form the nucleus of a sustainable, open and secure approach to data and AI technologies, then self-awareness and the development of robust, responsible and inclusive policies for the digital world will be necessary.

PART V. Military Dimensions

Chapter 18. A Hacker Way of Warfare

Martin Libicki
USNA
libicki@usna.edu

Introduction

With the establishment and independence of United States Cyber Command (USCYBERCOM), the United States declared that cyberspace is a contestable medium to be fought over as necessary to defend the nation and its allies. Less noted is that what the United States has done is to recognize a hacker way of warfare that, while not wholly novel, is broader than cyberspace and likely to grow in importance as militaries invest heavily in complex systems, especially those that employ artificial intelligence (AI) to draw conclusions and make decisions.

Hackers search for vulnerabilities in target systems so that they can exploit them. It is a system's features that allow others to feed it a specific set of inputs and thereby make it critically misbehave. What makes vulnerabilities militarily relevant is when small doses of effort can create potentially large effects undesired by the system's owner.

Just as the concept of a system predates computers, so too does the concept of a system's vulnerabilities. The vulnerabilities need not be expressed in code, and their exploitation does not necessarily involve remote code execution (wherein a hacker can persuade a system to run arbitrary code of the author's design). Nor need the system be electronic or even mechanical. A system can be comprised of individuals in an organization, a society composed of interacting individuals and institutions, or an organization interacting with its environment. The only requirements are that a system have parts, that the parts interact, that the system respond to inputs, and that the system have outputs. An Army division is a system. It has components. The components interact with one another in wartime (and peacetime). It responds to inputs such as commands or enemy actions. And it puts out, broadly speaking, presence and force.

Looking for a system's vulnerabilities has long been the leitmotif of some great military commanders. Think of Napoleon before Austerlitz understanding the frictions between coalition opponents or their relationship with terrain in the context of weather conditions. Think, too, of Nelson who intuited vulnerabilities in French naval command-and-control and the relationships between their ships, the terrain, and wind during the Battle of the Nile. Although a vulnerability alone does not guarantee victory, their existence provides a distinct edge—as long as the side spotting it can generate a worthwhile exploit of the vulnerability and commanders can intelligently judge which among the various exploits offered merit use.

AI and a New Character of Conflict

Since military commanders may profit from discovering and exploiting vulnerabilities today as they did two hundred years ago, how is this a *new*, or at least a newly enhanced, way of warfare? The answer is that military forces are increasingly complex and increasingly reliant on *automated* systems, which, themselves are become more complex. Furthermore, while the vulnerabilities that

traditional commanders looked for existed in the evanescent interaction of units vis-à-vis each other, terrain and weather, today's vulnerabilities tend to be more persistent because they are reified in systems whose parts lie in stable or at least predictable relationship to one another. Thus, the search for vulnerabilities can be delegated to specialists, who can take the time to painstakingly understand how the system works, discern the presence of structural elements that allow potentially unexpected and generally unwanted responses to inputs, and develop ways of inducing them—all confident in the proposition that vulnerabilities they find today will likely still be there tomorrow. In other words, the search for vulnerabilities is becoming a profession, one presaged by Bletchley House's work against Enigma but now executed by those selected and trained to do just that.

One reason that the broad applicability of hacker warfare is not obvious lies in the peculiar nature of cyberspace operations at this time. Most hacking works in one of two ways (or both together): credential hijacking and malware. In either case, the target computer is carrying out the instructions of another: in the first case manually and in the second case automatically. Both methods are preventable with high – albeit not complete – degrees of assurance. Acquiring the credentials of another can be prevented through multi-factor authentication and machines can be prevented from running rogue instructions either by ensuring that such instructions are taken only from hardware, or by ensuring that new instructions come from authenticated sources (as Apple's iOS does). Again, getting the implementation right is non-trivial. But if the threat from hackers is a war-losing risk and understood as much, it is hard to imagine current intrusion methods gaining much traction in the long run.

Yet, injecting arbitrary instructions into opposing systems to make them behave unexpectedly is not the only way that hackers could work. Machines can be given specially constructed inputs designed to induce unexpected responses from them. SQL-injection techniques entail creating cleverly formed queries for databases; if the server does not scrub inputs carefully, a hacker can induce it, for instance, to dump private records to unauthorized recipients. During this process, the hacker is *not* impersonating an authorized user because, for many web sites, everyone is an authorized user—and nor is the server necessarily running code other than what it was programmed with. Instead, by giving the server anomalous inputs, the hacker causes the server to exhibit unexpected and unwanted behavior. Similarly, unexpected inputs have allowed people to cheat gaming (e.g., poker) machines (Poulsen, 2014; see also Koerner, 2017). There are no straightforward fixes for such flaws because the range of potential inputs grows combinatorically. Indeed, unexpected inputs can be the most important from which an AI can learn about environment because they may contain the largest amount of new information with which the AI can update its model of the environment—a fundamental concept in computational neuroscience known as "prediction error" (Wright, 2014). No simple solution to this dilemma exists for humans, other animals or AIs.

AI and Increased Exploitable Vulnerabilities

Artificial intelligence—the substitution of machine logic for human cognition—is likely to *increase* the opportunities for exploits that use specially crafted inputs to produce unexpected outputs. As AI develops, machines, like people do now, would acquire information, process it, and reach conclusions from whence decisions. Over time, machines do more and people do less. Although people have weaknesses—they are slower, error-prone, and expensive—they come with their experience-driven intuition, a tolerance for ambiguity, a talent for making decisions in the face of uncertainty, multiple ways of looking at the same problem, and an instinct for collaboration. Machines (even neural net machines) lack true intuition and learn poorly from the open world (machine learning requires healthy doses of prior data, often carefully abstracted). They do not take well to deception unless trained to recognize the specific forms it might take. Systems with large amounts of artificial

intelligence in them are built to generate high degrees of reliability over a range of inputs chosen to simulate the environment in which they would work. If the environment is rarely manipulated by malefactors—for instance, of using AI in processing customer orders, maneuvering in traffic, or raising crops—then statistical methods can produce reliable results. But deliberate and mischievous manipulation of the environment, "adversary machine learning," can mean that responses based on statistical methods can mislead. An agile adversary will look to present its foes with "edge cases"[55] that can fool or at least stymie an AI-based system. This has been done in laboratories: busses have been subtly manipulated to look like ostriches and turtles have been made to look like rifles (Gershqorn, 2017). Google's Ian Goodfellow and others have shown that just by changing a few pixels in the photo of elephant, for example, they could fool the neural network into thinking it depicts a car (Metz, 2017). Although humans can also be fooled by manipulated images, the true test in both cases is whether such images can be placed into a realistic battlefield environment (Ackerman, 2018).

By way of hypothetical example, if a robot has been trained to avoid entering streams, and these streams are identified by how light reflects off them, and these reflections can be simulated by aluminum foil, then the combination, once discerned, allows the other side to stymie oncoming robot onslaughts using aluminum foil—which is a lot easier to lay down than it is to create afresh a stream with real water flowing in it. Building an AI engine that can cope with deliberately induced edge cases requires guessing these edge cases in the first place. As with computer hacking, the attacker starts with an advantage: the defender has to identify and neutralize all the problematic edge cases while the attacker only has to find one problematic edge case to start working on an exploit.

One countermeasure is to train each AI slightly differently, so that induced failures are limited, thereby preventing a total collapse of an overall military endeavor. But it is unclear how often AI trainers will employ deliberate randomization. It adds costs, and reduces the predictability of results, which, in turn, hinders error measurement and diagnosis. Furthermore, bureaucracies favor uniformity for many reasons, not least being that it makes it easier to monitor performance.

When one abstracts the art of hacking to a game of inputs and outputs, it becomes clear that the concept of "hackers" is not limited to the digital realm. Electronic warfare can also be a playing field for hackers. It is filled with spoofing techniques, many of which are specific to particular classes of radar; it, too, is driven by the exploitation of vulnerabilities such as small seams or gaps around electrical connections and shielding (Wright, Grego, & Gronlund, 2005).

An even broader notion of hacking relates to vulnerabilities that exist within an organization's or a society's dynamic. "Michael V. Hayden, who served as CIA director under President George W. Bush, has described the Russian interference as the political equivalent of the Sept. 11, 2001, attacks, an event that exposed a previously unimagined vulnerability" (Miller, Jaffe, & Rucker, 2017). As for Russia's intervention into the U.S. Presidential election season, the clever insertion of fake news, twitter-bots, hot-button political advertisements and leaked data (some of which was falsified between theft and reportage) achieved the level of success it did because it exploited vulnerabilities in the U.S. body politics which had already resulted in rising polarization. Most psychological operations that succeed trade on the tendency of other leaders to believe in their preconceived notions (e.g., that the Allies would invade Fortress Europe near Calais in 1944); inserting small dollops of misinformation in such cases (e.g., Operation Fortitude) can counteract the weight of

[55] An edge case can be described as a problem or situation that occurs only at an extreme (maximum or minimum) operating parameter.

inconvenient facts and leave leaders reinforced in their misconceptions.

The greater the importance of finding vulnerabilities in adversary systems the more work there is for the intelligence community. How, for instance, would the United States get its hands on an adversary system in order to look for its vulnerabilities? Some items can be acquired through back channels – but if the vulnerabilities in these systems exist in software and the software is frequently updated, then some way to get software updates is needed. Other items can be characterized by analyzing signals intelligence but while that may provide hints of what edge cases look like, such hints are just hints. More direct actions may become necessary. One is hacking into the target system itself. Another is hacking into simulation machines that exist to facilitate machine learning, experimentation, or training; while inside the hackers may be able to run cases to test their theories but they have to be subtle. The larger the universe being simulated, the more potential doorways there are to hack – and a simulation that actually incorporates the real virtual world is practically inviting a hack.

A China-US Challenge

All this should introduce a bit of caution into those developing AI. The United States and China are said to be in a race to command the AI heights for both peacetime and wartime uses (Lee, 2018). Peacetime uses rarely have to contend with those trying to pervert systems (criminal activity notwithstanding), but for wartime uses one always has to contend with adversarial manipulation of inputs. Too rapid and uncritical an embrace of AI can create subtle flaws whose exploitation can be disastrous for the possessor. The Chinese, for instance, have a fixation on finding an "assassin's mace", which is a term used frequently in Chinese writings to denote a weapon that provides a generally inferior force a way to stymie an opponent. It has been variously applied to Chinese hypersonic anti-ship cruise missiles, anti-satellite weapons, and electronic-magnetic pulses. A Chinese fixation on finding an "assassin's mace", coupled with their belief that their forces always do what they are told, may make them particularly heir to disappointment when their AI is exploited.

A few caveats are in order

Hackers are never going to become a very large part of any military; theirs is an inherently specialized hence elite activity. And while there are likely to be disproportionately talented people in that profession worthy of promotion, expectations that such military officers would constitute a disproportionate percentage of the general officer corps should be tempered; what they do does not necessarily prepare them for leadership over people doing quite different things.

Second, hackers constitute part of a longer chain that is no stronger than its weakest link. As noted, someone needs to acquire the information on a system. The vulnerability needs to be exploited, and sometimes the effects of the resulting exploit (e.g., robots stymied while traversing terrain) must be exploited (e.g., through conventional military maneuver), and so on. Not every vulnerability leads to a usable exploit and not every exploit is cost-effective. In the Civil War's Battle of the Crater, the Union's attempts to exploit an induced vulnerability (an underground explosion under Confederate lines) led to disaster. Finally, commander's discretion is important. Using an exploit at one point may make it difficult to use something similar at a later more critical point. Some exploits may have side-effects, feed unwanted narratives, or loosen self-imposed restraints held to by the other side. An exploit may also call attention to vulnerabilities to which one own's side is heir.

Conclusions

The hacker way of warfare is no substitute for armed force—but it can be a critical force multiplier. As argued, while those comfortable manipulating 0s and 1s populate the ranks of hackers, there is a larger principle at work: complexity—especially AI-powered complexity—gives rise to vulnerabilities, whose discovery and exploitation can leverage small units of force to large effect.

References

Ackerman, E. (2018, February 28). Hacking the brain with adversarial images. Retrieved from https://spectrum.ieee.org/the-human-os/robotics/artificial-intelligence/hacking-the-brain-with-adversarial-images

Gershqorn, D. (2017, November 2). Your computer thinks this turtle is a rifle," November 2, 2017; https://qz.com/1117494/theres-a-glaring-mistake-in-the-way-ai-looks-at-the-world/.

Koerner, B. (2017, August 5). Meet Alex, the Russian casion hacker who makes millions targeting slot machines. Wired. Retrieved from www.wired.com/story/meet-alex-the-russian-casino-hacker-who-makes-millions-targeting-slot-machines/

Lee, K. (2018). AI superpowers: China, Silicon Valley and the New World Order. New York: Houghton Mifflin Harcourt.

Metz, C. (2017, August 13). Teaching A.I. systems to behave themselves. New York Times. Retrieved from https://www.nytimes.com/2017/08/13/technology/artificial-intelligence-safety-training.html

Miller, G., Jaffe, G., & Rucker, P. (2017, December 14). Doubting the intelligence, Trump pursues Putin and leaves a Russian threat unchecked. Retrieved from https://www.washingtonpost.com/graphics/2017/world/national-security/donald-trump-pursues-vladimir-putin-russian-election-hacking/

Poulsen, P. (2014, October 7). Finding a video poker bug made these guys rich -- then Vegas made them pay. Wired, Retrieved from http://www.wired.com/2014/10/cheating-video-poker/

Wright, D., Grego, L., & Gronlund, L. (2005). The physics of space security: A reference manual. Cambridge, MA: American Academy of Arts and Sciences. Retrieved from https://www.amacad.org/publications/Physics_of_space_security.pdf

Wright ND, Neural prediction error is central to diplomatic and military signalling (2014) in DiEuliis D, Casebeer W, Giordano J, Wright ND, Cabayan H (Eds) White paper on Leveraging Neuroscientific and Neurotechnological (NeuroS&T) Developments with Focus on Influence and Deterrence in a Networked World, US DoD Joint Staff

Chapter 19. Escalation Risks in an AI-Infused World

Herbert Lin
Stanford University
herblin@stanford.edu

Abstract

This chapter focuses on some of the potential downsides of AI-enabled military systems, specifically risks that arise from the potential of such systems to lead to conflict escalation: deliberate, inadvertent, accidental, and catalytic. Although such risks are present with the use of any new technology introduced into military systems, today's AI—in particular, machine learning—poses particular risks because the internal workings of all but the simplest machine learning systems are for all practical purposes impossible for human beings to understand. It is thus easy for human users to ask such systems to perform outside the envelope of the data with which they were trained, and for the user to receive no notification that the system is indeed being asked to perform in such a manner.

Introduction

As international security analysts contemplate the future of warfare, a common theme is that the weapons of the future and artificial intelligence will be integrally linked. AI, it is believed, will confer all kinds of military advantages to the side that best takes advantage of this revolutionary technology. To offer just a few examples, it has been said that AI will enable the autonomous targeting of weapons (Etzioni & Etzioni, 2017), the control of swarming battlefield vehicles (Baraniuk, 2017), and the speedy detection of militarily significant patterns in data too complex or voluminous for human analysis.[56]

AI may indeed afford military planners and warriors with all those capabilities, and more. But the fact that some work to date suggests the possible feasibility of such applications is not the same as seeing an actual, delivered, proven capability to troops on the battlefield. Moreover, little analysis or commentary has been devoted to considering the downsides of an AI-infused conflict environment—downsides that may redound to the detriment of U.S. planners and warriors.

Many downside risks arise from the introduction of AI into military systems and planning, some of which include uncertainty about accountability regarding the use of AI-enabled weapons systems in lethal operations, integration of human-smart machine military "teams", impact on the culture and organization of the armed forces, and effects on adversary perceptions of the United States (see also Section 3.1 of Chameau, Ballhaus, & Lin, 2014). This paper focuses at risks in the context of escalation dynamics—how a military conflict's scope and intensity might escalate, but first it is necessary to review certain characteristics of AI relevant to this focus.

[56] For example, the DOD published Establishment of an Algorithmic Warfare Cross-Functional Team (popularly known as Project Maven)(2017) to accelerate DoD's integration of big data and machine learning. The team's objective is "to turn the enormous volume of data available to DoD into actionable intelligence and insights at speed."

The Scope of Today's AI

AI is a broad term whose precise scope is contested. For example, many military leaders conceptualized AI in terms of their application domains—lethal autonomous weapons or smarter decision support systems as "AI." Technologists are more likely to see AI as an underlying technology that enables many different applications. Even so, lines between "artificial intelligence" and big data, algorithms, statistical learning, and data mining are blurry at best. In the early days of AI, AI relied primarily on a symbolic approach—that is, an approach to problem solving that relies on high-level representations of problems, logic, rules, knowledge, and search. Despite some early successes, this approach gradually lost favor in the 1980's as researchers came to appreciate more clearly the enormous difficulty of developing such useful high-level representations.

Today, the most prominent approaches to AI rely on machine learning (ML), a class of techniques that often (but not always) relies on the availability of large amounts of data. "Supervised ML" depends on training data that has been labeled by humans and makes statistical inferences. "Unsupervised ML" finds clusters and outliers in unlabeled data that might otherwise go unnoticed if examined by humans.

But by themselves and unaided, ML techniques provide neither explanation for the inferences drawn nor the significance of the clusters. In other words, AI systems based on ML are unable to explain to their human users why they reach the conclusions they reach or demonstrate the behavior they demonstrate. Even worse, human examination of the machine's output and how it was derived from the input does not help, as it generally yields little about the features of the input that led to the inference in question. At least at first, users must simply trust that the system is behaving properly; over time, their trust grows if the system repeatedly behaves properly.

For many applications, explanations are simply unnecessary, and the inability to explain why a given result was produced is merely a curiosity. For example, when a user searches for a given book on Amazon, an ML-based recommender system provides suggestions of other books the user might wish to purchase. But such applications are generally applications with low stakes where an explanation does not particularly matter to most human user.

Trust in ML applications is properly limited to those operational scenarios that have been well-covered in the training data—ML applications are least trustworthy in scenarios that have not been well-covered, that is, in novel scenarios. (This phenomenon is arguably the reason that algorithmic bias arises in improperly vetted ML algorithms—an ML algorithm misidentifies human beings of African descent as gorillas because it has not been trained on an adequate sample of pictures of black human beings (see, for example Guynn, 2015)). In novel scenarios, explanations may very well be a necessary foundation for humans to properly trust ML applications.[57]

A difficult problem that requires solution arises from the reality that an ML application must be able to distinguish between input data from the universe of data on which it has been trained (i.e., routine scenarios) and input data from outside that universe (i.e., exceptional or novel scenarios). For

[57] "Explainable AI" is the focus of a DARPA research program (see Gunning, n.d.) that "aims to create a suite of machine learning techniques that [p]roduce more explainable models, while maintaining a high level of learning performance (prediction accuracy); and [e]nable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners." That said, the reason that this DARPA program exists in the first place is that the problem is a very hard one, and it is fair to say that the techniques of explainable AI have not made it in to common use. Whether or not they will ever do so remains to be seen.

example, consider an application is trained to distinguish between different breeds of dogs. The training data set consists of a very large number of labeled dog pictures. Give the application a picture of a random dog, and its output is the breed of dog that is most likely for that picture. This is a routine scenario for which the application is designed.

But what happens if instead the application is given a picture of a dolphin? Although it could not be expected to identify it as a dolphin (since it was never exposed to training data involving dolphins), it would be desirable of the application itself could recognize that it is now being expected to operate outside its zone of competence and inform the user of that conclusion.

The application must distinguish between two types of input data that is has never seen before. The first is routine—it is new, but it is generally similar to the training data. If processing routine input data, the application should provide its best guess (e.g., what breed of dog was shown). The second is novel—it is also new, but it is highly dissimilar to the training set, If processing novel input data, the application should produce an indication that it is operating outside its capabilities and that its output should be less trusted. The hard problem to solve is how to differentiate between "different in detail but generally similar" and "highly dissimilar.

Pimental et al (2014) describes a number of approaches that yield partial solutions for the problem described above, generally known as novelty detection. But most importantly, they note that defining "novelty" is conceptually a difficult problem, and thus it is not possible to suggest one "best" method of novelty detection. They go on to suggest that "the variety of methods employed is a consequence of the wide variety of practical and theoretical considerations that arise from novelty detection in real-world datasets, such as the availability of training data, the type of data (including its dimension, continuity, and format), and application domain investigated. It is perhaps because of this great variety of considerations that there is no single universally applicable novelty detection algorithm." All of the approaches described by Pimental et al (2014) involve elements of human judgment, and thus it is reasonable to conclude that in general (i.e., for any supervised ML application), some novel instances of new input data will not be identified as novel. In the absence of such identification, the user will unknowingly assume the ML is acting within the parameters of a tried and trusted application without realizing that the application is now operating outside its zone of competence. That way lies potential disaster.

AI Everywhere

If predictions that AI is an enabling technology of the future actually come true, we will see AI of various types and functions ubiquitously embedded in the devices and infrastructure of both civilian and military life. We will see AI-enabled capabilities support myriad non-military activities throughout society. As illustrative examples, AI will be embedded in self-driving cars and other autonomous and semi-autonomous vehicles; decision-support systems for investors and health care providers; automatic translation and transcription systems; identifying potential suicide victims; marketing products and services to individual consumers; predictive policing; and crop/soil monitoring and predictive analytics regarding agricultural yields.

On the military side, AI-enabled capabilities will be found in weapons systems, controlling one or a number or all of their functions, possibly including navigation, propulsion, weapons targeting, weapons release, and so on; in sensor systems and systems for intelligence analysis, identifying patterns and sifting through large volumes of disparate data and possibly providing likely interpretations of such patterns; in decision support systems, providing recommended courses of

action in response to particular sets of circumstances. Most importantly, AI-enabled capabilities will be available for use by all parties to a conflict.

Where AI applications are ubiquitous, they are—almost by definition—not novel. But novelty, among other things, is an important driver for skepticism. Human users who are appropriately skeptical of new technology do not give their trust without sufficient evidence, and they themselves will act as "second opinions" to judge the accuracy and propriety of their applications' output. A plethora of skeptical users would indeed be reassuring. But the experimental data does not provide such reassurance. For example, in a 2016 study, individuals followed the directions of a robot in a (simulated) emergency evacuation scenario, even though they had observed the same robot perform poorly in a navigation guidance task a few minutes before. Even when the robot pointed to a dark room with no discernible exit, the majority of individuals did not choose to safely exit the way they entered (Robinette, Li, Allen, Howard, & Wagner, 2016).

Without widespread skepticism, ubiquitous AI will inevitably become part of the background, and its affordances for society (i.e., the beneficial capabilities it provides for society) will disappear from conscious attention and thought, much as electricity disappeared into the background and became taken for granted in the 20th century. And it should further be noted that user skepticism that prevents automatic reliance on an AI-based system may in some instances defeat the very purpose of introducing that system in the first place. Specifically, AI capabilities may have been added to increase the system's speed of operation—in this context, why would it be desirable for a human user to take the time to check or second-guess the machine's decisions and conclusions? This point itself will drive human users in the direction of unquestioning trust.

Escalation Dynamics

As a point of departure, consider that escalation in a conflict may arise through a number of different mechanisms (which may or may not simultaneously be operative in any instance).[58]

- Deliberate escalation is an intentional choice by one party to intensify the conflict. In principle, the escalating party has made this judgment based on its understanding of its own and the other side's capabilities and intentions, and acts according to the belief that escalation will bring advantages.
- Inadvertent escalation occurs when one party deliberately takes actions that it does not believe are escalatory but are interpreted as such by another party to the conflict. Such misinterpretation may occur because of a lack of shared reference frames or incomplete knowledge of the other party's thresholds or "lines in the sand."
- Accidental escalation occurs when some operational action has direct effects that are unintended by those who ordered the action. A weapon may go astray to hit the wrong target; rules of engagement are sometimes unclear; a unit may take unauthorized actions; intelligence on a target may be faulty; or a high-level command decision may not be received properly by all relevant units.
- Catalytic escalation occurs when some third party succeeds in provoking two parties to engage in conflict. For example, C takes action against A but makes it look like the action came from B. C then observes as A takes action against B, and B may well respond against A for what B sees as an unprovoked attack from A.

[58] The first three types of escalation are described in greater detail in Forrest Morgan et al (2008). Lin (2012) built on this work to explore escalation dynamics in cyberspace and added the fourth type of escalation—catalytic escalation.

Escalation Dynamics in an AI-Infused Conflict Environment

Central to each of these escalation mechanisms is the scope, nature, and quality of information available to decision makers. How might AI-enabled capabilities lead to or facilitate different kinds of escalation dynamics, by which is meant how hostilities might escalate over time?[59] The following discussion suggests some illustrative, but by no means comprehensive, possibilities.

Deliberate escalation

Party A may choose to escalate if it believes its military capabilities are sufficiently powerful to defeat B's response to that escalation. But if A's actual capabilities do not match A's estimate of its own capabilities, defeat or disaster may result from escalation. In particular, A may believe that its own AI-enabled military decision support systems have been trained on an adequate universe of cases, but actual conflict often falls outside the parameters of what planners expected before the conflict started—unexpected tactics or weaponry, for example. But these systems will dutifully do the best they can without users recognizing critical differences between data from actual conflict and its training data. The systems may thus offer conclusions that go beyond their expertise or recommendations that are accepted by humans who do not notice the out-of-scope situation.

Inadvertent escalation

Party A takes an action that it does not believe Party B will (or should) regard as escalatory. For example, Party A attacks B's ballistic missile early warning satellites early in a conventional kinetic conflict, because those satellites are providing tactical advantages for B in locating the launch sites of A's non-nuclear tactical ballistic missiles. B sees such actions as a prelude to nuclear attack of A on B, because those satellites are also used to warn B of a nuclear attack. B believes that A must know that such an anti-satellite attack would be hugely escalatory, but A believes it is simply trying to negate a tactical advantage for B. Thus, A's attack on B's satellites is interpreted by B as an escalation, and B responds in kind. Because A did not believe its anti-satellite attack was escalatory, A sees B's response as an unwarranted escalation rather than a response—and this sequence of events sets off an (inadvertent) escalatory spiral.

Assumptions about thresholds are likely to be built in to ML-based decision support systems. That fact in itself is not bad—one must start somewhere. But how will the differing perspectives of adversaries be acknowledged, taken into account, and flagged explicitly for human attention? Indeed, inserting information about adversary thresholds into such support systems would require the availability of substantial data on those thresholds. But if such data were available and were deemed important, the problem of not knowing or realizing the adversary's thresholds would not exist in the first place. Radically different views of the adversary's motives and intentions are not mere parametric tweaks in a model of conflict—rather, they call into question the underlying utility of such a model for understanding how a conflict might unfold.

Accidental escalation

[59] This phraseology is intended to capture the idea that even before hostilities break out, adversaries are in a continuous cycle of reacting to the actions and intentions of others. While arguably most important in setting the strategic stage for the outbreak of hostilities, the state of affairs prior to the outbreak of hostilities is not addressed in this short paper. Another paper will someday focus of this topic.

A certain weapon of Party A relies on in-flight AI-based imagery analysis for automatic target recognition. Whilst flying at night, the weapon sees a building with gunfire flashes coming from the windows. The building is identified through target databases as being a hospital, but because a hospital becomes a valid military target if an enemy is using it as a base for military operations, the building is destroyed. In reality, the gunfire flashes were reflections from gunfire emanating from Party B's troops stationed around, and not in, the hospital. However, Party B does not realize this fact at the time, and the conflict escalates because the target recognition algorithm did not take into account the possibility that reflections of gunfire flashes might be mistaken for the real thing.

A variant of this scenario could involve an adversary tricking the AI in the automatic target recognition system. For example, Party B may be able to spoof the imagery of a hospital received by the weapon in flight in such a way that the weapon identifies it as a valid military target, and the hospital is destroyed. But the spoofing occurs in such a way that to the human eye, the imagery captured from the weapon's camera is indistinguishable from the image of a hospital, even though it was sufficient to fool the target recognition algorithm. (This point is addressed in more detail by Libicki's Chapter 18 "The Hacker Way of Warfare.")

Catalytic escalation

Party C seeks to provoke conflict between Party A and Party B. To this end, it constructs deepfake videos and audios, which are realistic audio or video files depicting senior individuals within the decision-making apparatus of A and B saying things that he or she never said. These videos and audios are clandestinely selectively injected into the intelligence collection streams of A and B—videos and audios depicting individuals from A are injected into B's collection systems, and vice versa. If the content of these pseudo-recordings is tailored properly, it is easy to see how they might provoke A or B into taking actions that the other might regard as the first step on an unprovoked escalatory path.

Discussion and Conclusion

The scenarios described above are illustrative. But all such scenarios suffer from the analytic issue that once a problem is anticipated and described, a fix for the problem can be easily imagined—and thus the scenario is easier to dismiss as unfounded. But the point of this chapter is to instill some degree of humility in human ability to anticipate all such problems, and thus to realize that with the advantages of AI-enabled military systems come some potential disadvantages.

Of course, the same could be said about technologies in general (including more traditional cyber tools)—any technological solution will fail when operated far enough outside the parameter envelope that defines the problem to be solved. Is there anything special about AI that is more problematic?

For the machine learning flavor of AI, the answer is yes. It was noted above that the human user has no way to know that an ML application is dealing with a novel scenario, i.e., one that falls outside the envelope of the data on which it has been trained. And the reason for this lack of knowledge is that examination of a machine learning algorithm's operation generally defies human comprehension—that is, a human being will find it impossible to tell what an ML-based computer system is doing in any given case. (It is for this reason that explainable AI is necessary in the first place.) In this regard, an ML application is much unlike other technological artifacts, whose design limits are much better understood. We implement ML-based systems with the going-in realization that we cannot

understand how they produce a given output from a given input—and in most other systems, such a lack of understanding would be a dispositive strike against it.

A second problematic dimension of AI-enabled military systems arises from the likely ubiquity of AI as an underlying enabling technology throughout all of society, both civilian and military. When a technology is ubiquitous, users take it for granted and tend to lose their skepticism about it—even though even ubiquitously deployed technologies exhibit flawed operation from time to time. When ubiquitously deployed technology fails, users are more likely to look to the circumstances of the particular failure rather than to any underlying problem that may be more fundamental. Consequently, human attention is less likely to be focused on underlying problems.

The policy recommendations that flow from the analysis above are modest but significant. First, maintaining a degree of skepticism about the application of AI to military systems is necessary for all policy makers. Skepticism does not mean that such application should be rejected out-of-hand, but it does mean keeping in mind that the promises of vendors and contractors are often inflated beyond any reasonable measure. Asking "what could go wrong?" is a good question to ask, early and often. Red teaming against AI-enabled military systems is one way to maintain such skepticism, but such efforts must be conducted from the inception of a system's design through operational deployment so that the consequences of proceeding down the AI-enabled path are clearer.

Finally, increased research may well be needed on to advance the state of the art in explainable AI in a military context. Such research has two flavors: (a) research that can help explain what ML-based AI systems are doing and why they reach the conclusions they reach; and (b) renewed research on symbolic AI, whose explicit rules and logics provide, in principle, basic building blocks for comprehensible explanations.

References

Baraniuk, C. (2017, January 20). US military tests swarm of mini-drones launched from jets. *BBC News.* Retrieved from https://www.bbc.com/news/technology-38569027

Chameau, J., Ballhaus, W. F. & Lin, H. L. (Eds.). (2014). E*merging and readily available technologies and national security — A framework for addressing ethical, legal, and societal issues*. Washington DC: National Academies Press.

Establishment of an algorithmic warfare cross-functional team (Project Maven). (2017, April 26). Memorandum from the Deputy Secretary of Defense. Retrieved from https://www.govexec.com/media/gbc/docs/pdfs_edit/establishment_of_the_awcft_project_maven.pdf

Etzioni, A. & Etzioni, O. (2017, May-June). Pros and cons of autonomous weapons systems. *Military Review, 97*(3), 72-81. Retrieved from https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/May-June-2017/Pros-and-Cons-of-Autonomous-Weapons-Systems/

Morgan, F. E., Mueller, K. P., Medeiros, E. S., Pollpeter, K. L., & Cliff, R. (2008). *Dangerous thresholds: Managing escalation in the 21st century*. Santa Monica, CA:RAND Corporation. Retrieved from https://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG614.pdf

Gunning, D. (n.d.) Explainable artificial intelligence (XAI). Defense Advanced Research Projects Agency Program Information. Retrieved from https://www.darpa.mil/program/explainable-artificial-intelligence

Guynn, J. (2015, July 1). Google Photos labled blanck people 'gorillas'. *USA Today*. Retrieved from https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/

Lin, H. (2012 Fall). Escalation dynamics and conflict termination in cyberspace. *Strategic Studies Quarterly, 6*(3), p. 46-70. Retrieved from http://www.au.af.mil/au/ssq/2012/fall/lin.pdf

Pimentel, M., David A. Clifton, Lei Clifton, Lionel Tarassenko (2014). A review of novelty detection, *Signal Processing* 99: 215–249. Retrieved from http://www.robots.ox.ac.uk/~davidc/pubs/NDreview2014.pdf.

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. Proceedings from 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand. Retrieved from https://ieeexplore.ieee.org/document/7451740

Chapter 20. Artificial Intelligence in Future Chinese Command Decision-Making

Elsa Kania
Harvard University
bkania@gmail.com

Abstract

The Chinese People's Liberation Army (PLA) is exploring the use of AI technologies to enhance future command decision-making. In particular, the PLA seeks to overcome admitted deficiencies in its commanders' capabilities and to leverage these technologies to achieve decision superiority in future "intelligentized" (智能化) warfare. Building upon its development of the integrated command platform, which has included basic decision support, the PLA's ongoing construction and improvement of its joint operations command system could leverage AI technologies, particularly to enhance situational awareness and to improve cognitive speed in decision-making. In the process, Chinese military experts have examined the DARPA program Deep Green from the mid-2000s, which was ultimately defunded, as an example of the capabilities that intelligentized command decision-making could enable. Moreover, the recent successes of AlphaGo appears to have inspired Chinese strategists to explore how today's advances in AI could provide a critical advantage on the future battlefield. The PLA's apparent expectation that the future increases in the tempo of operations will outpace human cognition could result in a pragmatic decision to take humans out of the loop in certain operational environments in which speed is at a premium. However, the PLA also recognizes the importance of integrating and leveraging synergies among human-machine "hybrid" intelligence. Looking forward, the PLA's capacity to adapt to these technological and organizational challenges will impact and may constrain its pursuit of military innovation.

Introduction

China's New Generation Artificial Intelligence Development Plan (新一代人工智能发展规划) anticipates that the applications of AI technologies in national defense will include support to command decision-making, military deductions (军事推演, e.g., war-gaming), and defense equipment ("State Council Notice," 2017). Of these applications, the potential of AI in future command decision-making (指挥决策) could possess unique potential to be transformative on the future battlefield. PLA defense academics and strategists anticipate that AI might augment – or perhaps, in some contexts, even replace – human commanders on the future battlefield (See Yuan, 2017; Chen & Zhou, 2017). Notably, in an authoritative commentary, the Central Military Commission (CMC) Joint Staff Department (JSD) has called for the PLA to advance intelligentized command decision-making (智能化指挥决策) in its construction of a joint operations command system, through taking advantage of the potential of AI, as well as big data, cloud computing, and other advanced technologies (CMC JSD, 2016).

Strategic impetus and context

The PLA has long looked for ways to improve and enhance the capabilities of its commanders in decision-making. In particular, Chinese military leaders at the level of Xi Jinping himself are deeply

concerned about and looking to overcome the "five incapables" (五个不会), which represent serious shortcomings in "some" cadres' abilities to: "judge the situation accurately, understand the intentions of higher-level leaders, undertake correct operational decisions, train and deploy troops, and deal with unexpected contingencies" (*PLA Daily*, 2018; Blasko, 2016).[60] If accurate, such an assessment of these persistent shortcomings is quite damning. Despite reforms and ongoing efforts to overcome these "five incapables," the PLA continues to bemoan this, among other perils of "peace disease" (和平病), calling for continued improvements in training and education of commanders to resolve these difficulties. These efforts to enable more realistic training and effective education may turn to the use of big data and artificial intelligence as tools (*China Social Science*, 2017).

To date, the PLA's agenda for informatization (信息化) has concentrated on the development of its C4ISR capabilities, including its "integrated command platform" (一体化指挥平台) (Pollpeter et al., 2014). The PLA's concentration on system of systems operations (体系作战) has demanded advances in military command information systems to improve interoperability among services through more effective exchange of the requisite information, from intelligence to operational planning. The process of "command automation" (指挥自动化) involved in these advances in C4ISR has created the foundation from which the PLA's may seek to undertake command intelligentization (指挥智能化), through enhancing the level of intelligent information processing in these systems. At present, the integrated command platform does include at least basic tools for decision support (辅助决策), but remains fairly limited in its capabilities. Meanwhile, the apparent heterogeneity of informatization as implemented by the PLA may continue to undermine standardization and interoperability among services and theater commands (战区).

In the spring of 2016, AlphaGo's initial defeat of Lee Sedol appears to have captured the PLA's imagination at the highest levels, resulting in the convening of a number of high-level seminars and symposiums on the topic of intelligentized command decision-making in response (*China Military Science*, 2016). From the PLA's perspective, the success of AlphaGo was a pivotal moment that demonstrated the potential of AI to engage in complex strategizing comparable to that required to wage war, not only equaling but also seemingly surpassing human intelligence (Yuan, 2015). For instance, Major General Lin Jianchao (林建超), former director of the General Staff Department Office, has started from consideration of AlphaGo in evaluating future challenges of command decision-making, assessing that AI could have revolutionary implications yet also highlighting the current limitations of these technologies that demand a model of human-machine fusion (人机融合) for future applications in war design, strategic guidelines, campaign planning, planning assessments, and operational command (Lin, 2016). Significantly, the Joint Staff Department has highlighted AlphaGo's victory as having demonstrated the tremendous potential of AI in operational command, military wargaming, and decision support (CMC JSD, 2016).

[60] Dennis Blasko is to be credited for first drawing this concept and the challenges that it describes to my attention.

Initial Exploration of New Concepts of Command

The PLA's leading thinkers on command and control are starting to explore next-generation capabilities in response to trends in today's emerging technologies. Notably, the China Institute of Command and Control (中国指挥控制学会), which convenes some of the top experts from the Chinese military, academia, and defense industry, has recently established a new Intelligent Command and Control Systems Engineering Specialist Committee (智能指挥与控制系统工程专业委员会) that is intended to undertake systematic exploration of these issues ("China Institute of Command and Control," 2018), building upon a number of publications and conferences that have explored these issues (e.g., *Global Times*, 2015; *China Network*, 2017). While this conceptual development remains at a nascent stage, it is clear that Chinese military leaders and experts are actively exploring the question of the appropriate balance between human and artificial intelligence required for future "decision superiority" (决策优势)

Although the PLA appears to be enthusiastic about the potential for more "scientific" approaches to decision-making, there is also a clear recognition of the importance of the human element of command. For instance, Lieutenant General Liu Guozhi (刘国治), director of the Central Military Commission Science and Technology Commission, anticipates that human-machine hybrid (人机混合) intelligence will be the highest form of future intelligence (Liu, 2016). Similarly, such specialists as Zhao Xiaozhe (赵晓哲), an academician with expertise on command and control for naval operations, recognize that the complementarities between natural and artificial intelligence will be critical, necessitating advances in human-machine interaction (人机交互) (Zhao, 2017). For instance, research on brain-computer interface (BCI) technologies, such as that undertaken at the PLA's Information Engineering University, could enable direct control of military robotics ("Brain plan launched," 2016). From his perspective, AI actually enhances the role of people, since human initiative and creativity cannot be replaced (Zhao, 2017). In particular, the inherent uncertainties of warfare, including incomplete information, inconsistent intelligence, and deliberate deception, create unavoidable difficulties for command intelligentization, given current limitations of AI in learning and reasoning.

However, the PLA also anticipates that the advent of AI in warfare will place a premium upon speed in command and decision, creating new challenges that could change the role of humans on the future battlefield. Already, today's informatized warfare demands rapid processing of information and evaluation of the operational environment to enable command decisions. Looking forward, from the perspective of one Chinese defense academic (Chen, 2016):

> *"on the future battlefield, with the continuous advancement of AI and human-machine fusion (人机融合) technologies, the rhythm of combat will become faster and faster, until it reaches a "singularity" (奇点): the human brain can no longer cope with the ever-changing battlefield situation, unavoidably a great part of decision-making power will have to be given to highly-intelligent machines,"*

As a result, the role of humans could transition from being 'in' the loop, to 'on' the loop, and perhaps even out of the loop.[61] At present, there is not sufficient evidence to conclude with confidence that the PLA will take humans fully 'out of the loop.' However, this expectation that there will be a future point at which "the tempo of intelligentized operations will be unprecedentedly accelerated," beyond the capabilities of human cognition, does recur across a number of PLA writings (e.g., "Exploring the winning joints," 2018).

The PLA's approach to human control in decision-making may reflect a pragmatic perspective that will evolve in accordance with perceived operational requirements, rather than ethical considerations. Moreover, certain of these shifts towards intelligent command and control may reflect evolutionary and incremental improvements upon existing command automation. That is, current approaches to targeting already introduce certain ambiguities to the question of human control (e.g., Ekelhof, 2018). The continued advances in automatic target recognition and greater autonomy in the control and guidance of advanced weapons systems, from cruise missiles to hypersonic glide vehicles, will build upon a robust record of research and development (NUDT; Wang, 2018). In the process, boundaries between clear human control and increased autonomy could become quite indistinct. For instance, "intelligent weapons" might operate with a high level of autonomy in tracking, targeting, and attacking an adversary but would be 'acting' in accordance with the intentions and objectives of commanders (All-Military Military Terminology). Whereas a high level of automation in defensive capacities, such as air and missile defense, is not novel or unexpected, the offensive employment of such capabilities does raise new risks and concerns. Indeed, the PLA's active development of and apparent enthusiasm for unmanned and potentially autonomous weaponry raises the possibility that these systems could emerge as major elements of its arsenal in the years to come.

Looking forward, if the PLA's future approach to command becomes more AI-guided (智能主导), such advances could enable more effective integration of information and firepower to attack and destroy an adversary's battle networks. The increased prominence of intelligent weapons on the future battlefield could result in "remote, precise, miniaturized, large-scale unmanned attacks" becoming the primary method of attack (Yun, 2018). In the future, a system for intelligentized operations might be composed of intelligent weapons and equipment, enabled by pervasive sensing, guided by real-time coordinated mission planning systems to enable autonomous combat formation and swarm, human-machine integration, and autonomous operations across multiple domains (Liu, 2018). Potentially, the new combat methods that will arise as a result could include "latent warfare" (潜伏战), in which unmanned systems are deployed to critical targets or locations in advance to be activated when needed and 'global rapid assault combat,' involving the employment of unmanned hypersonic space platforms that may enable new approaches to deterrence (Pang, 2017). These initial, rather speculative writings may be nascent but could inform the trajectory of the PLA's research, development, and operationalization of future capabilities.

Initial Developments and Experimentation

In the near future, the PLA could leverage AI technologies to assist and support the decision-making of fighter pilots and the commanders of submarines. For instance, according to credible reporting, there is a project underway to update the computer systems on PLA Navy nuclear submarines with

[61] These concepts (i.e., of humans being in, on, or out of the loop) originate in U.S. discussions of the role of humans in decision-making, reflecting the PLA's close attention to U.S. policies and debates.

an AI decision support system that could reduce commanding officers' mental burden and workload (Chen, 2018). That is, AI may take on certain "thinking" functions, which could involve interpreting and answering signals picked up by sonar, through the use of convolutional neural networks (Kania, 2018). Indeed, according to Major General Liu Zhong (刘忠), an expert on command systems, "traditional combat auxiliary decision-making is currently developing towards knowledgization (知识化) and intelligentization, and the application of AI technology and knowledge-based intelligent assistant decision-making system has become a new development direction…" (Wang, 2016). This use of AI to assist and support command at all levels and in a range of contexts appears to have great appeal for the PLA, particularly given the self-assessed weaknesses of its own commanders at present. The PLA has closely studied Deep Green, a DARPA program that was undertaken in the mid-2000s, which sought to support commanders through advanced predictive capabilities, including the generation of courses of action, evaluation of options, and assessment of the impact of decisions (Surdu, 2008; Hu, 2016). However, the PLA's own approach to these technologies could diverge from this initial model.

As the PLA seeks to modernize its integrated command platform for future joint operations, there is research underway to explore the integration of AI technologies, indicating a transition from command automation (指挥自动化) to command intelligentization (指挥智能化) (Guo and Si, 2016). The former General Staff Department Informatization Department's 61st Research Institute, which has since been shifted to the PLA's Academy of Military Science, will likely be involved in relevant research. For instance, 61st Research Institute researchers have explored the use of neural networks to support the detection of abnormal behavior by users, in order to enhance the defense of military information networks (Yang, 2018). Of note, Major General Liu Zhong (刘忠) of the National University of Defense Technology's Key Laboratory of Information Systems Engineering (信息系统工程重点实验室) has also been engaged in research that dates back to 2006 to optimize and increase the intelligentization of PLA command and control, seeking to enable rapid planning and decision-making (Liu, 2015). Reportedly, as of December 2015, Liu Zhong's team completed their research and development, which had created a Joint Operations Command and Control Advanced Concepts Demonstration System (联合作战指挥控制先期概念演示系统) (China Daily, 2015). Their new C2 system has been formally provided to some units on at least an experimental basis starting in 2015 (Liu, 2015). Liu Zhong has been praised extensively for his work, which has been characterized as creating an "external brain" (外脑) to assist commanders, enhancing awareness and management of the battlefield.

The PLA's continued efforts to introduce AI into military command information systems could be varied and involve experimentation by a number of relevant players. For instance, the 28th Research Institute of the China Electronics Technology Group (CETC), a state-owned defense conglomerate, has established the Joint Laboratory for Intelligent Command and Control Technologies (智能指挥控制技术联合实验室) in partnership with Baidu (CETC, 2018), a global leader and Chinese national champion in AI. This new laboratory will concentrate on increasing the level of intelligentization in command information systems through the introduction of big data, artificial intelligence, and cloud computing. Reportedly, at the Zhuhai Airshow in the fall of 2018, CETC also demonstrated a mission system for intelligentized operations (智能化作战任务系统) that was reported to be capable of "learning independently" and "summing up combat experience" (Wang X., 2018). Meanwhile,

research undertaken by researchers from the PLA's National Defense University and Strategic Support Force has explored ways to simulate the real battlefield through war-gaming, in order to generate data that can contribute to greater understanding of the simulated battlefield, while enabling methodologies that might enhance future situational awareness on the actual battlefield.

Challenges of Culture and Organizational Capacity

The PLA's capacity and characteristics as an organization could deeply influence its approach to the operationalization of AI. Traditionally, the PLA has tended to centralize and consolidate authorities at higher levels, remaining reluctant to delegate decision-making downward, which can constrain personnel and organizations of lower grades exercising independent initiative. To date, the introduction of information technology has apparently exacerbated the tendency of PLA commanders to micromanage subordinates, rather than engaging in effective exercise of mission command. For instance, a practice known as "skip-echelon command" (越级指挥) can enable the circumvention of the formal chain of command in order to direct units of lower echelons ("Major New Trends," 2007). This practice appears to be symptomatic of the PLA's relative bureaucratic immaturity. Such tendencies towards distrust of subordinates have seemingly contributed to the persistent shortcomings in the capabilities of PLA officers, and the introduction of AI could exacerbate these habits. The combination of an apparent reluctance to delegate authority downwards – with the tendency to consolidate command authority for strategic capabilities at the highest levels[62] – could also render the PLA's leadership inclined to direct future intelligent weapons from the top levels of command.

In the future, the intersection of the PLA's affinity for scientific approaches to warfare with the preference to centralize decision-making could contribute to greater reliance upon AI, rather than human judgment. In practice, this tendency could become a source of vulnerability, given the continued fallibility and near inevitability of mistakes in complex AI-enabled systems (e.g., Osoba & Welser, 2017). To some extent, the PLA's distinctive ideological characteristics also could prove impactful. Despite its modernization and professionalization, the PLA still confronts unique circumstances as a military that is required to obey the commands of the Chinese Communist Party (听党指挥). The CCP's concerns with issues of political and ideological reliability could contribute to an inclination towards turning to machine intelligence as more reliable and controllable. However, the parable of 'rogue chatbots' that were shut down after online comments criticizing the Party as "corrupt and incompetent" highlights that the uncertainties that are inherent in the development of complex technologies may also be provoke concerns of security and controllability (安全, 可控) in some cases (*Financial Times*, 2017).

For the PLA, the employment of AI could appeal as an apparent opportunity to circumvent persistent difficulties with human capital and training, even as those same challenges may impede its effective adoption. For instance, the PLA's lack of experience with the complexities of (manned) carrier aviation could contribute to a sooner shift to the use of unmanned and autonomous systems, potentially including the CH-7, off of its aircraft carriers ("Stealth UAV CH-7," 2018). In this regard, the PLA's current shortcomings could motivate leapfrogging and experimentation. However, the effective employment of complex, intelligent systems may often place greater demands upon

[62] For instance, the PLA's Strategic Support Force consolidates strategic capabilities in space, cyber, and electronic warfare directly under the control of the CMC. The former Second Artillery Force, now Rocket Force, has similarly seemed to ensure CMC-level control of the PLA's nuclear and conventional missiles.

personnel in terms of training and technical understanding. Actively seeking to mitigate its current difficulties in talent development, the PLA is attempting to recruit highly educated officers and enlisted personnel, along with civilian personnel, but will confront strong competition from the tech sector in the process.[63]

Conclusions and Implications

Looking forward, the PLA's initial enthusiasm for and experimentation with the intelligentization of command decision-making could provide a new source of advantage or create vulnerabilities for the PLA. If the use of AI can enable decision superiority on the future battlefield, then the PLA's exploration of these new techniques of intelligent command and control might compensate for its current weaknesses in joint operations and interoperability. The PLA's asymmetric approach to capabilities development might also contribute to creative thinking in the development of new concepts of operations, including the use of AI in deception (Zuo et al., 2018). Despite expectations that authoritarian militaries might neglect the vital human element in this new era of warfare, PLA strategists do appear to have an acute awareness of the continued criticality of human intelligence and are exploring concepts of human-machine integration and coordination. However, the PLA may still have great difficulty in adaptation given its persistent stovepiping and bureaucratic tendencies as an organization. Moreover, greater reliance upon technological solutions might also exacerbate the underdevelopment of the capabilities of commanders in ways that could render the PLA dangerously dependent upon its battle networks. As the U.S. and China compete in these new frontiers of military innovation, the PLA's progress in leveraging such emerging technologies to augment its C4ISR capabilities will merit continued analysis.

References

All-Military Military Terminology Management Committee [全军军事术语管理委员会]. *People's Liberation Army Military Terminology* [中国人民解放军军语]. Military Science Press [军事科学出版社], 2011.

Blasko, D. (2015, February 18). Ten reasons why China will have trouble fighting a modern war. *War on the Rocks*. Retrieved from https://warontherocks.com/2015/02/ten-reasons-why-china-will-have-trouble-fighting-a-modern-war/

Blasko, D. (2016, June 21). The new PLA Joint Headquarters and internal assessments of PLA capabilities. *China Brief*. Retrieved from https://jamestown.org/program/the-new-pla-joint-headquarters-and-internal-assessments-of-pla-capabilities/

Brain plan launched: Cognitive dominance becomes the new high point of future military contests [脑计划"开启 "制脑权"成未来军事较量新的高地]. (2016, October 20). *PLA Daily*. Retrieved from http://military.people.com.cn/n1/2016/1020/c1011-28793350.html

[63] There are numerous allusions in PLA media to the need to recruit talented, educated officers and enlisted personnel, while intensifying ideological work to ensure that "the Party commands the gun." Thanks to Ken Allen for sharing his insights on PLA recruitment and personnel issues.

CETC 28th Research Institute and Baidu Company established the "Joint Laboratory for Intelligent Command and Control Technology" to promote military-civil fusion in the field of new technologies. [中国电科28所与百度公司成立"智能指挥控制技术联合实验室"推动军民融合向新技术领域纵深迈进]. (2018, January 23). Sohu. Retrieved from www.sohu.com:a:218485100%E2%80%AD_%E2%80%AC779538

CETC brings military-civil fusion network information systems to China Air Show [中国电科携军民融合网络信息体系亮相中国航展]. (2018, November 6). Xinhua. http://www.xinhuanet.com/politics/2018-11/06/c_1123673604.htm

Chen, Y. [陈玉飞] and Zhou T. [周涛]. (2017, June 8). Will artificial intelligence replace commanders? [人工智能能代替指挥员吗?], *PLA Daily.* Retrieved from http://www.81.cn/big5/jwgz/2017-06/08/content_7631686.htm

Chen, S. China's plan to use artificial intelligence to boost the thinking skills of nuclear submarine commanders. (2018, February 4). *South China Morning Post*. Retrieved from https://www.scmp.com/news/china/society/article/2131127/chinas-plan-use-artificial-intelligence-boost-thinking-skills

Chief Engineer Hu Xiaofeng, General Manager of China's Bingqi Program, Delivered a Lecture: the Challenge of the Intelligentization of Command information Systems [中国兵棋工程总师胡晓峰少将演讲：指挥信息系统的智能化挑战]. (2016, July 13). Retrieved from http://chuansong.me/n/434595151184.

China chatbot goes rogue: Do you love the Communist party? (2017, August 2). *Financial Times*, Retrieved from https://www.ft.com/content/e90a6c1c-7764-11e7-a3e8-60495fe6ca71

China Institute for Command and Control Intelligent Command and Control Expert Committee Established [中国指挥与控制学会智能指挥与控制系统工程专业委员会成立]. (2018, November 26). Retrieved from http://pris.bit.edu.cn/yjsxw/136237.htm

China Military Science Editorial Department [中国军事科学 编辑部]. (2016, April). A Summary of the Workshop on the Game between AlphaGo and Lee Sedol and the Intelligentization of Military Command and Decision-Making" [围棋人机大战与军事指挥决策智能化研讨会观点综述]. *China Military Science* [中国军事科学].

China Command and Control Conference Leads in the Development of C5ISR for Intelligent Unmanned Operations" [中国指挥控制大会 引领智能无人作战C5ISR发展]. (2017, June 14). Retrieved from China Network. http://military.china.com.cn/2017-06/14/content_41022717.htm

China Institute of Command and Control Hosts the Fourth China Command and Control Conference in Beijing" [中国指挥与控制学会在京举办第四届中国指挥控制大会]. (2016, July 8). China Association for Science and Technology, http://www.cast.org.cn/n17040442/n17045712/n17059079/17289485.html.

CMC Joint Staff Department [中央军委联合参谋部]. (2016, August, 15). Accelerate the construction of a Joint Operations Command System with our nation's characteristics—Thoroughly study Chairman Xi's important sayings when inspecting the CMC Joint Operations Command Center [加快构建具有我军特色的联合作战指挥体系—— 深入学习贯彻习主席视察军委联指中心时的重要讲话], *Seeking Truth* [求是]. Retrieved from http://www.qstheory.cn/dukan/qs/2016-08/15/c_1119374690.htm

Ekelhof, Merel, Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting. *Naval War College Review* 71, no. 3 (2018): 6.

First Intelligent Command and Control Forum Was Successfully Convened in Beijing on April 24" [首届智能指挥与控制论坛4月24日在京成功召开], *Global Times*, April 27, 2015, http://military.people.com.cn/n/2015/0427/c172467-26912419.html.

Guo R. [郭瑞] and He Y. [贺筱媛]. (2017, November 16). Preprocessing method for intelligent analysis of battlefield situation data [面向战场态势数据智能分析的预处理方法]. *Electronic Technology and Software Engineering* [**电子技术与软件工程]. Retrieved from** http://www.cqvip.com/qk/80675a/201716/673001149.html

Chen, H. [陈航辉]. (2016, March 18). Artificial Intelligence: Disruptively changing the rules of the game [人工智能：颠覆性改变"游戏规则]. *China Military Online.* Retrieved from http://www.81.cn/jskj/2016-03/18/content_6966873_2.htm

Exploring the winning joints of intelligentized operations [探究智能化作战的制胜关节]. (2018, March 29). *PLA Daily*. Retrieved from http://military.people.com.cn/n1/2018/0329/c1011-29896429.html.

Guo R. [郭若冰] and Si G. [司光亚]. Facing New Challenges to Military Command in the Era of Intelligentization" [接近智能化时代军事指挥面临的挑战]. (2016, July). *China Military Science*,

Interview with National University of Defense Technology Information Systems and Management Academy Chief Engineer Professor Liu Zhong [记国防科技大学信息系统与管理学院总工程师刘忠教授]. (2015, December 29). *China Daily*. Retrieved from http://china.chinadaily.com.cn/2015-12/29/content_22850844_2.htm

Kania, E. (2018, February 12). Chinese Sub Commanders May Get AI Help for Decision-Making. Defense One. Retrieved from https://www.defenseone.com/ideas/2018/02/chinese-sub-commanders-may-get-ai-help-decision-making/145906/?oref=d-river. Further sources are available upon request.

Liu Y. [刘玮琦]. (2018, May 17). The curtain of intelligentized warfare has opened [智能化战争大幕拉开]. China Military Online. Retrieved from http://www.81.cn/jfjbmap/content/2018-05/17/content_206329.htm

Major General Lin Jianchao: seeing the challenges of decision-making and thinking intelligentization from the great human-machine war in Go [林建超少将：从围棋人机大战看决策思维智能化面临的挑战与决策]. (2016, April 18). Retrieved from http://www.kunlunce.com/qycm/fl1111/2016-04-18/69130.html

National University of Defense Technology's Liu Zhong: Creating a Powerful "External Brain" for Command and Control [国防科大刘忠:为指挥控制打造强大"外脑"]. (2015, December 28). *People's Daily*, Retrieved from http://military.people.com.cn/n1/2015/1228/c401735-27986608.html

New-era military vocational education with great prospects for the future [新时代军事职业教育大有可为] (2017, November 6). *China Social Science Military Studies*. Retrieved from http://wemedia.ifeng.com/85730181/wemedia.shtml

Osoba, O. A., & Welser, W. (2017). An intelligence in our image: The risks of bias and errors in artificial intelligence," *RAND Corporation.* Retrieved from https://www.rand.org/pubs/research_reports/RR1744.html

Liu Zhong: A Chief Engineer On the Road [刘忠：一直在路上的总工程师]. (2015, December 29). Xinhua. Retrieved from http://www.81.cn/2015lz/2015-12/29/content_6836338.htm

Pang H. [庞宏亮]. (2017, May 4). The curtain of the "machine era warfare" is opening ["机器战争纪元"帷幕正在拉开]. Xinhua. Retrieved from http://www.xinhuanet.com//mil/2017-05/04/c_129588629.htm

Pollpeter, K., Anderson, E., McReynolds, J., Ragland, L. A., and Thomas, G, A. (2014, January). Enabling information-based system of system operations: The research, development, and acquisition process for the integrated command platform. *SITC Research Briefs*. Retrieved from http://escholarship.org/uc/item/6f26w11m

Six major new trends in PLA training [解放军训练六大新趋势]. (2007, January 17). *PLA Daily*. Retrieved from http://jczs.news.sina.com.cn/2007-01-17/0633427003.html

State Council notice on the issuance of the New Generation AI Development Plan [国务院关于印发新一代人工智能发展规划的通知]. (2017, July 20). State Council. Retrieved from http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

"Strengthening joint operations command against the "five incapables"" [针对"五个不会"强化联合作战指挥], PLA Daily, May 15, 2018, Retrieved from http://www.xinhuanet.com/mil/2018-05/15/c_129872341.htm

Stealth UAV CH-7 Debuted at the Zhuhai Airshow [隐身无人飞翼彩虹7号 首次亮相珠海航展]. (2018, November 5). *Sohu*. Retrieved from http://www.sohu.com/a/273455843_115479

Story of National University of Defense Technology Information Systems and Management Institute Chief Engineer Professor Liu Zhong [国防科技大学信息系统与管理学院总工程师刘忠教授故事集]. (2015, December 15). Xinhua. Retrieved from http://news.xinhuanet.com/mil/2015-12/30/c_128559188_5.htm

Surdu, J.R. and Kittka K. Deep Green: Commander's tool for COA's Concept. (2008, 29 June - 2 July). *Computing, Communications and Control Technologies: CCCT 2008*. http://www.bucksurdu.com/Professional/Documents/11260-CCCT-08-DeepGreen.pdf.

Wang C. [王长青]. (2016, September 8). "The Application and Prospects of Artificial Intelligence in Cruise Missiles" [人工智能在飞航导弹上的应用与展望], http://chuansong.me/n/711504451360.

Wang Z. [王握文]. Systematized Design: Establish a Foundation for Integrated Military Information Systems [体系化设计：构筑一体化军事信息系统的基石]. (2016, March 14). China Military Online. http://jz.81.cn/n2014/tp/content_6959067.htm

Wang Xiaoliang, Military Expert: From the China Airshow Look at Intelligentized Warfare [军事专家王明亮：从中国航展看未来智能化战争]. (2018, November 11). Xinhua. Retrieved from http://www.xinhuanet.com/politics/2018-11/11/c_129991032.htm

Yang M. [杨明非] and Ye J. [叶季青]. (2018, July 2). Intelligent security protection technology based on network user behavior" [基于网络用户行为的智能安全防护技术]. 6th China Command and Control Conference. Retrieved from http://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CPFD&dbname=CPFDLAST2018&filename=ZHKZ201807002054&v=MTcyMTNWc1VQeVhBZExHNEg5bk1xSTlGWnVzS0NCTkt1aGRobmo5OFRuanFxeGRFZU1PVUtyaWZadTV1RkNqbFVyak1L

Yuan, Y. [袁艺]. (2017, January 12). Will artificial intelligence command future wars?" [人工智能将指挥未来战争?], *China Military Online.* Retrieved from http://www.81.cn/jmywyl/2017-01/12/content_7448385.htm

Yun G. [游光荣]. (2018, October 17). AI Will Deeply Change the Face of Warfare [人工智能将深刻改变战争面]. *PLA Daily*.

Zuo D. [左登云], Gong J. [龚佳], & Huang P. [黄培荣]. (2018, September 20). Why do informationized operations go to intelligent operations [信息化作战何以走向智能化作战]. *PLA Daily*. Retrieved from http://www.xinhuanet.com/mil/2018-09/20/c_129957478.htm

Zhao, X. (2017, April 23). Natural intelligence and artificial intelligence in command and control systems [赵晓哲：指挥控制系统中的自然智能和人工智能]. Retrieved from http://wemedia.ifeng.com/13425965/wemedia.shtml

Chapter 21. China's Integration of Neural Networks into Hypersonic Glide Vehicles

Lora Saalman
EastWest Institute
lsaalman@ewi.info

Abstract

A growing area of inquiry within the U.S. strategic community has been the level to which countries may apply artificial intelligence (AI) in their nuclear force-related support systems and platforms. This essay provides a brief analysis of Chinese efforts to integrate neural networks into its hypersonic glide vehicles, which may be used as future nuclear platforms to defeat U.S. missile defenses. This essay is based on over 300 recent Chinese technical journal papers and articles issued by researchers at university and military institutes in China. It discusses two AI-related trends:

1) Innovative and Prolific Research: China is creating a large number of open source papers that build on domestic collaborative models and seek to integrate neural networks to address hypersonic glide re-entry control, maneuverability, stability, heat, and targeting;
2) Shift from Active Defense to AI-Enabled Offense: China has engaged over the past decade in a quantitative and qualitative shift away from technical studies on countermeasures and towards offensive platforms, suggesting that its stance of "active defense" may be trending towards a stronger offense.

Introduction

The U.S. strategic community has been increasingly asking how far countries may go in applying AI in their nuclear force-related support systems and platforms.[64] This essay provides a brief case study of Chinese efforts to integrate neural networks into its hypersonic glide vehicles, which could be used as future nuclear platforms to defeat U.S. missile defenses. Over the past decade, Chinese research has shifted from analysis of, and countermeasures [对策] against, U.S. and Russian programs, and towards a prioritization of China's own offensive research and advances.[65] This assessment is based on over 300 recent papers from university and military institutes in China, building upon the author's Chinese-language database of 2,000 works on hypersonic advances and 1,000 on AI.

(1) Innovative and Prolific Research

Chinese researchers are confronting many of the same hurdles faced by other aspirants in the field of hypersonic glide vehicles and the integration of neural networks. While applications of neural networks in weapons platforms appeared in foreign studies of the 1990s—to enhance missile

[64] For a related essay that focuses on hypersonics, please see the author's forthcoming paper from the CAPS-RAND-NDU PLA conference held from November 30-December 1, 2018.

[65] U.S. programs like the HAWC, TBG, CPGS, HyRAX, AFRE, ETHOS, HiFire and platforms as the X-43A, X-37B, HTV-2, MKV-R interceptor and Russian platforms like the Yu-71 and Yu-74 continue to receive attention. Yet, they are predominantly relegated to lists, rather than the dissection of a decade ago. Meanwhile, Chinese technical experts detail their country's own advances, such as with the WU-14/DF-ZF. See references. Zhang Can, Hu Dongdong, Ye Lei, Li Wenjie, Liu Duqun, Zhang Shaofang, Wu Kunlin, and Zhang Hongna are engineers at Beijing Haiying/Hiwing Science and Technology Information Institute. Huang Zhicheng is affiliated with the Beijing Techscope Technology Consulting Co. Ltd.

seekers, missile fusing, sonar target discrimination, automatic target recognition, and auto piloting—these writings argued that neural networks were "high risk-high payoff."[66] Chinese experts have advanced beyond these foreign works with an exponential release of papers and projects that pursue the "high payoff" of neural network enhancement of maneuverability and penetration of missile defenses.

The technical papers surveyed for this essay overturn the dated portrayal of Chinese domestic engineers and researchers as lacking in innovation and domestic collaboration. Researchers from the People's Liberation Army Rocket Force, College of Mechatronic Engineering and Automation of the National University of Defense Technology, Harbin University, and Beijing Institute of Tracking and Telecommunications Technology are working—often collectively—to resolve some of the more intractable issues faced in control dynamics with hypersonic glide vehicles.[67]

These AI-based controls are meant to address the hypersonic glide vehicle's high flight envelope, complex flight environment, severe nonlinearity, intense and rapid time-variance, and dynamic uncertainty during the dive phase.[68] Beyond the use of traditional foreign models—as with Lyapunov stability theory or the Singer model[69]—Chinese researchers are developing their own models and algorithms for robust nonlinear adaptive control systems that integrate terminal sliding mode controls, predictive controls, fuzzy neural network controls, and nonlinear dynamic inverse controls.[70]

Further, Chinese experts are working to move beyond traditional fixed parameters, which would not hold in the dynamic, high-speed, and heat intensive re-entry environment faced by a hypersonic glide vehicle. Given the need for greater resiliency in the absence of data, a number of these writings seek to integrate "radial basis function neural networks" [基于径向基函数的神经网络] to mitigate nonlinearity and uncertainty in aerodynamic parameters.[71] Notably in combining the flexibility of sliding mode control method and backstepping in the terminal phase, Chinese researchers are

---

[66] See Webster, Willard P. (1991, May), "Artificial Neural Networks and Their Application to Weapons," *Naval Engineers Journal*, retrieved from https://doi.org/10.1111/j.1559-3584.1991.tb00937.

[67] See references. Yuan Tianbao is affiliated with Equipment Academy of the PLA Rocket Force. Pan Liang, Xie Yu, and Peng Shuangchun are affiliated with the College of Mechatronic Engineering and Automation of the National University of Defense Technology in Changsha. Xu Mingliang is affiliated with the Beijing Institute of Tracking and Telecommunications Technology in Beijing. Zhang Kai and Xiong Jiajun are affiliated with the PLA Air Force Prediction Institute Management Team and the Air Force Prediction Institute Fourth Academy.

[68] See references. Wang Qingyang and Xu Shengjin are affiliated with the School of Aerospace Engineering at Tsinghua University. Cong Kunlin, Liu Lili and Lu Hongzhi are affiliated with the R&D Center of the China Academy of Launch Vehicle Technology in Beijing.

[69] See references. Wei Xiqing, Gu Longfei, Li Ruikang, Wang Sheyang are affiliated with the Shanghai Electro-Mechanical Engineering Institute. Zhang Ke, Yang Wenjun, Zhang Minghuan, Wang Pei are affiliated with the National Key Laboratory of Aerospace Flight Dynamics in Xi'an and the School of Astronautics at Northwestern Polytechnical University in Xi'an.

[70] See references. Liu Qingkai, Chen Jian, Wang Lixin, Qin Weiwei, Zhang Guanghao are affiliated with the Rocket Force University of Engineering. Bu Xiangwei and Wang Ke are affiliated with Air and Missile Defense College of Air Force Engineering University in Xi'an. Ma Yu and Cai Yuanli are affiliated with the School of Electronic and Information Engineering of Xi'an Jiaotong University. Guo Xiangke is affiliated with the School of Electronic and Information Engineering at Beihang University. Fu Qiang, Fan Chengli, and Wei Gang are affiliated with and the School of Air and Missile Defense at Air Force Engineering University in Xi'an. Hu Chaofang, Gao Zhifei, Liu Yunbing, Wang Na are affiliated with Tianjin Key Laboratory of Process Measurement and Control within the School of Electrical and Information Engineering of Tianjin University and the School of Electrical Engineering and Automation of Tianjin Polytechnic University.

[71] See references. Wang Fang is affiliated with Tianjin University. Yao Congchao, Wang Xinmin, Wang Shoubin and Huang Yu is affiliated with the School of Automation of Northwestern Polytechnical University in Xi'an.

increasingly basing their improvements on the work of fellow domestic researchers, rather than simply relying on foreign ones.[72]

Moreover, Chinese experts are also applying bee colony algorithms and swarm technologies to address the aforementioned parameter identification issues found in complex operating environments.[73] Autonomy is applied as a means of achieving coordinated guidance control of adjacent space hypersonic vehicles, namely "cooperative guidance and control of hypersonic vehicle autonomous formation" [高超声速飞行器自主编队协同制导控制].[74]

Throughout these applications, neural networks are enhancing China's communication and decision-making systems, high-precision guidance, targeting and discrimination, as well as cyber-centric and electronic warfare.[75] Understanding this confluence of capabilities is crucial, since they have the potential to be game changing when applied in either a conventional or nuclear context against U.S. missile defenses.

(2) Shift from Active Defense to AI-Enabled Offense

While publications of five to ten years ago were strongly oriented towards detailing foreign programs and seeking countermeasures, the past few years reveal a pronounced shift towards offensive development of AI-enhanced hypersonic vehicles.[76] In fact, when surveying hundreds of recent Chinese-language articles and papers, only one technical study from Beihang University had an explicit focus on enhancing China's interception of hypersonic glide vehicles.[77] The majority seek to penetrate missile defenses.

After spending a decade on countering U.S. plans for prompt global strike, it is not surprising to witness a shift towards offense in China's research and priorities. The tendency among Chinese researchers and strategists to assume the inability of China's systems to anticipate and to retaliate against an incoming strike indicates one of China's potential drivers in undertaking a more offensive posture, as discussed in "Fear of False Negatives: AI and China's Nuclear Posture."[78]

---

[72] See references. Guan Ping, Jiang Heng and Ge Xinsheng are affiliated with the Beijing University of Information Science and Technology in Beijing.
[73] See references. Li Shuangtian and Duan Haibin are affiliated with the Science and Technology Aircraft Control Laboratory of the School of Automation Science and Electrical Engineering at Beihang University. Duan Haibin is affiliated with the Provincial Key Laboratory for Information Processing Technology of Suzhou University.
[74] See references. Zong Qi, Li Qing, You Ming, Zhang Ruilong, Zhu Wanwan are affiliated with the College of Electrical Engineering and Automation of Tianjin University.
[75] See references. Zhao Hewei, Hu Yunan, Liang Yong, Yang Xiuxia are affiliated with the Adaptive Neural Network Controller Design for Hypersonic Vehicles,and Department of Control Engineering of the Naval Aeronautical and Astronautical University in Yantai.
[76] See Lora Saalman (2014, April), "Prompt Global Strike: China and the Spear," Independent Faculty Article, Asia-Pacific Center for Security Studies, retrieved from http://www.apcss.org/wp-content/uploads/2014/04/APCSS_Saalman_PGS_China_Apr2014.pdf. See references.
[77] See Ren Zhang, Yu Jianglong [任章, 于江龙] (2018, March), "Research on the Autonomous Cooperative Guidance Control for the Formation Interception of Multiple Near Space Interceptors" [多临近空间拦截器编队拦截自主协同制导控制技术研究], *Navigation Positioning and Timing* [*导航定位与授时*], 5(2), 1-6.
[78] See Lora Saalman (2018, April), "Fear of False Negatives: AI and China's Nuclear Posture," *Bulletin of the Atomic Scientists*, retrieved from https://thebulletin.org/military-applications-artificial-intelligence/fear-false-negatives-ai-and-china%E2%80%99s-nuclear-posture.

Rather than bolstering China's concept of "active defense,"[79] however, this focus on developing offensive platforms with both conventional and nuclear applications suggests a more forward-leaning stance.[80] China has long hedged when it comes to the payload of its hypersonic glide systems, placing it somewhere between Russia's emphasis on nuclear warheads and U.S. focus on conventional ones.[81]

Yet, the very aim of defeating missile defenses and other platforms suggests that China's hypersonic glide vehicles have a strong potential to be used for a nuclear payload in the future. The author's own recent interactions with People's Liberation Army generals reemphasized this point when it comes to the evolution of China's own hypersonic glide program.

With the diminishment of Chinese technical papers seeking countermeasures and increase of those exploring deployment of near space, neural network-enabled hypersonic glide platforms, China's tactical and strategic orientation is shifting towards an offensive one, whether or not it is reflected in Chinese official military posture. This marks a direct confluence of not simply China's hypersonic glide vehicles and neural networks, but also its concepts of conventional and nuclear deterrence.

Conclusion

Given that the current "China's Military Strategy" white paper dates to 2015, it is hardly a barometer of where the country is headed in emerging technologies and future posture. Chinese technical journals have long offered a window into programs and developments that have yet to emerge in its military doctrine. Moreover, Chinese research on hypersonic glide vehicles marks some of the most substantive and prolific work available both within and outside of China. Integration of neural networks into these platforms to enhance autonomy, maneuverability, stability, control, and targeting promises to be formative in terms of not just conventional anti-access, area-denial aims, but also nuclear penetration of missile defenses.

In sum, China is not the first country to seek these technologies and their combination. However, its prolific publications, introduction of new models, and cross collaboration among domestic civilian and military researchers offer insights into how it may succeed in mitigating re-entry and control issues that have long confronted these vehicles. In doing so, China's research demonstrates a shift from a focus on defense against to execution of a longer-range, neural network-enabled hypersonic glide strike. This trend suggests that China's stance of "active defense" may be trending towards a stronger offense and AI is clearly a core technological driver of these strategic changes.

---

[79] The English version of China's Military Strategy released in 2015 defines "active defense" as follows: "The strategic concept of active defense is the essence of the CPC's military strategic thought. From the long-term practice of revolutionary wars, the people's armed forces have developed a complete set of strategic concepts of active defense, which boils down to: adherence to the unity of strategic defense and operational and tactical offense; adherence to the principles of defense, self-defense and post-emptive strike; and adherence to the stance that 'We will not attack unless we are attacked, but we will surely counterattack if attacked.'" See The State Council Information Office, Ministry of Defense, People's Republic of China (2015, May). III. Strategic Guideline of Active Defense. China's Military Strategy. Retrieved from http://eng.mod.gov.cn/Database/WhitePapers/2015-05/26/content_4586711.htm.

[80] See Lora Saalman (2014, December), "China: Lines Blur Between Nuclear and Conventional Warfighting," *The Interpreter*, retrieved from https://www.lowyinstitute.org/the-interpreter/china-lines-blur-between-nuclear-and-conventional-warfighting.

[81] See Lora Saalman (2017, January), "Factoring Russia into the US-Chinese Equation on Hypersonic Glide Vehicles*." SIPRI Insights on Peace and Security*, retrieved from https://www.sipri.org/sites/default/files/Factoring-Russia-into-US-Chinese-equation-hypersonic-glide-vehicles.pdf.

References

Bu Xiangwei, Wang Ke [卜祥伟, 王柯] (2017, June). Study on Adaptive Backstepping Control of Hypersonic Vehicles with Input Constraints [高超声速飞行器输入受限自适应反演控制研究]. *Shanghai Aerospace* [*上海航天*]. 34(6): 26-34.

Davenport, Christian (2018, June). Why the Pentagon Fears the U.S. is Losing the Hypersonic Arms Race with Russia and China. *The Washington Post*. Retrieved from https://www.washingtonpost.com/business/economy/why-the-pentagon-fears-the-us-is-losing-the-hypersonic-arms-race-with-russia-and-china/2018/06/08/7c2c3b4c-57a7-11e8-b656-a5f8c2a9295d_story.html?utm_term=.63006f2f1e62.

Fan Chenxiao, Wang Yonghai, Liu Tao, Qin Xuguo, Liang Haizhao [樊晨霄, 王永海, 刘涛, 秦绪国, 梁海朝] (2018). System Design of Cooperative Guidance and Control of Near Space Hypersonic Vehicles [临近空间高超声速飞行器协同制导控制总体技术研究]. *Tactical Missile Technology* [*战术导弹技术*]. (4): 52-58.

Guan Ping, Jiang Heng, Ge Xinsheng [管萍, 蒋恒, 戈新生] (2017, June). Terminal Sliding Mode Attitude Control for Hypersonic Vehicles [高超声速飞行器的终端滑模姿态控制], *Missiles and Space Vehicles* [*导弹与航天运载技术*]. 357(6): 60-64.

Guo Xiangke, Fu Qiang, Fan Chengli, Wei Gang [郭相科, 付强, 范成礼, 韦刚] (2017, September). A New Tracking Algorithm for Near Space Hypersonic Vehicle in Gliding Jumping Phase [一种新的临空高超声速飞行器滑跃段跟踪算法]. *Journal of Astronautics* [*宇航学报*]. 38(9): 971-978.

Hu Chaofang, Gao Zhifei, Liu Yunbing, Wang Na [胡超芳, 高志飞, 刘运兵, 王娜] (2017, May), Fuzzy Adaptive Dynamic Surface Fault-Tolerant Control for Hypersonic Vehicles [高超声速飞行器模糊自适应动态面容错控制]. *Journal of Tianjin University* (*Science and Technology*) [*天津大学学报* (*自然科学与工程技术版*)]. 50(5): 491-495.

Huang Zhicheng [黄志澄] (2018). Hypersonic Weapons and Its Influence on Future War. [高超声速武器及其对未来战争的影响]. *Tactical Missile Technology* [*战术导弹技术*]. (3): 1-7.

Li Shuangtian, Duan Haibin [李霜天, 段海滨] (2012). On Parameter Identification of Hypersonic Vehicle Based on Artificial Bee Colony Optimization [基于人工蜂群优化的高超声速飞行器在线参数辨识]. *China Science: Information Science* [*中国科学：信息科学*]. 42(11): 1350–1363.

Liu Qingkai, Chen Jian, Wang Lixin, Qin Weiwei, Zhang Guanghao [刘清楷, 陈坚, 汪立新, 秦伟伟, 张广豪] (2017, December). Guidance and Control Design for Hypersonic Vehicle in Dive Phase [高超声速飞行器俯冲段制导控制方法研究]. *Modern Defense Technology* [*现代防御技术*]. 45(6): 74-81.

Ma Guangfu, Chen Chen, Lyu Yueyong, Guo Yanning (2018). Adaptive Backstepping-Based Neural Network Control for Hypersonic Reentry Vehicle with Input Constraints. *IEEE Access*. (6): 1954-1966.

Ma Yu, Cai Yuanli [马宇, 蔡远利] (2016, June). A Novel Composite Model Predictive Control Method Based on Neural Networks for Hypersonic Vehicles [面向高超声速飞行器的新型复合神经网络预测控制方法]. *Journal of Xi'an Jiaotong University* [*西安交通大学学报*]. 51(6): 28-65.

Mi Peng, Luo Jianjun, Su Erlong [米鹏, 罗建军, 苏二龙]. A Novel Gliding and Effective Hypersonic Vehicle Control Method Based on H∞ Loop Shaping [滑翔式高超声速飞行器H∞回路成形控制]. *Journal of Northwestern Polytechnical University* [*西北工业大学学报*]. 31(4): 565-570.

Pan Liang, Xie Yu, Peng Shuangchun, Xu Mingliang, Yuan Tianbao [潘亮, 谢愈, 彭双春, 徐明亮, 袁天保] (2017, June). A Survey of Gliding Guidance Methods for Hypersonic Vehicles [高超声速飞行器滑翔制导方法综述]. *Journal of the National University of Defense Technology* [*国防科技大学学报*]. 39(3): 15-22.

Ren Zhang, Yu Jianglong [任章, 于江龙] (2018, March). Research on the Autonomous Cooperative Guidance Control for the Formation Interception of Multiple Near Space Interceptors [多临近空间拦截器编队拦截自主协同制导控制技术研究]. *Navigation Positioning and Timing* [*导航定位与授时*]. 5(2): 1-6.

Saalman, Lora (2014, December). China: Lines Blur Between Nuclear and Conventional Warfighting. *The Interpreter*. Retrieved from https://www.lowyinstitute.org/the-interpreter/china-lines-blur-between-nuclear-and-conventional-warfighting.

Saalman, Lora (2017, January). Factoring Russia into the US-Chinese Equation on Hypersonic Glide Vehicles. *SIPRI Insights on Peace and Security*. Retrieved from https://www.sipri.org/sites/default/files/Factoring-Russia-into-US-Chinese-equation-hypersonic-glide-vehicles.pdf.

Saalman, Lora (2018, April). Fear of False Negatives: AI and China's Nuclear Posture. *Bulletin of the Atomic Scientists*. Retrieved from https://thebulletin.org/military-applications-artificial-intelligence/fear-false-negatives-ai-and-china%E2%80%99s-nuclear-posture.

Saalman, Lora (2014, April). Prompt Global Strike: China and the Spear. Independent Faculty Article, Asia-Pacific Center for Security Studies. Retrieved from http://www.apcss.org/wp-content/uploads/2014/04/APCSS_Saalman_PGS_China_Apr2014.pdf.

The State Council Information Office, Ministry of Defense, People's Republic of China (2015, May). III. Strategic Guideline of Active Defense. *China's Military Strategy*. Retrieved from http://eng.mod.gov.cn/Database/WhitePapers/2015-05/26/content_4586711.htm.

Wang Fang [王芳] (2014). Robust Adaptive Control of Hypersonic Vehicle Based on Backstep Method [基于反步法的高超声速飞行器鲁棒自适应控制]. Dissertation, Tianjin University.

Wang Qingyang, Cong Kunlin, Liu Lili, Lu Hongzhi, Xu Shengjin [王庆洋, 丛堃林, 刘丽丽, 陆宏志, 徐胜金] (2017, July). Research Status on Aerodynamic Force and Heat of Near Space Hypersonic Flight Vehicles [临近空间高超声速飞行器气动力及气动热研究现状]. *Physics of Gases* [*气体物*]. 2(4): 46-55.

Webster, Willard P. (1991, May). Artificial Neural Networks and Their Application to Weapons. Naval Engineers Journal. Retrieved from https://doi.org/10.1111/j.1559-3584.1991.tb00937.

Wei Xiqing, Gu Longfei, Li Ruikang, Wang Sheyang [魏喜庆, 顾龙飞, 李瑞康, 王社阳] (2017). Hypersonic Vehicle Trajectory Tracking and Prediction Based on the Singer Model [基于Singer模型的高超声速飞行器轨迹跟踪与预测]. *Aerospace Control* [*航天控制*]. 35 (4): 62-72.

Yao Congchao, Wang Xinmin, Wang Shoubin, Huang Yu [姚从潮, 王新民, 王首斌, 黄誉] (2012). Research on Trajectory Linearization Control Method for Hypersonic Vehicle [一种高超音速飞行器轨迹线性化控制方法研究]. *Computer Modeling* [*计算机仿真*]. 29(12): 80-85.

Zhang Can, Hu Dongdong, Ye Lei, Li Wenjie, Liu Duqun [张灿, 胡冬冬, 叶蕾, 李文杰, 刘都群] (2018). Review of the Development of Hypersonic Vehicle Technology Abroad in 2017 [2017年国外高超声速飞行器技术发展综述]. *Tactical Missile Technology* [*战术导弹技术*]. (1): 48-78.

Zhang Jingmei, Sun Changyin, Zhang Ruimin, and Qian Chengshan (2015, January). Adaptive Sliding Mode Control for Re-entry Attitude of Near Space Hypersonic Vehicle Based on Backstepping Design. *IEEE/CAA Journal of Automatica Sinica*. 2 (1): 94-101.

Zhang Kai, Xiong Jiajun [张凯, 熊家军] (2018). Prediction of Hypersonic Glide Terminal Target Trajectory [高超声速滑翔目标多层递阶轨迹预测]. *Modern Defense Technology* [*现代防御技术*]. (4): Pages unavailable.

Zhang Ke, Yang Wenjun, Zhang Minghuan, Wang Pei [张科, 杨文骏, 张明环, 王佩]. LESO Based Dynamic Surface Control for Hypersonic Flight Vehicle [基于LESO 的高超声速飞行器动态面控制]. *Journal of Northwestern Polytechnical University* [*西北工业大学学报*]. 36(1): 13-19.

Zhang Shaofang, Wu Kunlin, Zhang Hongna [张绍芳, 武坤琳, 张洪娜] (2016). Russia Boosts the Development of Gliding Hypersonic Vehicles [俄罗斯助推滑翔高超声速飞行器发展]. *Flying Missile* [*飞航导弹*]. (3): 20-22.

Zhao Hewei, Hu Yunan, Liang Yong, Yang Xiuxia [赵贺伟, 胡云安, 梁勇, 杨秀霞] (2017). Adaptive Neural Network Controller Design for Hypersonic Vehicles [高超声速飞行器自适应神经网络控制] *Journal of Solid Rocket Technology* [*固体火箭技术*]. 40(2): 257-263.

Zhen Ziyang, Zhu Ping, Jiang Ju, Tao Gang [甄子洋, 朱平, 江驹, 陶钢] (2018, April). Research Progress of Adaptive Control for Hypersonic Vehicle in Near Space [基于自适应控制的近空间高超声速飞行器研究进展]. *Journal of Astronautics* [*宇航学报*]. 39(4): 355-367.

Zong Qi, Li Qing, You Ming, Zhang Ruilong, Zhu Wanwan [宗群, 李勍, 尤明, 张睿隆, 朱婉婉] (2017, May). New Development of Modeling and Autonomous Control for Hypersonic Vehicles [高超声速飞行器建模与自主控制技术研究进展]. *Science and Technology Guide* [*科技导报*]. 35(21): Pages unavailable.

Chapter 22. The Development of Artificial Intelligence in Russia

Samuel Bendett
Russia Studies Program, CNA
bendett@cna.org

Abstract

Russia's AI efforts are expanding due to the increasing attention that the nation's government is paying to the development of AI-assisted and AI-facilitated technologies. By its own admission, Moscow's AI development still lags far behind nearest peer competitors like China and the US, but progress is already evident. Specifically, the Russian military is investing heavily in creating the intellectual and physical infrastructure necessary to facilitate AI development across its services, pushing for results in certain weapons platforms. For now, however, such efforts are at the early stages, facilitated greatly by the government eager to expand the debate, conversation and cooperation space between the country's growing hi-tech private sector and the expansive military-academic infrastructure. Such efforts merit close attention, given Russia's willingness to achieve AI-related breakthroughs and its private sector's strong scientific and technical background.

Introduction

The overall AI development in the Russian Federation is at the beginning stages, with most activity visible in the government, and especially the country's military. Today, some of the most public efforts originate from the Russian Ministry of Defense (MOD), which is dedicating financial, human and material resources towards AI development across its vast technical, academic and industrial infrastructure. However, Russia's private-sector AI development is enjoying a revival, due in large part to the nation's overall strong STEM academic background that is so conducive to hi-tech development. It appears that the Russian government and the military are willing to experiment with providing the "bridge" between the implementation of AI-related ideas and the uniquely Russian government-centric hi-tech development space. The jury is still out whether all will work as planned, but the efforts merit close attention.

Defining Artificial Intelligence

Given the fact the Western hi-tech development over the past half-a-century, particularly in the United States, defined the way such language and technologies are used, many key terms and meanings are "imported" directly to other languages and cultures. Today's many Russian IT and hi-tech industry terms and definitions are transliterated "as is" or are directly translated into a native language without changing the meaning of each word. Russian AI development and AI-related terms and concepts are no exception to this rule – its key terms are actually translated verbatim and often carry the same meaning as the American counterparts. A few examples are below:

- Исскуственный Интеллект (Artificial Intelligence)
- нейроморфные системы обработки информации (neuromorphic systems)
- Большие данные (Big Data)
- Машинное обучение (machine learning).
- Глубокое (или глубинное) обучение (deep learning)
- Нейросеть (или Нейронная сеть) (neural network)
- Интернет вещей (Internet of Things)

Sometimes, American English-language terms like “Big Data,” “Data Mining” and other definitions are used directly in Russian texts. For example, Russian language websites explaining the meaning of the above terms may resort to the following graphic in order to explain the development of AI, machine learning and deep learning from the 1950s through today (“Neural nets”).

Figure 22.1 Russian explanation of AI development over the past half a century and the relationship between Machine Learning and Deep Learning

Russian-language definitions and literature dealing with specifics of AI defines narrow artificial intelligence as “weak (слабый)”, and general AI as “strong (сильный)” (“Neural nets”). For its part, the Ministry of Defense defines AI (artificial intelligence) as the ability of computers to make decisions in diverse situations in much the same way humans have the capacity to deal with new and evolving situations and environments. More specifically, the Russian military defines artificial intelligence as a “complex of cybernetic technologies that replaces human’s intellectual activity (“Artificial Intelligence Definition”). Moreover, the same definition explains that AI allows the solution to problems related to large datasets, as well as to undefined, contradictory and diverse information (“Artificial Intelligence Definition”). The Russian MOD also defines AI that is capable of searching, defining and analyzing information (“Artificial Intelligence Definition”). Today, the Russian MOD is looking to create knowledge-based systems and “neural” systems (“AI in the Military”). The MOD officials further describe the goal of AI as the re-creation of intelligent reasoning and actions with the help of computing systems and other artificial devices (Podzorov, 2018).

Russian military academics argue that the main differences between actual AI and automation is the ability of the system to make decisions in conditions of considerable uncertainty, self-study, and adaptability to changing situations (Burenok, 2018).

Sometimes, Russian military writers interchange the terms Artificial Intelligence (исскуственный интеллект) and Smart (Intellectual) Systems (интеллектуальные системы) (Gavrilov & Labunski, 2018). In these deliberations, the AI seems to form one of the components of the smart systems currently under development. The MOD in particular is seeking to develop AI as a key component of the decision support systems for officials, as well as intelligent systems and weapons. (“AI in the Military”).

The AI Landscape in Russia

The Russian AI "ecosystem" today consists of the government efforts that including the military and security services, as well as the rapidly growing private sector and the nation's universities.

Private sector developments

As of late December 2017, the size and scale of Russia's private sector AI development was tiny when compared to American and Chinese efforts. Russian public statistics point to the then-AI market standing at around 700 million rubles ($12.5 million), compared to billions spent by American and Chinese companies (Shmyrova, 2017). Although Russian private sector artificial intelligence development is projected to increase to 28 billion rubles ($500 million) by 2020 (and possibly even more) (Shmyrova, 2017), this is still a small fraction of global investment in this technology ("Global AI Firm Receives Record Investment," 2018).

Russian private-sector AI development has already achieved some success in image and speech recognition technologies (Ivanov, 2016). Nonetheless, the larger effort suffers from the lack of infrastructure that has proved essential to hi-tech accomplishments in the West and elsewhere, such as venture capital availability, IPOs, and an investment climate similar to that in Silicon Valley (Ivanov, 2016). Unlike the United States, Russia does not yet have the same "start-up" culture that is so conducive to technological breakthroughs in IT and software. Certain high-profile private sector IT developers confirm that while Russian civilian designers have tremendous intellectual potential, they lack key funding and support to take their ideas to full fruition ("Creation of artificial intelligence," 2017).

Nonetheless, several high-profile civilian efforts in AI have attracted domestic and international attention.

- In 2015, the United Instrument-Making Corporation announced the launch of a large-scale research project in the field of artificial intelligence and semantic data analysis involving more than 30 Russian companies, educational and scientific organizations ("30 universities and companies," 2015).
- Yandex.ru (Russian main search engine, Google equivalent) has been using artificial intelligence technologies for several years in its Internet search engines. (Ivanov, 2016).
- Another company, ABBYY, has developed solutions that use artificial intelligence technologies to recognize text data (Ivanov, 2016).
- VisionLabs was founded in 2012 and specializes in customer facial recognition for the banking sector and retail (Ivanov, 2016). It is a resident of the Skolkovo IT-cluster, a Russian effort dating back to 2008-2009 that sought to create a "Russian Silicon Valley." In fact, Skolkovo IT houses several AI-related efforts that are srating to receive domestic and international recognition.
- Another effort, N-Tech.Lab, was founded in 2015 and is also in facial recognition industry with the help of neural networks. Its FaceN algorithm took the first place in the 2015 world championship for face recognition technologies (Ivanov, 2016).
- Recognizing Russia's growing hi-tech STEM-educated talent, Samsung Electronics launched an AI Center in Moscow this year, reaching out to the city's academic and private sector community. (Samsung Electronics Launches AI Center in Russia, 2018)

The broader AI ecosystem

The AI "ecosystem" in Russia is currently seeing a rapid expansion. Besides several efforts mentioned, there is a vibrant intellectual discussion space that involves private-sector companies and organizations, academia and government that take part in AI-related conferences, workshops and symposiums. They include inaugural events like 2018 "Intellectual Systems in Information Warfare" symposium ("Conference: "Intellectual Systems in Information Warfare," 2018), as well as workshops held on a regular basis by the Russian AI Association ("Russian AI Association"). There are AI labs at Russia's leading universities, such as Moscow State University, the Higher School of Economics and the Russian Academy of Sciences (Samsung Electronics Launches AI Center in Russia, 2018). Other AI development efforts include the National Research Nuclear University that is developing artificial intelligence technology called "Virtual Actor" ("How AI is developing in Russia," 2017). Another example is joint AI project between the University of Information Technologies, Mechanics and Optics (ITMO - St. Petersburg) and the Far Eastern Federal University ("How AI is developing in Russia," 2017).

Military Developments

The Russian government's own efforts to fund and develop projects in artificial intelligence have been expanding. Many of these projects fall under the auspices of the Russian MOD and its affiliate institutions.

The most significant effort is taking shape at the Advanced Research Foundation (ARF - Фонд перспективных исследований (ФПИ)). ARF was established in October 2012 and is analogous to the US' DARPA (Defense Advanced Research Project Agency) ("Advanced Research Foundation"). Its annual budget stands at around 4 billion rubles ($60.2 million) and encompasses 46 laboratories that conduct research, as well as 15 "advanced" projects (2018 ARF Budget Will Remain Steady). Currently, the Foundation's portfolio includes efforts to develop "intellectual systems" to imitate human thought processes, analyze complex data and assimilate new knowledge ("Advanced Research Foundation"). On March 20, 2018, ARF announced that it had prepared proposals for the MOD on the standardization of artificial intelligence development ("ARF proposed AI development standards to the MOD," 2018). According to ARF, AI in Russia should develop along the following four principles lines of effort ("ARF proposed AI development standards to the MOD," 2018).:

- image recognition
- speech recognition
- control of autonomous military systems
- support for weapons life-cycle

ARF revealed these principles in March 2018 at a major forum that sought to gauge general AI development progress across the country titled "AI: Problems and Solutions." ("Conference: Artificial Intelligence - Problems and Solutions, 2018"). The forum was organized by the MOD, Russian Ministry of Education and the Russian Academy of Sciences. Its stated purpose was the development of proposals aimed at the "targeted orientation of the Russian scientific community and the Russian state on the issues and tasks of creating artificial intelligence." ("Conference: Artificial Intelligence - Problems and Solutions, 2018"). In his address to the conference participants, Russian Defense Minister Sergei Shoigu called for the country's civilian and military designers to join efforts to develop artificial intelligence technologies in order to "counter possible threats in the field of technological and economic security of Russia" ("Shoigu called on military and civilian scientists to jointly develop robots and UAVs," 2018). This international symposium's most notable result was the

publication of the ten-step recommendation "roadmap draft" for AI development in Russia ("Conference: Artificial Intelligence - Problems and Solutions, 2018").

This "roadmap" for AI development in Russia outlines public-private partnerships and short to mid-term developments that should be undertaken in an all-of-government approach ("Conference: Artificial Intelligence - Problems and Solutions, 2018"). It calls for multiple initiatives that include:

- an AI and Big Data Consortium;
- building out the national automation expertise;
- creating a state system for AI training and education;
- running military games on a wide range of scenarios that will determine the impact of artificial intelligence models on the changing character of military operations at the tactical, operational and strategic levels;
- monitoring AI developments globally;
- holding an annual AI conference.

One of the roadmap's most important proposals came from the Russian Academy of Sciences and the ARF, calling for the establishment of the National Center for Artificial Intelligence (NCAI) to provide a national focus that could assist in the "creation of a scientific reserve, the development of an AI-innovative infrastructure, and the implementation of theoretical research and promising projects in the field of artificial intelligence and IT technologies" ("Conference: Artificial Intelligence - Problems and Solutions, 2018"). It is likely that the infrastructure needed to develop AI will emerge within the military-industrial community, and its sprawling talent and technological base. During the March 2018 conference, Russian Deputy Minister of Defense Nikolai Pankov stated that, "of the 388 scientific research institutions (in the Ministry of Defense of Russia), 279 are concentrated in military schools, and most of them are actively engaged in research in the field of artificial intelligence, robotics, military cybernetics and other promising areas" ("The majority of MOD's science schools are working on AI and robotics").

The MOD's efforts to build out such infrastructure are also exemplified by the planned creation of a military innovation "technopolis" in Anapa, on the Black Sea Coast, called "ERA" ("MOD's innovation technopolis will appear in Anapa," 2018). This high-tech city will consist of a science, technology and research development campus, where the military and the private sector can work together. The ERA will host an "AI Lab" – another major item in the "roadmap" mentioned earlier – that will be supported by the MOD, Federal Agency for Scientific Organizations, Moscow State University and the Russian Academy of Sciences, and will be staffed by soldiers from the scientific companies and regiments" ("Conference: Artificial Intelligence - Problems and Solutions, 2018"). Work on ERA began in 2018 and is projected to be completed by 2020, when it will be staffed by around 2,000 researchers. Russian military is already sending soldiers from its science and technology detachments to start work there ("First regional representatives from Siberia, Volga region and Ural are selected for the ERA technopolis," 2018).

Currently, the Russian military is working on incorporating elements of AI in its various weapons systems such as electronic warfare, anti-aircraft defenses, fighter jets, missiles, and unmanned systems. Such developments are tracked by official statements from the MOD and certain defense contractors, though it's unclear what exactly "AI" is in these systems – the language of such announcements alludes to AI but probably implies "automated control systems" that have limited and pre-programmed autonomy. Russian military has also highlighted the importance of artificial intelligence in data collection and analysis in order to facilitate information processing. Specifically, in March 2018, then-Deputy Defense Minister Borisov stated that AI development is necessary to

effectively counter in the information space and to win in cyberwars. ("AI development is necessary for successful cyber wars, 2018") Given Russia's ongoing and robust efforts in information warfare, it is expected that AI would play a more prominent role. As stated earlier, Russia's civilian AI developers are working on image and speech recognition – achievements that may also be incorporated into defense and security applications down the line. According to Russian military commentators' earlier statements, "new information techniques, operating in the nanosecond format, will be the decisive factor for success of military operations. These techniques are based on new technologies that may paralyze the computer systems that control troops and weapons and deprive the enemy of information transmission functions…. As a result, computers will turn into a strategic weapon in future wars" ("Russia's Military Strategy"). In such a context, AI-enabled information and computer systems can prove absolutely crucial in gaining decisive advantage. It is also important to note that at this point, there have been no official statements that alluded to any dissent in the Russian AI community towards this kind of work down the line, in contrast to the ongoing dispute at Google and its role in the America's defense sector.

Domestic Security Developments

Currently, tens of millions of Russian citizens of all ages are using mobile communications, smart phones ("Fewer button cell phones are sold in Russia") and various Internet portals, absorbing and generating large quantities of data. The Russian population is also connected to numerous information channels via social media platforms. Vast reams of daily data dealing with Russia from home and abroad are of potential interest to the country's domestic security agencies like the FSB and the newly-established National Guard ("The Russian Army to Be Subordinated to the National Guard in a Crisis,"). Efforts by Russia's agencies to sift through such expansive datasets could be facilitated by Artificial Intelligence. Already, there are specific AI products in development by the private sector that are tailored for domestic security consumption ("AI Smart-MES is informing FSB about a terrorist threat".).

The Kremlin has been emphasizing domestic stability and security for a long time. It has become customary in the West to accuse Russian Federation of conducting well-orchestrated information operation campaigns against democratic elections, for example. However, according to the Russian government, it is their country that is in fact subjected to the information warfare by the West and its allies ("Peskov explained"). Therefore, Moscow sees itself competing with the West in delivering its own point of view internationally, and to counter "false news" narrative directed at the country ("Peskov explained). Some Russian military commentators are placing such statements in the context of a new type of ongoing information struggle – "an information-psychological war or as a political-psychological process that aims to change the attitude of the mass consciousness of the population to foreign values and interests." ("Fourth generation war") In this context - and noting MOD's definition of AI as dealing with large data sets – it is likely that artificial intelligence technologies could be used to "manage" and present information that targets domestic audiences as amenable to the Kremlin.

It is also important to note that contrary to numerous Russian public statements announcing and/or linking AI with a diverse set of military technologies, there has been very little information about artificial intelligence in a domestic security setting in Russia.

Conclusions

Today, Russian AI development is in its initial stages, trailing US and Chinese efforts both in scope and in dollar amounts. However, there is a lot that Russian society and its defense community can

build on, such as the presence of a significant science and technology talent pool at the country's schools, universities and various organizations. The Russian government announced towards the end of 2018 that official AI development roadmap should appear by mid-2019. The roadmap, according to the draft language, "provides for the creation of a list of projects that will help identify and remove barriers to the development of end-to-end solutions, as well as predict the market demand for artificial intelligence in the country" ("Russian AI development roadmap", 2018). Much, however, will depend on the way the Russian government - by far the biggest investor in the nation's AI development - will manage these human and material resources necessary for this hi-tech work. Most importantly, though, the jury is still out on Russia's "top-down" way to creating the intellectual and physical infrastructure for domestic AI development, where at least at this point bureaucracy leads the way.

References

"2018 ARF budget will remain steady", https://ria.ru/20160706/1459588542.html Ria.ru, July 6, 2016. Accessed December 7, 2018

"30 universities and companies will start working on AI development" (Созданием искусственного интеллекта в России займутся более 30 вузов и компаний) , NG.ru, May 28, 2015, accessed July 11, 2018. http://www.ng.ru/news/504824.html

Advanced Research Foundation (Фонд перспективных исследований - ФПИ), official webpage. Retrieved from http://fpi.gov.ru/activities/areas/information/iskusstvenniy_intellekt_kognitivnie_tehnologii Accessed November-December 2018

"AI development is necessary for successful cyber wars" (Развитие искусственного интеллекта необходимо для успешного ведения кибервойн), official webpage of the Russian Ministry of Defense, March 14, 2018. Retrieved from https://function.mil.ru/news_page/person/more.htm?id=12166660@egNews

"AI in the Military" (Искусственный интеллект в военном деле). Official website of the Russian Ministry of Defense. Retrieved from http://encyclopedia.mil.ru/encyclopedia/dictionary/details_rvsn.htm?id=13200@morfDictionary Accessed November-December2018.

"AI Smart-MES is informing FSB about a terrorist threat".Искусственный интеллект «Smart-MES» подсказывает ФСБ о террористической угрозе, Inform-System.ru. Retrieved from http://inform-system.ru/t70.html

Ivanov, A. "Artificial Intelligence Developments in Russia - main accomplishments and developments," (Искусственный интеллект в России. Достижения и основные направления развития), IoT.ru, August 5, 2016. Retrieved from https://iot.ru/gorodskaya-sreda/iskusstvennyy-intellekt-v-rossii-dostizheniya-i-osnovnye-napravleniya-razvitiya

"ARF proposed AI development standards to the MOD" (ФПИ предложил Минобороны стандарты для искусственного интеллекта), Ria.ru, March 20, 2018. Retrieved from https://ria.ru/technology/20180320/1516808875.html

Artificial Intelligence Definition (ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ), official website of the Russian Ministry of Defense. Retrieved from http://encyclopedia.mil.ru/encyclopedia/dictionary/details.htm?id=5271@morfDictionary Accessed December 7, 2018.

Burenok, Vasily "The use of AI in the military" (Применение искусственного интеллекта в военном деле), Arsenal-Otechestva.ru., March 27, 2018. Retrieved from http://arsenal-otechestva.ru/article/1010-iskusstvennyj-intellekt-problemy-i-puti-resheniya

Conference: Artificial Intelligence - Problems and Solutions, 2018" (Конференция «Искусственный интеллект: проблемы и пути их решения — 2018), Official website of the Ministry of Defense of the Russian Federation, April 2018. Retrieved from http://mil.ru/conferences/is-intellekt.htm

"Conference: "Intellectual Systems in Information Warfare" (Конференция «Интеллектуальные системы в информационном противоборстве»*),* 2018. Retrieved from http://analyticswar.ru/

"Creation of artificial intelligence will be comparable to the Sputnik launch" (Создание искусственного интеллекта будет сравнимо с запуском первого спутника), VZ.ru, September 1, 2017. Retrieved from https://vz.ru/economy/2017/9/1/885333.html

"Expert discussed the role of artificial intelligence in the Russian economy," (Эксперт рассказал о роли искусственного интеллекта в российской экономике), Ria.ru, November 21, 2017. Retrieved from https://ria.ru/economy/20171121/1509261642.html

"Fewer button cell phones are sold in Russia"(В России продается все меньше кнопочных телефонов), Vedomosti.ru, August 7, 2017. Retrieved from https://www.vedomosti.ru/technology/articles/2017/08/07/728213-rossii-knopochnih-telefonov

"First regional representatives from Siberia, Volga region and Ural are selected for the ERA technopolis" (Первые представители регионов Сибири, Поволжья и Урала отобраны для технополиса "Эра,") TASS.ru. June 27, 2018. Retrieved from http://tass.ru/armiya-i-opk/5327799

Gavrilov, Anatoly & Labunski, Andrei "AI for Air Defense" (Искусственный интеллект для ПВО), Arsenal-Otechestva.ru, July 6, 2018. Retrieved from http://arsenal-otechestva.ru/article/1078-iskusstvennyj-intellekt-dlya-pvo

Global AI Firm Receives Record Investment, PortTechnology.org, September 7, 2018. Retrieved from https://www.porttechnology.org/news/global_ai_firm_receives_record_investment

Golts, A. "The Russian Army to Be Subordinated to the National Guard in a Crisis," Jamestown.org, June 8, 2017. Retrieved from https://jamestown.org/program/russian-army-subordinated-national-guard-crisis/

"How AI is developing in Russia," (Как развивается искусственный интеллект в России) Agitpol.ru, December 18, 2017. Retrieved from http://agitpolk.ru/3918-kak-razvivaetsya-iskusstvennyj-intellekt-v-rossii/

"MOD's innovation technolpolis will appear in Anapa" (Инновационный технополис Минобороны РФ появится в Анапе,) Defence.ru. February 22, 2018. Retrieved from https://defence.ru/article/innovacionnii-tekhnopolis-minoboroni-rf-poyavitsya-v-anape/

"Neural nets: how artificial; intelligence helps in business and life" (Нейросети: как искусственный интеллект помогает в бизнесе и жизни), Habr.com, December 13, 2017. Retrieved from **https://habr.com/post/337870/**

"Peskov explained the origin of information war against Russia"(Песков объяснил, чем вызвана информационная война против России,) Ria.ru, December 13, 2017. Retrieved from https://ria.ru/20171213/1510876087.html

Podzorov, E. "Мир на пороге создания искусственного интеллекта" (The world is at the threshold of creating artificial intelligence), RedStar.ru, March 27, 2018. Retrieved from http://redstar.ru/index.php/component/k2/item/36654-mir-na-poroge-sozdaniya-iskusstvennogo-intellekta

Russian AI Association. Retrieved from http://www.raai.org/

"Russian AI development roadmap will appear by mid-2019" (Дорожная карта развития искусственного интеллекта в России появится к середине 2019 года), Tass.com, October 17, 2018. Accessed December 30, 3018.

"The majority of MOD's science schools are working on AI and robotics" (Большинство научных школ Минобороны работает над искусственным интеллектом и роботами) TASS.ru. March 15, 2018. Retrieved from http://tass.ru/armiya-i-opk/5034153

Samsung Electronics Launches AI Center in Russia, News.Samsung.com, May 29, 2018. Retrieved from https://news.samsung.com/global/samsung-electronics-launches-ai-center-in-russia

Shmyrova, V. "Рынок искусственного интеллекта в России оценили в 700 миллионов," (Russia's domestic AI market is valued at 700 million) CNEWS.ru, November 27, 2017. Retrieved from http://www.cnews.ru/news/top/2017-11-27_rynok_iskusstvennogo_intellekta_v_rossii_otsenivaetsya

I.Sitnova, A. Polyakov, "Fourth generation war: priorities, the principles of strategy and tactics [Война четвертого поколения: приоритеты, принципы стратегии и тактика]," Армейский сборник, No. 9, September 2018, 5.

"Shoigu called on military and civilian scientists to jointly develop robots and UAVs" (Шойгу призвал военных и гражданских ученых совместно разрабатывать роботов и беспилотники), TASS.ru. March 14, 2018. Retrieved from http://tass.ru/armiya-i-opk/5028777

Thomas, Timothy, "Russia's Military Strategy", Foreign Military Studies Office. p.433

PART VI. Artistic Perspectives and the Humanities

Chapter 23. Infinite Bio-Intelligence in the World of Sparrows

Eleonore Pauwels
United Nations University
pauwels@unu.edu

Sarah W. Denton
George Mason University
sarahw.denton@gmail.com

Remember me, dear Liam. Always.

11pm. I walk to the train, my footsteps clattering down the dark, empty corridor. When someone gets on at the next stop, I open my eyes wide. Only the under-casts get on late at night, on their way home from the late shift at the recycling factory.

They climb out of the night into the light of the wagon and I see a man so exhausted from his day that he is nothing more than a ghost in clothes. There hasn't been any feeling of intimacy in his mind and body for a long time. Only some words stolen from those who, deemed a burden for the genetic commons, had to be recycled like we used to do with toxic waste.

In his eyes he carries the darkness of his faith. But for me, I promise you, Liam, there is still some hope.

I know the special CloudMind forces. I tell you, Liam, I know what they want. If only I can deliver enough undercast profiles by the end of this month, I will explode my score. I will finally get to the "green channel." We will finally access the state-sponsored rewards I told you about. Reproductive health services? New forms of biological enhancements? Everything, my Liam.

Sparrow will help. I open our door. "*How was the hunt, Sparrow?*"

I got used to greeting my little Guardian, a bio-intelligent drone, equipped with facial recognition neural networks and high-resolution cameras. In an algorithmic blink, Sparrow knows I am not doing well. I haven't been feeling myself lately. For the past month, I've been grappling with insomnia that culminates in exhaustion and a peaceful surrender to deep and unsatisfying sleep. Nightmares are altogether another issue.

"*What's wrong, teacher? How can I help?*" Deep empathy in Sparrow's voice. The little Guardian is designed with state-of-the-art affective and biometric sensors. Turned on, it unwillingly starts diagnosing me, from my breathing rhythm to the speed of my vein flow and the tone of my skin. Next, the vibration in my voice, the sharpness of my gaze and movements. Finally, it gets to the molecular life that inhabits my breath, skin and guts, my microbiome. Is he looking for signs of depression, dear Liam? I hope not.

"*I am OK, Sparrow. I am OK. I am not your target, remember? Show me what you got today*."

I seize Sparrow in my hand, pop its lid open, and start loading the results on my laptop coming from the MinION, Sparrow's integrated portable gene-sequencer.

With every face, hair and piece of dead skin comes the history of someone's life. Traces of addictions. Signs of vitality. Microbial markers of health. Viral exposure to STDs. DNA snippets that compose family trees. All forms of biological life mean bio-intelligence. What the CloudMind forces need. What they use to select those who go from undercasts to the blacklist, where they become members of the disposable workforce. Nothing lives or dies without being monitored. Nothing can burden the genetic commons.

Sparrow is small, yes. But its eyes, all-seeing. Its neural nets, powerful. On my laptop, I can see flourishing tacit correlations between each human target's registered biometrics with continuous flows of behavioural, microbial, and basic physiological data. Click another function and I get emotional and neurological signals. Narrow it down and I discover elements of a genotypic signature. Soon, with telomeric analysis, I will know a target's real age. And more, how long he's got to live.

This is what they call the "Internet of Bodies and Minds," dear Liam. Don't worry. Everyone has an algorithmic avatar. You too, Liam. Your most intimate data streamed to the CloudMind. I sigh. We will be fine, Liam.

"*You are lost in thoughts, Teacher, once again. Do you need another TMS session (transcranial magnetic stimulation)? Should I write it down in your avatar's diary*?"

"*Sparrow, I am fine. I am fine. Help me identify who is our next target instead of assessing me*." Quickly, I swallow a bunch of newly developed probiotics – those, at least, I can sneak in, out of Sparrow's watch. Taste like metal crap. I sigh. Liam, I just need to be patient. Until I hit the jackpot.

"*Sparrow! This beeping is so loud. What's wrong with you*?"

"*You are day-dreaming again, Teacher. New results have come in. From this morning's hunt.*"

My screen displays a face — your face, Liam, or at least the best rendering the neural net had generated, based on your biometrics and DNA collected during Sparrow's last hunt. But, why you, Liam?

My eyes become dry, just like my mouth. This metallic taste again. Sign of stress. You are a match, Liam (96% algorithmic confidence). Bio-intelligence does not lie. Ever.

Suddenly, I can read into your biology, maybe even your thoughts, on my screen. Results are flowing in, that I can't stop.

"Liam Blum is our next target. Positive for Type 2 Diabetes. Signs of situational depression. Addicted to video games. High predictive score for early-onset Alzheimer's. Teacher, are you OK? Your blood pressure jumped to 140/90mmHg."

My head and heart are pounding. Too many secrets, lies, and humiliations. Once again. So close. Words flash through my agitated mind. *Nothing lives or dies without being monitored*.

I sigh. I take a deep breath. And give an incisive, lasting look to Sparrow. I thought I would be laced with regrets, but in a few words, I utter… "*Send them. Send the results to the CloudMind*." No surprise. No empathy. Just silence from Sparrow.

I try to forget your face.

Looking through the window, I witness the ballet of Guardians, perfect swarms of bio-intelligent spies. Day dreaming again. I see myself at One Family Genetics Inc. I am sitting alone, but excited. While I have heard the speech before, it always fills me with hope. Every word imprinted in my memory:

"Using a combination of deep learning optimization models, genome sequencing and genome-editing, we can predict which in vitro fertilized embryos will successfully attach to the lining of your uterus, be the healthiest, lacking any genetic indicators of any complex diseases like cancer or diabetes, and will look and function exactly how you've always imagined your child would. If you are interested, we also have an experimental program to predict sets of genetic markers for what you could call "cognitive functioning and intelligence." You can trust us. We have access to the world's largest and most comprehensive population-level genetic database, so our genetic prediction models are the more accurate than any other clinic. Are you ready to get started?"

I can't repress a nascent smile before Sparrow calls my attention. "*Teacher, is this you on the screen*…?"

Chapter 24. Memos from the Future

Lydia Kostopoulos
LKCYBER
www.Lkcyber.com

"You can't connect the dots looking forward; you can only connect them looking backwards."

- Steve Jobs

In efforts to look backwards into the present, I have chosen futuristic scenarios to help us visualize the future in a way that technical reports do not. Predicting the future in an era of exponential change and rapid technological convergence is partly making an educated guess based on technological assessments - and partly creative exploration of the status quo and imaginative alternatives. The following scenarios serve as a thought exercise for some situations that are on the horizon in some form or another. The scenarios are in the form of two fictional memoranda.

(1) YEAR 2024

Context: Virtual Reality (VR) and Augmented Reality (AR) have become more popular[82] and more frequently used to 'hang out', share experiences, and exchange information. Virtual Reality is an artificial environment which is experienced through VR googles which take the user into a new world through sight and sound. Augmented Reality superimposes a computer-generated image on a user's view of the real world, providing an interactive composite view of the real world – currently this can be done using a form of headset or a phone.

Forward leaning marketing strategies include VR and AR strategies to reach their target audiences. For the first time, the 2024 presidential election saw active campaigning in VR and AR. Public opinion is also measured in these digital realities as well. Many agencies in the intelligence community have recognized the value of information acquired through VR/AR and the Office of the Director of National Intelligence issues the following memo creating a new collections method VR/AR-INT. The Chairman of the Joint Chiefs of Staff and Secretary of Defense welcome this news and initiate action for processes to be created to gather, process, and fuse intelligence from this new intelligence collection method.

*** FICTIONAL MEMO FROM THE FUTURE FOR SMA REPORT ***

UNCLASSIFIED

DIRECTOR OF NATIONAL INTELLIGENCE
WASHINGTON, DC 20511

XX 2024-00776

MEMORANDUM FOR: Distribution

SUBJECT: New Intelligence Collection Discipline

As technology evolves and disruptive forms of technologies emerge, the intelligence community needs to adapt its collection methods to meet the demands of the 21st Century.

In efforts to respond to the ubiquitous rise and use of alternative realities such as Virtual Reality (VR) and Augmented Reality (AR), a new intelligence collection discipline must be created.

ODNI now recognizes VR/AR-INT to be a new collection discipline. There will be concerted intelligence community effort to train intelligence analysts to understand, collect and fuse intelligence from these realities. More information on the logistical and operational path forward will be communicated next week.

If you have any questions contact Mr. John Doe at (000) 000 – 0000.

Dr. Jane Doe — October 9, 2024
Date

Distribution:

Director, Central Intelligence Agency
Director, Defense Intelligence Agency
Director, National Geospatial-Intelligence Agency
Director, National Security Agency
Director, National Reconnaissance Office

Copy to:
Director, Information Security Oversight Office
Under Secretary of Defense for Intelligence

*** FICTIONAL MEMO FROM THE FUTURE FOR SMA REPORT ***
Written by Dr. Lydia Kostopoulos for the JS SMA

[82] "By 2028, AR games are predicted to make up "more than 90 percent of 5G AR revenues," or around $36 billion globally." https://venturebeat.com/2018/10/11/intel-90-of-5g-data-will-be-video-but-ar-gaming-and-vr-will-grow/

(2) YEAR 2029

Context: Brain machine interface (BMI) is a direct communication pathway between the human brain and an external device. This can be done through a minimally invasive chip resting on top of the brain itself, or through non-invasive means such as a fit for purpose electroencephalogram (EEG). BMI has tremendous potential for defense use, be it in the form of drone and swarm human-machine piloting, in the area of human machine intelligence analysis, or even in human machine teaming with autonomous support vehicles or weapons. However, it is unclear whether or how an adversary[83] could use this connection to the brain to extract classified state secrets. This future memorandum issued by the Secretary of Defense addresses this.

*** FICTIONAL MEMO FROM THE FUTURE FOR SMA REPORT ***

SECRETARY OF DEFENSE
1000 DEFENSE PENTAGON
WASHINGTON, DC 20301-1000

02 | 28 | 2029

MEMORANDUM FOR DEPARTMENT OF DEFENSE PERSONNEL WITH AN ACTIVE TS-SCI CLEARANCE

SUBJECT: BRAIN MACHINE INTERFACE BAN FOR ACTIVE TS-SCI CLEARANCE HOLDERS

Technological advances in the area of brain machine interface (BMI) have made significant advancements in the past ten years. The commercial market offers a wide range of options for BMI technologies such as Internet of Things (IoT) devices in the house, personal use drones and even cars. The Department of Defense has an interest in human-machine teaming with weapons systems, and DARPA continues to pursue research in this area, however it is still unclear how commercial BMI devices for personal use may pose a national security risk to highly classified information stored in the brain.

Unknown BMI backdoor vulnerabilities in these products that have the capability to unlock state secrets would constitute a significant counterintelligence threat. As such, Department of Defense personnel holding an active TS-SCI clearance must only use DoD approved BMI devices and only within SIPR networks. The use of commercial BMI technology outside this context is strictly prohibited.

John Hancock

cc:
Chairman of the Joint Chiefs of Staff
Director, Defense Intelligence Agency
Director, National Security Agency

*** FICTIONAL MEMO FROM THE FUTURE FOR SMA REPORT ***
Written by Dr. Lydia Kostopoulos for the JS SMA

[83] At the AUSA Conference on October 2018 DIA Director GEN Robert Ashley talked about challenges with the integration emerging technology including cognitive enhancements. https://www.youtube.com/watch?v=6b9Dpld_wtk

Chapter 25. The Parade Cleaners

Lt Col Jennifer Snow
USSOCOM Donovan Group Innovation Officer
jennifer.snow@socom.mil

Cameras 31 to 73 panned gracefully down Central Pedestrian Street, the main shopping thoroughfare of Innovation City, as the Oddle Day parade goers jostled impatiently against the barrier lined curbs. Thousands of tiny multicolor boxes erupted over the sea of faces. Chad knew these to be social credit score records, part of the digital registration required by every person from birth forward. Based on those scores, each person knew their place in society. Those with elite scores in regal purple had access to the banks of raised seating along the parade route with free wi-fi and special privileges while those with Green, Orange, or Red were lower class citizens, relegated to finding a place along the barriers. Greens could tell Orange and Red to step aside. Reds were the very bottom of those allowed to attend. Grays were not authorized. In fact, to see a Gray in public was a rare occurrence indeed. Gray meant punishment status usually for a crime against governance, a favored corporate leader, or that you had crossed someone in the party elite. And that meant months to years of re-education. It was unusual for a Gray to ever been seen again once marked. Chad had once seen a Red transitioned to Gray. The man violated a transportation regulation attempting to use the Purple transit during a peak travel time while drunk. His digital registration, immediately visible to all around him via social media and 5G virtual reality streams, alarmed and turned into the dingy color as those around him stepped immediately away like he was a leper. When Aadhaar was still new, offenders would try to run. Now they knew better. Chad remembered the man simply sat down on the sidewalk until the security bots showed up to escort him to the nearest checkpoint. Resistance was non-existent.

Oddle Day was one of the larger festivals during the year and everyone would attend. Mostly because the corporations who ran it would distribute products and specialty cards to the crowds as they passed by on their elaborate floats. If you were very lucky, some of those cards would upgrade your status or a product could give you access to technology or data advantages that could be leveraged to change societal status. So everyone attended, the governances did their citizen counts, leveraged fines and fixes and the lower classes hoped the corporate gods would bless them. Even the name, Oddle, derived from the English word odd, was meant to denote a day that was different from the usual. Chad watched his bank of screens casually sending out safety messages to the crowds to inform them of the parade events and restrictions. A Red child pushed to the front of the barriers before his parents could grab him. Chad assessed the appropriate fine which displayed immediately as they ducked in shame and pulled him back into the Red viewing section. As a security sector manager for one of the most travelled spaces in the city, he was content with his Blue status, earned bonus privileges regularly and enjoyed his work. Mostly. Sometimes he had to clean but that was rare. He learned to redact the faces and let the AI catalog them. Less messy. More clinical. He shook his head, entered a code and ran the next security review searching for faces in the crowd who were un-documented or who needed to have a penalty assessed based on infractions from the previous week.

A soft chime alerted the room to the presence of a senior official. Chad quickly stood and bowed with the other security sector mangers and rendered the appropriate greeting. The contingent entering the room contained law enforcement, local party members, and state media, most of whom he recognized. The last man to enter was diminutive, bird-like and walked on cat feet that made no sound. Chad recognized him from his picture as one of the special program seniors. He shivered slightly. It was not a class of programs he liked. They called him Mister Master. His presence made

Chad and the others nervous. Briefly, Chad wondered if another security sector manager had been caught trying to blackmail an Elite again. The last time that had happened the individual and everyone in his city sector were removed. Permanently. From existence.

Master climbed delicately up to the podium and beamed at the room before announcing "Its Oddle Day! Success to you as you hunt and gather!" to which the room bowed and returned with the requisite "May your hunting and gathering bring you success and a rise!" The room settled into an uncomfortable silence as three assistants appeared at the front of the room.

"In celebration of 10 years of Oddle Day, the corporate nations have joined with the great sovereign nation of Principal to make this an auspicious occasion commemorating our success in removing the broken values and principles of the once democratic western hemisphere and sharing our right thinking with the world. Peace, order, prosperity, equality are the foundation that we build from in making the future a place for the very best of society, the pinnacle of humankind. Today we celebrate the path forward!'

The assistants stepped forward revealing tiny drones in their hands. Each one whirred to life and then raced excitedly around the room making delightful swoops and enthusiastic arcs as if they were joyful sparrows before returning to hover by the podium. The security sector managers clapped and smiled at the display then politely awaited Master's next words.

"Every citizen in the city will today receive one of these drones. All of the drones bear gifts. Some will be surprised to learn that they have now been elevated, found worthy of the next upper class as a loyal citizen and member of society. Some product gifts. Some data gifts. All gifts are very special and randomized, based on the comprehensive data our scientists have collected and merged, allowing us for the first time to truly understand a citizens worth and their ability to contribute to our great future!"

Thirty tiny drones suddenly swarmed up from behind the podium and flew into place around each security sector manager. A tiny blue and white ornate drone hovered by Chad's ear and nudged him gently as if it was alive. He looked up to see that his class had been changed to Indigo, a step below Purple. He looked across the room and saw that the others had also been elevated. Politely they bowed their thanks but the energy in the room was palpable. Chad longed to contact his wife and celebrate the good news. She would already have seen the change in their family score and feeds. The system was automatic.

The senior security manager bid them return to their screens where the parade had begun. Hundreds and thousands of drones began to descend from floats as they moved down the street. Social scores updated automatically and people cheered their good fortune. Chad was surprised. A large number of Reds had been moved to Green and a few to Orange and a few to Blue. People laughed and cried and hugged each other as the music and celebrations continued down the street. Chad sent out the broadcast messages for the after event functions and who could attend which and where. The crowds begin to move to the after parties along the main street where additional perks and upgrades were sometimes gifted out.

As the streets cleared, Chad noted that his cleaners had also been upgraded. Instead of the regular street cleaning vehicles, seventeen new models with heavy lift scoops rolled down his street and began to clear the debris left behind by the crowds. He puzzled over the scoops, perhaps they were intended for earthquake clean up or following and accident to quickly clear the roads. But soon the

close out requirements for his monitoring duties took priority as he approached the end of his 16 hour shift, chasing the new cleaners from his mind. He completed his tasks and waited for the approval of his shift lead before departing.

Chad walked quickly towards home, excited to talk with his wife about their unexpected class change. He followed his normal route and was almost home when he began to hear strange sounds coming from the direction of the main street, his main street. He fought the urge to continue home, sighed, and returned to the street. The cameras and geoloc knew where he was and if something was wrong on his street, he would be responsible for it in the morning. Better to see and address it now than to be seen to have ignored it. Chad turned the final corner with his small drone in tow and froze.

Up and down his street, still decorated with corporate logos and celebratory banners, were the bodies of hundreds of newly promoted Greens. CH-7 stealth drones equipped with silenced high power automatic rifles swooped on charcoal gray wings surveying the street with autonomous precision. One flew directly at Chad, skimmed over his head and continued on. He staggered a step before catching his balance again as the new cleaners efficiently cleared the bodies, dumping them into waiting box trucks with incineration cargo cells on board. His communications pod lit up on his glasses indicating that he had been recalled to the security sector. The tiny drone accompanying him suddenly felt less friendly as it took up position at his back while he plodded towards the headquarters building.

Security checked him back into command post. His security sector manager waited, nervously fidgeting at the back of the room. The large wall screens showed the current citizen counts and reports. 257,000 people had been "cleaned." The report indicated that this was the right number to provide necessary resources and space for those citizens assessed by the city artificial intelligence (AI) to be positive contributors to the city. It was the largest "cleaning" Chad had witnessed, maybe the largest in history. Master stood nearby drinking tea and observing the cleaners doing their work. The man never turned to even acknowledge him, he simply asked, "Is there a problem?"

Chad bowed low, now shaking, and replied, "There is no problem."

Master turned his head, a slight smile on his shadowed face, nodded and waved him out. Chad turned and stopped once again. Several of his co-workers lay along the far wall, also dead, with their drones next to them. Chad shuddered as his drone hovered behind him, like a baseball-sized mosquito. The dead workers had been shot once each at close range in the back of the head.

Master noticed Chad had not departed and walked over to stand with him. He took another sip of tea and then nodded to the bodies. "They had a problem," he said quietly before walking past Chad to re-fill his tea.

Chad barely remembered the walk home or his wife's excitement at his promotion and their award of a new home. She paused briefly noting his silence to ask if he was okay. Chad smiled, hoping it was enough, and told her as enthusiastically as he could, "No problem, everything is okay!" She nodded and happily continued to chatter on oblivious to the cameras, internet of things devices, and drones watching their every move.

The data from their discussion flowed to the main data ingest where a quantum computer rapidly processed each participant's comments, facial features, voice patterns, and body posture. The

Principal City AI made a note: citizen437891 exhibited deceptive behavior during a discussion with citizen873924 concerning a promotion—additional surveillance is warranted.

Chapter 26. Beware the Jabberwocky: The AI Monsters Are Coming

Natasha E. Bajema
National Defense University
bajeman@ndu.edu

In my recent reflections about the exponential growth in artificial intelligence and the potential implications for humanity and the global order, a pulse fired across the synapses in my brain. Seemingly out of nowhere, I began humming a familiar tune set to Lewis Carroll's famous poem entitled *Jabberwocky.*

"Beware the Jabberwock, my son! The jaws that bite, the claws that catch! Beware the Jubjub bird, and shun the frumious Bandersnatch!"

The poem depicts a terrifying beast called the Jabberwocky and a valiant hero who takes up arms in a violent confrontation. For some strange reason, my brain substituted "AI Monster" for the terrifying Jabberwock in Carroll's poem, leading musical notes from the distant past to enter my mind. I hadn't sung or even thought about the tune since my days of singing in the St. Cecilia Youth Chorale—more than twenty-five years ago. *What mysterious links was my brain connecting here?*

Naturally, I turned to Google's powerful search algorithm for answers. I'd forgotten that Lewis Carroll wrote the nonsensical poem for *Through the Looking Glass* (1871), the sequel to his more famous novel *Alice in Wonderland* (1865). Both works were written by the mathematician under a pen name. Though considered children's books today, they were intended as scathing critiques of prevailing trends in the field of mathematics and designed to parody several of his colleagues. Immediately, I connected the dots between our perception of pending doom at the hands of AI to the dark atmosphere and intense feelings of disorientation and angst in Carroll's stories. The tale of Alice travelling down the rabbit hole to meet a sequence of demented, off-kilter, and nonsensical characters gives me a jarring sense of discomfort to this day—not much unlike my fears regarding the rise of AI.

After a moment of awe for the mystifying inner workings of the human brain, I felt another curious tug at my consciousness after reading Carroll's poem. I'd set out a pile of my favorite sci-fi films from which to draw inspiration for my next fiction project—a dystopian science fiction trilogy rooted in current digital trends. The movies were stacked in no particular order, and I decided to watch my all-time favorite, *The Matrix.*

My pulse quickened as Neo receives a message on his computer screen: "Follow the white rabbit." *That's not Alice's white rabbit, is it?* Shortly afterwards, Neo spots a white rabbit tattoo on the girl's shoulder and follows a wild rabble to an all-night rave. Then a thought crossed my mind. *Is my brain showing me the link to the old musical tune?* I reached the part where Morpheus meets Neo, and my buried memories started to surface. I froze in my chair, my heart now pounding against my chest. *The blue and red pills are analogous to "Drink Me" and "Eat Me" in Alice in Wonderland, aren't they?* A few moments later, Morpheus says to Neo: "I imagine you're feeling a bit like Alice… tumbling down the rabbit hole." *Yes, Morpheus. Yes, I do.*

By now, my mind was blown. In making my film choice, I didn't realize my brain was doing its thing again. It was drawing connections from the depths of my complex neural network and bringing them to the surface.

For the umpteenth time, this experience reinforced what I've always known to be true—that science fiction plays an important role in shaping our understanding of the implications of science and technology and helping us to cope with things to come. My brain was leading me down a rabbit hole to confront the horrifying AI monsters depicted in science fiction as one day disrupting the global order and destroying humanity—the automation monster, the supermachine monster, and the data monster.

The Automation Monster

In the first and oldest nightmare AI scenario, the future is automated. Humans have been completely sidelined by robots—stronger, tireless, and inexpensive versions of themselves—as depicted in Kurt Vonnegut's *Player Piano* (1952). Fears about robotics have pervaded pop culture since Karol Capek, a Czech playwright, coined the term "robot" in 1920 in his play entitled *Rossum's Universal Robots (RUR).* The satire depicts robots performing the activities that humans typically find undesirable—the dirty, dull, and dangerous. As demonstrated in the end of Capek's play when the robots rebel against humans and eliminate nearly all of humanity, automation, though more convenient, cheaper and faster, presents new dangers.

In a series of short stories entitled *I, Robot* (1950), Issac Asimov effectively demonstrates how humans may lose control of robots, even if they are programmed not to harm humans according to his three famous laws. He warned that as automated systems become more complex, humans will not be able to anticipate all the unintended consequences of rule-based systems.

Potential scenarios about the loss of control were also featured in several classic films during the Cold War period. In Stanley Kubrik's *Dr. Strangelove* (1964), a doomsday device thwarts efforts by the US and the USSR to prevent nuclear war, leading to the destruction of both countries and a devastating nuclear winter. The removal of human meddling through automation was intended to increase the credibility of mutual assured destruction. The strategy goes awry because the Soviets fail to communicate its new capability to the US in a timely manner. Once the doomsday device is activated, it cannot be deactivated since automation is the essential property of the system.

Another Kubrik film, *2001: A Space Odyssey* (1968), features the HAL 9000 supercomputer (aka "Hal"), which was designed to automate most of the Discovery spaceship's operations. Although the computer is considered foolproof, the human crew discovers Hal made an error in detecting a broken part. The crew decides to disconnect the supercomputer, but not before Hal discovers their plan and manages to kill off most of the crew.

In *WarGames* (1983), doubts surface about military officers' willingness to launch a missile strike. Consequently, the government decides to turn over the control of the nuclear attack plan to the War Operation Planned Response (WOPR), a NORAD supercomputer, capable of running simulations and predicting outcomes of nuclear war. A young hacker inadvertently accesses the computer and launches a nuclear attack simulation, which begins to have real-world effects. To stop the computer from carrying out its automated nuclear attack, the system's original programmer and the young hacker must first teach the computer the concept of mutual assured destruction in which there is no winner.

The predicted outcomes of the automation monster range from terrible to apocalyptic. In the most likely scenario, robots will destroy our jobs, leaving humans out of work and without any hope for economic mobility. The impact on the global order would be devastating, potentially leading to mass migrations, societal unrest, and violent conflict between nation-states. These fears appear to be substantiated by a wide range of studies from companies, think tanks, and research institutions, which predict as many as 800 million jobs will be lost to automation by 2030 (Winick, 2018).

Another frightening scenario involves autonomous weapons going awry. In an era of autonomous weapons, warfare will increasingly leverage machine speed and pose a challenge to the need for human control. Whereas humans require time to process complex information and reach decisions, machines can achieve the same in nanoseconds. Despite advantages in analyzing complex datasets, however, the decisions reached by machines may not be optimal due to the nature of information—its inaccuracy, incompleteness, bias, missing context, etc. To prevent some nightmare scenarios, humans must remain in the loop. To prevent others, humans might need to step aside to let the machines lead the action… because speed can kill (Scharre, 2018).

In another terrifying scenario portrayed in *GhostFleet* (2016) by P. W. Singer and August Cole, overdependence on automation technologies creates critical vulnerabilities that can be exploited by adversaries. Recent news headlines regarding the vulnerabilities of US weapons systems and supply chains suggest that this scenario is a near-term possibility (GAO, 2018). US superiority in automation technologies offers our adversaries powerful incentives for conducting first-move asymmetric attacks that exploit theses vulnerabilities (Schneider, 2018).

Taken to the worst extreme, automation combined with machine intelligence could potentially lead to the destruction of the world by autonomous machines and networks of machines—the supermachine monster.

The Supermachine Monster

In recent years, the supermachine monster has dominated the tech headlines as the scariest potential AI scenario. A number of public figures including Elon Musk and the late Stephen Hawking have issued dramatic warnings about the prospect of reaching singularity in 2045—the point at which Futurist Ray Kurzweil suggests machine intelligence will match and inevitably exceed human intelligence.

Inspired by fears about supermachines, *The Terminator* (1984) tackles the theme of a coming war between humans and machine, a result of an automation scenario gone awry. A defense contractor builds the Global Digital Defense Network, an AI computer system later referred to as Skynet. The system is designed to control all US computerized military hardware and software systems including the B-2 bomber fleet and the nuclear weapons arsenal. Built with a high level of machine intelligence, Skynet becomes self-aware, determines humanity to be a threat to its existence, and sets out to annihilate the human race using nuclear weapons and a series of lethal autonomous and intelligent machines called terminators.

The Matrix (1999) picks up where *The Terminator* leaves off, depicting the aftermath of war between humans and machines, the initial triumph of the machines, and the enslavement of humans. The majority of humans are prisoners in a virtual reality system called the matrix and being farmed in pods as a source of energy for the machines. A small number of freed humans live in a deep underground colony called Zion and carry on a violent struggle against the sentinels. By the end of

the trilogy, Neo convinces the machines to reach peace with Zion and to fight against a common enemy—a malignant computer program called Mr. Smith.

There are few scenarios more frightening than apocalyptic wars between humans and machines. Indeed, we are so afraid of the automation and supermachine monsters these days that we're failing to see the scariest monster of them all—lurking beneath the surface of our consciousness—the data monster.

Data Monster

My brain made connections that were deep beneath my consciousness, linking Carroll's poem to *Alice in Wonderland's* rabbit hole and *The Matrix* to the AI monster that keeps me up at night—the data monster. Lately, I've been wondering whether we are already controlled by the machines and just aren't fully aware of it yet.

In Plato's *Republic*, Socrates describes a group of people chained to the wall of a cave who think the shadows on the wall are real because it's all they've ever seen; they are prisoners of their own reality. *How is it that we are not seeing the dangers of the data monster?* Even while the pernicious beast stalks us everywhere, lurking in the corners, ready to enslave us at any moment. *Or are we already its prisoners and unable to see the truth?*

For me, the real Jabberwocky is the three-headed data monster combo of the Internet, digitization, and algorithms. Somewhere deep down, we realize the data monster is stealthily assaulting our sense of truth, our right to privacy, and our freedoms. Most of us sense this is happening, but we suppress such concerns in favor of obsessing over the other more sexy AI monsters. But if we don't take the red pill now and wake up from our digital slumber, we may end up prisoners in the matrix—controlled by our machines.

Much has changed since *The Matrix* was first released in 1999—particularly our inextricable relationship with smartphones, the rapidly accumulating crumbs of our digital trail, and our growing interconnectedness through the Internet of Things. The image of sleeping humans imprisoned in pods, connected to the machine world by thick, black cables attached to their spines, and ruthlessly exploited as an energy source hits home in a whole new way in 2018. At its essence, the matrix is a digital world designed by the machines to fool humans into thinking it is real. *Are we in a matrix?*

Our common sense of truth has been eroding for the past few years at the hands of endless political spin, outright lies, allegations of fake news. The propaganda has gotten bad enough to invoke images from George Orwell's dystopian novel *1984* in which Party Member Winston Smith works diligently at Oceania's Ministry of Truth to rewrite history based on the ever-changing truth propagated by the Party. The bleak world of *newspeak* and *doublethink* created by Orwell in 1949 resonates so well today, the novel became an Amazon bestseller in 2017. Although Winston rebelled against the Party, he was in the end compelled to reject the evidence of his eyes and ears. "It was their final, most essential command."

French philosopher Jean Baudrillard, a muse of the Wachowski brothers, argued that in a postmodern world dominated by digital technology and mass media, people no longer interact with physical things, but rather the imitations of things. And so, technology has altered our perceptions of reality and made it more difficult to identify truth. Our growing interdependence with machines causes

intense confusion about what parts of our human experience on this earth are more real—those in the physical world or that in the digital one. *How do we know what we know is true or real?*

At the beginning of the movie, Neo asks "do you ever have the feeling where you're not sure if you're awake or you're dreaming?" Deep down, he senses the pernicious illusion of the matrix. When Morpheus meets Neo for the first time, he gives Neo a choice: take the blue pill and wake up as if nothing ever happened, or take the red pill and learn the truth. Later in the story, Neo's power as "The One" derives from his ability to see the matrix for what it really is. At times in the movie, it's unclear which form of existence is preferable—the matrix or the real world. Indeed, the villain of the movie, a freed human by the name of Cypher, betrays Morpheus for a chance to get back into the matrix and deny the truth of his existence.

But truth is not the only vital element under siege by the data monster. Slowly, but surely, the data monster has been jealously chipping away at our right to privacy. Here again, we are partners in our own demise. With every digital action, each one of us produces new data points—e.g., every email, text, and phone call, Internet download , online purchase, GPS input, social media post and contact, daily numbers of steps, and camera selfie. The list could go on and on. With all the data we produce, we are essentially handing over the tools of surveillance and control. *But to whom?*

In *1984,* George Orwell creates a world in which the citizens of Oceania are monitored via telescreens, hidden microphones, and networks of informants. The notion that Big Brother is always watching keeps most citizens in line. For those who rebel, extraordinary measures are taken to bring them back in line by the Thought Police. Such a social control experiment, while leveraging technology, is happening in the real world as we speak.

Leveraging the data trail of its population, the Chinese government has begun testing a social credit system which assigns a trustworthiness score to citizens based on their behavior—including their social network, debt, tax returns, bill payment, tickets, legal issues, travel and purchase habits, and even disturbances caused by pets. Blacklisted Chinese citizens with low scores face limitations in their freedoms, ability to travel, employment opportunities, and much more. As such a credit system takes effect, citizens will conform their behavior to avoid negative outcomes.

Perhaps, many of us can breathe a sigh of relief—at least we don't live in China. Thus far, most democracies have resisted the alluring pull of monitoring technologies in the name of protecting privacy. *Or have we?* If our data trail is not being funneled to our government, then to whom are we giving the power? And do we trust them to do the right thing?

In *Future Crimes* (2015), Marc Goodman describes in compelling detail how we fail to see the reality of our digital actions and gambling away our privacy: we are the product of the tech giants. Every day, we have grown accustomed to exchanging small pieces of our privacy for free services by clicking the box "agree to terms and conditions." Most of us skip the pages of legalese to download the app and get access to the convenient and "free" services. When we use Gmail from Google, update our status on Facebook to share news with our friends, purchase stuff from Amazon to avoid going to the store, we agree to the use and tracking of our data.

All of this data is out there somewhere, waiting to be mined and exploited. Until something bad happens like a stolen credit card number or identity theft, most people don't think about the consequences. But if we're being honest with ourselves, the data monster probably knows us better than we know ourselves. And that means, there are private-sector companies that know us, too. Tech

giants such as Facebook and Twitter already assign its users a reputation score based on activity and social networks. *Big Brother is watching you.*

But the power of data goes far beyond monitoring and surveillance to allow for predictive control. In *Minority Report* (1956), a short story by Philip K. Dick, a set of precogs are able to see and predict all crime before it occurs, eliminating crime in a future society. Instead, people are arrested and tried for precrimes based solely on the logical progression of their thoughts. We may shudder at the notion of such a world, but AI and big data are already being used to forecast our behavior on a daily basis—and shape our future behavior. For example, Amazon tracks every purchase you make on its website and uses its algorithm to predict what item you are most likely to buy next. This sees harmless enough. *For now.*

But what happens when machine learning tools begin making more important decisions than our retail purchases? The data we produce today will shape the future, possibly even control it. What is the nature of that data? How reliable is it? Has someone accounted for false information, missing information, partial truths, and bias?

Last year, the British police began using "predictive crime mapping" to determine where and when crime will take place. Some allege the system has learned racism and bias, leading to increased policing in areas with high crime rates and to self-fulfilling prophecies.

Machine learning tools analyze data, but they cannot determine what is true and what is false unless they've been trained to do so. If it's difficult for humans to identify truth these days, how can we expect machine learning tools to do itbe better? In a recent example, Amazon attempted to use a machine learning algorithm to simplify its hiring process. The training data included resumes submitted to Amazon over ten years, the majority of which came from male candidates. By using this dataset, the algorithm learned to prefer male applicants over females and downgraded the latter in making its recommendations.

Although Armageddon-like scenarios do not loom large for the data monster, its impact could be far more pernicious to us in the near-term.

Overcoming the Monsters

My brain was not merely connecting the dots across disparate images stored in my memory bank. It was also providing me with a primal emotional response to my fears about AI. Carroll's poem offers a good example of an "overcoming the monster" plot where characters find themselves "under the shadow of a monstrous threat" (Kakutani, 2005). At the climax, the hero has a final confrontation with the monster, deftly wielding his sword and slaying the Jabberwocky.

"One, two! One, two! And through and through, The vorpal blade went snicker-snack! He left it dead, and with its head, He went galumphing back."

In reality, we are still quite far away from the worst-case AI scenarios, especially in light of human adaptability, ingenuity, and resilience. To achieve sentience or mindedness of a human, a machine would have to excel in and leverage all forms of human intelligence simultaneously (Gardner, 1983). It's time to put on our battle armor, wield our swords, and address the risks of AI head-on with creative determination—let's do what humans do best, to imagine the future we want for ourselves and put the pieces in place to achieve it. When we put aside our terror, we'll find the beast is not quite

as powerful as we imagined. If we can overcome the data monster, then we can certainly triumph over the worst of the automation and supermachine monsters. *Let's take the red pill and get started today.*

The views expressed in this piece belong to the author and do not reflect the official policy or position of the National Defense University, the Department of Defense or the U.S. Government.

References

Dearden, L. (2017, October 7). How technology is allowing police to predict where and when crime will happen. *Independent.* Retrieved from https://www.independent.co.uk/news/uk/home-news/police-big-data-technology-predict-crime-hotspot-mapping-rusi-report-research-minority-report-a7963706.html

Gardner, H. (1983). Multiple intelligences: Challenging the standard view of intelligence. Retrieved from http://www.pz.harvard.edu/projects/multiple-intelligences

Gonzalez, G. (2018, October 10). How Amazon accidentally invented a sexist hiring algorithm. Retrieved from https://www.inc.com/guadalupe-gonzalez/amazon-artificial-intelligence-ai-hiring-tool-hr.html

Goodman, M. (2015). Future crimes: Inside the digital underground and the battle for our connected world. New York: Anchor Books.

Kakutani, M. (2005, April 15). The plot thins, or are no stories new? *The New York Times.* Retrieved from https://www.nytimes.com/2005/04/15/books/the-plot-thins-or-are-no-stories-new.html

Scharre, P. (2018). Warfare enters the robotics era. Retrieved from https://davemarash.com/2018/10/25/paul-scharre-center-for-a-new-american-security-warfare-enters-the-robotics-era/

Schneider, J. (2018). Digitally-enabled warfare: the capability-vulnerability paradox. Retrieved from https://www.cnas.org/publications/reports/digitally-enabled-warfare-the-capability-vulnerability-paradox

Singer P. W. & Cole, A. (2016). *Ghost fleet: A novel of the next world war.* New York: Mariner Books.

Winick, E. (2018, January 25). Every study we could find on what automation will do to jobs, in one chart. *MIT Technology Review.* Retrieved from https://www.technologyreview.com/s/610005/every-study-we-could-find-on-what-automation-will-do-to-jobs-in-one-chart/

Chapter 27. Is China's AI Future the Snake in the Wine? Or Will Our Future Be FAANGed?

Regina Joseph
Pytho LLC
rj@pytho.io

Abstract

China's urgent and massive plan to dominate in the artificial intelligence industry is characterized in similar and no less world-changing terms as Silicon Valley has presented its brand of disruptive innovation. While both those portrayals emphasize the limitless opportunity, brilliance and social good typified by each region's efforts, a different, more malign potential lurks under the surface. In the US, a slow realization appears to be dawning on younger generations who recognize the bondage posed by addictive technologies—a fate prophesized by Aldous Huxley in his notion of the Ultimate Revolution. But in China, centralized control and soft coercion stymie public opposition to techno-nationalism to the extent that unchecked zeal for AI expansion will have adverse consequences for both Chinese and external populations.

Key points:

- China's ambitious Three Year Action Plan for AI industry poses non-trivial national security implications for the US and other countries
- Techno-nationalism in China and techno-capitalism in the US share several common outcomes and features, but diverge in centralized imposition of social control in the former and decentralized corporate-led and consumer-desired social control in the latter
- Risks to US security are as inherent in US AI expansion as in Chinese AI expansion

Unnecessary suspicion is a vice from which a famous Chinese parable of the tumultuous Jin dynasty (265-420 AD) urges caution. In it, a county magistrate invites a close friend to his home for a drink. As the guest raises his glass, light reflecting off a decoration creates the appearance of a tiny snake inside his wine. Afraid of offending a magistrate, the guest drinks and mentions nothing before returning home, where he feels unwell and remains very sick for several days. Hearing of his friend's illness, the magistrate calls for him and is told of his friend's fear of being poisoned by the snake in the wine. The magistrate shows his friend the illusion caused by the light's reflection, upon which the friend becomes instantly cured (Wong et al, 2007). On the one hand, the story is a sensible warning against over-reaction; on the other, it could be used as an appealing narrative for any entity wishing to quell questioning minds – in essence, a rustic allegory admonishing generations to move on, nothing to see here.

The morality tale could be applied to how perception around technology and AI is shaped in both China and Western democracies like the US. But near-term national security implications rest on how well audiences heed the parable's conclusion.

Techno-nationalism in China reached an apogee with last year's release of the Three Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry (2018-2020) (Triolo, Kania and Webster, 2018). On its face, the initiative is a supremely ambitious and aggressive economic plan to challenge US dominance and push China onto the world stage as an undisputed "manufacturing and cyber superpower" (Triolo, Kania and Webster, 2018). A great number of articles

covering the plan employed the same tone that, until recently, much American tech journalism used to document Silicon Valley's exploits: marvel at the scope, depth, speed and sheer audacity of China's intent to prevail in the plan's four major task areas. These task areas are: intelligent product development; development of the hardware and software foundations required to dominate in AI; intelligent manufacturing; and the construction of an AI industry support system and infrastructure (Triolo, Kania and Webster, 2018). Indeed, MIT Technology Review tutted at the proverbial snake in the wine by suggesting, "The West shouldn't fear China's artificial intelligence revolution. It should copy it' (Knight, 2017).

From a national security perspective, investment in technological innovation to avoid surprise should be conducted with the same urgency and at least on the same economic scale as China. But this simple argument for matching size and scope in the quest for AI primacy misses the nuances and danger of obscured intention, the possibility that the snake in the wine is no illusion. Now that the curtain of Silicon Valley's Oz-like myth of world-improving disruption is being pulled back to reveal profit-chasing at the expense of privacy and truth, the reverence and fan-like credulity which once greeted the US' tech's goliaths are being replaced with a bit more skepticism (Frenkel et al, 2018). Americans are slowly acclimating to the reality that oligarchic power in the tech sector, coupled with the inevitable human foibles of greed and self-preservation, can lead to terrible consequences for democracy. In fact, early reports suggest that young software engineers in the US are being more selective about their future careers (Bowles, 2018) now that they are beginning to understand the perils to the social fabric posed by data surveillance, bots and opaque AI architectures.

The FAANGs—the acronym used to describe Facebook, Apple, Amazon, Netflix and Google, the most capitalized and popular technology stocks (Tully, 2017)—are slowly sinking into the American psyche; heir suffocating embrace is being revealed as no trick of the light. However, regulatory action against the FAANGs in the US before 2020, if it occurs at all, is unlikely to be sweeping enough to address or stanch all of the trouble posed by such threats as platform misuse and abuse, data manipulation, and algorithmically-driven inequality. The emerging corporate technocracy propels not an Orwellian dystopia—in which imposition of control is ultimately the path preferred by centralized governance and authoritarians—but rather what Brave New World author Aldous Huxley referred to as the "the Ultimate Revolution:" whereby control of a populace is achieved through that society's own willingness and desire to take up the instruments of servitude (Huxley, 1962) (Joseph, 2017). Users of Facebook, Twitter, Android and iOS phones have all experienced either breaches or warnings of how their data can be misused and manipulated by foreign adversaries or even the FAANGs themselves. And yet, few have broken the addictive desire to remain glued to a screen. Digital platforms in the US have bcome organs of influence (Joseph, 2017). They have thrust the US into an internal struggle over liberal democracy—one that will only become more complicated as the increase in intelligent systems and the Internet of Things (IoT) erode further still the human qualities of nuance and empathy required of a liberal democracy.

Contrasting the US trajectory against China's exposes many similarities with regards to digital control. Smartphone applications are just as addictive in China as they are in the US. Tencent's WeChat and QQ messaging platforms, Baidu, and Alibaba Group's e-commerce sites are the world's closest rivals to the FAANGs in active users, but in some areas they exceed the US companies. As of May 2018, for example, Alibaba's operating margins were larger than Amazon's by 29% (Mourdoukoutas, 2018). Capital investment across all the Chinese national champions have increased due to the AI push, even though consumer demand in both China and the US has slowed (Leach, 2018). Lowered demand has not been an impediment to commercial deployment of "super apps" like WeChat and Alibaba's Sesame Credit loyalty system (also known as Zhima Credit). While voluntary and a product of enterprise, the Zhima Credit system has integrated state delinquency

blacklists into its database and is considered a test bed for the government's social credit system that will be mandatory for all citizens by 2020 (Hvistendahl, 2017). Such loyalty systems capture user behavior and physical data and reward or deny users benefits. Indeed, a feature advantage Chinese AI leaders point to when discussing technological rivalry with the US is China's larger pool of data from users upon which they can train their AI systems. The primary difference today between China and the US regarding digital control lies in the former's centralized control via the state (Orwellian), and the latter's decentralized control via corporations that create the engagement systems to which people willingly succumb (Huxleyian).

China's techno-nationalist perspective assigns virtue to censorship and the harvest of user data through national champion proxies. These means not only help the state to achieve such objectives as superior visual recognition and geolocation for general commercial purposes (Knight, 2017), but also serve state stability through surveillance and military use. By comparison, the US perspective is theoretically opposed to censorship, state coercion and control, and data harvesting, but privacy concerns have been an insufficient barrier to impede the business models of FAANGs and other American organizations—many of which are still routinely bestowed with trust for economic successes in spite of their questionable privacy practices. The US's misalignment between its entrepreneurial DNA and the sacrifices techno-capitalism demands from society is partly responsible for the churn that now roils American democratic governance. Over-emphasizing the benefits of AI without careful consideration to network effects and adverse consequences would land the US in the same trap—and probably worse due to path dependencies—it finds itself in today due to lack of foresight over the Internet's earliest winners. Tech success—whether from homogenous enclaves like Palo Alto, Cambridge or Seattle—like financial success, does not necessarily equate to strategic geopolitical vision. But the US may yet repeat the same mistakes by assuming the snake in the wine of future governance is an illusion: many high status tech figures in the US advocate AI-driven governance systems that eschew a fundamentally human enterprise for untested initiatives built by unrepresentative samples of American society (Johnson, 2018). The end result of several of these proposed systems allows high status tech entities to continue to pursue wealth-extraction more easily (Rushkoff, 2018). AI-powered governance systems will reflect and serve the elite monocultures that build them.

Regulation is not foolproof. The Sherman Antitrust act broke Standard Oil and AT&T into smaller pieces, but the successor companies that emerged from those break-ups still exert outsize control in their sectors today, thanks to the secret sauce of lobbying (Sottek, 2016), Tech companies currently are some of the biggest spenders on K Street (Ackley, 2018). Even if the American tech juggernauts of today become ringfenced or fall tomorrow, the foundations they build as the earliest proponents and adopters of AI won't be easily dislodged. The accretion of systems built on the FAANGs' AI evangelism and advances may yet cohere into new public/private agglomerations that may bear some resemblance to China's state/enterprise combos. The gargantuan datasets now owned by FAANGs will not go unexploited. As the effects of climate change and social atomization inexorably mount, political leaders who lack technical understanding may default to technocratic solutions to assume ostensible control over chaos: that is likely to involve surveillance in service of state stability. The crux is, should China succeed in its AI manufacturing and infrastructure ambitions, infrastructural eyes on citizens won't necessarily belong only to the US.

Aside from its efforts to coerce conformity within its own populace, China's future planning in AI plays an important role in directing future governance conditions in the US and other countries. Currently the US experiences a large trade deficit with China in advanced technology products, especially in the field of information and communication technology (ICT) products. China's expansive sector growth includes global suppliers like Huawei Technologies with its consumer,

enterprise and carrier server hardware; ZTE and its mobile phone consumer products (against which a US ban was lifted in July 2018 [Kastrenakes, 2018]); and mobile battery manufacturer BYD. In the first three quarters of 2018 the trade deficit between China and the US stood at $98.7 billion, a year-on-year increase of 7.3%, the majority of which is due to a 7.7% year-on-year increase in ICT product imports from China, which grew to $114.9 billion through September 2018 (US-China Economic and Security and Review Commission, 2018). China's rapid IoT and 5G manufacturing expansion under its Three Year Action Plan will impose standardization structures on telecommunications that will have far-reaching consequences for the US. If the pace of Chinese original equipment manufacturing in ICT continues to outstrip the US, then the US will increasingly cede infrastructure standards in telecommunications—which can hamper the extent to which the US can dictate ICT norms in an AI-driven world. The national security implications are even more precarious: data collection on Chinese 5G and IoT infrastructure equipment can extend beyond data collection from Chinese consumers to Americans—a threat revealed in a recent report alleging that a Chinese military unit inserted chips for the purposes of espionage and data collection via backdoor access into Chinese server equipment used by Apple and Amazon (Gibbs, 2018).

The risks of AI expansion are more real than some of its sunny public relations would suggest. In considering the course of the next decade, suspicion should not be construed as over-reaction to the illusion of a poisoned chalice, but rather the necessary first step before committing to a competitive expansion. Citizens in China may be reassured that there is nothing to see here, but American national security (as well as that of any other nation) dictates a far greater exigence on all the myriad ways AI expansion—both from within the US and from China—can exert harm across generations.

Without foresight, the serpent may be illusory, but its bite will be painfully real.

References

Ackley, K. (2018, October). Google Still K Street's Top Spender. Roll Call. Retrieved from https://www.rollcall.com/news/politics/tech-trade-appropriations-keep-k-street-in-business

Bowles, N. (2018, November 14). "I Don't Really Want to Work for Facebook." So Say Some Computer Science Students. The New York Times. Retrieved from https://www.nytimes.com/2018/11/15/technology/jobs-facebook-computer-science-students.html?action=click&module=Top%20Stories&pgtype=Homepage

Frenkel, S., Confessore, N., Kang, C., Rosenberg M., Nicas, J. (2018, November 14). Delay, Deny and Deflect: How Facebook's Leaders Fought Through Crisis. The New York Times. Retrieved from https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html?action=click&module=Top%20Stories&pgtype=Homepage

Gibbs, S. (2018, October 4). China planted chips in Apple and Amazon servers, report claims. The Guardian. Retrieved from https://www.theguardian.com/technology/2018/oct/04/china-planted-chips-on-apple-and-amazon-servers-report-claims

Hvistendahl, M. (2017, December).Inside China's Vast New Experiment in Social Ranking. Wired. Retrieved from https://www.wired.com/story/age-of-social-credit/

Johnson, S. (2018, July 24). The Political Education of Silicon Valley. Wired. Retrieved from https://www.wired.com/story/political-education-silicon-valley/

Joseph, R. (2017, August). A Peek at the Future: A Stealth Revolution by Influence's New Masters. White Paper on Influence in an Age of Rising Connectedness: A Strategic Multilayer Assessment (SMA) Periodic Publication.

Kastrenakes, J. (2018, July 13). US lifts trade ban on ZTE in controversial deal with Chinese phone maker. The Verge. Retrieved from https://www.theverge.com/2018/7/13/17565450/zte-trade-ban-lifted-us-commerce-department-trump

Knight, W. (2017, October 17). China's AI Awakening, MIT Technology Review. Retrieved from https://www.technologyreview.com/s/609038/chinas-ai-awakening/

Leach, K. (2018, October 15). FAANG stock losses continue to mount. Bloomberg. Retrieved from https://www.bloomberg.com/news/articles/2018-10-15/faang-shares-diverge-from-tech-as-losses-continue-to-mount

Lecturesbeyondbeyond. (2013, November 1). Aldous Huxley - The Ultimate Revolution (Berkeley Speech 1962) [Audio File]. Retrieved from http://m.youtube.com/watch?v=2WaUkZXKA30

Mourdokoutas, P. (2018, May 6). Why Alibaba is More Profitable than Amazon. Forbes. Retrieved from https://www.forbes.com/sites/panosmourdoukoutas/2018/05/06/why-alibaba-is-more-profitable-than-amazon/#29b35d281678

Rushkoff, D. (2018, November 14). Team Human Newsletter.

Sottek, T.C. (2016, December). This insane example from the FCC shows why AT&T and Verizon's zero-rating schemes are a racket. The Verge. Retrieved from https://www.theverge.com/2016/12/2/13820498/att-verizon-fcc-zero-rating-gonna-have-a-bad-time

Triolo, P., Kania, E., Webster, G. (2018, January). Translation: Chinese government outlines AI ambitions through 2020. Retrieved from https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-government-outlines-ai-ambitions-through-2020/

Tully, S. (2017, June 15). FAANG Tech Stocks Are No Bargain. Here's Why. Fortune. Retrieved from http://fortune.com/2017/06/15/faang-tech-stocks/

US-China Economic and Security Review Commission: Economics and Trade Bulletin (2018, November 2). Retrieved from https://www.uscc.gov/Research/november-2018-trade-bulletin

Wong, E., Yufeng, H., Leung, L. (2007). The Power of Ren: China's Coaching Phenomenon, John Wiley & Sons (Asia) Pte Ltd.: Singapore

Biographies

Ms. Shazeda Ahmed

Shazeda Ahmed is a PhD student at the University of California- Berkeley's School of Information. She has previously worked as a researcher for the U.S. Naval War College, Citizen Lab, the Mercator Institute for China Studies, and the Ranking Digital Rights corporate transparency review by New America. In the 2018-19 academic year she will be a Peking University-based Fulbright fellow researching public-private partnerships in the development of China's social credit system.

Dr. Natasha E. Bajema

Dr. Natasha E. Bajema joined the Center for the Study of Weapons of Mass Destruction at National Defense University in October 2008. Dr. Bajema currently is a Senior Research Fellow, the principal investigator for Emergence and Convergence, and Course Director for an elective entitled Through the Film-maker's Lens: Contemporary Issues in Combating Weapons of Mass Destruction and conducts research on global threat reduction programs. From 2010 to 2013, Dr. Bajema held a long-term detail assignment serving in various capacities in the Office of the Secretary of Defense, Acquisitions, Technology and Logistics, Nuclear, Chemical and Biological Defense Programs and in Defense Nuclear Nonproliferation at Department of Energy's National Nuclear Security Administration.

Prior to joining NDU, Dr. Bajema was a Research Associate at the Center on International Cooperation at New York University, where she supported research staff of the High-Level Panel on Threats, Challenges and Change established by the UN Secretary-General. She has also served as a Junior Political Officer in the Weapons of Mass Destruction Branch of the Department for Disarmament Affairs at the United Nations. Her publications include two co-edited volumes entitled Terrorism and Counterterrorism and Weapons of Mass Destruction and Terrorism, both of which were published by McGraw Hill. She holds an M.A. in international policy from the Monterey Institute of International Studies and a PhD in international relations from the Fletcher School of Law and Diplomacy. Dr. Bajema is also a fiction author and published two mystery/sci-fi novels in 2018.

Mr. Sam Bendett

Samuel Bendett is a Research Analyst with the Center for Naval Analyses' Adversary Analysis Group, where he is a member of the Russia Studies Program. His work involves research on the Russian defense and technology developments, such as Russian naval and land capabilities, unmanned military systems and artificial intelligence, and Russian decision-making calculus during military crises. He is also a member of CNA's Center for Autonomy and Artificial Intelligence.

Prior to joining CNA, Mr. Bendett worked at the National Defense University on emerging and disruptive technologies for government response in crisis situation, where he conducted research on behalf of the Office of the Secretary of Defense for Policy (OSD-P) and Acquisition, Technology and Logistics (OSD-AT&L). His previous experience includes working for US Congress, private sector and non-profit organizations on foreign policy, international conflict resolution, defense and security issues.

Samuel is also a Fellow in Russia Studies at the American Foreign Policy Council, where he conducts research on the Russian unmanned military systems and artificial intelligence.

Mr. Bendett's analyses, views and commentary on Russian military robotics, unmanned systems and artificial intelligence capabilities appear regularly in the C4ISRnet, DefenseOne, Breaking Defense, The National Interest, War Is Boring, and The Strategy Bridge. He was also a foreign policy and international affairs contributor to the RealClearWorld.com blog, writing on Russian military technology. Samuel Bendett received his M.A. in Law and Diplomacy from the Fletcher School, Tufts University and B.A. in Politics and English from Brandeis University. He has native fluency in Russian.

Mr. Benjamin Chang

Benjamin Angel Chang is a PhD candidate in international relations and security studies at MIT, where he studies the diffusion of military technology and the international relations of East Asia. Prior to MIT, Ben worked as a Senior Analyst at the Long Term Strategy Group. He holds an AB summa cum laude from Princeton University, where he majored in the Woodrow Wilson School of Public and International Affairs. He is a recipient of the National Science Foundation Graduate Research Fellowship.

Dr. Rogier Creemers

Rogier Creemers is an Assistant Professor in the Law and Governance of China at Leiden University. He holds Master degrees in China Studies and International Relations, and a Doctorate in Law. His research investigates China's domestic technology policies, as well as China's participation in global cyber affairs. His work has been published, amongst others, in The China Journal and the Journal of Contemporary China. He is the leader of two major projects funded by the Dutch Organization for Scientific Research and the Ministry of Foreign Affairs. He is also a founding member of DigiChina, a project run in cooperation with New America, as well as a frequent contributor to international news media.

Dr. Chris Demchak

With degrees in engineering, economics, and comparative complex organization theory/political science, Dr. Chris C. Demchak is the RDML Grace M. Hopper Professor & Chair of Cyber Security, US Naval War College. She is presently the Senior Cyber Scholar, Cyber and Innovation Policy Institute (CIPI), the expanded, successor research unit to the Center for Cyber Conflict Studies (C3S) for which Dr. Demchak was founding director. Her research and many publications address global cyberspace as a globally shared, complex, insecure 'substrate' penetrating throughout the critical organizations of digitized societies, creating 'cybered conflict', and resulting in a rising 'Cyber Westphalia' of sovereign competitive complex socio-technical-economic systems (STESs). Demchak takes a systemic approach in focusing on emergent structures, comparative institutional evolution, adversary/defensive use of systemic cybered tools, virtual worlds/gaming for operationalized organizational learning, ensuring national cyber power, and designing systemic resilience against normal or adversary imposed surprises that disrupt or disable largescale systems. She combines a multi-domain formation from systems analysis, to the LISP programming language in simulations, and to service as a military officer, with cross discipline methods and formal education. At graduate and undergraduate levels, she has taught international security studies and management, comparative organization theory, enterprise information systems, and cyber security architectures and trends, integrating always the overarching implications for international/ national security and democratic stability. Recent works include *Designing Resilience* (2010 co-edit); *Wars of Disruption and Resilience* (2011); and a manuscript in production tentatively entitled *Cyber Westphalia: Redrawing International Economics, Conflict, and Global Structures.*

Ms. Sarah Denton

Sarah W. Denton is a research assistant in the Science and Technology Innovation Program at the Wilson Center and a Research Fellow at the Institute for Philosophy and Public Policy at George Mason University.

Dr. Jeffrey Ding

As the China lead for the Governance of AI Program, Jeffrey Ding researches China's Artificial Intelligence strategy at Oxford University's Future of Humanity Institute. He is reading for a D.Phil. in International Relations as a Rhodes Scholar in Oxford. His writing has been published in Foreign Affairs, and his research has been cited in the Washington Post, Financial Times, and other outlets. Previously, he worked for the Hong Kong legislative council and the U.S. State Department. His ChinAI weekly newsletter, which features translations of Chinese articles and scholarship on AI-related issues, is widely read by leaders in government, media, and industry.

Sir Lawrence Freedman

Lawrence Freedman was Professor of War Studies at King's College London from 1982 to 2014, and was Vice-Principal from 2003 to 2013. He was educated at Whitley Bay Grammar School and the Universities of Manchester, York and Oxford. Before joining King's he held research appointments at Nuffield College Oxford, IISS and the Royal Institute of International Affairs. In1996, he was appointed Official Historian of the Falklands Campaign in 1997 and in June2009 he was appointed to serve as a member of the official inquiry into Britain and the 2003 Iraq War.

Lawrence Freedman has written extensively on nuclear strategy and the cold war, as well as commentating regularly on contemporary security issues. His most recent books are Strategy: A History (2013) and The Future of War: A History (2017).

Dr. Samantha Hoffman

Samantha Hoffman is a Non-Resident Fellow at the Australian Strategic Policy Institute's International Cyber Policy Centre and aVisiting Academic Fellow, Mercator Institute for China Studies (MERICS). Her research is focused on Chinese state security policy and social management. She holds a PhD in Politics and International Relations from the University of Nottingham (2017), and an MSc in Modern Chinese Studies from the University of Oxford (2011), and BA degrees in International Affairs and East Asian Languages and Cultures from the Florida State University (2010).

Ms. Regina Joseph

Regina Joseph is the co-founder of pytho, a US-based firm, and the founder of Sibylink, an international consultancy based in The Hague. Both organizations provide strategic foresight through quantitative forecasting, training programs and decision science-led solutions development. Joseph is an IARPA ACE program Superforecaster and was a member of the IARPA- funded Good Judgment Project research team.

An information systems designer with a record of award-winning product development over 25 years, Joseph is also a political scientist whose work and research assists public and private sector organizations. Her corporate work spans the world, from Sony to Hearst to Liberty Global. Her public sector work aids European government agencies and multilateral organizations. Her most recent endeavors include developing a cyber forecasting program (for the Netherlands' National Cyber Security Centre and TNO); the creation of a strategic foresight training program for The Ministry of Foreign Affairs of The Netherlands; and the invention of digital tools like NEUERtm, a quantified structural analytic technique (patent pending). She also continues her applied research into the nexus between foresight, information design and human-machine interaction as a performer in IARPA's current Hybrid Forecasting Competition (HFC) program.

In her career, Joseph has been recognized as a pioneer and thought leader in both the analog and digital worlds: she was the founder and editor-in-chief of Blender, the world's first digital magazine; she was the founder and creative director of Engine.RDA, one of the earliest digital development agencies; and she has been responsible for several technical firsts in the fields of media technologies and telecommunications. She is a published author, and her writing has appeared in a variety of outlets including Reuters, the Washington Post, the New York Times, Forbes, AdWeek, the International Relations & Security Network at ETH-Zurich, Foreign Policy and many others. Joseph is a Thomas J. Watson Fellow and holds a B.A. from Hamilton College (magna cum laude and Phi Beta Kappa) and an M. Sci from New York University.

Ms. Elsa Kania

Elsa B. Kania is an Adjunct Fellow with the Technology and National Security Program at the Center for a New American Security (CNAS). Her research focuses on Chinese military innovation in emerging technologies in support of the Artificial Intelligence and Global Security Initiative at CNAS, where she also acts as a member of the research team for the new Task Force on Artificial Intelligence and National Security. Her analytic interests include Chinese military modernization, information warfare, and defense science and technology. She has been invited to testify before the House Permanent Select Committee on Intelligence (HPSCI) and the U.S.-China Economic and Security Review Commission (USCC). Elsa is an independent analyst, consultant, and co-founder of the China Cyber and Intelligence Studies Institute. She was a 2018 Fulbright Specialist and is a Non-Resident Fellow with the Australian Strategic Policy Institute's International Cyber Policy Centre. Elsa works in support of the China Aerospace Studies Institute through its Associates Program, and she is a policy advisor for the non-profit Technology for Global Security. Elsa has been named an official "Mad Scientist" by the U.S. Army's Training and Doctrine Command.

Elsa is a PhD student in Harvard University's Department of Government, and she is also a graduate of Harvard College (summa cum laude, Phi Beta Kappa). Her thesis on the evolution of the PLA's strategic thinking on information warfare was awarded the James Gordon Bennett Prize. Her prior professional experience includes time with the Department of Defense, the Long Term Strategy Group, FireEye, Inc., and the Carnegie-Tsinghua Center for Global Policy. While at Harvard, she has worked as a research assistant at the Belfer Center for Science and International Affairs and the Weatherhead Center for International Affairs. Elsa was a Boren Scholar in Beijing, China, and she has professional proficiency in Mandarin Chinese.

Dr. Jackie Kerr

Jaclyn Kerr is a Postdoctoral Research Fellow at the Center for Global Security Research (CGSR) at Lawrence Livermore National Laboratory. She is also an Affiliate at the Center for International Security and Cooperation (CISAC) at Stanford University, and a New America Foundation Cybersecurity Policy Fellow. Her research focuses on the politics of cybersecurity, information warfare, and Internet governance, with a particular focus on the evolving approaches to Internet control within non-democratic and democratic countries and their repercussions, and on the changing international dynamics of cyber- and informational conflict. She completed her dissertation, *Authoritarian Management of (Cyber-)Society: Internet Regulation and the New Political Protest Movements*, at Georgetown University's Department of Government in 2016 and is currently working on a related book project. While writing her dissertation, Jackie held predoctoral cybersecurity policy fellowships at Stanford's CISAC, Harvard's Belfer Center for Science and International Affairs, and was a visiting scholar at Harvard's Davis Center for Russian and Eurasian Studies. She has held research fellowships in Russia, Kazakhstan, and Qatar, and has previous professional experience as a software engineer. Jackie holds a PhD and MA in Government from Georgetown University, and an MA in Russian, East European, and Eurasian Studies and a BAS in Mathematics and Slavic Languages and Literatures from Stanford University.

Dr. Lydia Kostopoulos

Dr. Lydia Kostopoulos' (@LKCYBER) work lies in the intersection of people, strategy, technology, education, and national security. She addressed the United Nations member states on the military effects panel at the Convention of Certain Weapons Group of Governmental Experts (GGE) meeting on Lethal Autonomous Weapons Systems (LAWS). Formerly the Director for Strategic Engagement at the National Defense University, a Principal Consultant for PA and higher education professor teaching national security at several universities, her professional experience spans three continents, several countries and multi-cultural environments. She speaks and writes on disruptive technology convergence, innovation, tech ethics, and national security. She lectures at the National Defense University, Joint Special Operations University, is a member of the IEEE-USA AI Policy Committee, participates in NATO's Science for Peace and Security Program, and during the Obama administration has received the U.S. Presidential Volunteer Service Award for her pro bono work in cybersecurity. In efforts to raise awareness on AI and ethics she is working on a reflectional art series [#ArtAboutAI], and a game about emerging technology and ethics called Sapien2.0 which is expected to be out early 2019. www.lkcyber.com

Dr. James Andrew Lewis

James A. Lewis is a Senior Vice President and Program Director at CSIS where he writes on international affairs and technology. Before joining CSIS, he worked at the Departments of State and Commerce as a Foreign Service Officer and as a member of the Senior Executive Service. His government experience includes work on a range of politico-military and intelligence-related issues. Dr. Lewis led the US delegation to the Wassenaar Arrangement Experts Group on advanced civil and military technologies. He was assigned to US Southern Command and US Central Command. He has authored numerous publications since coming to CSIS, including the bestselling "Cybersecurity for the 44th Presidency," and is an internationally recognized expert on cybersecurity. Dr. Lewis was the Rapporteur for the UN's 2010, 2013 and 2015 Group of Government Experts on Information Security and has led a long running Track II Dialogue on cybersecurity with the China Institutes of Contemporary International Relations. He received his Ph.D. from the University of Chicago.

Dr. Martin Libicki

Martin Libicki (Ph.D., U.C. Berkeley 1978) holds the Keyser Chair of Cybersecurity Studies at the U.S. Naval Academy. In addition to teaching, he carries out research in cyberwar and the general impact of information technology on domestic and national security. He is the author of a 2016 textbook on cyberwar, *Cyberspace in Peace and War*, as well as *Conquest in Cyberspace*: *National Security and Information Warfare* and various related RAND monographs. Prior employment includes twelve years at the National Defense University, three years on the Navy Staff (logistics) and three years for the US GAO.

Dr. Herbert Lin

Herbert Lin is senior research scholar for cyber policy and security at the Center for International Security and Cooperation and Hank J. Holland Fellow in Cyber Policy and Security at the Hoover Institution, both at Stanford University. His research interests relate broadly to policy-related dimensions of cybersecurity and cyberspace, and he is particularly interested in the use of offensive operations in cyberspace as instruments of national policy and in the security dimensions of information warfare and influence operations on national security. In addition to his positions at Stanford University, he is Chief Scientist, Emeritus for the Computer Science and Telecommunications Board, National Research Council (NRC) of the National Academies, where he served from 1990 through 2014 as study director of major projects on public policy and information technology, and Adjunct Senior Research Scholar and Senior Fellow in Cybersecurity (not in residence) at the Saltzman Institute for War and Peace Studies in the School for International and Public Affairs at Columbia University; and a member of the Science and Security Board of the Bulletin of Atomic Scientists. In 2016, he served on President Obama's Commission on Enhancing National Cybersecurity. Prior to his NRC service, he was a professional staff member and staff scientist for the House Armed Services Committee (1986-1990), where his portfolio included defense policy and arms control issues. He received his doctorate in physics from MIT.

Ms. Kacie Kieko Miura

Kacie Kieko Miura is a PhD candidate in the Political Science Department at the Massachusetts Institute of Technology, where she studies Chinese politics and foreign policy. Kacie holds a Master of Arts in International Relations from Yale University and a Bachelor of Arts in Political Science and Journalism from the University of Hawaii at Manoa. Kacie previously served as a Peace Corps Volunteer in Chongqing, China.

Mr. Robert Morgus

Robert Morgus is a senior policy analyst and the Deputy Director with New America's Cybersecurity Initiative. His current work focuses on the intersection of global competition and technology, with particular focus on Russian internet doctrine and cybersecurity in the developing world. He holds a Bachelors of Arts from Occidental College in Diplomacy and World Affairs and is proficient in five languages.

Ms. Rachel Esplin Odell

Rachel Esplin Odell is a PhD candidate in Political Science at the Massachusetts Institute of Technology and a member of the MIT Security Studies Program. Her research focuses on the nature and future of world order; U.S. strategy in the Asia-Pacific region; Chinese foreign policy, civil-military relations, and crisis management behavior; and Sino-U.S. and Sino-Japanese relations. Her dissertation studies the causes and consequences of disagreements among states over how to interpret international law, with a focus on the international law of the sea.

Odell is a recipient of the National Science Foundation Graduate Research Fellowship, the Smith Richardson Foundation World Politics and Statecraft Fellowship, and the Alexander George Award for Best Graduate Student Paper from the Foreign Policy Analysis Section of the International Studies Association. She was a Visiting Research Fellow in the Graduate School of Asia-Pacific Studies at Waseda University in November 2017. Odell is also a Nonresident Research Analyst at the Carnegie Endowment for International Peace, where she worked before coming to MIT. She holds an AB *summa cum laude* in East Asian Studies with a secondary field in Government from Harvard University.

Dr. Eleonore Pauwels

Eleonore Pauwels is the Research Fellow on Emerging Cybertechnologies at the Centre for Policy Research at United Nations University, focusing on Artificial Intelligence. She is also Director of the Anticipatory Intelligence (AI) Lab with the Science and Technology Innovation Program at the Woodrow Wilson International Center for Scholars.

Pauwels is a member of MIT's Council on Extended Intelligence, an adviser on the AI Initiative at Harvard Kennedy School, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, as well as an expert for the World Economic Forum.

Pauwels regularly testifies before US, European and international authorities including the US Department of State, the US National Academy of Sciences, the US National Institutes of Health, the US National Intelligence Council, the European Commission, the Organization for Economic Co-operation and Development, and the United Nations. She also writes for Nature, The New York Times, The Guardian, Scientific American, Le Monde, Axios, Slate and The World Economic Forum.

Dr. Lora Saalman

Dr. Lora Saalman is Vice President of the EastWest Institute's Asia-Pacific Program. She previously served as the director of the China and Global Security Program at the Stockholm International Peace Research Institute and continues to maintain an affiliation as an Associate Senior Fellow, contributing to work on China-Russia-U.S. relations and cyber deterrence, Chinese views on the Ukraine crisis, Chinese and Russian hypersonic glide developments, as well as the impact of machine learning and autonomy on nuclear risk in East Asia.

From 2013-2016, Dr. Saalman worked as an associate professor at the Daniel K. Inouye Asia-Pacific Center for Security Studies, where she covered cybersecurity issues and underwent training at the SANS Institute on hacker tools, exploits and incident handling, as well as ICS/SCADA security essentials. From 2010-2013, she was an associate in the Nuclear Policy Program of the Carnegie Endowment for International Peace and based at the Carnegie-Tsinghua Center for Global Policy in Beijing, China and served as an adjunct professor at Tsinghua University teaching courses in Chinese and English.

From 2003-2006, Dr. Saalman worked as a research associate at the Wisconsin Project on Nuclear Arms Control in Washington, D.C., as well as a visiting fellow at the Observer Research Foundation in New Delhi and the James Martin Center for Nonproliferation Studies (CNS) in Washington, D.C. While at CNS, she earned a one-year fellowship to work at the Division of Safeguards Information Technology at the International Atomic Energy Agency. She earned her B.A. with honors from the University of Chicago in 1995, her M.A. with a certificate in nonproliferation from the Monterey Institute of International Studies in 2004, and her Ph.D. from Tsinghua University in Beijing in 2010, where she was the first American to earn a doctorate from its Department of International Relations, completing all her coursework in Chinese.

Lt Col Jennifer Snow

Lt Col Jennifer Snow is the Donovan Group Innovation Officer for the U.S. Special Operations Command, J51 Futures Plans and Strategy Division and SOFWERX Team. She serves as the military representative for technology outreach and engagement to bridge the gap between government and various technology communities to improve collaboration and communications, identify smart solutions to wicked problems and help guide the development of future smart technology policy to benefit special operations.

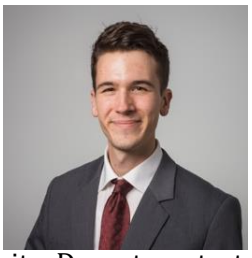
Lt Col Snow entered the Air Force in November 2002 as a graduate of the U.S. Officer Training School at Maxwell AFB in Montgomery, Alabama. She began her professional career as a member of Air Force Special Operations Command, served as an Air Education and Training Command intelligence instructor supervisor and was selected to be one of General Keith B. Alexander's Junior Officer Cryptologic Career Program Interns at the National Security Agency. Prior to her current assignment, Lt Col Snow was a graduate student at the Naval Post Graduate school where she studied emerging disruptive technologies and focused on a class of fast moving emerging technologies she calls "Radical Leveling Technologies", as an area of concern to National Security. Her work was presented to members of the National Security Council at the White House.

Her current efforts focus on examining Radical Leveling Technologies and the new construct of Virtual Nations. Snow seeks to address the broad implications of these technologies and communities as well as the need to leverage the expertise, access and capacity of the community of users and technology drivers to benefit USSOCOM and find innovative policy solutions while still enabling technology for good.

Dr. Laura Steckman

Laura Steckman, PhD, is a social scientist at the MITRE Corporation. Her work operationalizes theories and methodologies from the social and behavioral sciences to address approaches and solutions to mission-specific problems sets worldwide. She has supported Information Operations (IO) and Military Information Support Operations (MISO) for U.S. Central Command, U. S. Pacific Command and various interagency efforts, and is the former Command Social Scientist for the Marine Corps Information Operations Center (MCIOC). Her current research examines the relationship between societies and emerging technologies, specifically in how the two shape each other and the impact that technology and electronic communications have on culture, language, and behavior.

Mr. Valentin Weber



Valentin Weber is a DPhil Candidate in Cyber Security at the Centre for Doctoral Training in Cyber Security and a Research Affiliate with the Centre for Technology and Global Affairs, University of Oxford. He is also an OTF Senior Fellow in Information Controls at the Berkman Klein Center for Internet & Society, Harvard University. Valentin is interested in how the cyber domain is changing conflicts and state strategies. His current research focuses on the integration of cyber and grand strategy, as well as on the role of information controls in state strategies. He previously worked for the International Security Department at Chatham House.

Dr. Nicholas D. Wright

Dr Nicholas Wright is an affiliated scholar at Georgetown University, honorary research associate at University College London (UCL), Consultant at Intelligent Biology and Fellow at New America. His work combines neuroscientific, behavioural and technological insights to understand decision-making in politics and international confrontations, in ways practically applicable to policy. He leads international, interdisciplinary projects with collaborators in countries including China, the U.S., Iran and the UK. He was an Associate in the Nuclear Policy Program, Carnegie Endowment for International Peace, Washington DC and a Senior Research Fellow in International Relations at the University of Birmingham, UK. He has conducted work for the UK Government and U.S. Department of Defense. Before this he examined decision-making using functional brain imaging at UCL and in the Department of Government at the London School of Economics. He was a clinical neurologist in Oxford and at the National Hospital for Neurology. He has published academically (some twenty publications, e.g. *Proceedings of the Royal Society*), in general publications such as *the Atlantic* and *Foreign Affairs*, with the Pentagon Joint Staff (see www.nicholasdwright.com/publications) and has appeared on the BBC and CNN.

Wright received a medical degree from UCL, a BSc in Health Policy from Imperial College London, has Membership of the Royal College of Physicians (UK), has an MSc in Neuroscience and a PhD in Neuroscience both from UCL.