

# ARTIFICIAL INTELLIGENCE, STRATEGIC STABILITY AND NUCLEAR RISK

VINCENT BOULANIN, LORA SAALMAN,  
PETR TOPYCHKANOV, FEI SU AND  
MOA PELDÁN CARLSSON

**June 2020**

**STOCKHOLM INTERNATIONAL  
PEACE RESEARCH INSTITUTE**

SIPRI is an independent international institute dedicated to research into conflict, armaments, arms control and disarmament. Established in 1966, SIPRI provides data, analysis and recommendations, based on open sources, to policymakers, researchers, media and the interested public.

The Governing Board is not responsible for the views expressed in the publications of the Institute.

**GOVERNING BOARD**

Ambassador Jan Eliasson, Chair (Sweden)  
Dr Vladimir Baranovsky (Russia)  
Espen Barth Eide (Norway)  
Jean-Marie Guéhenno (France)  
Dr Radha Kumar (India)  
Ambassador Ramtane Lamamra (Algeria)  
Dr Patricia Lewis (Ireland/United Kingdom)  
Dr Jessica Tuchman Mathews (United States)

**DIRECTOR**

Dan Smith (United Kingdom)



**STOCKHOLM INTERNATIONAL  
PEACE RESEARCH INSTITUTE**

Signalistgatan 9  
SE-169 72 Solna, Sweden  
Telephone: + 46 8 655 9700  
Email: [sipri@sipri.org](mailto:sipri@sipri.org)  
Internet: [www.sipri.org](http://www.sipri.org)

# Artificial Intelligence, Strategic Stability and Nuclear Risk

VINCENT BOULANIN, LORA SAALMAN,  
PETR TOPYCHKANOV, FEI SU AND  
MOA PELDÁN CARLSSON



June 2020



# Contents

<i>Preface</i>	v
<i>Acknowledgements</i>	vi
<i>Abbreviations</i>	vii
<i>Executive Summary</i>	ix
<b>1. Introduction</b>	1
Box 1.1. Key definitions	6
<b>2. Understanding the AI renaissance and its impact on nuclear weapons and related systems</b>	7
I. Understanding the AI renaissance	7
II. AI and nuclear weapon systems: Past, present and future	18
Box 2.1. Automatic, automated, autonomous: The relationship between automation, autonomy and machine learning	15
Box 2.2. Historical cases of false alarms in early warning systems	20
Box 2.3. Dead Hand and Perimetr	22
Figure 2.1. A brief history of artificial intelligence	10
Figure 2.2. The benefits of machine learning	11
Figure 2.3. Approaches to the definition and categorization of autonomous systems	14
Figure 2.4. Benefits of autonomy	16
Figure 2.5. Foreseeable applications of AI in nuclear deterrence	24
<b>3. AI and the military modernization plans of nuclear-armed states</b>	31
I. The United States	33
II. Russia	44
III. The United Kingdom	52
IV. France	59
V. China	67
VI. India	78
VII. Pakistan	87
VIII. North Korea	93
Box 3.1. The artificial intelligence race	32
Figure 3.1. Recent policy developments related to artificial intelligence in the United States	34
Figure 3.2. Recent policy developments related to artificial intelligence in Russia	45
Figure 3.3. Recent policy developments related to artificial intelligence in the United Kingdom	53
Figure 3.4. Recent policy developments related to artificial intelligence in France	60
Figure 3.5. Recent policy developments related to artificial intelligence in China	68

Figure 3.6. Recent policy developments related to artificial intelligence in India	79
Figure 3.7. Recent policy developments related to artificial intelligence in Pakistan	88
Table 3.1. Applications of artificial intelligence of interest to the US Department of Defense	38
Table 3.2. State of adoption of artificial intelligence in the United States nuclear deterrence architecture	42
Table 3.3. State of adoption of artificial intelligence in the Russian nuclear deterrence architecture	50
Table 3.4. State of adoption of artificial intelligence in the British nuclear deterrence architecture	58
Table 3.5. State of adoption of artificial intelligence in the French nuclear deterrence architecture	66
Table 3.6. State of adoption of artificial intelligence in the Chinese nuclear deterrence architecture	77
Table 3.7. State of adoption of artificial intelligence in the Indian nuclear deterrence architecture	85
Table 3.8. State of adoption of artificial intelligence in the Pakistani nuclear deterrence architecture	92
Table 3.9. North Korean universities conducting research and studies in artificial intelligence	96
Table 3.10. State of adoption of artificial intelligence in the North Korean nuclear deterrence architecture	99
<b>4. The positive and negative impacts of AI on strategic stability and nuclear risk</b>	<b>101</b>
I. The impact on strategic stability and strategic relations	101
II. The impact on the likelihood of nuclear conflict: Foreseeable risk scenarios	113
Box 4.1. Machine learning and verification of nuclear arms control and disarmament: Opportunities and challenges	104
<b>5. Mitigating the negative impacts of AI on strategic stability and nuclear risk</b>	<b>123</b>
I. Mitigating risks: What, how and where	123
II. Possible technical and organizational measures for risk reduction	127
III. Possible policy measures for risk reduction	130
Figure 5.1. Risks and challenges posed by the use of artificial intelligence in nuclear weapons	124
<b>6. Conclusions</b>	<b>136</b>
I. Key findings	136
II. Recommendations	140
Figure 6.1. Possible risk reduction measures and how they can be implemented	142
Figure 6.2. Four key measures to deal with the negative impact of AI on strategic stability and nuclear risk	143
<b>About the authors</b>	<b>144</b>

# Preface

The current period sees the post-cold war global strategic landscape in an extended process of redefinition. This is the result of a number of different trends. Most importantly, the underlying dynamics of world power have been shifting with the economic, political and military rise of China, the reassertion under President Vladimir Putin of a great power role for Russia, and the disenchantment expressed by the current United States administration with the international institutions and arrangements that the USA itself had a big hand in creating. As a result, the China–US rivalry has increasingly supplanted the Russian–US nuclear rivalry as the core binary confrontation of international politics. This pair of dyadic antagonisms is, moreover, supplemented by growing regional nuclear rivalries and strategic triangles in South Asia and the Middle East.

Against this increasingly toxic geopolitical background, the arms control framework created at the end of the cold war has deteriorated. Today, the commitment of the states with the largest nuclear arsenals to pursue stability through arms control and potentially disarmament is in doubt. The impact of coronavirus disease 2019 (COVID-19) is not yet clear but may well be a source of further unsettling developments.

All of this is the volatile backdrop to considering the consequences of new technological developments for armament dynamics. The world is going through a fourth industrial revolution, characterized by rapid advances in artificial intelligence (AI), robotics, quantum technology, nanotechnology, biotechnology and digital fabrication. The question of how these technologies will be used has not yet been answered in full detail. It is beyond dispute, however, that nuclear-armed states will seek to use these technologies for their national security.

The SIPRI project ‘Mapping the impact of machine learning and autonomy on strategic stability’ set out to explore the potential effect of AI exploitation on strategic stability and nuclear risk. The research team has used a region-by-region approach to analyze the impact that the exploitation of AI could have on the global strategic landscape. This report is the final publication of this two-year research project funded by the Carnegie Corporation of New York; it presents the key findings and recommendations of the SIPRI authors derived from their research as well as a series of regional and transregional workshop organized in Europe, East and South Asia and the USA. It follows and complements the trilogy of edited volumes that compile the perspectives of experts from these regions on the topic.

SIPRI commends this study to decision makers in the realms of arms control, defence and foreign affairs, to researchers and students in departments of politics, international relations and computer science, as well as to members of the general public who have a professional and personal interest in the subject.

Dan Smith  
Director, SIPRI  
Stockholm, June 2020

## Acknowledgements

The authors would like to express their sincere gratitude to the Carnegie Corporation of New York for its generous financial support of the project. They are also indebted to all the experts who participated in the workshops and other events that SIPRI organized in Stockholm, Beijing, Colombo, New York, Geneva and Seoul. The content of this report reflects the contributions of this international group of experts.

The authors also wish to thank the external reviewer, Erin Dumbacher, as well as SIPRI colleagues Sibylle Bauer, Mark Bromley, Tytti Erästö, Shannon Kile, Luc van de Goor, Pieter Wezeman and Siemon Wezeman for their comprehensive and constructive feedback. Finally, we would like to acknowledge the invaluable editorial work of David Cruickshank and the SIPRI editorial department.

Responsibility for the views and information presented in this report lies entirely with the authors.



# Abbreviations

A2/AD	Anti-access/area-denial
AGI	Artificial general intelligence
AI	Artificial intelligence
ATR	Automatic target recognition
AURA	Autonomous Unmanned Research Aircraft
CAIR	Centre for Artificial Intelligence and Robotics
CBM	Confidence-building measure
CCW	Certain Conventional Weapons Convention
CD	Conference on Disarmament
DARPA	Defense Advanced Research Projects Agency
DBN	Deep belief network
DCDC	Development, Concepts and Doctrine Centre
DGA	Direction générale de l'armement (director general of armaments of France)
DIU	Defense Innovation Unit
DOD	Department of Defense
DRDO	Defence Research and Development Organisation
GAN	Generative adversarial network
ICBM	Intercontinental ballistic missile
ICT	Information and communications technology
ISR	Intelligence, surveillance and reconnaissance
IT	Information technology
JAIC	Joint Artificial Intelligence Center
LAWS	Lethal autonomous weapon systems
MAF	Ministry of the Armed Forces
MOCI	Ministry of Commerce and Industry
MOD	Ministry of Defence
NATO	North Atlantic Treaty Organization
NC3	Nuclear command, control and communications
New START	Treaty on Measures for the Further Reduction and Limitation of Strategic Offensive Arms
NFU	No-first-use
NPT	Non-Proliferation Treaty
PIAIC	Presidential Initiative for Artificial Intelligence and Computing
R&D	Research and development
RGB	Reconnaissance General Bureau
SCCSS	Strategic Command and Control Support System
SLBM	Submarine-launched ballistic missile
SSBN	Nuclear-powered ballistic missile submarine
UAV	Unmanned aerial vehicle
UCAV	Unmanned combat aerial vehicle
UN	United Nations

USAF	United States Air Force
USV	Unmanned surface vehicle
UUV	Unmanned underwater vehicle
WMD	Weapon of mass destruction
XLUUV	Extra-large unmanned underwater vehicle

# Executive Summary

The world is undergoing a fourth industrial revolution that is characterized by rapid and converging advances in many technological areas. Few of these technologies are expected to have as profound an impact on relations among nuclear-armed states as artificial intelligence (AI). While the field of AI has been around since the 1950s, it has experienced a renaissance since the beginning of the 2010s. The recent advances in AI could have an impact on the field of nuclear weapons and posture, with consequences for strategic stability and nuclear risk reduction. Nuclear-armed states and international organizations must thus consider a spectrum of options to deal with the challenges generated by AI.

The two major technological developments of the current AI renaissance are machine learning and autonomy. Machine learning—at the core of the renaissance—is an approach to software programming that now enables the development of increasingly capable AI applications. Autonomy—a key by-product of the renaissance—refers to the ability of a machine to execute tasks without human input, using interactions of computer programming with the environment. Autonomous systems have been around for a long time, but recent advances in machine learning have made them more sophisticated and useful.

From a technical perspective, it is beyond dispute that the AI renaissance will have an impact on nuclear weapons and postures. Advances in machine learning and autonomy could unlock new and varied possibilities for a wide array of nuclear force-related capabilities, ranging from early warning to command and control and weapon delivery. The key question is therefore not if, but when, how and by whom these recent advances of AI will be adopted in nuclear force architectures. However, these technological developments are still only a few years old and little detailed information is available in official sources about how nuclear-armed states see the role of AI in their nuclear force development or modernization plans.

Nonetheless, there is already clear evidence that all nuclear-armed states have taken notice of the AI renaissance and have made the pursuit of AI a priority. The ability to harness the recent advances in AI is typically presented as an essential enabler of national and military power in the years to come. AI is also systematically presented as a stake in the great power competition, and official sources show that nuclear-armed states are determined to be world leaders in this field. In this context, while it is too early to determine the net effect of recent advances in AI on strategic stability and nuclear risk, some informed speculation is possible.

Given the typical way in which military technology is adopted, the incorporation of AI into nuclear weapon systems is likely to be slow and steady. This development could have both stabilizing and destabilizing effects on strategic stability, depending on the country and the regional context. For example, if a dominant power were to use AI to enhance its nuclear force structure, this could further compromise the deterrence capability of a weaker country, weakening strategic stability. Alternatively, if the weaker power is able to harness AI to improve its

own nuclear forces, then it may be able to redress existing asymmetries, thereby enhancing mutual vulnerability and strategic stability. In cases of multilateral nuclear deterrence relations, calculations of AI-driven strategic stability become even more complex.

At the same time, advances in AI can have an impact on strategic stability relations among nuclear-armed states even before they are fully developed, much less deployed. For example, a state may perceive that an adversary's investment in AI, even non-nuclear-related, could give that adversary the ability to threaten the state's future second-strike capability. This could be sufficient to generate insecurity and lead that state to adopt measures that could decrease strategic stability and increase the risk of a nuclear conflict.

Throughout these and other scenarios, AI could fail or be misused in ways that could trigger an accidental or inadvertent escalation of a crisis or conflict into a nuclear conflict. However, for these scenarios to become reality, a number of destabilizing dynamics would need to align. In the current geopolitical context, it is hard to imagine how AI technology alone could be the determining trigger for nuclear weapon use. Geopolitical tensions, lack of communication and inadequate signalling of intentions are all variables that would play an equally important if not greater role than AI technology in triggering an escalation of crisis or conflict to the nuclear level.

While it might be hard to predict the exact impact that AI may have, it is not too early to start discussing options that nuclear-armed states and the international security community could explore to prevent and mitigate the risks that military applications of AI, including nuclear weapon systems, pose to peace and stability. Some solutions already exist. Existing arms control instruments include a number of proven technical, organizational and policy measures that could be discussed and implemented, unilaterally, bilaterally or multilaterally.

However, political pragmatism is required to determine which measures and adoption processes will be adequate, implementable and effective. The main challenge is that the political and institutional conditions required for a constructive discussion among nuclear-armed states on arms control have worsened dramatically in recent years, while the conversation on AI-related risks is still new and speculative.

In this light, states and international organizations should take a number of measures—sequentially or simultaneously—to deal pragmatically with the strategic challenges that AI raises. One measure would be to support awareness-raising measures that will help the relevant stakeholders—governmental practitioners, industry and civil society, among others—gain a realistic sense of the challenges posed by AI in the nuclear arena. Another measure would be to support transparency and confidence-building measures that can help to reduce misperception and misunderstanding among nuclear-armed states on AI-related issues. An additional measure would be to support collaborative resolution of the challenges posed by AI and the exploration of beneficial use of AI for arms control. A final possible measure would be to discuss and agree on concrete limits to the use of AI in nuclear forces.

# 1. Introduction

The nuclear order that was inherited from the cold war is under great stress as its political, institutional, geopolitical and technological foundations are being called into question in unprecedented ways.

The political will of the Russian Federation and the United States—the states with the world’s largest nuclear arsenals—to pursue stability through arms control and disarmament seems to have diminished. In line with its 2018 Nuclear Posture Review, in early 2020 the USA announced the deployment of submarine-launched ballistic missiles (SLBMs) with low-yield nuclear warheads.<sup>1</sup> Russia is investing in new strategic weapons, reportedly including the Poseidon, a nuclear-capable, nuclear-powered unmanned underwater vehicle (UUV), and the Burevestnik, a nuclear-powered long-range cruise missile.<sup>2</sup>

Accompanying these shifts in doctrine and armaments, the arms control framework created by the Soviet Union and the USA during the cold war is disintegrating. In August 2019 the USA withdrew from the 1987 Treaty on the Elimination of Intermediate-Range and Shorter-Range Missiles (INF Treaty) following years of deadlock over alleged Russian non-compliance.<sup>3</sup> The 2010 Russian-US Strategic Arms Reduction Treaty (New START) will also expire in 2021 unless both parties agree to extend or replace it.<sup>4</sup> However, neither the USA nor Russia has officially begun to negotiate deeper nuclear arms reductions beyond the levels of New START.

The global strategic landscape has also changed with an expansion in the number of declared nuclear-armed states over the past quarter of a century to include India, Pakistan and the Democratic People’s Republic of Korea (DPRK, or North Korea).<sup>5</sup> The nuclear order is increasingly multipolar.<sup>6</sup> The East versus West nuclear binary that characterized the cold war has now been complicated by regional nuclear rivalries and even strategic triangles, such as that between China, Russia and the USA.<sup>7</sup> Moreover, China, India, Pakistan are continuing to

<sup>1</sup> US Department of Defense (DOD), *Nuclear Posture Review* (DOD: Washington, DC, Feb. 2018), pp. 54–55; and US Department of Defense, ‘Statement on the fielding of the W76-2 low-yield submarine launched ballistic missile warhead’, 4 Feb. 2020.

<sup>2</sup> TASS, ‘Key stage of Poseidon underwater drone trials completed, says Putin’, 2 Feb. 2019; ‘Russia begins testing of “Poseidon” underwater nuclear drone’, PressTV, 26 Dec. 2018; and Ramm, A., [Winged ‘Burevestnik’: What is known about Russia’s secret weapon], *Izvestia*, 5 Mar. 2019 (in Russian). See also Hwang, I. and Kim, J., ‘The environmental impact of nuclear-powered autonomous weapons’, ed. L. Saalman, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. II, *East Asian Perspectives* (SIPRI: Stockholm, Oct. 2019), pp. 86–90.

<sup>5</sup> Sood, R. (ed), *Nuclear Order in the Twenty-First Century* (Observer Research Foundation: New Delhi, 2019).

<sup>6</sup> Legvold, R., ‘The challenges of a multipolar nuclear world in a shifting international context’, S. E. Miller, R. Legvold and L. Freedman, *Meeting the Challenges of the New Nuclear Age: Nuclear Weapons in a Changing Global Order* (American Academy of Arts and Sciences: Cambridge, MA, 2019), pp. 28–61.

<sup>7</sup> Kaspersen, A. and King, C., ‘Mitigating the challenges of nuclear risk while ensuring the benefits of technology’, ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, *Euro-Atlantic Perspectives* (SIPRI: Stockholm, May 2019), pp. 119–27; and Arbatov, A. and Dvorkin, V., *The Great Strategic Triangle* (Carnegie Moscow Center: Moscow, Apr. 2013).

increase the size of their nuclear arsenals. China and India are also expanding their range of delivery platforms to include nuclear-powered ballistic missile submarines (SSBNs), intercontinental ballistic missiles (ICBMs) armed with multiple independently targetable re-entry vehicles (MIRVs), dual-capable cruise missiles and hypersonic glide vehicles.<sup>8</sup>

These shifts in arms control frameworks and the strategic landscape have been accompanied by new technological developments in armaments. The world is undergoing a fourth industrial revolution that is characterized by rapid and converging advances in many technological areas, such as artificial intelligence (AI), robotics, quantum technology, nanotechnology, biotechnology and digital fabrication technologies.<sup>9</sup> There are strong incentives for nuclear-armed states to use these technologies to build military capabilities that they perceive as important for their national security—and a high likelihood that they will do so. In turn, the adoption of these technologies could affect the way in which these states think of the adequacy of their nuclear deterrent. Advances by one nuclear-armed state in a specific emerging technology have knock-on effects on other countries depending on how advanced the latter's nuclear and conventional military capabilities are. States that find themselves lacking in one or both of these arenas are likely to feel compelled to boost or complement their nuclear arsenals, which in turn could have a detrimental impact on strategic stability and increase the likelihood of nuclear conflict.

The impact that new technological developments could have on strategic stability and nuclear risk is the focus of this report. It draws particular attention to the suite of technologies that fall under the category of AI. Few of the technological developments that characterize the fourth industrial revolution are expected to have as profound an impact on relations between nuclear-armed states as AI. This impact may also be wide-ranging since AI is not a definite, unified or even verifiable technology in the way that nuclear weapons or missile technology are. Rather, it is a portfolio or umbrella term with a wide variety of enabling applications that may be used to give some form of cognitive capability to many types of technology.<sup>10</sup>

Some AI experts have compared the transformative power of AI to that of electricity: 'Just as everything became more useful when it was "electrified", everything will become useful when it is "cognified"'.<sup>11</sup> In military contexts, this means that AI could make any type of weapon system—whether conventional or nuclear—smarter and more autonomous. One area in which AI is likely to have a

<sup>8</sup> ChinaPower, 'Does China have an effective sea-based nuclear deterrent', Center for Strategic and International Studies, 2 May 2019; and Kile, S. N. and Kristensen, H. M., 'World nuclear forces: Overview', *SIPRI Yearbook 2019: Armaments, Disarmament and International Security* (Oxford University Press: Oxford, 2019), pp. 319–20.

<sup>9</sup> On the impact of these emerging technologies on arms control see Brockmann, K., Bauer, S. and Boulanin, V., *Bio Plus X: Arms Control and the Convergence of Biology and Emerging Technologies* (SIPRI: Stockholm, Mar. 2019).

<sup>10</sup> Boulanin, V., 'Artificial intelligence: A primer', ed. Boulanin (note 7), pp. 13–25.

<sup>11</sup> Ng, A, cited in Kostopoulos, L., 'AI, emerging tech and national defence @SIPRI Stockholm Security Conference', Medium, 23 Sep. 2018.

significant impact is human decision-making. Since AI enables the development of increasingly intelligent decision-support systems, it has the potential to deeply transform the way in which military commanders take and execute decisions in or prior to an armed conflict.

The transformative potential of AI brings both promise and concerns. The net effect that AI could have on military affairs is the focus of a growing literature. Much of this is connected to the ongoing intergovernmental discussion on lethal autonomous weapon systems (LAWS) that is taking place in the framework of the 1981 Convention on Certain Conventional Weapons (CCW Convention).<sup>12</sup> This discussion focuses on the humanitarian risks posed by the use of autonomous weapon systems in armed conflict.<sup>13</sup> The effects that military use of AI could have outside the specific context of armed conflict has not yet received the attention it deserves—partly because it falls outside the scope of the CCW Convention. However, military AI could have an impact on international peace and security in many ways. For example, it could redistribute the balance of power and thereby increase or decrease states' sense of security. In fact, AI seems to be already affecting the way that the leaders of various countries think about power and pursue their national security interests.

Major military powers have come to consider data to be the new 'oil'—as essential for military operations as the fuel needed for aircraft and tanks.<sup>14</sup> According to this view, a lack of access to digital data to train AI systems combined with a paucity of AI talent and companies that can drive innovation in this field may result in an imbalance between countries' abilities to exploit AI for military advances. Between 2017 and early 2018 no fewer than 17 countries released a national strategy or made a strategic policy announcement on AI, which indicates that the development of national capabilities in AI is a top priority.<sup>15</sup>

In the case of strategic relations among states, whether nuclear-armed or non-nuclear-armed, the quest for military AI could be contributing to the existing strategic balance—or imbalance—between offensive and defensive weapons. The adoption or perceived adoption of new AI capabilities by any state could make a nuclear-armed state (or states) fear for the survivability and reliability of its nuclear deterrent. This would thereby motivate the concerned state to respond with measures that could undermine the current status quo and even increase the

<sup>12</sup> Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects (CCW Convention, or 'Inhumane Weapons' Convention), opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983. Examples of this literature include Asaro, P., 'On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making', *International Review of the Red Cross*, vol. 94, no. 886 (summer 2012), pp. 687–709; Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017); Marchant, G. E. et al., 'International governance of autonomous military robots', *Columbia Science and Technology Law Review*, vol. 12 (2015), pp. 272–315; Horowitz, M. C., 'The ethics & morality of robotic warfare: Assessing the debate over autonomous weapons', *Daedalus*, vol. 145, no. 4 (fall 2016), pp. 25–36; and Cummings, M. L., *Artificial Intelligence and the Future of Warfare* (Chatham House: London, Jan. 2017).

<sup>13</sup> Cummings (note 12).

<sup>14</sup> Horowitz, M. C. et al., *Strategic Competition in the Era of Artificial Intelligence* (Center for New American Security: Washington DC, July 2018).

<sup>15</sup> Dutton, T., 'An overview of national AI strategies', Medium, 18 June 2018.

risk of a nuclear conflict. Such activities could include increasing the alert status of nuclear weapons or automating nuclear launch policies. A similar mechanism can already be seen in the reactions to missile defence programmes among both nuclear-armed and non-nuclear-armed states.<sup>16</sup>

The incorporation of AI into military systems could also pose new challenges as far as nuclear risk reduction is concerned, as it could increase the risk of escalation into a nuclear conflict either accidentally (i.e. as a result of technical failure or unauthorized use) or inadvertently (i.e. as ‘the unintended consequence of a decision to fight a conventional war’).<sup>17</sup> In politically sensitive situations, such as an intentional or unintentional collision at sea, increasing reliance on AI systems could turn out to be highly problematic. Such a dependence could lead humans to take premature or misguided decisions and actions that could trigger an armed conflict, and possibly even a nuclear war.

In other words, the impact that AI could have on strategic stability and nuclear risk merits greater scrutiny. It should be acknowledged that there is an emerging discourse on the topic: an increasing number of scholars working on nuclear-related issues have turned their attention since 2017 to the topic of AI.<sup>18</sup> However, the conversation remains nascent. The number of publications is limited and many of them are Western-centric and take an abstract approach to the topic.

This report aims to make an original and detailed contribution to the AI and nuclear literature with a thorough exploration of empirical evidence and analysis that considers risk scenarios in varied geographical contexts. The primary hypothesis is that the effect of militarization of AI on strategic stability and nuclear risk will differ from one region to another. Factors that matter include the size and sophistication of conventional and nuclear arsenals, the speed of technological adoption, geographic and geopolitical tensions, technological symmetry or asymmetry, and the status and maturity of the strategic relationships.

This report aims to offer the reader a concrete understanding of how the adoption of AI by nuclear-armed states could have an impact on strategic stability and nuclear risk and how related challenges could be addressed at the policy level. The analysis builds on extensive data collection on the AI-related technical and strategic developments of nuclear-armed states. It also builds on the authors’

<sup>16</sup> Pifer, S., ‘Missile defence—would the Kremlin pitch a deal?’, Order from Chaos, Brookings Institution, 2 June 2016; and Gomez, E., ‘Russia claims its new nuclear weapons are a response to US missile defense’, *National Interest*, 15 Mar. 2020.

<sup>17</sup> On accidental and inadvertent escalation see Posen, B. R., ‘Inadvertent nuclear war? Escalation and NATO’s northern flank’, *International Security*, vol. 7, no. 2 (fall 1982), pp. 28–54, p. 29. For a broader typology of conflict escalation see also Woodhams, G. and Borrie, J., *Armed UAVs in Conflict Escalation and Inter-State Crisis*, United Nations Institute for Disarmament Research (UNIDIR) Resources (UNIDIR: Geneva, 2018); and the discussion in chapter 4, section II, in this volume.

<sup>18</sup> Notable studies exploring this issue among the sparse examples include Altmann, J. and Sauer, F., ‘Autonomous weapon systems and strategic stability’, *Survival*, vol. 59, no. 5 (Nov. 2017), pp. 117–42; Payne, K., ‘Artificial intelligence: A revolution in strategic affairs?’, *Survival*, vol. 60, no. 5 (Oct.–Nov. 2018), pp. 7–32; Horowitz et al. (note 14); Horowitz, M. C., ‘Artificial intelligence, international competition, and the balance of power’, *Texas National Security Review*, vol. 1, no. 3 (May 2018), pp. 37–57; Geist, E. and Lohn, A. J., *How Might Artificial Intelligence Affect the Risk of Nuclear War?* (Rand Corporation: Santa Monica, CA, 2018); and Lieber, K. A. and Press, D. G., ‘The new era of counterforce: Technological change and the future of nuclear deterrence’, *International Security*, vol. 41, no. 4 (spring 2017), pp. 9–49.



conclusions from a series of regional workshops that SIPRI organized in Sweden (on Euro-Atlantic dynamics), China (on East Asian dynamics) and Sri Lanka (on South Asian dynamics), as well as a transregional workshop in New York. At these workshops, AI experts, scholars and practitioners who work on arms control, nuclear strategy and regional security had the opportunity to discuss why and how the adoption of AI capabilities by nuclear-armed states could have an impact on strategic stability and nuclear risk within or among regions.<sup>19</sup>

The report is structured around four questions.

1. What is the state of AI and what types of capability could nuclear-armed states derive from the recent, current and foreseeable advances in AI?
2. Why and to what extent are nuclear-armed states currently investing in AI? Have they articulated a concrete plan around how AI could be used in future nuclear modernization or developments plans? Are there notable regional differences?
3. What impact might the adoption of AI for military purposes by nuclear-armed states have on strategic stability and nuclear risk? What differences are visible among regions?
4. How should the strategic risk posed by AI be mitigated or even prevented, both regionally and transregionally?

The following four chapters tackle these questions in turn.

Chapter 2 describes the state of AI and maps its past, present and possible future applications in nuclear weapon systems and related systems (see the definition in box 1.1). The chapter shows how AI, at least in some form, is already used for a number of purposes in nuclear and non-nuclear systems and why its importance could expand in the coming decades with advances in machine learning and autonomous systems. It then describes the spectrum of capabilities that may, realistically, emerge from these technologies.

Chapter 3 provides an overview of AI developments by the eight declared nuclear-armed states: China, France, India, North Korea, Pakistan, Russia, the United Kingdom and the United States.<sup>20</sup> In doing so, these case studies offer a better understanding of strategic visions, national policies, state of adoption and key capabilities of each of these states that are shaping the integration of AI into their military modernization plans and may have an impact on nuclear risk.

Chapter 4 builds on this foundation to explore the risks and benefits associated with the integration of the recent advances in AI, particularly machine learning-based applications and autonomous capabilities, into nuclear weapon systems and non-nuclear strategic systems. It discusses a variety of regional scenarios in which the adoption and use of AI capabilities could destabilize relations among

<sup>19</sup> The view of the experts who attended these workshops have been compiled in 3 edited volumes. ed. Boulanin (note 7); ed. Saalman (note 2); and Topychkanov, P. (ed.), *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk, vol. III, South Asian Perspectives* (SIPRI: Stockholm, Apr. 2019).

<sup>20</sup> Israel is not a declared nuclear-armed state and so is not discussed here. Israel continues to maintain its long-standing policy of neither confirming nor denying that it possesses nuclear weapons.

**Box 1.1. Key definitions****Artificial intelligence**

Artificial intelligence is a catch-all term that refers to a wide set of computational techniques that allow computers and robots to solve complex, seemingly abstract problems that had previously yielded only to human cognition.<sup>a</sup>

**Nuclear weapon systems**

Nuclear weapon systems should be understood in the broadest sense. They include not only the nuclear warheads and the delivery systems but also all nuclear force-related systems such as nuclear command and control, early-warning systems and intelligence, reconnaissance and surveillance systems. Relevant non-nuclear strategic weapons include long-range high-precision missiles, manned combat aircraft, unmanned combat aerial vehicles (UCAVs) and ballistic missile defence systems.

**Strategic stability**

Strategic stability has many definitions. It is understood here in the narrowest sense as a concept that describes ‘the absence of incentive to use nuclear weapons first (crisis stability) and the absence of incentive to build up a nuclear force (arms race stability)’.<sup>b</sup> It is ‘a state of affairs in which countries are confident that their adversaries would not be able to undermine their nuclear deterrent capability’ using nuclear, conventional or other non-conventional means.<sup>c</sup>

<sup>a</sup> See the detailed definition in Boulanin, V., ‘Artificial intelligence: A primer’, ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, Euro-Atlantic Perspectives (SIPRI: Stockholm, May 2019), pp. 13–25.

<sup>b</sup> Edward Warner cited in Acton, J. M., ‘Reclaiming strategic stability’, eds E. A. Colby and M. S. Gerson, *Strategic Stability: Contending Interpretations* (US Army War College Press: Carlisle Barracks, PA, Feb. 2013), pp. 117–46, p. 117.

<sup>c</sup> Podvig, P., ‘The myth of strategic stability’, *Bulletin of the Atomic Scientists*, 31 Oct. 2012.

Source: Adapted from Boulanin, V., ‘Introduction’, ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, *Euro-Atlantic Perspectives* (SIPRI: Stockholm, May 2019), pp. 3–9, p. 4.

nuclear-armed states and could be a factor in nuclear escalation in a politically sensitive or crisis situation.

Chapter 5 discusses how the strategic risks posed by AI might be prevented or mitigated, regionally and globally. It reviews the policy options that states and international organizations already have at their disposal. In doing so, it discusses the types of additional risk-mitigation measure that may be required, particularly to tackle the risk scenarios identified in chapter 4.

Chapter 6 concludes this report by summarizing its key findings and recommendations.

## 2. Understanding the AI renaissance and its impact on nuclear weapons and related systems

In order to understand why nuclear-armed states and military powers in general see AI as a key enabler of their current and future military modernization plans, it is useful to start with a brief overview of the current state of AI (section I) and the types of capability that nuclear-armed states could derive from recent, current and foreseeable advances in this field (section II).

### I. Understanding the AI renaissance<sup>21</sup>

#### **What is artificial intelligence?**

The concept of AI was coined in the mid-1950s by John McCarthy, who defined it broadly as the ‘science and engineering of making intelligent machines’.<sup>22</sup> Researchers tend to use AI as a catch-all term that refers to a wide set of computational techniques that allow computers and robots to solve complex, seemingly abstract problems that had previously yielded only to human cognition—for example, observing the world through vision, processing natural language and learning.<sup>23</sup>

Some AI researchers differentiate between so-called narrow AI and artificial general intelligence (AGI), which is general-purpose AI. AGI is the AI that would match—if not outperform—the ability of a human to make sense of the world and to develop an understanding of his or her environment. It is the kind of AI that is typically depicted in popular culture in films such as *The Terminator*, *Blade Runner* or *2001: A Space Odyssey*. AGI has always fascinated AI researchers, but its design remains an unresolved technical challenge. There are, in fact, strong disagreements as to whether AGI will ever be possible. Even the most optimistic AI researchers admit that AGI programs are likely to remain in the realm of science fiction for the foreseeable future.<sup>24</sup>

Narrow AI has been around for decades and is the type of AI that is widely used today. Narrow AI systems are complex software programs that can execute discrete ‘intelligent’ tasks such as recognizing objects or people from images, translating language or playing games. Narrow AI systems execute complex calculations, but they are brittle in nature: they are limited by the boundaries of their programming and they only work, at least reliably, for the intended tasks and

<sup>21</sup> This section is largely based on Boulanin (note 10).

<sup>22</sup> Pearl, A., ‘Homage to John McCarthy, the father of artificial intelligence (AI)’, *Artificial Solutions*, 2 June 2017.

<sup>23</sup> International Panel on the Regulation of Autonomous Weapons (IPRAW), *Focus on Computational Methods in the Context of LAWS*, ‘Focus on’ Report no. 2 (German Institute for International and Security Affairs: Berlin, Nov. 2017).

<sup>24</sup> Ready, C., ‘Kurzweil claim that singularity will happen by 2045’, *Futurism*, 5 Oct. 2017.

operating environment. This is the type of AI to which this report refers when it discusses AI.

A key definitional element to bear in mind is that narrow AI is not a definite, unified technology. Instead, it is a portfolio technology that encompasses a wide variety of enabling applications which may be used to ‘cognify’ (i.e. give some form of cognitive capability to) multiple types of technology, including weapon technologies.

### **Genesis of the current AI renaissance**

AI is often depicted as a new or emerging technology. However, as an academic discipline, AI is three-quarters of a century old. Narrow AI applications have been used for civilian and military purposes since the 1960s.<sup>25</sup> A constant in the public debate has been the use of the concept of AI to refer to the newest computer technologies. Once these technologies have been widely deployed and adopted, they are no longer thought of or depicted as AI. In other words, the frontier of AI is always moving. What is considered AI today may be deemed as normal software technology in the near future.

Since the 1950s the field of AI has gone through several ‘hype cycles’: each period of major success (an ‘AI summer’) was inevitably followed by a period of disillusionment (an ‘AI winter’) as the new and promising approach of AI eventually failed to match its early expectations (see figure 2.1).<sup>26</sup> These AI winters typically resulted in funding cutbacks.

During the 2010s, the field of AI has been experiencing a new ‘summer’ due to a breakthrough in machine learning. This approach to AI software development has been around since the beginning of AI research but has greatly benefited in the past decade from the progress of computer power and the increasing availability of digital data.<sup>27</sup>

As with previous AI summers, success stories about what current AI systems can achieve have channelled major interest in and investment towards the most promising approaches to AI engineering—currently machine learning—but also towards concrete applications that could be derived from it, such as autonomy.

These two features of the AI renaissance—the enabler (machine learning) and the by-product (autonomy)—are the key technologies that are discussed in this report. The rest of this section introduces them further.

<sup>25</sup> Dale, R., ‘An introduction to artificial intelligence’, ed. A. M. Din, SIPRI, *Arms and Artificial Intelligence: Weapons and Arms Control Applications of Advanced Computing* (Oxford University Press: Oxford, 1987); Russel, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, 3rd edn (Pearson Education: Harlow, 2014); Kit, P., ‘What should we learn from past AI forecasts?’, Open Philanthropy Project, May 2016; and Armstrong, S. and Sotala, K., ‘How we’re predicting AI—or failing to’, eds J. Romportl et al., *Beyond AI: Artificial Dreams*, Proceedings of the International Conference ‘Beyond AI 2012’, Pilsen, Czechia, 5–6 Nov. 2012 (University of West Bohemia: Pilsen, 2012), pp. 52–75.

<sup>26</sup> On hype cycles see Gartner, ‘Gartner hype cycle’, [n.d.]; and Kit (note 25).

<sup>27</sup> Knight, W., ‘There is a big problem with AI’, *MIT Technology Review*, 11 Apr. 2017.

## Machine learning: A key enabler of the AI renaissance

### *What is machine learning?*

Machine learning is an approach to software development that first builds systems that can learn and then teaches them what to do using a variety of methods (i.e. supervised learning, reinforcement learning or unsupervised learning). When used in non-technical contexts, the term ‘learning’ can sometimes be a source of confusion, as it invites an anthropomorphic interpretation. However, the way in which machine learning works has nothing to do with the way that humans learn: machines learn by finding statistical relationships in past data. Engineers use the term ‘learning’ for practical reasons since it is a concise and memorable way of describing a complex computing process.<sup>28</sup> The main advantage of this approach is that it removes the need for hand-coded programming, whereby humans hard-code software features into the systems.<sup>29</sup>

Machine learning has been around since the beginning of AI research. However, it remained a marginal subfield of AI for a long time as it was of limited practical use. It became briefly popular in the 1980s and 1990s. The digitalization of many industries and the emergence of large data sets—on which machine learning systems can be trained—reignited interest and inspired the development of new techniques. Among these were refined versions of a method known as ‘artificial neural networks’, which drew on knowledge of the human brain, statistics and applied mathematics. However, this technique failed to deliver the expected concrete applications. By the early 2000s, the field of machine learning and artificial neural networks had again become marginalized and unfunded.

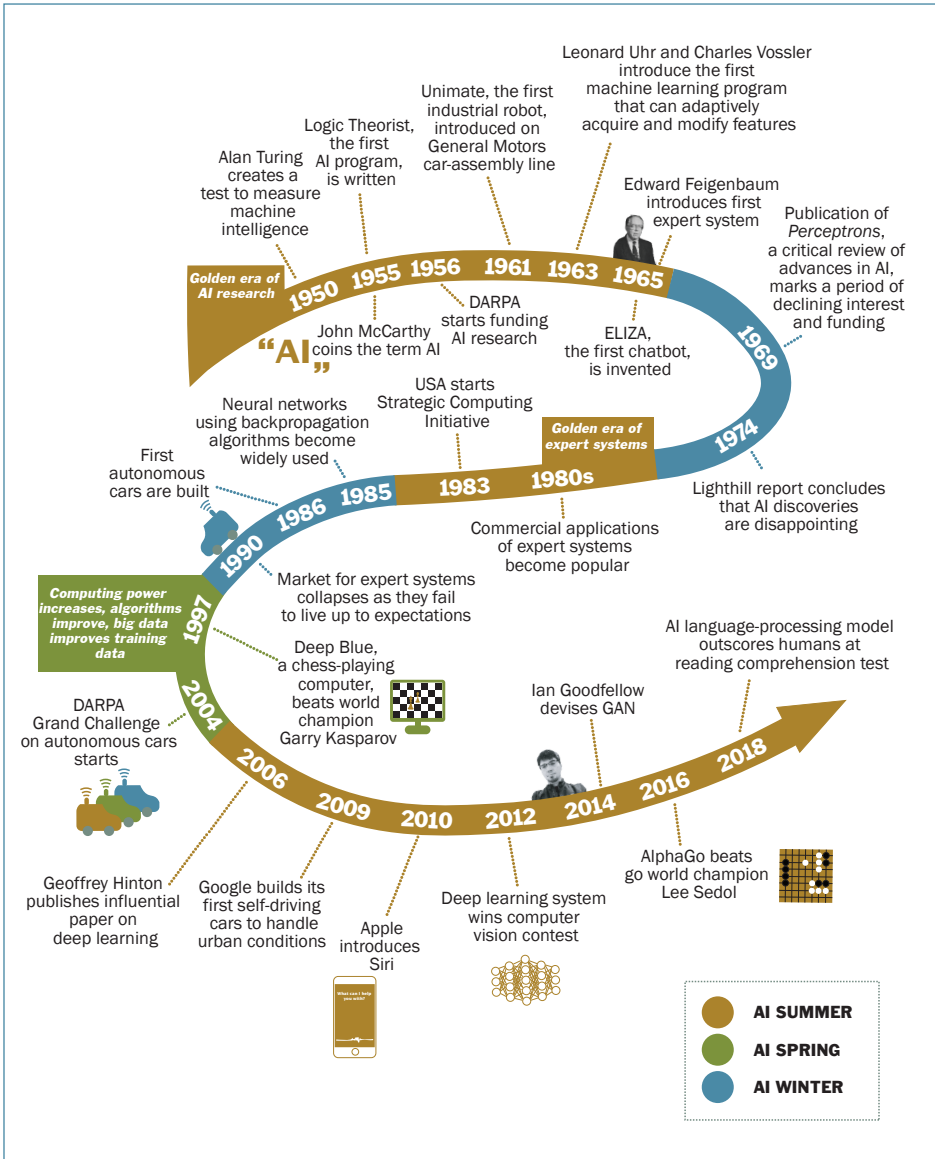
By the 2000s, Geoffrey Hinton, a cognitive psychologist and computer scientist at the University of Toronto, Canada, was one of the last few scholars in the AI community who specialized in neural networks. Hinton came up with a technical tweak that would eventually not only transform the field of AI from the inside, but also reignite massive interest in machine learning and the field of AI.<sup>30</sup> The tweak involved combining different layers of neural networks and a process known as ‘backpropagation’. That modification made artificial neural networks outperform traditional AI programming techniques by a wide margin at any task involving pattern recognition such as recognizing images or speech, classifying data, predict behaviour, translating language, or playing games such as chess (see figure 2.2).<sup>31</sup> Industry and governments alike soon identified major opportunities and started to invest in deep learning and machine learning more broadly. In fact, when companies and governments talk about investing in AI, they primarily talk about investing in machine learning and what fuels it: data.

<sup>28</sup> Boulanin and Verbruggen (note 12).

<sup>29</sup> Knight (note 27).

<sup>30</sup> Somers, J., ‘Is AI riding a one-trick pony’, *MIT Technology Review*, 29 Sep. 2017.







<sup>31</sup> Gershgorin, D., ‘The data that transformed AI research—and possibility the world’, *Quartz*, 26 July 2017; and Allen, K., ‘How a Toronto professor’s research revolutionized artificial intelligence’, *Toronto Star*, 17 Apr. 2015.



**Figure 2.1.** A brief history of artificial intelligence

AI = artificial intelligence, DARPA = United States Defense Advanced Research Projects Agency, GAN = generative adversarial network, LISP = List Processor (software)

Source: Russel, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, 3rd edn (Pearson Education: Harlow, 2014).

KEY ADVANTAGE OF MACHINE LEARNING	WHAT IT MEANS
<b>MACHINE PERCEPTION</b> 	<ul style="list-style-type: none"> <li>▪ Computer vision</li> <li>▪ Natural language processing</li> </ul>
<b>DATA CLASSIFICATION</b> 	<ul style="list-style-type: none"> <li>▪ Text</li> <li>▪ Images</li> <li>▪ Objects</li> <li>▪ Radar signals</li> </ul>
<b>PREDICTION</b> 	<ul style="list-style-type: none"> <li>▪ Predict future behaviour</li> <li>▪ Make recommendations</li> </ul>
<b>ANOMALY DETECTION</b> 	<ul style="list-style-type: none"> <li>▪ Identify new malware</li> <li>▪ Identify new behaviour</li> </ul>
<b>OPTIMIZATION</b> 	<ul style="list-style-type: none"> <li>▪ Improve efficiency</li> </ul>
<b>CREATIVE DATA GENERATION</b> 	<ul style="list-style-type: none"> <li>▪ Create video</li> <li>▪ Create audio</li> <li>▪ Deep fakes</li> </ul>

**Figure 2.2.** The benefits of machine learning

*Opportunities and challenges of machine learning: A one-trick pony?*

The key strength and the key weakness of machine learning both lie in its ability to process data.

Machine learning is good at finding connections between data, which makes it a powerful tool for automating any tasks that require advanced pattern recognition. These can range from machine perception tasks to data management tasks (see figure 2.2).<sup>32</sup> Machine perception tasks can include computer vision (recognizing objects, people or situations) and natural language processing (voice and speech recognition) and any other type of signal recognition (acoustic or electromagnetic signatures). Data management tasks can include data classification, predictive analysis, anomaly detection, synthetic data manipulation and generation, and optimization. From this standpoint the possibilities that machine learning offers in the military realm are wide ranging. It could improve the perceptual capabilities of virtually any type of military system and increase the possibility that a military system can be made to operate autonomously in a dynamic and unstructured environment. It could also improve the performance of existing data management systems that the armed forces use for intelligence,

<sup>32</sup> Hagström, M., ‘Military applications of machine learning and autonomous systems’, ed. Boulanin (note 7), pp. 119–27; and Scharre, P and Horowitz, M. C., *Artificial Intelligence: What Every Policymaker Needs to Know* (Center for New American Security: Washington DC, June 2018).

surveillance and reconnaissance (ISR), logistics, battle management, training and simulation. Cybersecurity is also an area where machine learning finds obvious applications such as pattern recognition, spoofing and forensics.

Machine learning's relationship to data is also its vulnerability. Machine learning algorithms are well suited to tasks that involve processing large volumes of unstructured data, but they are only as good as the data on which they are trained.<sup>33</sup> To be taught, a machine learning system needs to be provided with large volumes of real world examples or training data, as well as rules about the data relationships. In order to recognize a type of object such as a car, a bus or a dog in an image, a computer vision system may need to be trained with millions of pictures of that type of object. The quality of the data on which the systems are trained is equally important. If the training data set is not representative, then the system may fail, may perform poorly, or may misinform human decisions and actions by reinforcing existing human biases or creating new ones.<sup>34</sup> This poses a major challenge for the development of machine learning applications in the military context. Machine learning capabilities are useless for tasks that do not generate data with which the systems can operate.

The fact that machines learn by finding a statistical relationship within data also makes their functioning opaque and their behaviour potentially unpredictable. Unlike traditional hand-coded AI systems that work according to clear rules and logic, machine learning systems operate like a black box. The input and the output of such a system are observable, but the computational process leading from one to the other is difficult for the system's designers to understand. It is particularly difficult for engineers to understand what such a system has learned and hence how it might react to input data that is different from that used during the training phase.<sup>35</sup> This creates a problem of predictability: a machine learning system might fail in ways that were unthinkable to its designers because they do not have a full understanding of its inner working. In the context of weapon systems, this unpredictability could have dramatic consequences.

It is also worth stressing that, while AI systems trained with machine learning may outperform humans for many tasks, they still lack what humans understand as basic common sense.<sup>36</sup> Computer vision systems, for instance, do not perceive a pattern at an abstract level, as a human would. They just see a correlation between a group of pixels. One study has demonstrated that variations in an image that are imperceptible to the human eye could cause an image-recognition system to completely mislabel the object or people in the image (e.g. mistaking a lion for a

<sup>33</sup> Gershgorin (note 31); and Hao, K., 'We analyzed 16,225 papers to figure out where AI is headed next', *MIT Technology Review*, 25 Jan. 2019.

<sup>34</sup> Knight, W. and Hao, K., 'Never mind killer robots—here are six real AI dangers to watch out for in 2019', *MIT Technology Review*, 7 Jan. 2019; and Hao, K., 'This is how AI bias really happens—and why it's so hard to fix', *MIT Technology Review*, 4 Feb. 2019.

<sup>35</sup> Righetti, L., 'Emerging technology and future autonomous systems', International Committee of the Red Cross (ICRC), *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons*, Expert meeting, Versoix, Switzerland, 15–16 Mar. 2016 (ICRC: Geneva, Aug. 2016), pp. 36–39.

<sup>36</sup> Mitchell, M., *Artificial Intelligence: A Guide for Thinking Humans* (Farrar, Straus and Giroux: New York, 2019).



car or a building).<sup>37</sup> Another study has demonstrated that it is easy to produce images that are completely unrecognizable to humans but that computer vision software identifies as a recognizable object with over 99 per cent confidence.<sup>38</sup> In other words, machine learning systems can easily be fooled or they may fail in unpredictable ways according to human standards, which is a major weakness when it comes to military applications such as automated target recognition.

Thus, recent advances in machine learning have created important opportunities for the development of highly efficient military applications of AI, but it remains, in many regards, an immature technology. There are many technical challenges associated with the use of machine learning methods that could make their adoption in a military context problematic, particularly in the case of systems and applications that are safety critical (e.g. cars, aeroplanes and weapons) or for deployment in combat operations. The fundamental question that developers, users and regulators need to resolve is how to ensure the responsible adoption and use of this technology through procedures such as testing, training of operators and regulation.

### **Autonomy: A key by-product of the AI renaissance**

#### *What is autonomy?*

As a technology area, autonomy is related to but distinct from AI. While machine learning can be depicted as a key ingredient of the current renaissance of AI and the associated hype, autonomy can be portrayed as one of its key by-products. Autonomous systems ranging from virtual assistants such as Amazon's Alexa or Apple's Siri via self-driving cars and auto-piloted unmanned aerial vehicles (UAVs) to autonomous weapons are among the most debated technological developments derived from the current AI renaissance and they receive the highest level of media attention.<sup>39</sup>

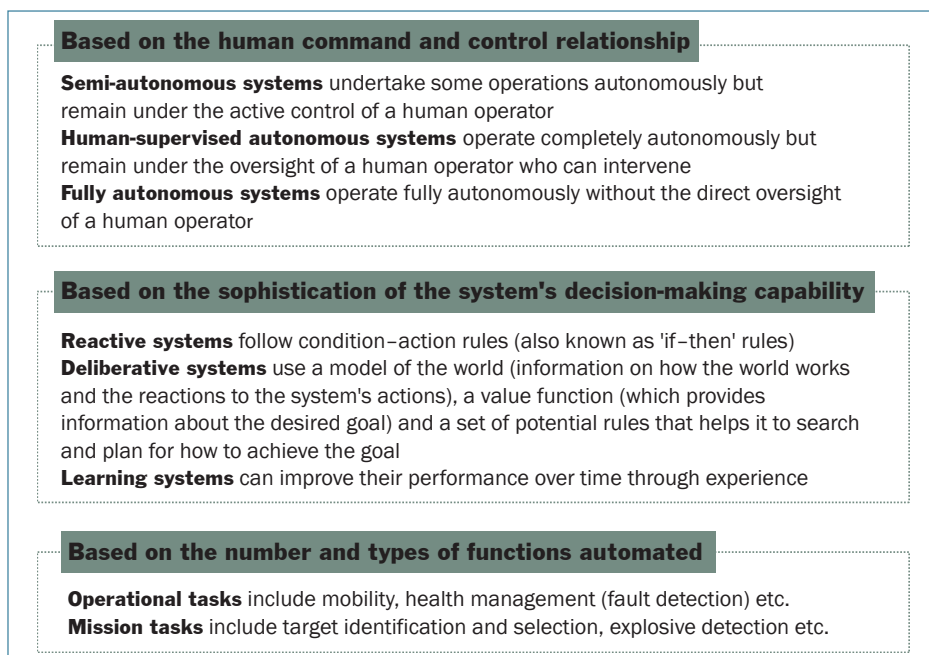
Autonomy or 'machine autonomy' can be defined as the ability of a machine to execute a task or tasks without human input, using interactions of computer programming with the environment.<sup>40</sup> An autonomous system is, by extension, usually understood as a system—whether hardware or software—that, once activated, can perform some tasks or functions on its own. Autonomous systems can be further divided between systems that use autonomy at rest (e.g.

<sup>37</sup> Szegedy, C. et al., 'Intriguing properties of neural networks', arXiv, 1312.6199, version 4, 19 Feb. 2014.

<sup>38</sup> Nguyen, A., Yosinski, J. and Clune J., 'Deep neural networks are easily fooled: High confidence predictions for unrecognizable images', *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, Proceedings, 7–12 June 2015 (Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, 2015), pp. 427–36.

<sup>39</sup> Hao, K., 'One day your voice will control all your gadgets, and they will control you', *MIT Technology Review*, 11 Jan. 2019; 'Autonomous weapon and the new laws of war', *The Economist*, 17 Jan. 2019; and Salesky, B., 'A decade after DARPA: Our view on the state of the art in self-driving cars', Medium, 16 Oct. 2017.

<sup>40</sup> This definition is based on one previously proposed by Andrew Williams. Williams, A., 'Defining autonomy in systems: Challenges and solutions', eds A. P. Williams and P. D. Scharre, *Autonomous Systems: Issues for Defence Policymakers* (NATO Headquarters Supreme Allied Commander Transformation: Norfolk, VA, 2015), pp. 27–62.



**Figure 2.3.** Approaches to the definition and categorization of autonomous systems

Source: Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017). Reproduced from Boulanin, V., 'Artificial intelligence: a primer', ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, *Euro-Atlantic Perspectives* (SIPRI: Stockholm, May 2019), pp. 13–25, figure 2.1.

AI assistants) and those that use autonomy in motion (e.g. self-driving cars and auto-piloted UAVs).

It should be stressed that the concept of autonomy and autonomous systems can mean different things to different people, primarily because autonomy is a relative notion that can be interpreted in several ways. The level of autonomy of a system can be analysed from three different and independent perspectives (see figure 2.3): (a) based on the extent to which humans are involved in the execution of the task carried out by the system; (b) based on the extent to which the system can exercise control over its own behaviour and deal with uncertainties in its operating environment; and (c) based on the number and types of functions that are automated.<sup>41</sup>

This definitional uncertainty appears clearly in public debate on autonomous systems: journalists, experts and policymakers sometimes use different terminology and metrics to talk about autonomy. Some people, for instance, make a distinction between automatic, automated and autonomous systems, while others use these terms interchangeably (see box 2.1). The distinction between the three terms can be conceptually useful to describe in layman's

<sup>41</sup> As identified by Scharre, P., 'The opportunity and challenge of autonomous systems', eds Williams and Scharre (note 40), pp. 3–26. See also Thrun, S., 'Toward a framework for human–robot interaction', *Human–Computer Interaction*, vol. 19, nos 1–2 (June 2004), pp. 9–24; and Boulanin and Verbruggen (note 12).

**Box 2.1. Automatic, automated, autonomous: The relationship between automation, autonomy and machine learning**

**Automatic**

The label ‘automatic’ is usually reserved for systems that mechanically respond to sensory input and step through predefined procedures, and whose functioning cannot accommodate uncertainties in the operating environment. An example of this is a robotic arm used in the manufacturing industry.

**Automated versus autonomous**

Machines that can cope with variations in their environment and exercise control over their actions can be described as either automated or autonomous. What distinguishes an automated system from an autonomous system is a contentious issue.

Some experts, such as David Mindell, see the difference in terms of the degree of self-governance. They view autonomous systems merely as more complex and intelligent forms of automated systems.<sup>a</sup>

Others see value in making a clear distinction between the two concepts. A report from the US Defense Science Board presents an automated system as a system that is governed by ‘prescriptive rules that permit no deviations’.<sup>b</sup> This means that the system logically follows a pre-defined set of rules in order to provide an outcome; its output is predictable if the set of rules under which it operates is known. In contrast, an autonomous system is able to ‘independently compose and select among different courses of action to accomplish goals based on its knowledge and understanding of the world, itself, and the situation’.<sup>c</sup>

**Automation, autonomy and machine learning**

What then is the relationship between automation, autonomy and machine learning? Autonomy can be described as a complex form of automation that allows a machine to execute a task or tasks using explicit or implicit programming rules and without human intervention. Automation is a feature of machine learning in two ways. First, it is characteristic of how machine learning works. In technical terms, machine learning involves ‘automatic reparameterization and partial reprogramming’.<sup>d</sup> Second, machine learning can be used to design systems that work in an autonomous way, that is, without human control or intervention.

<sup>a</sup> Mindell, D. A., *Our Robots, Ourselves; Robotics and the Myths of Autonomy* (Viking: New York, 2015), p. 12.





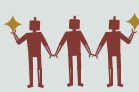
<sup>b</sup> US Department of Defense (DOD), Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy (DOD: Washington, DC, June 2016)*, p. 4.

<sup>c</sup> US Department of Defense (note b), p. 4.

<sup>d</sup> Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017), box 4.1.

Source: Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017). Adapted from Boulanin, V., ‘Artificial intelligence: a primer’, ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, *Euro-Atlantic Perspectives* (SIPRI: Stockholm, May 2019), pp. 13–25, box 2.3.

terms the sophistication of a system’s decision-making capability. Automatic and automated systems are generally used to describe ‘reactive systems’ (as described in figure 2.3), while autonomous systems use deliberative or learning-based decision-making algorithm. In practice, however, it is difficult to determine to which of the three categories a system belongs. Moreover, the definitions of and boundaries between these three categories remain contested within and between the expert communities.

KEY ADVANTAGE OF AUTONOMY	WHAT IT MEANS	UTILITY FOR MISSIONS
 <p><b>SPEED</b></p>	<p><b>Faster implementation of the OODA loop</b></p>	<ul style="list-style-type: none"> <li>▪ Air defence</li> <li>▪ Cyber-defence</li> <li>▪ Electronic warfare</li> </ul>
 <p><b>AGILITY</b></p>	<p><b>Reduced reliance on command and control</b></p>	<ul style="list-style-type: none"> <li>▪ ISR</li> <li>▪ Cyberwarfare</li> <li>▪ Electronic warfare</li> <li>▪ Submarine and mine hunting</li> <li>▪ Logistics operations</li> </ul>
 <p><b>PERSISTENCE</b></p>	<p><b>Constant performance of unmanned systems for dull, dirty and dangerous missions</b></p>	<ul style="list-style-type: none"> <li>▪ Logistics operations in enemy territory</li> <li>▪ Casualty evacuation</li> <li>▪ Countermine operations</li> <li>▪ Long-range ISR</li> <li>▪ Air defence</li> </ul>
 <p><b>REACH</b></p>	<p><b>Access to communications-denied environments</b></p>	<ul style="list-style-type: none"> <li>▪ Casualty evacuation</li> <li>▪ Submarine and mine hunting</li> <li>▪ ISR in A2/AD environments</li> <li>▪ Strikes in A2/AD environments</li> <li>▪ Logistics operations in A2/AD environments</li> </ul>
 <p><b>COORDINATION</b></p>	<p><b>Ability to coordinate weapon systems in structured and strategic way</b></p>	<ul style="list-style-type: none"> <li>▪ ISR in complex and cluttered environments</li> <li>▪ Combat operations in A2/AD environments</li> <li>▪ Force protection</li> </ul>

**Figure 2.4.** Benefits of autonomy

A2/AD = anti-access/area-denial, ISR = intelligence, surveillance and reconnaissance, OODA = observe, orient, decide and act.

Sources: Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017); and US Department of Defense (DOD), Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy* (DOD: Washington, DC, June 2016).

For the analytical purposes of this report, autonomy is understood a complex form of automation that allows a machine to execute a task or tasks using explicit or implicit programming rules and without human intervention.

*Opportunities and challenges of autonomy*

Advances in autonomy are generating great expectations in both civilian and military spheres as they enhance the usefulness and reliability of computer and robotics systems, which in turn could generate significant economic and operational benefits. Companies, governmental institutions and armed forces alike could achieve greater manpower efficiency by increasing their reliance on robotic systems.<sup>42</sup> Advances in autonomy could also allow them to overcome a

<sup>42</sup> US Department of Defense (DOD), Defense Science Board, *The Role of Autonomy in DoD Systems*, Task Force Report (DOD: Washington, DC, July 2012); and Scharre, P., *Robotics on the Battlefield*, part II, *The Coming Swarm* (Center for New American Security: Washington, DC, Oct. 2014).

number of operational challenges associated with manned operations or the use of teleoperated systems (see figure 2.4).<sup>43</sup>

Autonomy in a robotic system can execute some tasks much faster than any human or human-operated robot ever could, which for the armed forces is particularly attractive for time-critical missions or tasks such as air defence, air-to-air combat or cyber-defence. Autonomy can make robotic systems far more agile from a command-and-control perspective and reduce the need to maintain a constant communications link—that can be jammed—between the robot and the military command as is currently necessary with remotely controlled systems. It can also allow the armed forces to reduce the number of human operators and analysts needed to oversee the system and process information. Autonomy is also useful for so-called dull, dirty and dangerous (3D) missions as it removes limitations (e.g. fatigue, boredom, hunger or fear) that may make human performance deteriorate over time. Autonomy also gives systems greater reach. It grants access to operational theatres that were previously inaccessible to remote-controlled systems or too risky for manned operations. These include areas protected by anti-access/area-denial (A2/AD) systems and areas with harsh operating environments for humans (and where communication is limited), such as deep water, the Arctic and, potentially, outer space. Finally, autonomy also provides new opportunities for collaborative operations as it permits weapon systems to operate in large groups, or ‘swarms’, in a much more coordinated, structured and strategic way than if each were individually controlled by a human operator.

However, the advances in autonomy are raising a wide spectrum of ethical, legal and security concerns, which apply to both civilian and military applications.<sup>44</sup> From an ethical standpoint, the development of autonomy in safety-critical systems (e.g. cars or weapons) raises the difficult question of whether, and to what extent, autonomous systems should be trusted to operate outside direct human control and supervision. The ongoing intergovernmental discussion on LAWS in the CCW framework and the debate around car accidents involving semi-autonomous cars have shown that there is no simple answer to that question.<sup>45</sup> The question of the balance between autonomy and human control also has profound and complex legal implications, particularly with regards to the attribution of individual criminal responsibility.<sup>46</sup> Advances in autonomy also create new security risks. In addition to the increasing vulnerability to cyberattacks, the limitations of existing autonomous systems in terms of perceptual and decision-

<sup>43</sup> eds Williams and Scharre (note 40).

<sup>44</sup> Cath, C. et al., ‘Governing artificial intelligence: Ethical, legal and technical opportunities and challenges’, *Philosophical Transactions of the Royal Society A*, vol. 376, no. 2133 (Nov. 2018).

<sup>45</sup> Boulanin, V., ‘Mapping the debate on LAWS at the CCW: Taking stock and moving forward’, EU Non-proliferation Paper no. 49, EU Non-proliferation Consortium, Mar. 2016; and Bhuyian, J., ‘Uber’s semi-autonomous cars detected the pedestrian six seconds before the fatal crash, a federal agency says’, *Recode*, 25 May 2018.

<sup>46</sup> Docherty, B., *Mind the Gap: The Lack of Accountability for Killer Robots* (Human Rights Watch: New York, 2015).

making intelligence could easily be exploited by a malevolent actor who could defeat a system by simply spoofing the sensors or control systems.<sup>47</sup>

More broadly, the increasing adoption of and reliance on autonomous systems is bound to ignite organizational changes.<sup>48</sup> Among other effects, it will change the way that companies, governmental agencies and the armed forces operate. Taking the case of an air force as an example, replacing manned combat aircraft with autonomous unmanned aerial systems will necessitate a change in the way that personnel are selected, trained and meant to operate—with control moving from a pilot to a remote operator and then to a systems supervisor—which in turn could cause major changes in professional culture.<sup>49</sup>

The key takeaway from this overview is that the field of AI is going through a renaissance as a result of a breakthrough in the area of machine learning. Advances in machine learning have unlocked numerous possibilities, including that of creating increasingly autonomous systems. However, developers and users alike have only just begun to work out how to make the best use of it, and also how they should not use it.<sup>50</sup>

## II. AI and nuclear weapon systems: Past, present and future

AI technology is still immature and thus any attempt to forecast its impact will remain speculative. Nonetheless, the key question of how recent advances in AI may have an impact on nuclear weapons can still be addressed. As a start, it is useful to recall that automation—the central feature of machine learning and autonomy (see box 2.1)—has been used in nuclear weapon systems for decades, in particular early-warning, command-and-control and delivery systems. This section describes how automation has been used so far and then discuss what might change with the adoption of machine learning capabilities and autonomous systems.

### **Past and present applications of automation in the nuclear domain**

Looking at how automation has been used helps to contextualize current developments and trends. This allows a better understanding of why the use of machine learning and autonomous systems may be an attractive prospect for nuclear-armed states.

<sup>47</sup> Versprille, A., 'Army still determining the best use for driverless vehicles', *National Defense*, June 2015; and Endsley, M. R., *Autonomous Horizons: System Autonomy in the Air Force—A Path to the Future*, vol. 1, *Human-Autonomy Teaming* (US Air Force, Office of the Chief Scientist: Washington, DC, 2015), p. 5.

<sup>48</sup> Wright, N., 'Three distinct AI challenges for the UN', *AI & Global Governance*, United Nations University, Centre for Policy Research, 7 Dec. 2018.

<sup>49</sup> For detailed discussion see Boulanin and Verbruggen (note 12), pp. 69–73; and Bronk, J., 'The impact of unmanned combat aerial vehicles on strategic stability', ed. Boulanin (note 7), pp. 99–104.

<sup>50</sup> Vogt, H., 'Artificial intelligence rules more of your life. Who rules AI?', *Wall Street Journal*, 13 Mar. 2018; and Hao, K., 'Why AI is a threat to democracy—and what we can do to stop it', *MIT Technology Review*, 26 Feb. 2019.

*How AI became part of the nuclear deterrence architecture during the cold war*

A fundamental pillar of nuclear deterrence during the cold war was mutually assured destruction between the USA and the USSR: the concept is based on the logic that, as each side maintains nuclear forces that could survive a first strike and inflict retaliatory damage that the aggressor would consider unacceptable, nuclear war became irrational.<sup>51</sup> It is often credited with reducing the likelihood of a nuclear first strike and contributing to strategic stability.<sup>52</sup> This imperative bolstered the development of highly survivable nuclear triads of strategic bombers, ICBMs and SLBMs, which were protected by hardened silos or their increasing mobility. It also required nuclear-armed states to develop early-warning and agile command-and-control systems that would allow the strategic command to identify a threat and an adequate response within a limited time frame—from minutes to a few hours. These states were thus also compelled to develop elaborate and resilient systems for communications, control and response.<sup>53</sup>

The USA and the USSR, which devoted the most attention and resources to maintaining a retaliatory nuclear capability during the cold war, saw already in the 1960s that greater automation could have a role to play in nuclear weapon systems.<sup>54</sup> In command and control, they pursued the development of automated systems that could assist with a number of tasks including logistical planning for transmission of launch orders, in-flight refuelling of bombers, and missile targeting and guidance. As they both developed ‘launch-on-warning’ postures—in which authorities may initiate a nuclear second strike based on the early detection of the enemy first strike—the USA and the USSR also saw the need to develop automated and pre- or semi-automated systems for early-warning systems.<sup>55</sup> Automating the detection of threats was perceived as a way to provide radar information more quickly to decision makers and thereby give them more time to decide about whether to launch a counterstrike.

Early on, nuclear decision makers identified the appeal of automation for nuclear deterrence, but also its limitations.<sup>56</sup> Automated detection technology during the cold war suffered from numerous problems. While electronics and information and communications technology (ICT) made great strides between the 1960s and the end of the 1980s, there were significant inadequacies in enabling technologies such as sensors and computer chips. Early-warning systems, on both the Soviet

<sup>51</sup> Brooks, L., ‘Can the United States and Russia reach a joint understanding of the components, prospects and possibilities of strategic stability?’, *Revitalizing Nuclear Arms Control and Non-Proliferation* (International Luxembourg Forum on Preventing Nuclear Catastrophe: Moscow, 2017), pp. 80–95, p. 82.

<sup>52</sup> Brodie, B., *Strategy in the Missile Age* (Rand Corporation: Santa Monica, CA, 1959), pp. 264–305.

<sup>53</sup> Geist and Lohn (note 18).

<sup>54</sup> Borrie, J., ‘Cold war lessons for automation in nuclear weapon systems’, ed. Boulanin (note 7), pp. 41–52; Yarynich, V. E., *C3: Nuclear Command, Control, Cooperation* (Center for Defense Information: Washington, DC, May 2003), pp. 137–49, 193–202; and Blair, B. G., *The Logic of Accidental Nuclear War* (Brookings Institution: Washington, DC, 1993).

<sup>55</sup> Blair, B. G., ‘Loose cannons: The president and US nuclear posture’, *Bulletin of the Atomic Scientists*, vol. 76, no. 1 (Jan. 2020), pp. 14–26, p. 15.

<sup>56</sup> Yarynich (note 54); Blair (note 54); and Hoffman, D. E., *The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy* (Anchor Books: New York, 2009).

**Box 2.2. Historical cases of false alarms in early warning systems****The 1979 training tape incident<sup>a</sup>**

On 9 November 1979 computers from three different United States nuclear command posts showed that a massive Soviet strike was underway. Senior officers from the three posts immediately convened a threat-assessment conference. In the meantime, the intercontinental ballistic missile (ICBM) force received a preliminary warning and the entire North American air defence interceptor force was put on alert.

The officers reviewed the raw data from the satellites and radars spread around the country. The data showed that there was no sign that an attack was underway, so the alert was cancelled. It was later determined that the false alarm had been caused by a software simulation of a Soviet missile attack that had been ‘inexplicably transferred’ into the normal early-warning display at the command headquarters.

**The 1980 computer chip incident<sup>b</sup>**

On 3 June 1980 the US early-warning systems reported that that Soviet Union had launched a nuclear strike. However, the systems displayed inconsistent and changing attack patterns. The number of detected missiles kept changing. The ICBM force and bomber crews were nonetheless put on alert while a threat-assessment conference was underway.

Here again the raw data showed that this was a false alarm. An investigation later showed that it had been caused by the failure of a single computer chip.

**The 1983 Petrov incident<sup>c</sup>**

On 26 September 1983, at about 00.30, the Soviet early-warning system Oko detected that a US attack was underway. The computer display showed that five missiles originating from the USA were heading towards the USSR. Within minutes, Lieutenant Colonel Stanislav Petrov, who was in charge of supervising the operation of early-warning system, and his team cross-checked the intelligence information, but they could not determine with certainty whether it was a false alarm.

Petrov nonetheless decided to report the incident as a false alarm to his superiors. He reportedly trusted his gut instinct. He knew the technical limitation of the systems and it seemed to him that the US attack would not make sense: ‘when people start a war, they don’t start it with only five missiles. You can do little damage with just five missiles.’d He was right. The satellite had wrongly identified the missiles. Petrov’s decision to rely on his common sense was later hailed as having saved the world from a nuclear war.

<sup>a</sup> Burr, W. ‘The 3 AM phone call: False warning of Soviet missile attacks during 1979–1980 led to alert actions for US strategic forces’, National Security Archive Electronic Briefing Book no. 371, 1 Mar. 2012; Fordon, G., ‘False alarms in the nuclear age’, Nova, Public Broadcasting Service, 5 Nov. 2001; and Sagan, S. D., *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons* (Princeton University Press: Princeton, NJ, 1993), pp. 228–29, 238.

<sup>b</sup> Fordon (note a).

<sup>c</sup> Aksekov, P., ‘Stanislav Petrov: The man who may have saved the world’, BBC, 26 Sep. 2013; Hoffman, D., ‘“I had a funny feeling in my gut”’, *Washington Post*, 10 Feb. 1999, p. A10; and Likhanov, D., [40 minutes before World War III], *Rossiiskaya Gazeta*, 1 Sep. 2017 (in Russian).

<sup>d</sup> (note a).

and US sides, were marked by crude perception intelligence and suffered from numerous faults and false alarms.

Due to these limitations, Soviet and US nuclear policymakers were reluctant to hand over higher-order assessments and decision-making capabilities to automated systems. Humans had to remain ‘in the loop’ in nuclear command and control in order to verify and analyse the information provided by the systems and



deal with technical problems as they arose and, more importantly, make nuclear launch decisions. There were several cases in which a machine's errors could have led to the use of nuclear weapons if human decision makers had not intervened (see box 2.2).

The only circumstances in which automation of command and control was considered included situations in which the decision makers would physically be unable to make assessments and decisions. There were two such scenarios: (a) active operation of a missile defence system—these were generally designed to include a 'pre-authorized' mode for situations in which humans would not have time to respond; and (b) the exceptional case of a decapitating attack on the political and military leadership, which would annihilate their ability to take decision on retaliation.

To deal with this second possibility, in 1985 the USSR deployed an automated command-and-control system known in Russian sources as Perimetr and as the Dead Hand in the Western literature. The purpose of this system, whose existence was classified until the end of the cold war, has been widely debated (see box 2.3).<sup>57</sup> Reportedly, the system was intended to guarantee mass retaliation to a US attack and thereby keep an overeager Soviet military or civilian leader from premature launch during a crisis.<sup>58</sup> The USA—out of fear of malfunction that could lead to nuclear catastrophe—did not develop an equivalent.<sup>59</sup> However, it did develop a signal rocket system equivalent to the one used by the Perimetr system, which could be used to disseminate pre-recorded launch orders to nuclear missiles if the usual means of communication were down.<sup>60</sup>

### *How AI is currently used in nuclear weapon systems*

The field of nuclear weapons development has historically been conservative and has relied on aged infrastructure.<sup>61</sup> For safety and security reasons, it has been slow to integrate some major ICT developments as they could introduce new vulnerabilities or affect reliability. This is particularly the case for nuclear command and control, which continues to rely on old but reliable cold war-era technology. As one example, a 2016 US Government Accountability Office report noted that the US armed forces still used 8-inch floppy disks to coordinate nuclear force operations.<sup>62</sup> When asked about it, the US Department of Defense (DOD) explained that 'The system remains in use because, in short, it still works'.<sup>63</sup>

<sup>57</sup> Hoffman (note 56); Borrie (note 54); and Topychkanov, P., 'Autonomy in Russian nuclear forces', ed. Boulanin (note 7), pp. 68–75.

<sup>58</sup> Hoffman (note 56).

<sup>59</sup> Yarynich (note 54), pp. 181–202.

<sup>60</sup> 'Emergency Rocket Communications System', National Museum of the United States Air Force, 27 May 2015.

<sup>61</sup> Futter, A., 'The double-edged sword: US nuclear command and control modernization', *Bulletin of the Atomic Scientists*, 29 June 2016.

<sup>62</sup> US Government Accountability Office (GAO), *Information Technology: Federal Agencies Need to Address Aging Legacy Systems*, GAO-16-468 (GAO: Washington, DC, May 2016), p. 60.

<sup>63</sup> 'US nuclear force still uses floppy disks', BBC, 26 May 2016.

**Box 2.3. Dead Hand and Perimetr**

In 1985 the Soviet Union deployed a semi-automated retaliation system, known in the Western literature as the Dead Hand and as Perimetr in Russian sources. The system was designed to guarantee a Soviet response to a nuclear strike from the United States. To launch any retaliatory strike, the system would have had to run through an if-then propositional formula:

IF the system was turned on

AND

IF a nuclear weapon had hit Soviet soil

AND

IF there was no communications link to the war room of the Soviet General Staff,

THEN the system would initiate the process of launching a retaliatory strike.

If the communications link remained, then the system would assume that high-level decision makers who could order an attack were still alive and remained in full control of the decision. It would then do nothing. If the link was dead, then the system would infer that these decision makers were no longer alive and transfer the decision authority to whoever was then manning the system deep inside a protected bunker.

*Source: Hoffman, D. E., *The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy* (Anchor Books: New York, 2009).*

However, in more recent years, there are indications that countries such as Russia and the USA have taken steps to modernize their nuclear forces, including nuclear command-and-control systems, by retiring some of these legacy systems and adopting state-of-the-art digital technologies.<sup>64</sup> In the case of Russia, there have been some reports that Perimetr has been upgraded, but it is impossible to determine to what extent these upgrades involve AI technology.<sup>65</sup> The US DOD also reportedly replaced the floppy disks with solid state storage systems in 2019, suggesting that upgrades have begun (although public information on nuclear command and control remains limited).<sup>66</sup> It is also clear that other nuclear-armed states intend to use AI technology to further develop or maintain their nuclear capability (see chapter 3). This raises the question of what the impact will be.

<sup>64</sup> Deptula, D. and LaPlante, W. A. with Haddick, R., *Modernizing US Nuclear Command, Control, and Communications* (Mitchell Institute for Aerospace Studies: Arlington, VA, Feb. 2019); Bennet, M., 'Projected costs of U.S. nuclear forces, 2019 to 2028', US Congressional Budget Office, 24 Jan. 2019; Reif, K., 'US nuclear modernization programs', Arms Control Association, Aug. 2018; Woolf, A. F., *US Strategic Nuclear Forces: Background, Developments, and Issues*, Congressional Research Service (CRS) Report for Congress RL33640 (US Congress, CRS: Washington, DC, 3 Sep. 2019); Podvig, P., 'Russia's current nuclear modernization and arms control', *Journal for Peace and Nuclear Disarmament*, vol. 1, no. 2 (2018), pp. 256–67; and Woolf, A. F., *Russia's Nuclear Weapons: Doctrine, Forces, and Modernization*, Congressional Research Service (CRS) Report for Congress R45861 (US Congress, CRS: Washington, DC, 5 Aug. 2019).

<sup>65</sup> Valagin, A., [Assured retaliation: How the Russian 'Perimetr' system works], *Rossiiskaya Gazeta*, 22 Jan. 2014 (in Russian).

<sup>66</sup> Mizokami, K., 'US Air Force finally ditches 8-inch floppy disks', *Popular Mechanics*, 22 Oct. 2019.

## Potential applications of machine learning and autonomy in the nuclear deterrence architecture<sup>67</sup>

Since automation has been part of the nuclear deterrence architecture for decades, the question of whether recent advances in AI will have a transformative impact deserves exploration. The remainder of this section reviews how recent advances in machine learning and autonomy may be applied in the context of nuclear forces. It also discusses the extent to which these potential applications may differ from historical uses of automation. In doing so, it looks at four key areas of the nuclear deterrence architecture: early warning and ISR; command and control; nuclear weapon delivery systems; and non-nuclear operations (see figure 2.5).

### *Early warning and intelligence, surveillance and reconnaissance*

Machine learning and autonomy hold major promise for early warning and ISR. The potential of machine learning in this area is derived from three abilities.

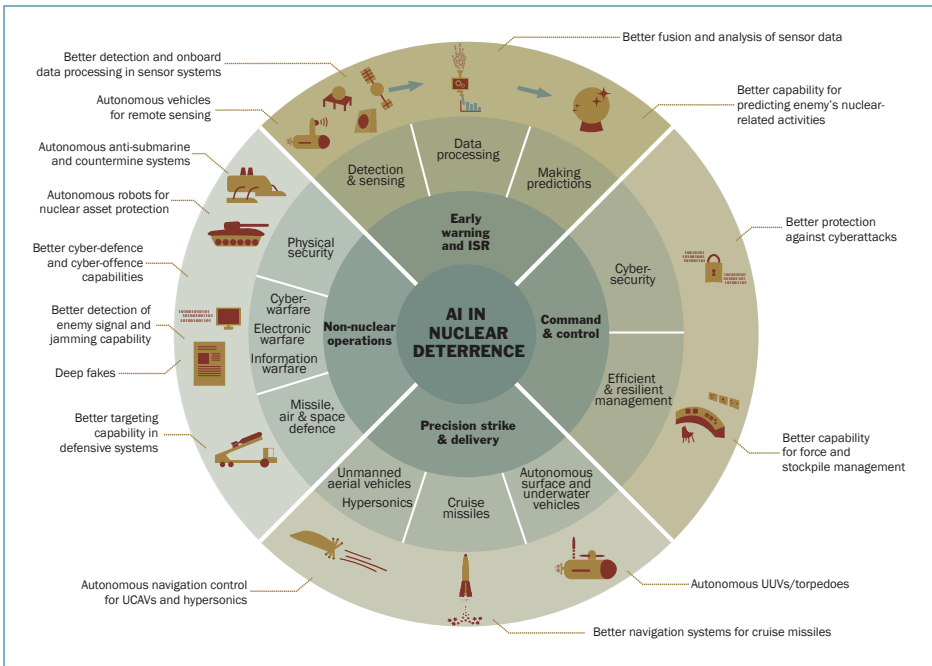
1. *Improving capabilities of early-warning and ISR systems.* Machine learning can be used to give any type of ISR system more perceptual intelligence. One foreseeable development would be a mobile ISR platform (e.g. a surveillance UAV or UUV) that could process data on-board and identify by itself not only signals or objects, but also situations of interest, such as an unusual movement of troops, weapons or equipment
2. *Searching through and deciphering large sets of intelligence data.* Machine learning can be used to find correlations in large and potentially heterogeneous sets of intelligence data.
3. *Making predictions.* Data-processing capability can be used to help military commanders to predict developments related to nuclear weapons, including the possible production, commissioning, deployment and use of nuclear forces by an adversary.<sup>68</sup> The cross analysis of intelligence data using machine learning algorithms could help the armed forces to identify more quickly and reliably if preparations for a nuclear attack are or may be underway.

In sum, machine learning could give a human military commander better situational awareness and potentially more time to make decisions.

Autonomy and autonomous systems, in contrast, have the primary value of improving the remote-sensing capabilities of nuclear-armed states—be it for early-warning or ISR missions. The main advantages of autonomous systems compared to remotely controlled and manned systems are that they can achieve greater reach, greater persistence and greater mass. They can be deployed in such operational theatres as deep water or areas protected by A2/AD systems that are dangerous for humans. They can conduct extended mission over days or, in

<sup>67</sup> This subsection is largely based on Boulanin, V., 'The future of machine learning and autonomy in nuclear weapon systems', ed. Boulanin (note 7), pp. 53–62.

<sup>68</sup> US Department of Defense (DOD), Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy* (DOD: Washington, DC, June 2016).



**Figure 2.5.** Foreseeable applications of AI in nuclear deterrence

AI = artificial intelligence, ISR = intelligence, surveillance and reconnaissance, UCAV = unmanned combat aerial vehicle, UUV = unmanned underwater vehicle.

the case of underwater systems, even months. Moreover, they can potentially be deployed in great number as they can be relatively inexpensive.<sup>69</sup>

These attributes are particularly attractive in the conduct of nuclear-related ISR operations, particularly airborne and space surveillance and submarine reconnaissance. Among the many types of platform that could be used for submarine reconnaissance missions are autonomous unmanned surface vehicles (USVs), UUVs and UAVs. Most nuclear-armed states are developing such platforms (as described in chapter 3).

### *Command and control*

In the near term, recent progress in machine learning and autonomy is unlikely to have a major transformative impact on nuclear command-and-control systems. There are two reasons for this. First, command-and-control systems already rely on a great degree of automation. Second, the types of algorithm underlying machine learning-driven applications and complex autonomous systems remain too unpredictable due to the problems of transparency and explainability. Nuclear command-and-control systems are too safety critical to be left to algorithms

<sup>69</sup> On the benefits of autonomy see Boulanin and Verbruggen (note 12). On A2/AD systems see Simon, L., 'Demystifying the A2/AD buzz', War on the Rocks, 4 Jan. 2017.

that engineers and operators cannot fully understand.<sup>70</sup> Moreover, relatively traditional rule-based algorithms would be sufficient to further automate command and control. There seems to be a general agreement among nuclear-weapon experts that machine learning and autonomy should not be integrated into nuclear command and control, even if technological developments would permit it.<sup>71</sup> A notable exception occurred in August 2019 when two US experts called on the USA to develop fully automated machine learning-based nuclear command and control.<sup>72</sup>

Even if they do not revolutionize nuclear command, control and communications (NC3) systems, however, advances in machine learning and autonomous systems could bring some qualitative improvements in the nuclear command-and-control architecture. They could be used to enhance protection against cyberattacks and jamming attacks. Machine learning could also help planners to more efficiently manage their forces, including their human resources. Similarly, autonomous systems could be used to enhance the resilience of the communications architecture. Long-endurance UAVs could, for example, be used to replace signal rockets in forming an alternative airborne communications network in situations where satellite communication is impossible.

### *Nuclear weapon delivery*

Advances in machine learning and autonomy are likely to have an impact on nuclear weapon delivery in different ways. In the case of machine learning, the impact is likely to result primarily in a qualitative improvement in delivery systems. Machine learning could be used to make nuclear delivery systems capable of navigating to their target more autonomously and precisely, with less reliance on humans setting navigation and guidance parameters. A number of countries are currently exploring the use of machine learning to develop control systems for hypersonic vehicles.<sup>73</sup> It could also make them more resilient to countermeasures by adversary electronic warfare, since this would interrupt the communications or data link between vehicle and operator.

In the case of autonomy, systems such as UAVs, in particular unmanned combat aerial vehicles (UCAVs), and UUVs also have a role to play. Unmanned vehicles—whether remotely controlled or autonomous—can conduct much longer missions than their manned counterparts. This is particularly notable for unmanned aircraft, which can stay in flight for several days, particularly if in-flight refuelling or the use of solar power is possible. The endurance record for an unmanned aircraft

<sup>70</sup> Loss, R. and Johnson, J., ‘Will artificial intelligence imperil nuclear deterrence’, *War on the Rocks*, 19 Sep. 2019.

<sup>71</sup> Borrie (note 54); and Sauer, F., ‘Military application of artificial intelligence: Nuclear risk redux’, ed. Boulanin (note 7), pp. 84–90, p. 88.

<sup>72</sup> Lowther, A. and McGiffin, C., ‘America needs a “Dead Hand”’, *War on the Rocks*, 16 Aug. 2019. For comments and discussion see Field, M., ‘Strangelove redux: US expert propose having AI control nuclear weapons’, *Bulletin of the Atomic Scientists*, 30 Aug. 2019.

<sup>73</sup> Saalman, L., ‘China’s integration of neural networks into hypersonic glide vehicles’, ed. N. D. Wright, *AI, China, Russia, and the Global Order: Technological, Political, Global, and Creative Perspectives*, White Paper (US Department of Defense and Joint Chiefs of Staff: Washington, DC, Dec. 2018), pp. 153–60.

of 26 days was set by a solar-powered UAV from Airbus in 2018.<sup>74</sup> Increased endurance also means greater reach. An unmanned platform can cover a much larger area and, in the case of an underwater system, reach greater depths than a manned vehicle. The extended endurance of unmanned platforms potentially increases their ability to survive countermeasures and decreases policymakers' fear of a nuclear decapitation.<sup>75</sup>

The ability to alter or cancel a UAV or UUV mission also sets them apart from missiles and torpedoes and offers policymakers new tools for managing escalation in a crisis or conflict. The decision to send an unmanned bomber or spaceplane out on patrol is not equivalent to the decision to launch a one-way device such as a nuclear ICBM or torpedo (although some of these systems may be aborted after launch). Recoverability gives decision makers greater flexibility in that they would have more time to make a decision and, potentially, to recall the system.

The added value of autonomous systems lies, in other words, less in the degree of automation or autonomy than in the physical properties of the delivery platforms. In practice, ICBMs and SLBMs, once launched, already operate autonomously since they rely on automation to set their flight trajectory and navigate to their target. Once launched, manned bombers and submarines also operate autonomously from the command centre due to the risks of countermeasures and of communications being intercepted. While autonomy enhances the strategic value of robotics platforms, it is not an essential requirement. This is with the notable exception of long-range, deep-water systems, which cannot be operated remotely.

At least two nuclear-armed states are considering the possibility of using UAVs or UUVs for nuclear weapon delivery. In 2015 a Russian television report revealed that Russia was developing a large nuclear propulsion UUV, Poseidon (also known as Status-6).<sup>76</sup> The system, which has been described as both a long-range torpedo and an unmanned submarine, reportedly has a range of 10 000 kilometres and a speed of 56 knots and can descend to a depth of 1000 metres.<sup>77</sup> It will reportedly operate autonomously but, as explained above, that is primarily a requirement of its operating environment. The USA is also building a dual-capable bomber, the B-21 Raider, which would reportedly be 'optionally-manned'.<sup>78</sup> The USA has not specified whether it would be prepared to operate the bomber remotely while carrying nuclear weapons, but a 2013 US Air Force (USAF) report suggests that it is unlikely, stating that 'certain missions [for unmanned aircraft], such as nuclear strike, may not be technical feasible unless safeguards are developed and even

<sup>74</sup> Airbus, 'Airbus Zephyr Solar High Altitude Pseudo-Satellite flies for longer than any other aircraft during its successful maiden flight', 8 Aug. 2018.

<sup>75</sup> Horowitz, M. C., Scharre, P. and Velez-Green, A., 'A stable nuclear future? The impact of automation, autonomy, and artificial intelligence', arXiv, 1912.05291, version 2, 13 Dec. 2019.

<sup>76</sup> Oliphant, R., 'Secret Russian radioactive doomsday torpedo leaked on television', *Daily Telegraph*, 15 Nov. 2015. See also Hwang and Kim (note 2).

<sup>77</sup> Insinna, V., 'Russia's nuclear underwater drone is real and in the Nuclear Posture Review', *Defense News*, 12 Jan. 2018.

<sup>78</sup> Majumdar, D., 'USAF leader confirms manned decision for new bomber', *Flight International*, 23 Apr. 2013. See also Gates, R., US Secretary of Defense, 'Statement on department budget and efficiencies', US Department of Defense, 6 Jan. 2011.

then may not be considered'.<sup>79</sup> It is thus hard to imagine that the USA is currently considering the use of autonomously piloted UAVs for nuclear weapon delivery. That being said, the technology exists, as with China's work on the DF-ZF manoeuvrable hypersonic glide vehicle, a dual-capable prototype that will in practice fly autonomously.

### *Non-nuclear operations*

Both nuclear-armed and non-nuclear-armed states could use machine learning and autonomy in non-nuclear applications for strategic purposes. These include conventional high-precision strikes, missile, air and space defences, cyberwarfare, electronic warfare, information warfare and physical security.

*Conventional high-precision strikes.* Advances in machine learning and autonomy could be critical to the application of conventional force to an adversary's high-value assets, including nuclear forces and command-and-control infrastructure. On the one hand, machine learning could be used to increase the on-board intelligence of manned and unmanned combat aircraft and make them more capable in penetrating enemy defences. On the other hand, autonomy could give access to operational theatres that were previously inaccessible to remotely controlled unmanned systems or too risky for manned systems. Such theatres include A2/AD bubbles and areas where there are harsh operating environments for humans and where communications are limited, such as deep water and potentially outer space.

*Missile, air and space defences.* Machine learning methods could significantly improve the detecting, selecting, tracking, targeting and intercepting capabilities of defensive systems. Missile and air defence systems have relied on automation for decades.<sup>80</sup> Since the 1970s, air defence systems have been using an AI-based technology known as automatic target recognition (ATR) that can detect, track, prioritize and select incoming air threats more rapidly and more accurately than a human possibly could. However, the progress of the target-identification capabilities of these systems has been slow, particularly due to the difficulties associated with the development of target libraries (i.e. the database of target signatures that an ATR system uses to recognize its target). With traditional AI programming methods, the designers of an ATR system have to upload a large and representative sample of data about the target in all conceivable variations of its operating environment (i.e. background and weather conditions). This is a challenging task for many target types and operational situations.<sup>81</sup>

<sup>79</sup> US Air Force, *RPA Vector: Vision and Enabling Concepts, 2013–2038* (Headquarters US Air Force: Washington, DC, 17 Feb. 2014), p. 54.

<sup>80</sup> Mindell, D. A., 'Automation's finest hour: Radar and system integration in World War II', eds A. C. Hughes and T. P. Huges, *Systems, Experts and Computers: The Systems Approach in Management and Engineering, World War II and After* (MIT Press: Cambridge, MA, 2000), pp. 27–56, pp. 40–44.

<sup>81</sup> Ratches, J. A., 'Review of current aided/automatic target acquisition technology for military target acquisition tasks', *Optical Engineering*, vol. 50, no. 5 (July 2011), article no. 072001.

Advances in machine learning, particularly deep learning and generative adversarial networks (GANs), could significantly simplify that process. With deep-learning methods, engineers could make ATR systems capable of learning independently not only the differences between types of target but also the differences between military and civilian objects (e.g. a commercial aeroplane and a strategic bomber).<sup>82</sup> With GANs, engineers could generate realistic synthetic data on which an ATR system can be trained and tested in simulation. An ATR system trained with these machine learning techniques would perform comparatively much better than an ATR system trained with traditional methods.

Equally, autonomous systems offer new defensive tools against incoming threats. Autonomous unmanned vehicles can be deployed as decoys or flying mines to complement traditional air defences.<sup>83</sup> Advances in autonomy for swarming and for multi-vehicle control could also enable autonomous unmanned systems to operate in a coordinated way and conduct advanced A2/AD manoeuvres.<sup>84</sup> Such systems would increase deterrence against both conventional and nuclear attack as they would increase the risks for an attack by both manned and unmanned platforms.

*Cyberwarfare.* Autonomy is not a new development in the cyber realm. Automation is already a key component of any cyber-defence architecture. Anti-malware programs are designed to automatically identify and neutralize malware. Cyberweapons generally need to operate autonomously—that is, outside direct human supervision—at least during key parts of their mission.<sup>85</sup> This was the case, for instance, for the Stuxnet worm.<sup>86</sup> However, recent advances in machine learning are changing the way that this automation or autonomy works as it changes the way in which cyberwarfare tools are designed and operated—whether for defensive or offensive purposes.

On the defensive side, machine learning methods have opened the possibility of locating previously unknown types of malware and detecting suspicious activities in a network.<sup>87</sup> On the offensive side, machine learning facilitates the identification of an adversary's zero-day vulnerabilities (i.e. undiscovered or unaddressed vulnerabilities in software). Machine learning in a nuclear context is a double-edge sword: it can boost the protection of nuclear command-and-control infrastructure against cyberattacks while expanding the enemy's capacity for cyberattacks against that same infrastructure. Machine learning

<sup>82</sup> Berlin, M. and Young, M., 'Automatic target recognition systems', *Technology Today*, no. 1 (2018), pp. 10–13.

<sup>83</sup> Hipple, M., 'Bring on the countermeasure drones', *Proceedings* (US Naval Institute), Feb. 2014.

<sup>84</sup> Scharre (note 42).

<sup>85</sup> Guarino, A., 'Autonomous intelligent agents in cyber offence', eds K. Podins, J. Stinissen and M. Maybaum, *2013 5th International Conference on Cyber Conflict*, Proceedings, Tallinn, 4–7 June 2013 (NATO Cooperative Cyber Defence Centre of Excellence: Tallinn, 2013), pp. 377–89.

<sup>86</sup> Kile, S. N., 'Nuclear arms control and non-proliferation', *SIPRI Yearbook 2011: Armaments, Disarmament and International Security* (Oxford University Press: Oxford, 2011), pp. 363–87, p. 384; and Sanger, D., 'Obama order sped up wave of cyberattacks against Iran', *New York Times*, 1 June 2012.

<sup>87</sup> Polyakov, A., 'Machine learning for cybersecurity 101', *Towards Data Science*, 4 Oct. 2018.



could enable so-called left-of-launch capabilities. According to one US DOD official, ‘The development of left-of-launch capabilities will provide US decision-makers additional tools and opportunities to defeat missiles. This will in turn reduce the burden on our “right-of-launch” ballistic missile defense capabilities. Taken together, left-of-launch and right-of-launch will lead to more effective and resilient capabilities to defeat adversary ballistic missile threats’.<sup>88</sup> Among those operations frequently described as ‘left-of-launch’ are cyber-offensive operations that would defeat the threat of a nuclear ballistic missile before it is launched.<sup>89</sup>

*Electronic warfare.* Machine learning can bring major improvements to the field of electronic warfare in the same ways as for cyberwarfare. On the defensive side, machine learning enhances anti-jamming capabilities as it opens the possibility to automate analysis and defence against new enemy signals.<sup>90</sup> In 2016 the US Defense Advanced Research Projects Agency (DARPA) launched a public challenge to develop systems with the capability to identify and analyse new enemy signals on the fly—that is, during the operation of the systems rather than afterward as is currently the case.<sup>91</sup> On the offensive side, machine learning can be used to develop new jamming tools that could also play a role in a left-of-launch operation.

*Information warfare.* Machine learning and—to a lesser extent—autonomy could also have strategic impact on information warfare. Machine learning offers new tools to directly or indirectly manipulate nuclear decision makers. An example of direct use would be using GANs to create lifelike fake orders—in audio or video—that would trick nuclear weapon operators into launching a nuclear weapon or not responding to an attack. Higher command-and-control decision makers could also be indirectly tricked into doing or not doing something if their normal sources of information were tainted with fake information or fake opinion from people who would normal seem to be sensible.<sup>92</sup> Should a nuclear-armed state decide to use machine learning algorithms for collection and processing of ISR information, this would open the possibility for an opponent to use a method known as data poisoning to undermine or manipulate the performance of early-warning systems.

*Physical security.* Finally, advances in AI could play a role in the security of nuclear weapons. Nuclear-armed states could combine the advances in machine learning

<sup>88</sup> McKeon B. P., US Principal Deputy Under Secretary of Defense for Policy, Statement before the US Senate Armed Services Subcommittee on Strategic Forces, 13 Apr. 2016. See also US Department of Defense and Joint Chiefs of Staff, ‘Declaratory Policy, concept of operations, and employment guidelines for left-of-launch capability’, Report to the US Congress, 10 May 2017, p. 1.

<sup>89</sup> Ellison, R., ‘Left of launch’, Missile Defence Advocacy Alliance, 16 Mar. 2015.

<sup>90</sup> Freeberg, S. J., ‘Jammer not terminators: DARPA & the future of robotics’, *Breaking Defense*, 2 May 2016.

<sup>91</sup> US Defense Advanced Research Projects Agency (DARPA), ‘New DARPA Grand Challenge to focus on spectrum collaboration’, 23 Mar. 2016.

<sup>92</sup> This scenario is further discussed in chapter 4 in this volume. See also Avin, S. and Amadae, S. M., ‘Autonomy and machine learning at the interface of nuclear weapons, computers and people’, ed. Boulanin (note 7), pp. 105–18.

and autonomy to automate the protection of their nuclear forces against physical attacks by saboteurs or terrorists. Autonomous robots—whether land, aerial or maritime—trained by machine learning are well suited for dull surveillance missions. Machine learning gives robots advanced detection capabilities, while autonomy guarantees that they can keep a sharp and unblinking eye on the perimeters under protection.

Russia already includes anti-saboteur robots in the formations accompanying the mobile Yars ICBMs. These include both unarmed UAVs and combat land systems that are meant ‘for conducting field reconnaissance, for identification and elimination of stationary and mobile targets, for providing fire support for military units, and for patrolling and protecting of sensitive facilities in combination with automated security systems’.<sup>93</sup> Armed automated surveillance systems have also been developed for border and perimeter protection. One example is the robotic sentry weapon Super aEgis II, produced by DoDaam of the Republic of Korea (South Korea), although this is not possessed by a nuclear-armed state. Super aEgis II is a gun turret equipped with sensors and an ATR system that can automatically detect, track and (potentially) attack targets—the system is designed to operate under human control, but it includes a ‘fully autonomous’ mode.<sup>94</sup>

In sum, advances in machine learning and autonomy could have numerous applications in the realm of nuclear forces. The question in that context is not if these will be adopted by nuclear-armed states but when and by whom, as outlined below.

<sup>93</sup> TASS, ‘Russia’s strategic missile force to test mobile robot at forthcoming exercise’, 30 Mar. 2016; and ‘Russia’s new Yars ICBM system entering service with Irkutsk missile formation’, TASS, 31 Mar. 2016.

<sup>94</sup> Boulanin and Verbruggen (note 12), pp. 44–46.

### 3. AI and the military modernization plans of nuclear-armed states

The concept of an AI arms race—or, more aptly, capability race—is increasingly used in the literature to describe how great powers compete with each other on AI (see box 3.1).<sup>95</sup> It is a powerful analogy. The concept of a capability race is easily understood and can be extremely effective in highlighting concerns about the risks posed by the militarization of AI. However, two questions remain: Is there evidence of a capability race specific to AI? If there is such a capability race, what does the evidence say about its nature and status?

The answers to these questions matter because concepts such as that of an AI capability race have an impact on the way in which policymakers think about their national security needs. Stating that AI is the focus of an arms race contributes to elevating AI on the agenda of national security professionals around the globe—a process that international relations scholars would describe as the securitization of AI.<sup>96</sup> It leads policymakers and their armed forces to make policy decisions and initiate measures that can further militarize AI but also fragment the field of AI along national lines. Words have a performative power—they can condition the way in which a problem is seen and how solutions are explored. Political sociology has shown how they can trigger self-fulfilling prophecies.<sup>97</sup> It is therefore important to ensure that they reflect reality so that they can be used in an appropriate manner.

This chapter offers an empirically informed analysis of the state of great power competition in the field of AI, focusing on the declared nuclear-armed states: the United States, Russia, the United Kingdom, France, China, India, Pakistan and North Korea (in sections I–VIII, respectively).<sup>98</sup> For each of these eight states, two sets of questions are systematically explored.

1. Regarding vision and policies on AI: Is AI on the country's political agenda? What is its specific vision for military AI?
2. Regarding adoption of military AI: What is the country's capability to adopt the most recent advances in AI for military purposes? What is the state of adoption of AI by its armed forces?

For a number of reasons, there are limits to the clarity of the available information on how far the integration of AI technology into national military applications has advanced. The first, obvious reason is the lack of transparency on technical details of strategic importance. Moreover, the technical information that is available should be treated with scepticism: declared advances may be intentionally exaggerated or downplayed. A second, more technical reason is that,

<sup>95</sup> Simonite, T., 'For superpower, artificial intelligence fuels new global arms race', *Wired*, 9 Aug. 2017.

<sup>96</sup> On the concept of securitization see Buzan, B., Wæver, O. and de Wilde, J., *Security: A New Framework for Analysis* (Lynne Rienner: Boulder, CO, 1998), p. 25.

<sup>97</sup> Austin, J. L., *How to Do Things with Words* (Clarendon Press: Oxford, 1962).

<sup>98</sup> As noted above (see note 20), since Israel is not a declared nuclear-armed state, it is not discussed here.

**Box 3.1.** The artificial intelligence race

The term ‘arms race’ was originally coined by Lewis Fry Richardson to describe developments in armament in the run-up to World War I and World War II and was often used to describe developments during the cold war. It seems to have become the term of choice for journalists, scholars and practitioners when talking about great power competition in artificial intelligence (AI).<sup>a</sup>

It is a powerful yet imperfect analogy. Historically, the concept of an arms race has a specific meaning, which could lead to misunderstanding with regards to what is currently happening with AI. During the cold war, arms races were about increasing the number or physical capabilities (e.g. speed, range, kinetic effect, precision) of weapon systems in order to seek or maintain a balance of power.<sup>b</sup> The parameters of the AI race differ from the actual arms race that occurred during the cold war. AI is not a weapon per se; it is not even a definite unified technology.<sup>c</sup> It is more accurate to compare AI to electricity than to a specific weapon technology such as missiles or nuclear weapons.<sup>d</sup>

Advances in AI are also mainly driven by civilian needs. Cutting-edge innovations in the field of AI are currently not coming from military research and development (R&D) laboratories, but from civilian ones and are primarily destined for the consumer market. This is not to say that military R&D institutions, such as the United States’ Defense Advanced Research Projects Agency (DARPA), are not playing a role, particularly when it comes to initiating and funding high-risk blue-sky R&D projects; but in terms of volume of investments and tangible outputs, the civilian sector is now outpacing the military. The leaders of the race are neither states nor arms-producing or military services companies but private civilian companies that are not under the direct or indirect control of the state—with the notable exception of China.<sup>e</sup>

Another important difference is that the competition between major powers in the field of AI does not (yet) seem to be about acquiring large numbers of high-quality AI-powered weapons. Rather, it is about acquiring the resources necessary to develop or maintain cutting-edge AI capabilities in all areas. These include AI talent, large volumes of high-quality data to train machine learning systems and computer power resources. In that regard, it seems more correct to talk about a ‘capability race’ than an ‘arms race’ to discuss the technological competition between China, Russia, the USA and other countries that are active in the field of AI. The concept of an AI arms race emphasizes only one aspect of this competition.

<sup>a</sup> Simonite, T., ‘For superpower, artificial intelligence fuels new global arms race’, *Wired*, 9 Aug. 2017.

<sup>b</sup> Buzan, B. and Herring E., *The Arms Dynamics in World Politics* (Lynne Rienner: Boulder, CO, 1998).

<sup>c</sup> See chapter 2 in this volume.

<sup>d</sup> Boulanin, V., ‘Artificial intelligence: A primer’, ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, *Euro-Atlantic Perspectives* (SIPRI: Stockholm, 2019), pp. 13–25.

<sup>e</sup> Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017).

as an enabling technology rather than a single and defined technology, AI can take many forms and can have various degree of sophistication depending on the type of system or capability it supports. Presenting the state of a country’s military AI in a single description is difficult because the description depends on the types of task, system and environment that are of interest. In describing the state of adoption of AI by the armed forces, the following sections look at the capability areas listed in chapter 2: (a) early warning and ISR, (b) command and control, (c) precision strike and delivery, and (d) various non-nuclear operations.

## I. The United States

### **Vision and policies**

#### *AI on the political agenda*

The United States is in many regards the birthplace of artificial intelligence. The discipline was born in 1956 in a workshop at Dartmouth College, New Hampshire, where the term artificial intelligence was first coined.<sup>99</sup> US universities and scholars pioneered the field and were behind its most remarkable technological achievement. The US Government itself has been instrumental in funding research and development (R&D) in this area. Government interest in AI has risen and fallen over time as the field has gone through various periods of success and failure, but it never completely stopped. Moreover, the USA has one of the highest degrees of transparency on the ways it has applied AI.

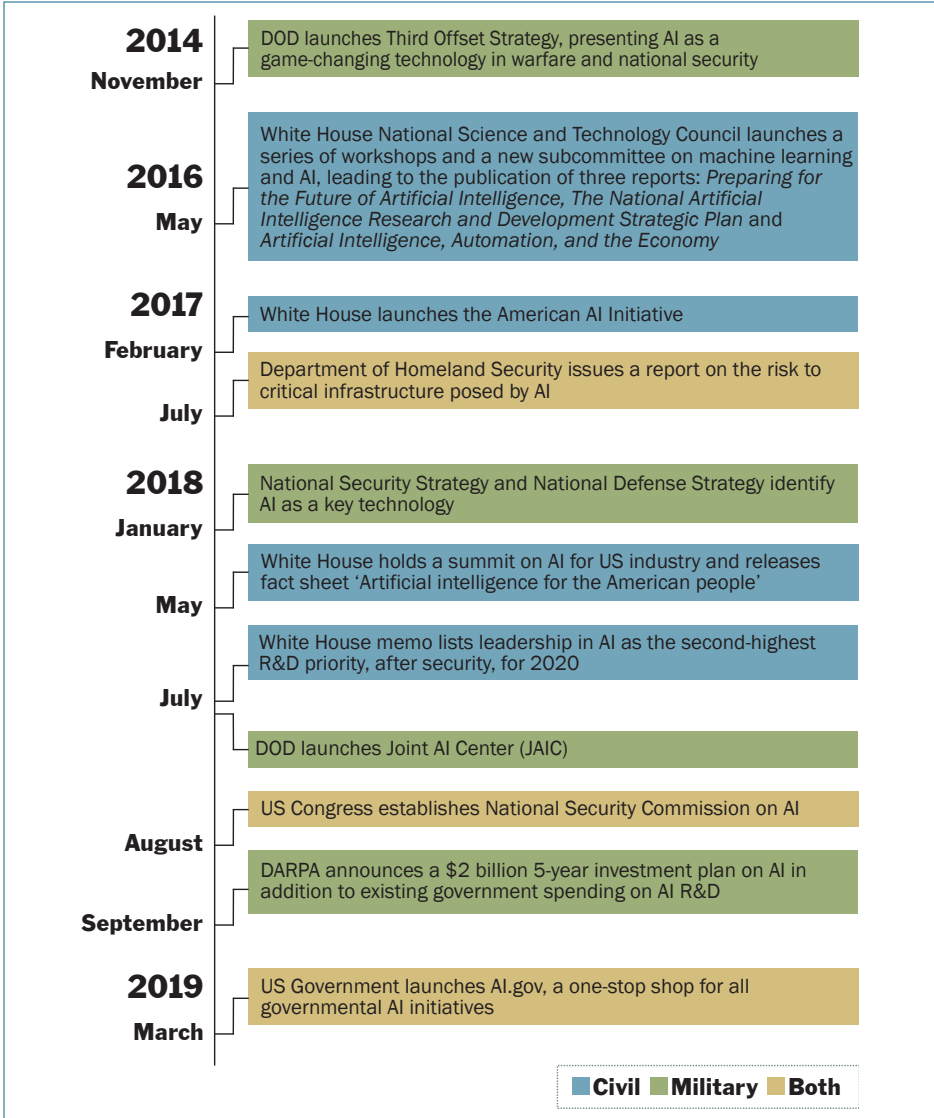
The breakthrough that the field of AI has been experiencing since the beginning of the 2010s has moved AI to the top of the political agenda. Since 2014, the US Government has published multiple policy documents that outline how it intends to harness the current revolution in machine learning—the most recent being the American AI Initiative of February 2019 (see figure 3.1).<sup>100</sup> These documents address a wide spectrum of issues, including how to derive the best economic benefits while limiting the potential negative effects on the workforce; how to foster responsible innovation through ethical guidelines and regulate the use of new technologies such as self-driving cars and delivery UAVs; how to adopt and use AI technology for governmental administration; and how to adopt AI for military purposes while ensuring that the USA can retain leadership in this area.

Some of the policies have already led to the implementation of concrete measures, including the designation of AI and autonomous and unmanned systems as R&D priorities for the US administration and increased government funding for R&D on AI. Notably, they have led to the creation of the Defense Innovation Unit (DIU) and the Joint Artificial Intelligence Center (JAIC) in the US Department of Defense. The DIU is in charge of fostering military-relevant AI innovation through close cooperation with the private sector, including small and large civilian companies. The JAIC is responsible for informing policy and synchronizing AI activities with the DOD.

The US Government's interest in AI has always been dual-use in nature—both civilian and military—but overall the armed forces seem to have been the driving force in governmental efforts. Many of the technical breakthroughs in AI over the past 50 years were connected in some way with the DOD. The US armed forces have played an important role in setting priorities for and funding research efforts in AI. For example, AI virtual assistants and self-driving cars are applications that

<sup>99</sup> Pearl (note 22).

<sup>100</sup> White House, Office of Science and Technology Policy, 'Accelerating America's leadership in artificial intelligence', 11 Feb. 2019; and White House, 'Maintaining American leadership in artificial intelligence', Executive order, 11 Feb. 2019.



**Figure 3.1.** Recent policy developments related to artificial intelligence in the United States

AI = artificial intelligence, DARPA = Defense Advanced Research Projects Agency, DOD = Department of Defense; R&D = research and development.

Sources: White House, National Science and Technology Council, 'Charter of the Subcommittee on Machine Learning and Artificial Intelligence', 6 May 2016; US Department of Homeland Security (DHS), Office of Cyber and Infrastructure Analysis, *Artificial Intelligence*, Narrative analysis (DHS: Washington DC, July 2017); White House, 'Artificial intelligence for the American people', Fact sheet, 10 May 2018; Mulvaney, M. and Kratsios, M., 'FY2020 administration research and development budget priorities', Memorandum for the heads of executive departments and agencies, Executive Office of the President, 31 July 2018; White House, 'Artificial intelligence for the American people', Mar. 2019; and this volume: Boulanin et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, June 2020).

were initially researched and developed with the support of the DOD, specifically DARPA.<sup>101</sup>

The US policy documents published since 2014 do not represent a shift in this regard. They aim to ensure that the USA can keep its leadership in both the civilian and military arenas. The two most strategic recent policy documents on AI—*Preparing for the Future of Artificial Intelligence* (2016) and the American AI Initiative (2019)—discuss both civilian and military uses of AI.<sup>102</sup> The DOD also remains an instrumental actor as it continues to support many R&D efforts that could benefit both arenas.

### *What vision for military AI?*

AI has been part of US strategic calculations for a long time. In the 1980s the DOD was already portraying AI in official documents as a technology that could change the character of warfare.<sup>103</sup> It undertook ambitious research projects that were intended to prepare the US armed forces at the technical and doctrinal levels for the era of the automated battlefield.<sup>104</sup> The interest of the US armed forces in AI fell in the 1990s and early 2000s as the projects initiated in the early 1980s (notably the 1983 Strategic Computing Initiative) and the field of AI more generally failed to deliver on the original promise.<sup>105</sup> The DOD continued to carry out some R&D on specialized AI applications, such as pilot assistants and autonomous vehicles, but AI was no longer central to conversations in the US military community about the future of warfare and future military modernization plans.

This changed around the beginning of the 2010s with the increasing use of robotic systems in the US military interventions in Afghanistan and Iraq, which renewed interest in autonomous technologies, and the breakthrough in deep learning. By 2014 AI had returned as the central concept in the debate on the future of warfare among US military planners.<sup>106</sup> The publication of the US DOD's Defense Innovation Initiative in 2014, also known as the Third Offset Strategy, was the clearest evidence that the USA once again considered AI to be a game-changing technology in warfare and national security.<sup>107</sup> AI and machine learning were described by Robert Work, US Deputy Secretary of Defense, as the key technological ingredients for US military superiority in the future.<sup>108</sup> The DOD

<sup>101</sup> Davis, A., 'Inside the races that jump-started the self-driving car', *Wired*, 11 Oct. 2017.

<sup>102</sup> US Executive Office of the President and National Science and Technology Council (NSTC) Committee on Technology, *Preparing for the Future of Artificial Intelligence* (White House: Washington, DC, Oct. 2016); and White House, 'Maintaining American leadership in artificial intelligence' (note 100).

<sup>103</sup> The history of AI and robotics in the USA is discussed in Boulanin and Verbruggen (note 12), pp. 58–59.

<sup>104</sup> For a contemporary analysis of these projects see ed. Din (note 25).

<sup>105</sup> Roland, A. with Schiman, P., *Strategic Computing: DARPA and the Quest for Machine Intelligence, 1983–1993* (MIT Press: Cambridge, MA, 2002).

<sup>106</sup> Boulanin and Verbruggen (note 12), pp. 58–61.

<sup>107</sup> Hagel, C., US Secretary of Defense, 'The Defense Innovation Initiative', Memorandum, US Department of Defense, 15 Nov. 2014.

<sup>108</sup> Work, B., US Deputy Secretary of Defense, 'Remarks by Deputy Secretary Work on Third Offset Strategy', US Department of Defense, 28 Apr. 2016.

views AI as the technology that will allow the USA to offset the advantages of adversaries in other areas and maintain strategic stability.

For the DOD, the fundamental challenge will be less to maintain leadership in technological development than to find the most innovative and effective way to do so in the military context. In fact, US official documents acknowledge that other military powers, notably China, have the resources to develop AI technology that is just as advanced as that of the USA, if not more so.<sup>109</sup> Like the United Kingdom (see section III), the US military establishment places great importance on human-machine teaming (i.e. collaboration). For the USA, AI technology will deliver best value when intelligently combined with human capabilities.<sup>110</sup> According to that narrative, the future of warfare will not be a battlefield full of fully autonomous robots and weapons fighting each other. Humans will continue to play a key role, notably because the limitations of AI technology require them to continue to have a crucial function as receiver and arbitrator of tactical information on the battlefield.

The narrative in the government defence community is that AI technology will bring many operational benefits.<sup>111</sup> These will include benefits for the non-combat part of warfare by making command and control and logistics more independent of humans. AI will also provide new ways to manage the battlefield, for instance by helping commanders to find new ways to anticipate enemy tactics.<sup>112</sup> Perhaps more importantly, the US military community also believes that AI could bring humanitarian benefits: it could allow the armed forces to apply force with greater precision and thereby reduce the risk of collateral damage to civilians and also reduce the exposure of military personnel to danger. This one of the reasons why the US Government has not welcomed the idea of a preventive ban on Lethal Autonomous Weapons Systems (LAWS). Since the debate on LAWS under the CCW Convention started in 2014, the US delegation has repeatedly stressed that the development of autonomy in weapon system is not problematic—from both a legal and ethical standpoint—as long as ‘appropriate levels of human judgement’ are exercised.<sup>113</sup> For the USA, existing international law provides sufficient limitations on how militaries should use autonomous weapon systems and military application of AI more generally.<sup>114</sup>

<sup>109</sup> Saylor, K M., *Artificial Intelligence and National Security*, Congressional Research Service (CRS) Report for Congress R45178 (US Congress, CRS: Washington, DC, 30 Jan. 2019), pp. 19–22.

<sup>110</sup> Freedberg, S. J., ‘Iron Man, not Terminator: The Pentagon’s sci-fi inspirations’, *Breaking Defense*, 3 May 2016.

<sup>111</sup> US Department of Defense (note 42); US Department of Defense (note 68).

<sup>112</sup> Carter, W. A., Kinnucan, E. and Elliot, J., *A National Machine Intelligence Strategy for the United States* (Center for Strategic and International Studies: Washington, DC, Mar. 2018), p. 17.

<sup>113</sup> Certain Conventional Weapons Convention, Group of Governmental Experts on Lethal Autonomous Weapon Systems, ‘Human-machine interaction in the development, deployment and use of emerging technologies in the area of lethal autonomous weapons systems’, Working Paper submitted by the United States, 28 Aug. 2018, CCW/GGE.2/2018/WP.4, paras 8–15; and McKendrick, K., ‘Banning Autonomous Weapons is not the answer’, Chatham House, 2 May 2018.

<sup>114</sup> Certain Conventional Weapons Convention, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, ‘Implementing international humanitarian law in the use of autonomy in weapon systems’, Working paper submitted by the United States, CCW/GGE.1/2019/WP.5, 28 Mar. 2019.



However, the US establishment is concerned by how other actors might use AI technology, notably autonomous systems, in different ways, particularly in ways that would disregard obligations of international law.<sup>115</sup> Notably, it fears that non-state actors could develop new asymmetric warfare tactics that could lead to more deadly and destructive use of force against US personnel.<sup>116</sup>

The US armed forces and the government more broadly, like the other major military powers, are concerned with the USA's ability to maintain leadership in the field of AI.<sup>117</sup> The measures that the US Government has prioritized in various policy documents since 2016 include investing in R&D; educating, recruiting and retaining AI engineers; developing appropriate data sets for training and testing; and developing methods to test and certify the reliability and safety of complex AI technology.<sup>118</sup>

In terms of concrete capabilities, the DOD is interested in using AI in almost all vital mission areas (see table 3.1).<sup>119</sup> Most applications are relevant for nuclear weapon systems. The US strategic documents related to US nuclear capability and posture give little information on how AI fits into future nuclear force modernization plans. The 2018 US Nuclear Posture Review makes no direct reference to AI, machine learning or autonomous systems other than a brief reference to Russia's plans to develop an autonomous nuclear-powered UUV (i.e. Poseidon).<sup>120</sup>

The USA is currently engaged in the modernization of its nuclear triad and plans to replace or upgrade all of its nuclear delivery systems. It intends to develop a new class of SSBNs (the Columbia class), a nuclear-capable strategic bomber (the B-21 Raider) and dual-capable combat aircraft (the F-35A), an ICBM (the Ground Based Strategic Deterrent) and nuclear-capable cruise missiles (the air-launched Long Range Standoff missile and a ground-launched modification of the Tomahawk).<sup>121</sup> The modernization plans also include the upgrade of command and control and creation of a low-yield warhead for the US Navy.<sup>122</sup> The budget for modernizing and operating the US nuclear arsenal will reach an estimated US\$494 billion in 2019–28.<sup>123</sup> There is currently no publicly available information on the full extent of the role that AI will play in the USA's nuclear modernization programme. However, it is likely that machine learning will play a role. Machine

<sup>115</sup> Sayler (note 109), pp. 32–36.

<sup>116</sup> US Department of Defense, *Summary of the 2018 National Defense Strategy of the United States of America* (DOD: Washington DC, 2018), p. 3.

<sup>117</sup> Sayler (note 109).

<sup>118</sup> US National Science and Technology Council, Select Committee on Artificial Intelligence, *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update* (Executive Office of the President: Washington, DC, June 2019); and US Department of Defense (DOD), *Summary of the 2018 Department of Defense Artificial Intelligence Strategy* (DOD: Washington, DC, 2018).

<sup>119</sup> US Executive Office of the President and National Science and Technology Council (NSTC) Committee on Technology (note 102), p 38; and US Department of Defense (note 68).

<sup>120</sup> US Department of Defense, *Nuclear Posture Review* (note 1), p. 9.

<sup>121</sup> Kristensen, H. M., 'US nuclear forces', *SIPRI Yearbook 2019* (note 8), pp. 294–95; and Kristensen, H. M. and Korda, M., 'United States nuclear forces, 2020', *Bulletin of the Atomic Scientists*, vol. 76, no. 1 (Jan. 2020), p. 49.

<sup>122</sup> Kristensen (note 121); and Kristensen and Korda (note 121), p. 49.

<sup>123</sup> Kristensen (note 121); and Kristensen and Korda (note 121), p. 49.

**Table 3.1.** Applications of artificial intelligence of interest to the US Department of Defense

Mission	Technology/application of interest	Military rationale
Force application	Automated target recognition systems Autonomous navigation systems for missiles and unmanned combat vehicles AI software for force operation planning Autonomous swarms that exploit large quantities of low-cost assets Autonomous vehicles for offensive mining, countermine operations and decoy delivery Tactical unmanned aerial vehicles for ground force support	Increase precision and autonomy of weapon systems Enable longer combat mission, facilitate operation in contested environment Increase speed and agility of force deployment Enable attrition attack that would overwhelm enemy defences Increase persistence, survivability and manpower efficiency Support rapid strike; provide immediate battlefield intelligence; provide cover
Battlespace awareness	On-board processing of sensing and intelligence data AI software for ISR data processing Autonomous swarms that exploit large quantities of low-cost assets	Reduce need to analyse ISR off-board and to maintain high-quality communication bandwidth with deployed systems Reduce need for human analysts; faster and more agile processing of ISR data Enhance situational awareness with larger and more effective geographical coverage
Force protection	Automated cybersecurity and cyber-defence systems Unmanned autonomous systems for lifesaving battlefield medical assistance and casualty evacuation	Reduce reaction time after cyberattacks Reduce risk to rescue personnel
Logistics	Machine learning-powered data analytics software for adaptive logistics Unmanned autonomous systems for delivery and maintenance	Increase efficiency of supply chain management Increase manpower efficiency; reduce risk to personnel

AI = artificial intelligence; ISR = intelligence, surveillance and reconnaissance.

Sources: US Executive Office of the President and National Science and Technology Council (NSTC) Committee on Technology, *Preparing for the Future of Artificial Intelligence* (White House: Washington, DC, Oct. 2016), p 38; US Department of Defense (DOD), Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy* (DOD: Washington, DC, June 2016); and US Department of Defense (DOD), *Summary of the 2018 Department of Defense Artificial Intelligence Strategy* (DOD: Washington, DC, 2018).

learning could be used to enhance the guidance capabilities of SSBNs and ICBMs, to enhance the detection capabilities of early-warning systems, and also to automate some aspects of maintenance of nuclear assets.<sup>124</sup>

## **Adoption of military AI**

### *Capability to adopt the most recent advances in AI for military purposes*

The USA is, and is likely to remain in the near future, the country that sets the benchmark for what AI can or could deliver in the military sphere. Over the 2010s and previous decades, it pushed the boundaries of what is technically feasible, notably thanks to unmatched levels of investment in relevant R&D. The rise of AI on the agenda of the military establishment in the 2010s indicates that the USA will continue to make significant efforts to fund and orient innovation in this area. Most notably, in 2018 DARPA announced the launch of a \$2 billion investment campaign to develop the next wave of AI technology.<sup>125</sup>

The USA also has the most well-developed innovation ecosystem for the development of military AI and a strong research base. US universities dominate the academic landscape in AI and related disciplines. According to the Microsoft Academic search index, the most influential research papers on AI and related disciplines (including machine learning and computer vision) originate from the USA.<sup>126</sup> Moreover, the USA currently has the strongest industrial base in the world for AI. It hosts the majority of the industrial giants that are currently shaping the future of AI and robotics: Alphabet (owner of Google), Facebook and Amazon in the civilian sphere and Lockheed-Martin, Northrop Grumman, Boeing and Raytheon in the military sphere.

Most importantly, the USA has the world's large military budget.<sup>127</sup> Its ability to invest money in the development of technology and capabilities is unmatched. China is the only state that could reach the same level of investment in the relatively near future.

The challenge for the DOD is that the civilian industry is leading innovation in AI and the DOD finds it difficult to make these civilian companies develop military applications.<sup>128</sup> The large companies that matter—Alphabet (Google), Facebook and Amazon—have limited incentive to work with the DOD: the DOD's acquisition process is cumbersome, the value of the contract might be less than can be gained on the civilian market and the contractual requirements in terms

<sup>124</sup> Stoutland, P. O., 'Artificial intelligence and the modernization of US nuclear forces', ed. Boulanin (note 7), pp. 63–67, p. 65.

<sup>125</sup> US Defense Advanced Research Projects Agency (DARPA), 'DARPA announces \$2 billion campaign to develop next wave of AI technologies', 7 Sep. 2018.

<sup>126</sup> Microsoft Academic; and Boulanin, V., *Mapping the Innovation Ecosystem Driving the Advance of Autonomy in Weapon Systems*, Working paper (SIPRI: Stockholm, Dec. 2016), p. 29.

<sup>127</sup> SIPRI Military Expenditure Database, <<https://www.sipri.org/databases/milex>>.

<sup>128</sup> Boulanin and Verbruggen (note 12), pp. 109–10; and Verbruggen, M., 'The role of civilian innovation in the development of lethal autonomous weapon systems', *Global Policy*, vol. 10, no. 3 (Sep. 2019), pp. 338–42, pp. 339–40.

of proprietary rights are too stringent.<sup>129</sup> In recent years the DOD has taken several initiatives to change this, notably through the creation of the DIU, which established offices in all the three big US centres of the technology industry: Silicon Valley in California, Boston in Massachusetts and Austin in Texas.<sup>130</sup> Some companies are also concerned about bad publicity that may arise from work with the US armed forces. The most publicized example of this was Google's decision in 2018 to discontinue its contract with the DOD to design software for automated analysis of video footage, known as Project Maven.<sup>131</sup> Google subsequently adopted a set of AI principles that limit its involvement in development of military application of AI.<sup>132</sup> A more generic challenge is that the big civilian companies are partly responsible for the shortage of skilled engineers in the military sector. They have recruited the top researchers in the fields of AI and robotics, and they can offer a new graduate a salary that universities and the military establishment cannot compete with. These companies also retain a large amount of training and test data to which the DOD does not have access.

These problems disadvantage the DOD in relation to its competitors, notably China.<sup>133</sup> There is no separation between the civilian and military AI sectors in China (see section V). In Russia the frontier between the two sectors is also more porous than in the USA (see section II). These two countries have reportedly less difficulty than the USA in adapting civilian innovation for military purposes.<sup>134</sup>

#### *State of adoption of AI by the armed forces*

As explained above, while the DOD's interest in AI has risen and fallen over time, its R&D in AI has never completely stopped. As a result of this, the USA is the most advanced country in terms of adoption of AI for military purposes.

The US armed forces already employ some technologies that rely on AI. These range from guided munitions and air-defence systems that use ATR technology, via voice-controlled pilot assistants in combat aircraft (e.g. the F-15) to unmanned vehicles capable of autonomous navigation.<sup>135</sup> Compared to other countries, a lot of open-source information is available about the AI technology that is used. This makes it clear that the technology that is currently in use does not rely on recent advances in machine learning but on traditional, hard-coded AI. The capabilities of this AI technology also need to be put in perspective: they remain brittle. The

<sup>129</sup> Boulanin and Verbruggen (note 12), p. 81.

<sup>130</sup> Pellerin, C., 'Carter opens DIUx outpost in Texas', US Department of Defense, 14 Sep. 2016.

<sup>131</sup> Pellerin, C., 'Project Maven to deploy computer algorithms to war zone by year's end', US Department of Defense, 21 July 2017; and Wakabayashi, D. and Scott, S., 'Google will not renew Pentagon contract that upset employees', *New York Times*, 1 June 2018.

<sup>132</sup> D'Offro, J., 'Google promises not to use AI for weapons or surveillance, for the most part', CNBC, 7 June 2018; and Google AI, 'Artificial intelligence at Google: Our principles', Google.

<sup>133</sup> Simonite, T., 'China is catching up to the US in AI research—fast', *Wired*, 13 Mar. 2019.

<sup>134</sup> Allen, G. C., *Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security* (Center for a New American Security: Washington, DC, Feb. 2019); and Horowitz et al. (note 14).

<sup>135</sup> Bell, G., Schultz, M. C. and Schultz, J. T., 'Voice recognition in fighter aircraft', *Journal of Aviation/Aerospace Education and Research*, vol. 10, no. 1 (fall 2000), pp. 17–27.

ATR systems can only recognize large, pre-defined types of military object.<sup>136</sup> The voice-controlled pilot assistants deployed in combat aircraft allow little interactivity and are mainly used to provide information to the pilot.<sup>137</sup> Similarly, most unmanned systems that are reported to be capable of autonomous navigation rely on way-point navigation: the systems merely follow a series of geodetic coordinates that are entered by a human operator.<sup>138</sup>

The DOD has been transparent in its R&D activities. The available information shows that efforts to further advance the military use of AI by the USA are well underway. These include many projects that are or could be relevant in the future for the US nuclear deterrence architecture (see table 3.2).

Many of these projects are well advanced in their development. Their capabilities, notably autonomous capabilities, have been showcased in operational tests. In the aerial domain, the most noteworthy cases include the X-47B and the MQ-25 Stingray UAV demonstrators developed for the US Navy that are capable of autonomous take-off and landing, navigation in communications-denied environments, and in-flight refuelling.<sup>139</sup> These could be used for such missions as suppression of enemy air defence (SEAD) and conventional strikes on strategic assets. Another UAV demonstrator, the XQ-58 Valkyrie, shows how far the USA has gone in the use of AI for collaborative aerial operations. It was designed for the USAF to operate as a wingman for manned aircraft. It can reportedly fly autonomously in formation with other systems and be tasked to conduct such tasks as scouting or protection against enemy fire.<sup>140</sup> Another notable achievement was the demonstration of the Perdix UAV swarm in 2017, a swarm of more than 100 mini UAVs that can be deployed from a combat aircraft to conduct collaborative ISR autonomously.<sup>141</sup> In the maritime domain, notable achievements include the *Sea Hunter*, an autonomous surface system capable of anti-submarine operations and mine countermeasures; the Orca, an extra-large UUV (XLUUV) designed for ISR missions, and the Control Architecture for Robotic Agent Command and Sensing (CARACAS) system, a software architecture that allows a swarm of USVs and UAVs to conduct A2/AD manoeuvres autonomously.<sup>142</sup>

It should be noted that of these new technologies, only the MQ-25 Stingray and the Orca XLUUV have been approved as a 'programme of record' or acquisition programme. The others have not yet been officially procured by the US armed forces. There are two reasons for this: first, the DOD is still exploring the potential of these technologies; and second, the DOD is also struggling to make

<sup>136</sup> Boulanin and Verbruggen (note 12), pp. 25–26.

<sup>137</sup> Bell et al. (note 135).

<sup>138</sup> Boulanin and Verbruggen (note 12), pp. 21–23.

<sup>139</sup> Northrop Grumman, 'X-47B: Program overview'; and Boeing, 'Boeing's MQ-25 is ready'.

<sup>140</sup> Axe, D., 'The Air Force's mysterious XQ-58 Valkyrie drone is almost ready', *National Interest*, 4 Nov. 2019.

<sup>141</sup> Condliffe, J., 'A 100-drone swarm, dropped from jets, plans its own moves', *MIT Technology Review*, 10 Jan. 2017.

<sup>142</sup> Boulanin V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, 2017). SIPRI data set on autonomy in military systems, 2016, available on request from SIPRI.

**Table 3.2.** State of adoption of artificial intelligence in the United States nuclear deterrence architecture

Application area	AI in use	Example or mention in official sources	Status	What is known about AI use
<i>Early warning and intelligence, surveillance and reconnaissance</i>				
AI for data collection and analysis	✓	Project Maven <sup>a</sup>	R&D	Uses machine learning to automatically analyse video surveillance footage gathered during counterinsurgency operations
ISR/remote sensing	✓	Perdix UAV swarm <sup>b</sup>	R&D	Allows up to 100 UAVs to conduct collaborative ISR operations autonomously
ISR/remote sensing	✓	X-47B unmanned aerial vehicle <sup>c</sup>	R&D	Capable of autonomous take-off and landing, navigation in communications-denied environments and in-flight refuelling
ISR/remote sensing	✓	Orca XLUUV <sup>d</sup>	Production	Would need to include some autonomous navigation capabilities
<i>Command and control</i>				
Stockpile management	✓	Reportedly use for nuclear weapon stockpile management <sup>e</sup>	R&D	..
Decision-support systems	✓	Deep Green <sup>f</sup>	R&D	Assists commanders to rapidly generate courses of action through evaluation of the options, development of alternatives and evaluation of the impact of decisions on other parts of the plan
<i>Precision strike and delivery</i>				
Air launched	..	AGM-183A Air-Launched Rapid Response Weapon (ARRW) <sup>g</sup>	R&D	Based on photographs of a B-52 bomber carrying a 'sensor-only' AGM-183A prototype; machine learning and autonomy may be used for guidance and manoeuvrability
Sea launched	..	Conventional Prompt Strike Glide vehicle <sup>h</sup>	R&D	Machine learning and autonomy may be used for guidance and manoeuvrability
Ground launched	..	Long-range Hypersonic Weapon (LRHW) <sup>i</sup>	R&D	Machine learning and autonomy may be used for guidance and manoeuvrability
Missile/air/space defence	✓	Aegis ballistic missile defence systems <sup>j</sup>	Deployed	Uses an active radar seeker

*Other*

Cyber/ electronic information warfare	✓	2016 DARPA Cyber Grand Challenge <sup>k</sup>	R&D	Competition on how to autonomously detect, evaluate, and patch software vulnerabilities
Physical security	✓	<i>Sea Hunter</i> autonomous surface vehicle <sup>l</sup>	R&D	Autonomously detects, tracks and trails submarines
Physical security	✓	Control Architecture for Robotic Agent Command and Sensing (CARACAS) <sup>m</sup>	R&D	Allows a swarm of surface vessels and UAVs to conduct anti-access/area-denial manoeuvres autonomously

. . = no or unclear, ✓ = yes, AI = artificial intelligence, ISR = intelligence, surveillance and reconnaissance, R&D = research and development, UAV = unmanned aerial vehicle, XLUUV = extra-large unmanned underwater vehicle.

<sup>a</sup> Weisgerber, M., 'General: Project Maven is the just the beginning of the military's use of AI', *Defense One*, 28 June 2018.

<sup>b</sup> Condliffe, J., 'A 100-drone swarm, dropped from jets, plans its own moves', *MIT Technology Review*, 10 Jan. 2017.

<sup>c</sup> Northrop Grumman, 'X-47B: Program overview'.

<sup>d</sup> Lockheed Martin, 'Orca XLUUV: Extra large unmanned undersea vehicle'.

<sup>e</sup> Martin, J., 'What role does AI have in the American nuclear arsenal?', *Defense News*, 7 Oct. 2019.

<sup>f</sup> 'DARPA's commander's aid: From OODA to Deep Green', *Defense Industry Daily*, 3 June 2018.

<sup>g</sup> Saylor, K. M., *Hypersonic Weapons: Background and Issues for Congress*, Congressional Research Service (CRS) Report for Congress R45811 (US Congress, CRS: Washington, DC, 4 Mar. 2020); and Trevithick, J., 'Behold the first flight of a B-52 bomber carrying the AGM-183A hypersonic missile', *The Drive*, 17 June 2019.

<sup>h</sup> Saylor (note g).

<sup>i</sup> Saylor (note g).

<sup>j</sup> Missile Threat, 'Aegis ballistic missile defense', Center for Strategic and International Studies.

<sup>k</sup> Frazee, D., 'Cyber Grand Challenge (CGC)', US Defense Advanced Research Projects Agency (DARPA).

<sup>l</sup> US Defense Advanced Research Projects Agency (DARPA), 'ACTUV "Sea Hunter" prototype transitions to Office of Naval Research for further development', 30 Jan. 2018.

<sup>m</sup> Tucker, P., 'Inside the Navy's secret swarm robot experiment', *Defense One*, 5 Oct. 2014.

the transition from development to operational implementation.<sup>143</sup> In order to be validated for acquisition, a system needs to solve all outstanding problems related to safety, reliability and cultural acceptance.<sup>144</sup>

<sup>143</sup> Cummings (note 12), p. 8; and Mindell, D., *Our Robots, Ourselves: Robotics and the Myths of Autonomy* (Viking: New York, 2015). See also Bronk (note 49).

<sup>144</sup> Cummings (note 12), p. 8.

## II. Russia

### Vision and policies

#### *AI on the political agenda*

In September 2017 Russian President Vladimir Putin declared ‘Artificial intelligence is the future, not only for Russia, but for all humankind. It comes with colossal opportunities, but also threats that are difficult to predict. Whoever becomes the leader in this field will become the ruler of the world.’<sup>145</sup> This sweeping statement on the impact that AI could have for Russia and the world made the headlines at home and abroad. Commentators around the globe generally saw it as evidence that great power competition on AI had started and that Russia was prepared for it.<sup>146</sup> However, it took a few years for Putin’s vision to translate into concrete policy action at the domestic level (see figure 3.2).

The first modest policy development came in March 2018, when the Russian Ministry of Defence (MOD) organized an expert conference at the end of which a 10-point plan on AI was presented (see below).<sup>147</sup>

AI became a more visible national policy priority in February 2019, when Putin proposed a series of measures on AI as part of a discourse on the digital economy in his annual address to the Russian Federal Assembly.<sup>148</sup> Soon after this address, Putin also commissioned the drafting of a national AI strategy for the government. The Russian Government would draft the strategy jointly with Sberbank, one of the Russia most prominent investment banks and a strong advocate of the potential of AI for the Russian economy.<sup>149</sup>

A first draft of the AI strategy was officially presented to the public in October 2019.<sup>150</sup> It presents a number of ideas about how to accelerate the development and application of AI in Russia. The aim of the strategy is twofold: on the one hand, ensure that Russia can become a leader in the field of AI; and, on the other, ensure Russia’s sovereignty in this area. The later point is a vital concern for Russia. In AI, and information technology (IT) more generally, Russia is highly dependent on foreign technology, which is a critical vulnerability from a Russian perspective. To address this, the strategy outlines six priorities: (a) supporting scientific research for ‘advanced’ development of AI; (b) building and developing software with AI; (c) increasing the accessibility and quality of the data needed for AI development; (d) increasing access to the computers and platforms needed for AI development; (e) increasing the number of AI professionals and informing

<sup>145</sup> President of Russia, [National open lesson ‘Russia focused on the future’], 1 Sep. 2017 (in Russian, author translation).

<sup>146</sup> Pecotic, A., ‘Whoever predicts the future will win the AI arms race’, *Foreign Policy*, 5 Mar. 2019; and Polyakova, A., ‘Weapons of the weak: Russia and AI-driven asymmetric warfare’, *A Blueprint for the Future of AI*, Brookings Institution, 15 Nov. 2018.

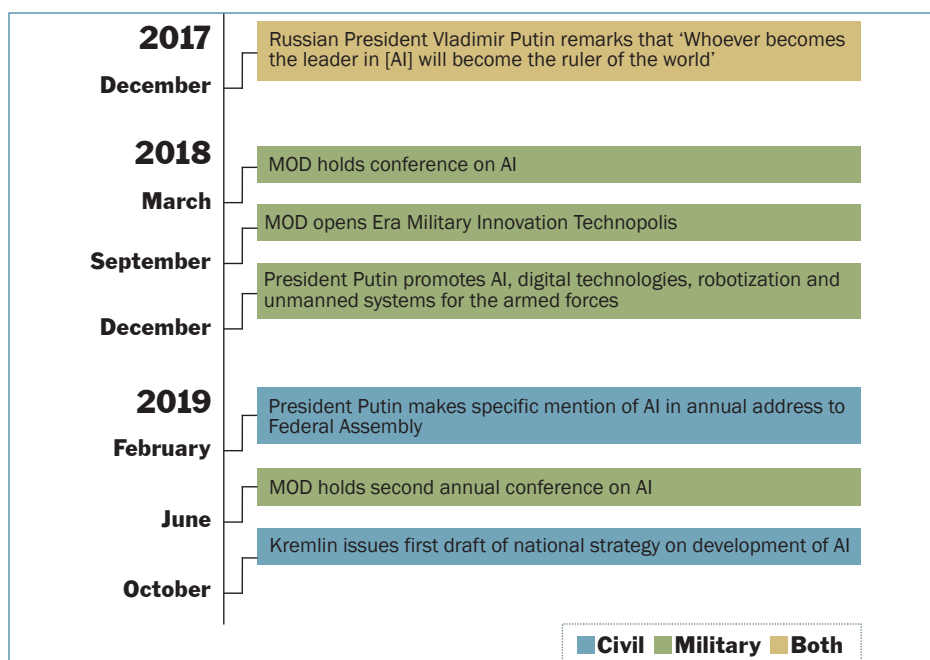
<sup>147</sup> Bendett, S., ‘Here is how the Russian military is organising to develop AI’, *Defense One*, 20 July 2018.

<sup>148</sup> President of Russia, ‘Presidential address to Federal Assembly’, 20 Feb. 2019.

<sup>149</sup> Yastrebova, S., [The government will work on artificial intelligence for Putin], *Vedomosti*, 27 Feb. 2019 (in Russian).

<sup>150</sup> [National strategy on development of artificial intelligence until 2030], Russian Presidential Decree no. 490, 10 Oct. 2019 (in Russian).





**Figure 3.2.** Recent policy developments related to artificial intelligence in Russia

AI = artificial intelligence, MOD = Ministry of Defence.

Sources: Russian Ministry of Defence, 'International Military and Technical Forum ARMY 2019: Business Programme'; and this volume: Boulanin et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, June 2020).

the public about the benefits of AI; and (f) creating a comprehensive system for the regulation of the social relations affected by the development and use of AI.<sup>151</sup>

The strategy has been described as ambitious, particularly in the light of its proposed timeline.<sup>152</sup> It aims to have implemented all of the listed measures by 2024 and to have achieved the level of advancement defined for each of these priorities by 2030.

The draft national strategy makes no reference to military applications of AI: its focus is purely civilian. The budgetary funds that are intended to support implementation of the national AI strategy will come from the national programme on the digital economy, which is estimated to be 1.2 trillion roubles (US\$18.4 billion) for the period 2019–24.<sup>153</sup> The areas of AI, big data and quantum technologies will receive 125.3 billion roubles (\$1.9 billion) from this budget. The civilian focus is remarkable given that Putin's 2017 statement on AI indicated that Russia saw the development of its civilian and military aspects as deeply interrelated.<sup>154</sup> For now, the government has kept its policy efforts on civilian and military AI separate.

<sup>151</sup> [National strategy on development of artificial intelligence until 2030] (note 150).

<sup>152</sup> E.g. Bendett, S., 'Sneak preview: First draft of Russia's AI strategy', *Defense One*, 10 Sep. 2019.

<sup>153</sup> Kantyshev, P., Bazanova, E. and Kodachigov, V., [The digital economy budget was estimated at 1.2 trillion roubles], *Vedomosti*, 20 Aug. 2018 (in Russian).

<sup>154</sup> President of Russia (note 145).

Russia's State Armament Programme for 2018–27 gives priority to AI-related technologies, such as automated command and control for the armed forces, battlefield control systems, battlefield visualization, and robotics.<sup>155</sup>

*What vision for military AI?*

President Putin's public statements on AI since 2017 have made clear that the upper echelons of the Russian Government see AI as a game-changing technology and that Russia will have to harness its potential, including in the military sphere, in order to uphold its great power ambitions. In 2018 Putin declared that 'digital technologies and artificial intelligence, robotisation, and unmanned systems—all this should be on the qualitative development agenda of our Armed Forces'.<sup>156</sup>

The 10-point plan on AI presented in March 2018 at an MOD-organized conference was primarily intended to identify R&D priorities and ways for Russia to explore them.<sup>157</sup> The 10 measures are to (a) form an AI and big data consortium; (b) gain expertise in automated systems; (c) create a state system for AI training and education; (d) build an AI laboratory at the Era Military Innovation Technopolis (a new campus in Anapa, Krasnodar Krai, for the development and implementation of hardware and software); (e) establish a national AI centre; (f) monitor global AI development; (g) hold AI war games; (h) assess whether new technologies comply with AI requirements; (i) discuss AI proposals at domestic military forums; and (j) hold an annual conference on AI.<sup>158</sup>

The 10 point-plan does not amount to a road map or strategy. It says nothing about how or whether Russia intends to use AI for military missions. The 10 measures are practical and focus on Russia's ability to develop its AI capabilities at the fundamental level. It is not known whether the MOD is currently working on a full military strategy on AI.

In these circumstances, without official sources, little can currently be said about the types of AI-enabled military capability that Russia finds of particular interest and how it plans to use them. However, as in the cases of China, India and Pakistan (see sections V, VI and VII), a review of the expert literature and of Russia's statements in the intergovernmental debate on LAWS provides some insights.

Russian military analysts believe that AI could be of critical importance for the following capabilities: battlefield and force management (i.e. to build mathematical models of tactical situations to plan operations and to calculate the amount of forces and resources necessary to implement tasks); integrated command, control, communications, computers, intelligence, surveillance and reconnaissance (C4ISR); remotely operated strike and reconnaissance; autonomous systems for

<sup>155</sup> Centre for Analysis of Strategies and Technologies (CAST), [State armament programmes of the Russian Federation: Problems of implementation and prospects for optimization] (CAST: Moscow, 2015) (in Russian), p. 22.

<sup>156</sup> President of Russia, 'Defence Ministry Board meeting', 18 Dec. 2018.

<sup>157</sup> Innovation Club, [Conference 'Artificial Intelligence: Problems and Solutions—2018'], 14–15 Mar. 2018 (in Russian); and Bendett (note 147).

<sup>158</sup> Bendett (note 147).

protection of certain high-value objects; battlefield security and force protection; and simulation and training.<sup>159</sup>

On the question of use, Russia's position on LAWS in the CCW context gives the impression that, like other nuclear-armed states, it wants to keep as much room for manoeuvre as possible to determine what constitutes responsible use of military AI. The Russian delegation at CCW meetings has, over the years, been one of the most vocal in—but also critical of—the intergovernmental discussion on LAWS. Russia has repeatedly criticized the premise on which the debate is held, noting that there cannot be meaningful discussion on LAWS unless these systems were clearly defined. This point was raised year after year to argue against proposals for elevating the discussion on LAWS to a more formal and politically binding level. In 2018 Russia eventually proposed a definition, which turned out to be rather ambiguous: 'an unmanned piece of technical equipment that is not a munition and is designed to perform military and support tasks under remote control by an operator, autonomously or using the combination of these methods'.<sup>160</sup> For Russia, the virtue of this—broad—formulation is that it distinguishes 'good' and 'bad' weapons.<sup>161</sup> At the same time, Russia saw the need to stress that the definition does not apply to existing systems with a high degree of automation and autonomy, including UAVs, since these comply with international humanitarian law—and so can be regarded as 'good' weapons.<sup>162</sup> Russia also repeatedly made the point that it believes that the use of LAWS can have both positive and negative humanitarian consequences and it would be impractical to define international standards, for instance on the concept of human control: 'it is doubtful whether criteria to determine a due level of "significance" of human control over the machine could be developed'.<sup>163</sup> For Russia, each state should develop its own standards. Russia's position in the discussion on human control is, in that regard, a guide to what it would consider to be a responsible standard for the use of AI weapon systems:

advanced as it may be, an autonomous system cannot perform its functions without a human behind it. Hence, the responsibility for the use of LAWS should be with the human who operates or programs the robot system and orders [the] use [of] LAWS.<sup>164</sup>

This position is similar to that of some other nuclear-armed states. Russia does not believe that direct human supervision and control is necessary to ensure the responsible use of weapon systems. The reason for this is that Russia, like other nuclear-armed states, has and is working on a number of weapon systems, notably strategic systems, that once activated operate without direct and continuous

<sup>159</sup> Burenok, V. M., Durnev, R. A. and Krukov, K. U., [Reasonable weapons: The future of artificial intelligence in military affairs], *Vooruzheniie i ekonomika*, no. 1(43) (Jan. 2018) (in Russian), pp. 4–13. Cited in Kashin, V., 'Artificial intelligence and military advances in Russia', ed. Saalman (note 2), pp. 39–42, p. 41.

<sup>160</sup> Certain Conventional Weapons Convention, Group of Governmental Experts on Lethal Autonomous Weapon Systems, 'Russia's approaches to the elaboration of a working definition and basic functions of lethal autonomous weapon systems in the context of the purposes and objectives of the Convention', Working paper submitted by Russia, CCW/GGE.1/2018/WP.6, 4 Apr. 2018, para. 2.

<sup>161</sup> Certain Conventional Weapons Convention, CCW/GGE.1/2018/WP.6 (note 160), para. 6.

<sup>162</sup> Certain Conventional Weapons Convention, CCW/GGE.1/2018/WP.6 (note 160), para. 9.

<sup>163</sup> Certain Conventional Weapons Convention, CCW/GGE.1/2018/WP.6 (note 160), para. 9.

<sup>164</sup> Certain Conventional Weapons Convention, CCW/GGE.1/2018/WP.6 (note 160), para. 11.

human supervision. One of them is the Poseidon UUV that is being developed and will be ready for the Russian Navy soon.<sup>165</sup>

In the absence of official sources that describe how Russia intends to use AI for nuclear deterrence-related purposes, revelations about the Poseidon programme received great attention within the expert communities on Russia and nuclear deterrence.<sup>166</sup> One of the questions of greater concern is whether Russia sees the possibilities offered by AI—be it for Russia or its adversaries—as a part of a more offensive nuclear posture, as happened in the cyber domain when new capabilities had an impact on Russia’s more offensive behaviour.<sup>167</sup> It was, in any case, perceived as evidence that Russia is looking into exploiting the recent advances in AI in its nuclear deterrence apparatus.

### **Adoption of military AI**

#### *Capability to adopt the most recent advances in AI for military purposes*

Russia has a number of assets that allow it to exploit the most recent advances in AI for military purposes. Official sources highlight three of these.<sup>168</sup> First, according to President Putin, is an information infrastructure that has allowed Russia to have one of the world’s highest penetration rates of ICT, but also one of the lowest costs for network access. Second is Russia’s strong education in mathematics, physics and software programming. Putin has highlighted that students from Russia have won the International Collegiate Programming Contest eight years in a row. The third asset is Russia’s innovative competitive software companies in areas such as computer vision, voice recognition and cybersecurity.

The national AI strategy foresees a number of challenges for Russia.<sup>169</sup> These challenges are in many ways similar to those identified by France, India and Pakistan (see sections IV, VI and VII). First, as Herman Gref, chief executive officer of Sberbank, puts it, Russia has a ‘massive shortage’ of human resources; according to Gref, it had only 6000–6500 AI researchers in 2017.<sup>170</sup> By way of comparison, the US company Amazon employs more than 10 000 people for just one AI product, the virtual assistant Alexa.<sup>171</sup> Second, Russia needs to improve its access to the data on which AI systems can be trained. Third, Russia is dependent on foreign technology for specialized hardware such as AI chips and 5G wireless, which is a fundamental weakness from a Russian perspective because it gives external

<sup>165</sup> See also Hwang and Kim (note 2).

<sup>166</sup> E.g. Horowitz, M. C., ‘Artificial intelligence and nuclear stability’, ed. Boulanin (note 7), pp. 79–83, pp. 81–82; and Topychkanov (note 57), pp. 74–75.

<sup>167</sup> Topychkanov (note 57), p. 74; and Connell, M. and Vogler, S., *Russia’s Approach to Cyber Warfare* (Center for Naval Analyses: Arlington, VA, Mar. 2017), pp. 8–9.

<sup>168</sup> President of Russia, [Meeting on artificial intelligence technology development], 30 May 2019 (in Russian).

<sup>169</sup> Meeting on artificial intelligence technology development (note 150).

<sup>170</sup> President of Russia, ‘Council for Strategic Development and Priority Projects meeting’, 5 July 2017; and President of Russia (note 168).

<sup>171</sup> President of Russia (note 168); and Hartmans, A., ‘Amazon has 10,000 employees dedicated to Alexa — here are some of the areas they’re working on’, *Business Insider*, 22 Jan 2019.

actors influence.<sup>172</sup> The MOD hopes to address some of these challenges with the Era Military Innovation Technopolis, which was opened in 2018. It is intended to host more than 800 laboratories, design centres and experimental facilities.<sup>173</sup> It is planned to carry out ‘complex applied and exploratory’ research and ‘advanced development’ in eight priority areas: information and telecommunications systems and AI systems; robotics; supercomputers; technical vision and pattern recognition; information security; nanotechnologies and nanomaterials; energy and life-support technologies and devices; and bioengineering, biosynthetic and biosensor technologies.<sup>174</sup>

It remains to be seen what impact the Era technopolis will have. A challenge that is rarely discussed openly in Russian official sources is that Russia continues to have difficulty in fostering a successful innovation ecosystem as a result of endemic corruption and the high level of state control—in both the civilian and the military spheres.<sup>175</sup> A recent demonstration of this was the failure of the Skolkovo Technopark. In 2009 Russia unveiled an ambitious plan to develop a Russian ‘Silicon Valley’ near Moscow. At first, the plan seemed a success: five years after the park was inaugurated, 30 000 people were reported to work on the campus, which received investment from foreign firms including Microsoft and IBM. However, it did not take long before key talent and investors started to leave—not only Skolkovo Technopark but Russia altogether—as problems of corruption emerged and cases of state interference multiplied.<sup>176</sup> The imposition of economic sanctions on Russia following the annexation of Crimea in 2014 further aggravated the dynamics.

In this context, it seems reasonable to conclude that, while Russia certainly has institutional resources to harness the recent advances in AI in the military, it faces a number of difficulties that currently makes it difficult to compete at the same level as China or the United States.

### *State of adoption of AI by the armed forces*

Russia has a long record of using software solutions to automate functions of military systems, include systems that are directly connected with its nuclear deterrent (see table 3.3). The most notable and discussed Russia technology is Perimetr (see box 2.3), the semi-automated command and control systems for nuclear retaliation, which was developed during the cold war by the Soviet Union. The system has reportedly been modernized recently.<sup>177</sup> However, generally speaking, it is hard to assess the level of sophistication and maturity of Russian military AI technology based on the publicly available information. What is

<sup>172</sup> President of Russia (note 168).

<sup>173</sup> Russian Ministry of Defence, ‘Russian Defence Ministry Board holds offsite meeting in Sevastopol’, 20 June 2018.

<sup>174</sup> Russian Ministry of Defence (note 173); and Russian Ministry of Defence, ‘Scientists of military technopolis to focus on developing supercomputer and seven more areas’, 14 Mar. 2018.

<sup>175</sup> Bateman, A., ‘Russia’s quest to lead the world in AI is doomed’, *Defense One*, 12 June 2019.

<sup>176</sup> Aspel, J., ‘The short life and speedy death of Russia’s Silicon Valley’, *Foreign Policy*, 6 May 2015.

<sup>177</sup> Khrolenko, A., [‘Perimeter’: How does the Russian system of retaliation work], RIA Novosti, 21 Aug. 2017 (in Russian); and Valagin (note 65).

**Table 3.3.** State of adoption of artificial intelligence in the Russian nuclear deterrence architecture

Application area	AI in use	Example or mention in official sources	Status	What is known about AI use
<i>Early warning and intelligence, surveillance and reconnaissance</i>				
AI for data collection and analysis	✓	National Defence Operations Centre <sup>a</sup>	Deployed	For collecting and organizing information
Remote sensing	✓	Kasatka avionics system for aircraft, helicopters and unmanned aerial vehicles <sup>b</sup>	Deployed	For automatic detection of submarines and surface, ground and air targets
<i>Command and control</i>				
Command and Control	✓	Perimetr automated reserve command-and-control system <sup>c</sup>	Deployed	For automated launch of a nuclear strike based on data from sensors, indicating a nuclear attack
<i>Precision strike and delivery</i>				
Air launched	✓	Autonomous on-board guidance and control systems for hypersonic cruise missiles <sup>d</sup>	Deployed	For on-board control for cruise missiles, and maintenance
Sea launched	✓	Poseidon nuclear-powered, nuclear-capable unmanned underwater vehicle <sup>e</sup>	R&D	Autonomous dual-use platform
Ground launched	✓	Burevestnik nuclear-powered, nuclear-capable cruise missile <sup>f</sup>	R&D	Long-range system, capable of automatically changing its path to avoid ballistic missile defence zones
Missile/air/space defence	✓	Land-based early warning radar <sup>g</sup>	R&D	AI-enabled radar to be a central part of the automated command and control of the Russian Aerospace Forces
<i>Other</i>				
Cyber/electronic information warfare	..	..	..	..
Physical security	✓	Nerekhta autonomous combat vehicle <sup>h</sup>	Deployed	Can select and eliminate targets in a fully automated mode

.. = no or unclear, ✓ = yes, AI = artificial intelligence, R&D = research and development.

<sup>a</sup> Russian Ministry of Defence, [National Defence Operations Centre of the Russian Federation], [n.d.] (in Russian); and Ramm, A. and Lavrov, A., [In the centre of the storm: How military operators protect national security], *Izvestiya*, 30 Dec. 2019 (in Russian).

<sup>b</sup> [‘Radar MMS’: Artificial intelligence system for aircraft and drones has been created in Russia], Radar MMS, 25 Aug. 2018 (in Russian).

<sup>c</sup> Valagin, A., [Assured retaliation: How the Russian ‘Perimetr’ system works], *Rossiiskaya Gazeta*, 22 Jan. 2014 (in Russian).

<sup>d</sup> Tactical Missiles Corporation JSC, Granit-Electron Concern, [Autonomous on-board control systems, guidance systems for supersonic and hypersonic missiles, system testing and control equipment], 2017 (in Russian).

<sup>e</sup> ‘Key stage of Poseidon underwater drone trials completed, says Putin’, TASS, 2 Feb. 2019; and ‘Russia begins testing of “Poseidon” underwater nuclear drone’, PressTV, 26 Dec. 2018.

<sup>f</sup> Ramm, A., [Winged ‘Burevestnik’: What is known about Russia’s secret weapon], *Izvestia*, 5 Mar. 2019 (in Russian).

<sup>g</sup> Kozachenko, A. and Ramm, A., [Tracking from far: Air defence forces receive a land-based AWACS], *Izvestia*, 15 Sept. 2019 (in Russian).

<sup>h</sup> [Combat robot ‘Nerekhta’ protected the ‘Topol-M’ at the exercises near Irkutsk], RIA Novosti, 31 Mar. 2016 (in Russian).

relatively certain is that AI technology in Russian systems, as in US technologies, is rather brittle.

Russia is in the middle of a nuclear modernization process. It has reportedly already replaced 82 per cent of the weapons and equipment of the Strategic Rocket Forces with new systems.<sup>178</sup> In addition to existing platforms, it has recently revealed several new offensive weapons, including the Burevestnik nuclear-powered long-range cruise missile, the Poseidon nuclear-powered UUV, the Kinzhal air-launched supersonic missile, the Sarmat silo-based heavy ballistic missile, and the Avangard boost-glide system.<sup>179</sup> Unlike the US case, the budget for the nuclear modernization of Russia is hard to evaluate.

Research, development, modernization and procurement programmes show that AI is intended to play an essential part of many future Russian military systems, including nuclear-related systems.<sup>180</sup> For instance, in August 2018 it was reported that the modernized Tupolev Tu-22M3M strategic dual-capable bomber has been equipped with AI. At the roll-out ceremony of the first upgraded bomber prototype, Lieutenant General Sergei Kobylash, commander of long-range aviation, said: ‘The capabilities of this aircraft are impressive and considerably surpass all similar foreign rivals. This plane has artificial intelligence’.<sup>181</sup> But here again, too little information is available to corroborate whether AI is actually used, for what purpose and to what extent.

Machine learning is a technology that Russian developers have started exploring only recently, partly in response to the increasing level of attention that it has received globally. The critical areas that the Russian armed forces expect to be reinforced by AI and machine learning in the near future are automated image recognition of satellite imagery and analysis of data from early-warning

<sup>178</sup> Kristensen, H. M. and Korda, M., ‘Russian nuclear forces, 2020’, *Bulletin of the Atomic Scientists*, vol.76, no. 2 (Mar. 2020), pp. 102–17, p. 102.

<sup>179</sup> Woolf, *Russia’s Nuclear Weapons* (note 64), p. 19.

<sup>180</sup> Stefanovich, D., ‘Artificial intelligence advances in Russian strategic weapons’, ed. Topychkanov (note 19), pp. 25–29.

<sup>181</sup> TASS, ‘Russia’s upgraded Tu-22M3 strategic missile-carrying bomber gets artificial intelligence’, 16 Aug. 2018.

radars and satellites.<sup>182</sup> This indicates that machine learning remains detached from strategic command and control and from nuclear force delivery.

### III. The United Kingdom

#### **Vision and policies**

##### *AI on the political agenda*

The United Kingdom has played a major role in the history of artificial intelligence. Many past and present leading scholars of AI and machine learning, from the pioneer Alan Turing to Geoffrey Hinton, the father of deep learning, were educated in the UK. AI has historically been a relatively well-established and well-funded field in the UK. However, taking note of the breakthroughs in AI in the early 2010s, the British Government has since 2015 identified the need to dedicate more resources and make a more concerted policy effort to ensure that the UK can remain at the global forefront in the field.

The government commissioned a series of reports and studies that were to explore how the UK should best embrace the current AI renaissance.<sup>183</sup> This led to the publication in 2018 of the AI Sector Deal, an industrial strategy that lays out a plan for how the UK can maintain its leadership in the field.<sup>184</sup> This is the highest-level and most mature policy document on AI that the British Government has published in recent years (see figure 3.3).

The AI Sector Deal indicates that the UK has bold ambitions: it should become ‘the best in the world’.<sup>185</sup> To that end, it articulates a series of concrete measures that the UK is to implement. These are primarily aimed at fostering responsible innovation and ensuring that the British economy is competitive in the area of AI. It includes an investment plan on R&D, as well as the creation of governmental institutions that try to increase cooperation between business, academia and government: the AI Council, the Office for Artificial Intelligence, and the Centre for Data Ethics and Innovation (CDEI).<sup>186</sup>

The AI Sector Deal and the parliamentary reports that preceded it deal primarily with civilian uses of AI. There is little or no mention of policy priorities related to military applications of AI. However, two issues are viewed through a national security lens: access to data—particularly the risk of monopolization of data by big foreign companies—and cybersecurity.<sup>187</sup>

<sup>182</sup> [Dual-use artificial intelligence], Era Techopolis (in Russian).

<sup>183</sup> British House of Commons, Science and Technology Committee, *Robotics and AI*, 5th report of session 2016–17 (House of Commons: London, 12 Oct. 2016); British House of Lords, Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able?*, Report of session 2017–19 (House of Lords: London, 16 Apr. 2018); and British Ministry of Defence, Development, Concepts and Doctrine Centre (DCDC), *Human–Machine Teaming*, Joint Concept Note 1/18 (DCDC: Swindon, May 2018).

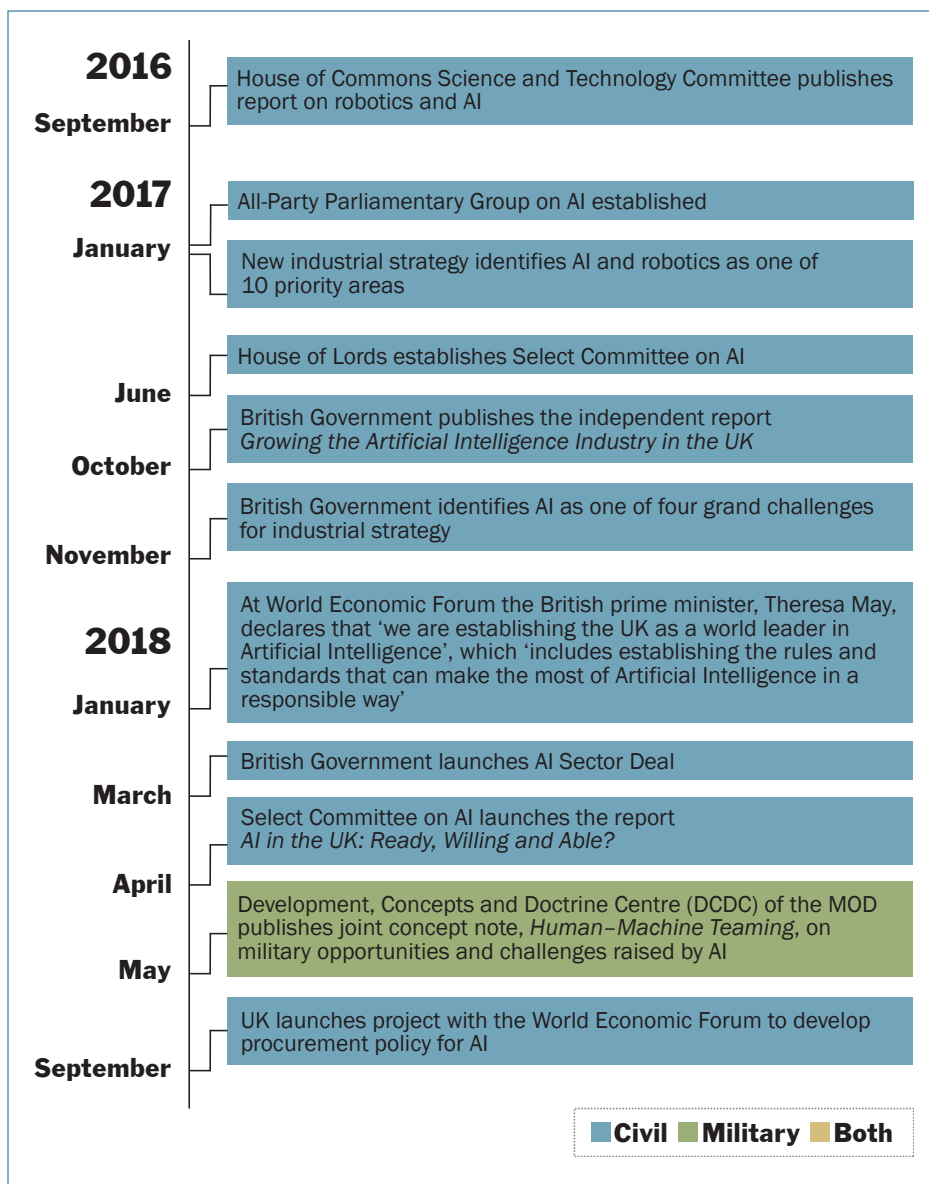
<sup>184</sup> British Government, *Industrial Strategy: Artificial Intelligence Sector Deal* (Department for Business, Energy and Industrial Strategy: London, 2018).

<sup>185</sup> British Government (note 184), p. 7.

<sup>186</sup> British Government (note 184), p. 31.

<sup>187</sup> British Government (note 184), pp. 11; British House of Commons (note 183), pp. 16–22; and British House of Lords (note 183), pp. 99–102.





**Figure 3.3.** Recent policy developments related to artificial intelligence in the United Kingdom

AI = artificial intelligence, MOD = Ministry of Defence.

Sources: Hall, W. and Presenti, J., 'Growing the artificial intelligence industry in the UK', 15 Oct. 2017; British Government, Industrial Strategy, Building a Britain Fit for the Future, White paper (Department for Business, Energy & Industrial Strategy, London, Nov. 2017); World Economic Forum, 'UK Government first to pilot AI procurement guidelines co-designed with World Economic Forum', 20 Sep. 2019; and this volume: Boulanin et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, June 2020).

The high-level defence policy documents that the British Government has published since 2015 also say little about AI. The 2015 National Security Strategy and Strategic Defence and Security Review does not mention the term AI, while the 2018 National Security Capability Review makes a brief reference to AI as a key capability area for the British armed forces.<sup>188</sup> The only concrete measure that is listed in the National Security Capability Review is a need for more experimentation to understand the opportunities and threats from autonomy and AI.<sup>189</sup>

In this light, it seems that the UK has not yet made AI a top or central priority in its military modernization plans. However, it is clear that AI is a technology that the UK intends to take into consideration to ‘maintain [its] competitive advantage in the immediate term and for the decades to come’.<sup>190</sup>

### *What vision for military AI?*

The relatively few mentions of AI in general defence policy does not mean that the UK has not already articulated a comparatively mature vision of the role that AI could play in the future of warfare. In 2018 the British Ministry of Defence published a number of reports and doctrinal documents that outline the UK’s view of the potential impact of AI on the future of warfare.<sup>191</sup>

Among these reports, a joint concept note, *Human–Machine Teaming*, by the MOD’s Development, Concepts and Doctrine Centre (DCDC) provides the most detail on what the UK sees as the military opportunities and challenges raised by AI. In terms of opportunities, it explains that ‘Robotics and [AI] offer the potential for another inflexion point in delivering military transformation and advantage’.<sup>192</sup> Development of AI and robotics will allow ‘the ability to scale physical mass and battlefield points of presence increasingly independent of the numbers and locations of human combatants; extending the reach and persistence of our [ISR] and weapon systems; and information advantage for understanding, decision-making, tempo of activity and assessment’.<sup>193</sup>

The report underscores that the real strategic challenge for the UK will not just be staying ahead of competitors in terms of technological advances, but also determining the best way to use AI in military contexts. As the title of the report indicates, the MOD believes that the best way to gain military advantage with AI is to combine its strengths with those of humans: ‘Developing the right blend of human–machine teams—the effective integration of humans and machines into our war fighting systems—is the key’.<sup>194</sup> As the report goes on to demonstrate, this

<sup>188</sup> British Government, *National Security Strategy and Strategic Defence and Security Review 2015: A Secure and Prosperous United Kingdom*, Command Paper no. 9161 (Her Majesty’s Stationery Office: London, Nov. 2015); and British Government, *National Security Capability Review* (Cabinet Office: London, Mar. 2018), p. 39.

<sup>189</sup> British Government, *National Security Capability Review* (note 188), p. 16.

<sup>190</sup> British Ministry of Defence (MOD), *Mobilising, Modernising & Transforming Defence*, A report of the Modernising Defence Programme (MOD: London, 2018), p. 13.

<sup>191</sup> E.g. British Ministry of Defence (note 190); and British Ministry of Defence (note 183).

<sup>192</sup> British Ministry of Defence (note 183), p. iii.

<sup>193</sup> British Ministry of Defence (note 183), pp. 1–2.

<sup>194</sup> British Ministry of Defence (note 183), p. iii.

is not only a technical challenge; it also raises political, ethical and institutional questions. The UK will have to determine not only how to develop and acquire the best technologies but also make the organizational and institutional changes required to ensure that the British armed forces can make the best human-machine combinations.<sup>195</sup>

The MOD's priority in capability development appears to be to retain and enhance the capability to develop military applications of AI, and robotics more broadly, within the UK. The DCDC report notes that the MOD faces three challenges.<sup>196</sup> First is the difficulty of accessing the necessary data sets to train and test MOD-specific AI solutions. Second is the shortage of skilled AI engineers who can get the necessary security clearance to design, test and use the technologies for the armed forces. Third is the problem of finding AI components that meet the necessary cybersecurity standards to be used in British military platforms. Resolution of these problems is deemed essential to the ability of the UK to use AI for military purposes in the future.

There are many AI capabilities and applications that the MOD views as being essential for future military capabilities, including the following.

1. *Improving coverage of the battlefield and automating information processing and management cycles.* These capabilities include unmanned automated ISR platforms and software that can 'pre-filter, fuse and classify all data flows, eliminate paralysing information overload, and accelerate the observe, orient, decide and act (OODA) loop of decision-makers'.<sup>197</sup>
2. *Making the logistics chain more agile and less manpower intensive.* These include self-driving transport vehicles and automated logistic monitoring software.<sup>198</sup>
3. *Increasing the persistence, reach, mass and precision of weapon systems.* Notably, these include loitering munitions, an automated technology that could deliver 'step changes in military capability'.<sup>199</sup>
4. *Enhancing the capability to fight cyberwar.*<sup>200</sup>

None of the open-source official documents give any indication of whether or how the MOD plans to develop, adopt or use AI specifically for nuclear force-related purposes. However, it can be deduced from the DCDC report that the MOD is likely to see great potential for AI in nuclear-related ISR tasks.<sup>201</sup> The report places a lot of emphasis on the opportunities that AI generates for information collection and analysis, and how this optimizes the time for decision-making. At the same time, the report also shows that the MOD is aware of the risks and depend-

<sup>195</sup> British Ministry of Defence (note 183), p. 3.

<sup>196</sup> British Ministry of Defence (note 183), pp. 7–8.

<sup>197</sup> British Ministry of Defence (note 183), p. 16.

<sup>198</sup> British Ministry of Defence (note 183), p. 22.

<sup>199</sup> British Ministry of Defence (note 183), p. 12.

<sup>200</sup> British Ministry of Defence (note 183), p. 26.

<sup>201</sup> British Ministry of Defence (note 183), pp. 14–17, 29–36.

encies that AI creates for command and control.<sup>202</sup> In this light, it seems unlikely that the UK will delegate higher-order nuclear-related decisions to machines or automate decision-making without tight human control. Given the strategic importance that the USA plays within the UK's Trident nuclear programme, it is likely that future US decisions on the integration of AI into nuclear weapon systems will influence those of the UK.

### **Adoption of military AI**

#### *Capability to adopt the most recent advances in AI for military purposes*

The UK has the capability to be among the countries that adopt AI for military purposes the most quickly and most intelligently. Statements from various governmental sources, including Theresa May when she was prime minister (2016–19), show that the UK perceives itself to be a global leader in the field of AI.<sup>203</sup>

However, these sources also indicate that the strengths of the UK lie primarily in its academic and research base: for example, assessing that 'The UK has a strong pedigree in the theory and algorithmic side of AI development'.<sup>204</sup> One illustration of this point is the UK's leading AI company, DeepMind (now a subsidiary of Google), arguably one of the world's most cutting-edge AI companies, is primarily known for its groundbreaking contributions to fundamental and applied research. A number of research projects currently being conducted by the MOD and British defence research laboratories (e.g. the MOD's Defence Science and Technology Laboratory, DSTL, and the private-sector company QinetiQ) indicate that the UK has the capability to research and develop complex military applications of AI. These include technologies that could help the British armed forces to access the capabilities that the DCDC identifies as being of interest: self-driving vehicles, software for designing swarming robotic systems, and on-board and off-board ISR analysis systems.<sup>205</sup> Many were funded through an £800 million (US\$1 billion) innovation fund launched in 2016.<sup>206</sup> Some of the capabilities developed in the framework of these projects were tested by the British Army in a four-week experiment in November 2018.<sup>207</sup>

The fundamental challenge for the UK will be to translate the products of AI research into marketable applications and operationally viable capabilities. The DCDC report explains that the 'industrial manufacturing base in the UK is weaker than the research base, partly due to . . . poor exploitation of research

<sup>202</sup> British Ministry of Defence (note 183), pp. 39–52.

<sup>203</sup> May, T., British Prime Minister, Address at the World Economic Forum, Davos, 25 Jan. 2018; British Government (note 184), p. 4; British House of Commons (note 183), pp. 30; British House of Lords (note 183), p. 129; and British Ministry of Defence (note 183), p. 7.

<sup>204</sup> British Ministry of Defence (note 183), p. 7.

<sup>205</sup> Boulanin (note 126), pp. 38–39.

<sup>206</sup> Defence Online, 'Defence Innovation Funds set to unearth defence and security pioneers', Defence Contracts Online, 30 Jan. 2017.

<sup>207</sup> British Ministry of Defence, 'Army start biggest military robot exercise in British history, Defence Secretary announces', 12 Nov. 2018.

base innovations in the past'.<sup>208</sup> One particular weak point of the British industry relates to the development of subcomponents, notably computer chips: most manufacturers are based in Asia or the United States. However, the UK does host a number of arms-producing companies, notably BAE Systems, that are well-established systems integrators capable of manufacturing highly sophisticated military technology.<sup>209</sup> Another important comparative advantage is that the UK is well positioned to access US technologies given that it is a close ally of the USA and that the British arms-producing companies are also well established in the US military-industrial base.

### *State of adoption of AI by the armed forces*

The UK is currently engaged in a nuclear modernization programme which includes replacing its Vanguard-class submarines with the new Dreadnought class. The first of this class will enter service in the early 2030s. It is expected that the Dreadnought class will constitute the backbone of the UK's Continuous At-Sea Deterrent until the 2060s. The total cost of the Dreadnought programme is estimated to be \$47.4 billion.<sup>210</sup> There is no official information on the role that AI plays, or could play, in the UK nuclear modernization program.<sup>211</sup> However, it is likely that AI will play some role given that, in some ways, the UK has already made significant progress in the adoption of AI technology.

The British armed forces' most sophisticated weapon platforms already include some subsystems or capabilities that have been enabled by AI research, such as ATR and voice command (see table 3.4). The UK is also working on research projects in a large number of areas including some that could be of direct relevance for its nuclear deterrence capacity.

In the area of ISR, for example, a recent achievement has been the Sensing for Asset Protection using Integrated Electronic Network Technology (SAPIENT) programme, which uses automation and AI to process ISR footage.<sup>212</sup> The UK has also developed several force-delivery platforms with a great degree of autonomy. A notable example is the Brimstone fire-and-forget missile, which can autonomously find a predefined target in a predefined area.<sup>213</sup> Another major notable achievement is the Taranis, a prototype stealth UCAV that would reportedly be capable of conducting strikes autonomously in a communications-denied environment. These air- and ground-launched systems do not play a role in the UK's nuclear deterrence architecture, which is entirely sea-based. However, they indicate that the UK has systems at its disposal that could allow it to conduct strategic conventional strikes.

<sup>208</sup> British Ministry of Defence (note 183), p. 7fn.

<sup>209</sup> Fleurant, A. et al., 'The SIPRI Top 100 arms-producing and military services companies, 2017', SIPRI Fact Sheet, Dec. 2018.

<sup>210</sup> Kile, S. N. and Kristensen, H. M., 'British nuclear forces', *SIPRI Yearbook 2019* (note 8), pp. 311–12.

<sup>211</sup> Kile and Kristensen (note 210).

<sup>212</sup> British Ministry of Defence, 'Streets ahead: British AI eyes scan future frontline in multinational urban experiment', 24 Sep. 2018.

<sup>213</sup> Missile Defense Project, 'Brimstone', Missile Threat, Center for Strategic and International Studies (CSIS), 18 Apr. 2019.

**Table 3.4.** State of adoption of artificial intelligence in the British nuclear deterrence architecture

Application area	AI in use	Example or mention in official sources	Status	What is known about AI use
<i>Early warning and intelligence, surveillance and reconnaissance</i>				
AI for data collection and analysis	✓	Sensing for Asset Protection using Integrated Electronic Network Technology (SAPIENT) <sup>a</sup>	R&D	Can autonomously process and select ISR information
Remote sensing	✓	XLUUV programme <sup>b</sup>	R&D	Would need to include some autonomous navigation capabilities
<i>Command and control</i>				
Command and control	..	..	..	..
<i>Precision strike and delivery</i>				
Air launched	✓	Brimstone fire-and-forget missile <sup>c</sup>	Production	Missile with a conventional payload that can autonomously find a predefined target in a predefined area
Missile/air/space defence	✓	Sea Ceptor air defence system <sup>d</sup>	Deployed	Uses an active radar seeker
<i>Other</i>				
Cyber/electronic information warfare	..	Darktrace <sup>e</sup>	Deployed	Commercial technology that can autonomously detect and respond to cyberattacks
Physical security	✓	C-Hunter semi-submersible autonomous surface vehicle <sup>f</sup>	Deployed	Includes autonomous navigation capability

.. = no or unclear, ✓ = yes, AI = artificial intelligence, ISR = intelligence, surveillance and reconnaissance, XLUUV = extra-large unmanned underwater vehicle.

<sup>a</sup> British Ministry of Defence, 'Streets ahead: British AI eyes scan future frontline in multinational urban experiment', 24 Sep. 2018.

<sup>b</sup> Kumar, H. and Hussein, T., 'UK MoD invite proposal for autonomous UUV', Naval Technology, 18 Apr. 2019.

<sup>c</sup> Missile Defense Project, 'Brimstone', Missile Threat, Center for Strategic and International Studies (CSIS), 18 Apr. 2019.

<sup>d</sup> MBDA Missile Systems, 'Sea Ceptor'.

<sup>e</sup> It is not known if the British armed forces are using this technology. See Darktrace.

<sup>f</sup> L3 Harris, 'Product information: C-Hunter'.

## IV. France

### Vision and policies

#### *AI on the political agenda*

Since 2017 artificial intelligence has emerged as a top priority area at the policy level in France (see figure 3.4). It started with the publication by the Ministry of Higher Education and Research and the Ministry of Economy and Finance of a national strategy for AI research and competitiveness.<sup>214</sup> The strategy set out the ambition and a course of action for France to become and remain a major player in AI. The strategy focuses solely on the civilian sphere, but the same year the defence minister, Jean-Yves Le Drian, announced publicly that AI would be a priority area of the forthcoming military planning law (*Loi de Programmation militaire*) for 2019–25—a multi-year plan that guides France’s military investment.<sup>215</sup>

France’s strategic interest in AI continued and strengthened following the election of Emmanuel Macron as president. Macron ordered a parliamentary report that could serve as the basis for a new strategy that would cover both the civilian and military sectors and would establish a framework to ensure that France can be an innovative but responsible champion of AI. The 150-page parliamentary report, *For a Meaningful Artificial Intelligence*, published in March 2018, represents France’s national AI strategy.<sup>216</sup> It presents a number of measures intended to reinforce France’s innovation ecosystems with the aims of attracting foreign companies and talent; developing a data policy; creating a regulatory and financial framework for AI research projects and start-ups; and giving thought to AI regulation and ethics. It identifies four key sectors for AI development in France: healthcare, the environment, transport mobility, and defence and security.

The national AI strategy and other sources indicate that France sees the pursuit of AI in the civilian and military spheres as fundamentally interlinked.<sup>217</sup> Given the dual-use nature of the technology, the development of a productive ecosystem for civilian AI innovation is perceived as being essential for the development of military applications. Policy challenges related to data accessibility, ethics, data-processing capacity, safety and security are also seen as cross-cutting.

However, the Ministry of the Armed Forces (MAF, as the defence ministry is now known) has also made it clear that it sees the pursuit of military AI as a critical enabler of France’s future strategic autonomy. In that regard, Florence Parly, the armed forces minister, announced in March 2018 a plan to increase spending on AI to €100 million (US\$123 million) annually as part of an innovation drive to develop future weapon systems.<sup>218</sup> About half of that amount would fund R&D,

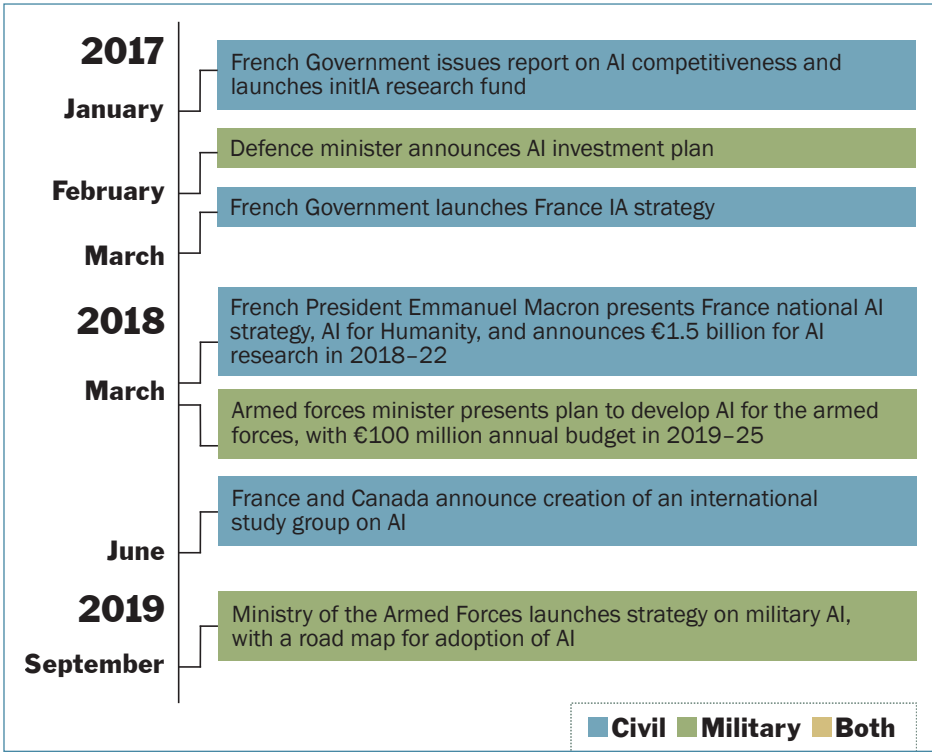
<sup>214</sup> French Government, *France intelligence artificielle: Rapport de synthèse* [France artificial intelligence: Synthesis report] (French Government: Paris, 2017).

<sup>215</sup> Barbaux, A., ‘« L’intelligence artificielle est un élément de souveraineté nationale », selon Jean-Yves Le Drian’, *L’Usine Nouvelle*, 16 Feb. 2017.

<sup>216</sup> Villani, C., *For a Meaningful Artificial Intelligence: Toward a French and European Strategy* (Conseil national du numérique: Paris, Mar. 2018).

<sup>217</sup> Villani (note 216).

<sup>218</sup> Tran, P., ‘France to increase investment in AI for future weapon systems’, *Defense News*, 16 Mar. 2018.



**Figure 3.4.** Recent policy developments related to artificial intelligence in France

AI = artificial intelligence.

Sources: French Government, ‘France IA, la stratégie française en intelligence artificielle’ [France AI, the French artificial intelligence strategy], 20 Mar. 2017; Bains, N. S., Canadian Minister of Innovation, Science and Economic Development, and Vidal, F., French Minister of Higher Education, Research and Innovation, ‘Canada–France statement on artificial intelligence’, 7 June 2018; and this volume: Boulanin et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, June 2020).

and €10 million each year would be dedicated to testing and integration of existing AI technology. The plan also included recruitment by the Direction générale de l’armement (DGA), the defence procurement agency, of 50 AI specialists by 2022.

Approximately a year after she announced the MAF’s investment plan, Parly outlined in a public speech at Saclay in April 2019 the vision for how the ministry intends to use AI in its future military modernization plans.<sup>219</sup> Parly described AI as a strategic element of great power competition and an ‘indispensable’ enabler of France’s future operational superiority: ‘[France] cannot risk missing this technological shift’.<sup>220</sup> At the same time, the speech also underlined that France sees AI as a young technology that is not yet sufficiently mature to be safely used in the critical systems of the armed forces.

<sup>219</sup> Parly, F., Minister of the Armed Forces, ‘Intelligence artificielle et défense’ [Artificial intelligence and defence], Speech, Saclay campus, French Ministry of the Armed Forces, 5 Apr. 2019.

<sup>220</sup> Parly (note 219) (author translation).



The MAF subsequently created a task force that would investigate how France could responsibly adopt AI for military purposes. Its report, published in September 2019, presents the MAF's strategy on military AI.<sup>221</sup> It elaborates on the MAF's priorities in that domain and present a road map for adoption of AI by the French armed forces. Key measures include the creation of an ethics committee at the ministerial level, sensitization measures to mainstream the use of AI by the armed forces, technical measures to enable the development of 'trusted' AI applications, and guidelines on data governance (including data collection, data protection, and data processing and storage).

*What vision for military AI?*

Parly's speeches, the report of the task force and the 2018 national AI strategy make clear that France sees a central role for AI in its future military capabilities. AI could enable swifter data and information management, greater ability to anticipate operational developments and needs, new possibilities to protect military personnel (notably through the use of robotic systems), and more efficient manpower and logistical management.<sup>222</sup>

These documents also indicate that the MAF sees a number of risks associated with the advance of AI in the military sphere. From an operational standpoint the MAF is worried by the fact that AI remains an immature technology whose vulnerabilities (e.g. brittleness, dependence on data) could be exploited by adversaries in ways that could dramatically incapacitate the ministry's decision-making and operational capabilities.<sup>223</sup> The MAF also see risks at the more strategic level. For example, the AI task force report points out that the use in influence and disinformation operation of AI deep fakes—exceptionally accurate but false still and moving images—could be politically destabilizing for democracy.<sup>224</sup> It also stresses that the advance of military AI capabilities could generate strategic insecurity among adversaries and give them incentives for more escalatory behaviour in time of crisis.

Official publications show that France sees a number of ethical challenges related to adoption of AI, particularly with regard to how it affects the role of humans in warfare. For France, AI cannot and should not displace humans as the central actors of military affairs: according to Parly, 'AI is not an end in itself, it must be a support for more informed, faster decisions, a tool for lucidity for strategic and tactical decision makers' and 'Regardless of the degree of automation or autonomy of our current and future weapon systems, they will remain subject to human command'.<sup>225</sup>

<sup>221</sup> French Ministry of the Armed Forces (MAF), *L'intelligence artificielle au service de la défense* [Artificial intelligence at the service of defence], Report of the AI task force (MAF: Paris, Sep. 2019).

<sup>222</sup> Tran (note 218); Parly (note 219); French Ministry of the Armed Forces (note 221); and Villani (note 216).

<sup>223</sup> French Ministry of the Armed Forces (note 221), pp. 6–7.

<sup>224</sup> French Ministry of the Armed Forces (note 221), p. 8. On deep fakes see also chapter 4 in this volume.

<sup>225</sup> Parly (note 219) (author translation). See also Villani (note 216), p. 126.

The development of LAWS is depicted in that regard as a redline that France does not want to cross. This was made clear by President Macron himself following the launch of the national AI strategy:

you always need a human check. And in certain ways, a human gateway. At a point of time, the machine can prepare everything, can reduce uncertainties, can reduce until nil the uncertainties and that's an improvement which is impossible without it, but at a point of time, the go or no-go decision should be a human decision because you need somebody to be responsible for it.<sup>226</sup>

However, France's definition of LAWS is narrow: it must be 'capable of moving, adapting to its land, marine or aerial environments and targeting and firing a lethal effector (bullet, missile, bomb, etc.) without any kind of human intervention or validation'.<sup>227</sup> Thus, France does not entirely preclude the use of autonomous systems but focuses on mobile weapon systems outside human supervision. One of the reasons for this is that the French armed forces believe that the development of autonomous systems and AI more generally can provide multiple operational and humanitarian benefits: it could allow French forces to increase force efficiency and reduce risks for military personnel while improving compliance with obligations of international law.<sup>228</sup> However, the French Government worries that other actors might exploit AI in ways that would totally disregard the standards of international law. That is reportedly why France put the issue of LAWS on the agenda of the CCW Convention in 2013.<sup>229</sup> It hopes that the CCW framework is the right one for states to discuss possible normative principles to govern advances in AI and autonomy in weapon systems and to limit the risks associated with their misuse.

Among the MAF's priorities with regard to adoption of AI, the foremost is to ensure that the armed forces have the 'fuel'—that is, the data and talent—necessary to use AI at the most fundamental level.<sup>230</sup> That requires investing in raw data-processing capability, notably through the development of internal cloud-based computing, storage and systems to facilitate the sharing of data between the different parts of the MAF, and by recruiting more AI experts. In terms of concrete capabilities and applications, the ministry has already identified six priority areas, which closely resemble those outlined by the United States, the United Kingdom and to some extent China (see sections I, III and V).<sup>231</sup>

1. *AI for planning decision making.* France expects that AI will allow better informed decisions to be made within increasingly short time frames. These include using AI-enabled speech interfaces for

<sup>226</sup> Thompson, N., 'Emmanuel Macron talks to Wired about France's AI strategy', *Wired*, 31 Mar. 2018.

<sup>227</sup> Certain Conventional Weapons Convention, Group of Governmental Experts on Lethal Autonomous Weapon Systems, 'Characterization of LAWS', Non-paper submitted by France, 11–15 Apr. 2016.

<sup>228</sup> Certain Conventional Weapons Convention, Group of Governmental Experts on Lethal Autonomous Weapon Systems, 'Human-machine interaction in the development, deployment and use of emerging technologies in the area of lethal autonomous weapons systems', Working paper submitted by France, 28 Aug. 2018, CCW/GGE.2/2018/WP3.

<sup>229</sup> Certain Conventional Weapons Convention, Meeting of Parties, 2013 Session, Final Report, CCW/MSP/2013/10.

<sup>230</sup> Parly (note 219) (author translation).

<sup>231</sup> French Ministry of the Armed Forces (note 221), p. 19.

better human–machine teaming in combat and transport aircraft and AI-enabled decision-support systems that can help a command-and-control centre to fuse data, to evaluate the state and potential of armed forces, and to anticipate the possible consequences of future manoeuvres.

2. *ISR collection and analysis.* France is particularly interested in the ability to automatically search, fuse and cross-reference data collected from various intelligence sources, from satellite imagery to content gathered from the Dark Web.
3. *Collaborative operations.* France believes that AI could help its armed forces to better integrate their various weapon systems. Notably, it is interested in the development of a command-and-control architecture that would automatically distribute target suggestions to various systems, and an architecture that would allow systems to operate as a swarm or coordinated groups of heterogeneous weapon systems.
4. *Robotics for dull, dirty and dangerous tasks.* These tasks include underwater demining and rescue missions. France wants to invest in technologies that will enhance the navigational autonomy of robotic systems in complex environments as well as their ability to detect targets and any object or situation of interest.
5. *Cyberwarfare.* AI is viewed as a critical tool to ensure that French digital assets can be protected against cyberattacks. For France, AI could enable the automated detection of cyberattacks, facilitate the identification of vulnerabilities, improve the analysis and anticipation of cyber-threats, and provide support for the conduct of cyberwarfare operations, both defensive and offensive.
6. *Logistics and maintenance.* France is particularly interested in AI applications, such as predictive maintenance, that would reduce the operating and maintenance costs of current and future equipment.<sup>232</sup>

The official documents make no explicit connection between AI and France's nuclear capability. However, French official sources describe AI as a cross-cutting technology that will benefit all the missions of the armed forces.<sup>233</sup> There is little doubt that France sees a role for AI in the future of its nuclear deterrent. In fact, nearly all the capabilities listed above have relevance to France's deterrence capability. The repeated references to the immaturity of AI technology and the safety risks associated with premature use and the various statements on the need to maintain sufficient human control indicate, however, that France may be cautious about the integration of AI technology into its nuclear apparatus.

<sup>232</sup> French Ministry of the Armed Forces (note 221), pp. 6–7; and French Ministry of the Armed Forces (MAF), *Imager au-delà: Document d'orientation de l'innovation de défense* [Thinking beyond: Guidance document on defence innovation] (MAF: Paris, July 2019).

<sup>233</sup> French Ministry of the Armed Forces (note 232), p. 13.

## Adoption of military AI

### *Capability to adopt the most recent advances in AI for military purposes*

France is relatively well equipped to adopt and use the most recent advances in AI for military purposes.<sup>234</sup> It has a strong academic and research base: every year it trains around 1000 students in 18 master programmes; it has 250 research groups and a total of 5300 researchers working on AI-related topics in different disciplines; and it is the fourth-most productive country in terms of academic articles on AI, and in some specific topics it is leading the field.<sup>235</sup> On the industrial side, it can count on a network of 80 small- and medium-sized enterprises and 270 start-ups that work on AI and a handful of large industrial groups, including arms-producing companies, that have strong expertise on AI.<sup>236</sup> The latter includes Atos, Dassault, Thales and the trans-European company MBDA. Thus, France has the expertise and the know-how necessary to be an important player in the field of AI. Following the publication of the national AI strategy in March 2018, the French Government engaged in a charm offensive to persuade big US technology companies to establish research laboratories and offices in France. The initiative was successful given that both Facebook and DeepMind (an AI subsidiary of Google) opened laboratories in France.<sup>237</sup>

France also continues to maintain great power ambitions and a defence policy that places a premium on strategic autonomy in technological, industrial and operational terms. It is important for the MAF to develop and use technology made in France.<sup>238</sup> For that reason, France continues to invest a significant share of its military budget in R&D. In that regard the DGA plays an important role in funding fundamental and applied research in France. In April 2019 Parly announced several measures that indicate a commitment to reinforce conditions for France to have some strategic autonomy in AI.<sup>239</sup> Notably, she repeated the commitment to invest €100 million (now \$110 million) in AI-specific R&D annually in 2019–25 and increased the MAF's 2018 objective to recruit AI specialists from 50 to 200 by 2023. These experts will lead work on AI in the MAF's newly formed Defence Innovation Laboratory.<sup>240</sup> The DGA will also develop an AI development guide for the researchers, engineers and companies willing to develop AI solutions for the French armed forces.<sup>241</sup> Meanwhile, the French Innovation Council has been working since 2018 on the creation of an AI certification label that will set specific

<sup>234</sup> French Government, *France intelligence artificielle: Rapport de synthèse—Groupes de travail* [France artificial intelligence: Synthesis report—working groups] (French Government: Paris, 2017), pp. 53–70.

<sup>235</sup> French Government, *La stratégie IA en France* [AI strategy in France] (French Government: Paris, 21 Mar. 2017), p. 9.

<sup>236</sup> French Government (note 235), p. 9.

<sup>237</sup> David, E., 'Google's DeepMind opens AI labs in Paris', *Silicon Angle*, 29 Mar. 2018.

<sup>238</sup> French Ministry of the Armed Forces (note 221), pp. 14–15.

<sup>239</sup> Parly (note 219).

<sup>240</sup> French Ministry of the Armed Forces, 'Florence Parly présente son plan en faveur de l'intelligence artificielle' [Florence Parly presents her plan for artificial intelligence], 19 Mar. 2018; and Parly (note 219).

<sup>241</sup> Parly (note 219).

safety and security standards for AI systems marketed to the French armed forces and to government agencies more generally.<sup>242</sup>

However, France has a number of notable weaknesses, beginning with its infrastructure.<sup>243</sup> Neither the state nor the private sector has data-processing capacities comparable to that found in China or the USA. French engineers also have access to a much smaller volume of data than their Chinese and US peers.<sup>244</sup> France is also not training sufficient numbers of engineers to meet the demands of the private sector. Its academic and private sectors are also known for being less well integrated than those of the USA—France reportedly has difficulties translating its academic research into commercial products.<sup>245</sup> It is further handicapped by the fact that many of the AI engineers trained in France are recruited abroad, primarily by big US groups such as Google. It is hard for the MAF to compete with the salaries offered on the other side of the Atlantic. France thus believes that increased cooperation and partnership with its European allies, notably Germany and the UK, are essential.<sup>246</sup>

#### *State of adoption of AI by the armed forces*

France is in the process of modernizing its nuclear arsenal, with a focus on SLBMs and associated warheads. It intends to replace its Triomphant-class submarines with an operational successor by 2035.<sup>247</sup> France is also taking into consideration emerging threats such as cyberattacks and anti-satellite capabilities.<sup>248</sup> It is difficult to determine whether AI currently plays a role in the French nuclear modernization programme—and, if it does, what that role might be—based on currently available sources.

In general, AI does not yet seem to play an important role in the French armed forces, notably for the safety reasons mentioned above. AI, in the form of basic rule-based automation, is reportedly operationally in use only for discrete tasks in the context of air-defence missions or ISR (see table 3.5). Notably, it is used to detect incoming threats and calculate missile trajectories. For example, the Artemis naval infrared search and track system is used on FREMM-class frigates to automatically detect moving targets.<sup>249</sup> Concrete applications in ISR include automated translation and transcript of foreign language sources.<sup>250</sup>

Like many of the other nuclear-armed states, France did not wait for the recent political focus on AI before conducting AI-related R&D work. Over the past decade the DGA has commissioned a number of R&D projects that demonstrate

<sup>242</sup> Parly (note 219).

<sup>243</sup> French Government (note 234), pp. 53–70, 161–72.

<sup>244</sup> Villani (note 216), pp. 20–22.

<sup>245</sup> Villani (note 216), p. 77.

<sup>246</sup> Villani (note 216), p. 49.

<sup>247</sup> Kile, S. N. and Kristensen, H. M., 'French nuclear forces', *SIPRI Yearbook 2019* (note 8), pp. 314–15.

<sup>248</sup> French Prime Minister's Office, Directorate of Legal and Administrative Information, 'Dissuasion nucléaire : Quel financement pour sa modernization?' [Nuclear deterrence: What funding for its modernization?], *Vie Publique*, 13 July 2017.

<sup>249</sup> Thales Group, 'ARTEMIS IRST—360° naval infrared search and track system'.

<sup>250</sup> Tran (note 218).

**Table 3.5.** State of adoption of artificial intelligence in the French nuclear deterrence architecture

Application area	AI in use	Example or mention in official sources	Status	What is known about AI use
<i>Early warning and intelligence, surveillance and reconnaissance</i>				
AI for data collection and analysis	✓	Research ongoing according to the MAF's AI strategy <sup>a</sup>	R&D	..
Remote sensing	✓	Thales stealthy underwater and surface vehicle for long-lasting and long-distance operations <sup>b</sup>	R&D	..
<i>Command and control</i>				
Autonomous wingman	✓	Man-machine teaming study <sup>c</sup>	R&D	Enable command and control of autonomous unmanned wingman for manned aircraft
<i>Precision strike and delivery</i>				
Air launched	✓	nEUROn unmanned combat aerial vehicle <sup>d</sup>	R&D	Capable of autonomous take-off and landing, navigation in communication-denied environments and in-flight refuelling
Sea launched	..	..	..	..
Missile/air/space defence	✓	Artemis infrared search and track system <sup>e</sup>	Deployed	Automatic passive detection of up to 200 targets
<i>Other</i>				
Cyber/electronic information warfare	✓	Research ongoing according to the MAF's AI strategy <sup>f</sup>	R&D	Competition on how to autonomously detect, evaluate, and patch software vulnerabilities
Physical security	..	..	..	..

.. = no or unclear, ✓ = yes, AI = artificial intelligence, MAF = Ministry of the Armed Forces, R&D = research and development.

<sup>a</sup> French Ministry of the Armed Forces (MAF), *L'intelligence artificielle au service de la défense* [Artificial intelligence at the service of defence], Report of the AI task force (MAF: Paris, Sep. 2019).

<sup>b</sup> Tran, P., 'Thales shows off autonomous stealth drone at Euronaval', *Defence News*, 19 Oct. 2016.

<sup>c</sup> Dassault Aviation, 'Launch of the Man Machine Teaming advanced study programme', 16 Mar. 2018.

<sup>d</sup> Dassault Aviations, 'nEUROn: Introduction', [n.d.].

<sup>e</sup> Thales Group, 'ARTEMIS IRST-360° naval infrared search and track system'.

<sup>f</sup> French Ministry of the Armed Forces (note a).

the ability of the French arms industry to design and develop viable prototypes of autonomous systems that operate in the air, on land, at sea or underwater.<sup>251</sup> Finding detailed information about the R&D projects funded by the DGA has proven difficult, with the exception of widely publicized projects, such as the launch of a man-machine teaming study in 2018. This study is intended to use AI to enhance the intelligence capability of combat aircraft and enhance the ability of machines to interact with pilots and operators.<sup>252</sup> A concrete objective of the study is to enable manned combat aircraft and UCAVs to fly together in ways that could evade an enemy's air defences. The study is also meant to explore how AI can be used on board systems to make sense of sensor data and suggest options to the pilot or operator.<sup>253</sup> Another notable achievement in the aerial domain is the nEUROn UCAV, a collaborative European programme led by Dassault Aviation, a French arms-production company. The nEUROn system is a prototype similar to the USA's X-47B and the UK's Taranis that has the ability to conduct air operations autonomously.<sup>254</sup> In the naval domain, Thales, another leading French arms producer, is reportedly developing a stealthy underwater and surface vehicle that could be capable of long-lasting and long-distance operations.<sup>255</sup>

## V. China

### Vision and policies

#### *AI on the political agenda*

China is determined to become a leading—if not the leading—state in the field of artificial intelligence. The uppermost echelons of China's leadership have made this clear, with AI-related policy pronouncements—which came to the forefront in 2015—and through cultivation of a wide array of public and private sector entities working on AI (see figure 3.5).<sup>256</sup>

A turning point in the articulation of the vision behind this vast network was the publication in July 2017 of the New Generation Artificial Intelligence Development Plan by the Chinese Government.<sup>257</sup> Like other military and industrial states such

<sup>251</sup> Boulanin and Verbruggen (note 12), pp. 98–99.

<sup>252</sup> Tran (note 218).

<sup>253</sup> Dassault Aviation, 'Launch of the Man Machine Teaming advanced study programme', 16 Mar. 2018.

<sup>254</sup> Dassault Aviations, 'nEUROn: Introduction', [n.d.].

<sup>255</sup> Tran, P., 'Thales shows off autonomous stealth drone at Euronaval', *Defence News*, 19 Oct. 2016.

<sup>256</sup> These entities include the State Council, the Ministry of Industry and Information Technology, the Ministry of Science and Technology, the AI Strategy Advisory Committee, the National Natural Science Foundation of China, the Leading Small Group on National Science and Technology Structural Reform and Innovation System Construction, the Central Military-Civilian Fusion Development Commission Office, the Central Military Commission (CMC) Science and Technology Commission, the CMC Equipment Development Department, the Beijing Frontier International AI Research Institute, the New Generation AI Strategic Advisory Commission, along with many Chinese companies and universities. Saalman, L., 'China and India: Two models for AI military acquisition and integration', eds K. Bajpai, S. Ho and M. Chatterjee Miller, *Routledge Handbook of China-India Relations* (Routledge: Abingdon, 2020); and Chinese State Council, '中国制造2025' [Made in China 2025], Order no. 28, 8 May 2015.

<sup>257</sup> Chinese State Council, '新一代人工智能发展规划' [New Generation Artificial Intelligence Development Plan], Order no. 35, 8 July 2017.



**Figure 3.5.** Recent policy developments related to artificial intelligence in China

AI = artificial intelligence, MIIT = Ministry of Industry and Information Technology, MOST = Ministry of Science and Technology, NDRC = National Development and Reform Commission, SASTIND = State Administration for Science, Technology and Industry for National Defence.

Sources: Webster, G., "Translation: Chinese AI alliance drafts self-discipline "joint pledge"", *New America*, 17 Jun. 2019; Center for Security and Emerging Technology (CEST), "Profiles of members of China's New Generation Artificial Intelligence Strategic Advisory Committee", Georgetown University, 13 Apr. 2020; Chinese Ministry of Education, "教育部办公厅关于成立教育部人工智能科技创新专家组的函" [Notice of the Office of the Ministry of Education regarding the establishment of the Ministry of Education Artificial Intelligence Technology Innovation Expert Group], 24 Aug. 2018; and this volume: Boulanin et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, June 2020).



as the United States and the United Kingdom, China has long been interested in and actively supported the development of AI technology, but it had not previously articulated its ambitions in this field with such clarity and comprehensiveness.

This plan provides the most consolidated view of China's policy on AI, notably because it binds China's AI ambitions in the civilian and military spheres. When viewed in the light of the other economic and industrial policy-related plans that mention AI (e.g. Made in China 2025 and China Standards 2035<sup>258</sup>), the New Generation AI plan shows that China is determined to move beyond its traditional approach of watching and waiting for US models and advances. China wants to lead the development of AI—to become a country that innovates and sets the technical and other standards domestically and internationally.<sup>259</sup> For the same reason, China also wants to reduce its vulnerable dependence on key foreign technologies and advanced equipment.<sup>260</sup> In this regard it is interesting to note that its rollout of strategic documents on AI and in some cases the development of the technologies themselves runs in tandem with—and sometimes even precedes—other countries' initiatives. For example, the Chinese Government released its 'Internet Plus' Artificial Intelligence Three-Year Action Implementation Plan in the same month, May 2016, as the USA announced the formation of a new subcommittee of its National Science and Technology Council on machine learning and AI.<sup>261</sup>

When it comes to implementation, China has accompanied its official pronouncements with a series of action plans that are targeted and concrete in terms of the roles of government, industry, academia and even the military. The Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry, under the Ministry of Industry and Information Technology (MIIT), provides guidelines to industry and other actors to pursue the development of applications such as autonomous vehicles, intelligent service robots, video- and image-identification systems, voice interactive systems, translation systems, and smart home products.<sup>262</sup> It also seeks breakthroughs in 'core foundational' technologies such as intelligent sensors, neural network chips and open-source platforms.

Beyond action plans and guidelines, the 13th Five-Year National Science and Technology Innovation Plan calls for China to seize the 'high ground' in international scientific development and launch a series of 15 Science and Technology Innovation 2030 Megaprojects that includes big data, intelligent

<sup>258</sup> Chinese State Council (note 256).

<sup>259</sup> Chinese Central Cyberspace Affairs Commission, '数字丝绸之路国际合作论坛' [Forum on International Cooperation Along the Digital Silk Road], 14 Mar. 2019.

<sup>260</sup> Chinese State Council (note 256); Chinese Central Cyberspace Affairs Commission (note 259); and Lewis, J. A., *Learning the Superior Techniques of the Barbarians: China's Pursuit of Semiconductor Independence* (Center for Strategic and International Studies: Washington, DC, Jan. 2019).

<sup>261</sup> Zhao, X., 'Development strategy analysis of "Internet Plus" artificial intelligence technology', International Conference on Network, Communication, Computer Engineering (NCCE 2018), *Advances in Intelligent Systems Research*, vol. 147 (May 2018); and White House, National Science and Technology Council, 'Charter of the Subcommittee on Machine Learning and Artificial Intelligence', 6 May 2016.

<sup>262</sup> Chinese Ministry of Industry and Information Technology, '促进新一代人工智能产业发展三年行动计划 (2018-2020年)' [Three-Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry (2018-20)], Order no. 315, 14 Dec. 2017.

manufacturing and robotics.<sup>263</sup> Highlighting the fundamental role of AI in achieving these technological breakthroughs, the Ministry of Science and Technology (MOST) announced the decision to add ‘AI 2.0’ as a 16th megaproject, with an emphasis on data intelligence, cross-media intelligence, swarm intelligence, hybrid-augmented intelligence and autonomous intelligent systems.<sup>264</sup>

On the whole, while there are many military applications, Chinese AI-related policy documents, projects and funds place a strong emphasis on the development of AI in the civilian sphere. There are several possible explanations for this. First, the civilian sector is driving AI innovation. China thus needs to focus its efforts on that sector to become a global leader in AI. Second, the economic opportunities are immense given that AI is poised to play a role in nearly all industrial sectors in which China is or intends to become a leader. Third, it is easier for civilian entities, universities and companies to collaborate with and learn from foreign actors.

Most importantly, AI is a dual-use suite of technologies that lends itself particularly well to China’s push towards ‘military–civil fusion’ (军民融合) in science and technology.<sup>265</sup> Advances in civilian technology can be adapted for military purposes and China is not timid about its intention to make the work done in the civilian sphere benefit the development of its military capabilities. For example, the wave of AI-related policy documents that China has released illustrate how political, military, industrial and academic communities take a multisectoral approach to the development of AI, particularly with regard to its military applications.<sup>266</sup>

#### *What vision for military AI?*

It is clear that the Chinese military establishment views the pursuit of military AI as being critical to its future military advances. This is encapsulated by a 2018 statement by Lieutenant General Liu Guozhi, director of the Central Military Commission’s Science and Technology Commission: ‘facing disruptive technology, [we] must . . . seize the opportunity to change paradigms (弯道超车); if you don’t disrupt, you’ll be disrupted!’<sup>267</sup> While the New Generation AI plan does not provide detail about the concrete role that China sees for AI in future warfare, there is some anecdotal evidence that China views automation and autonomy—key attributes of AI—as being decisive. As just one example, in 2018 a senior executive of NORINCO, one of China’s largest arms-producing companies,

<sup>263</sup> Chinese Ministry of Science and Technology, Department of International Cooperation, ‘13th Five-year Plan on Science, Technology and Innovation’, *China Science and Technology Newsletter*, 15 Sep. 2016.

<sup>264</sup> Ding, J., *Deciphering China’s AI Dream: The Context, Components, Capabilities, and Consequences of China’s Strategy to Lead the World in AI* (University of Oxford, Future of Humanity Institute: Mar. 2018).

<sup>265</sup> US–China Economic and Security Review Commission, ‘Technology, trade, and military–civil fusion: China’s pursuit of artificial intelligence, new materials, and new energy’, 7 June 2019.

<sup>266</sup> Chinese Ministry of Science and Technology (note 263); Xu, G., ‘Analysis of “Three-Year Implementation Plan of ‘Internet Plus’ Artificial Intelligence”’, *China Internet*, Dec. 2016, pp. 39–43; Triolo, P., Kania, E. and Webster, G., ‘Translation: Chinese government outlines AI ambitions through 2020’, *New America*, 26 Jan. 2018; ThreatConnect, ‘OPM breach analysis’, 5 June 2015; and Koerner, B. I., ‘Inside the cyberattack that shocked the US Government’, *Wired*, 23 Oct. 2016.

<sup>267</sup> Canadian Security Intelligence Service, ‘Chinese military innovation in emerging technologies’, 11 May 2018.

stated that ‘In future battlegrounds, there will be no people fighting’, lethal autonomous weapons would be commonplace by 2025 and the military use of AI is ‘inevitable’.<sup>268</sup>

Regarding autonomy and LAWS, Chinese AI-related strategic documents do not explicitly discuss how China intends to use or not use military AI, unlike British, French and US official sources (see sections III, IV and I). This discussion is notably absent with regard to the delicate question of delegation of authority from humans to autonomous systems. The 2017 New Generation AI plan aims to ‘strengthen the study of major international common problems’ and ‘deepen international cooperation on AI laws and regulations’ but does not elaborate on what China would see as responsible use of military AI and autonomy.<sup>269</sup> A year later, however, China articulated its definition of the technical characteristics that constitute LAWS in the ongoing intergovernmental discussion under the CCW Convention.<sup>270</sup> Its accompanying working paper also notes concern that, because of the suitability of LAWS for use in environments in which ‘threats of nuclear, biological and chemical weapons are involved’, they ‘would reduce the threshold of war, and the cost of warfare on the part of the user countries’.<sup>271</sup> Despite this, China’s proposed definition was received with scepticism by the non-governmental organizations that had called for a ban. There were two notable reasons for this.

First, the definition of LAWS that China proposed was narrow. To fall within the definition, a weapon would have to have the following five characteristics: (a) lethality, which ‘means sufficient pay load (charge) and . . . means to be lethal’; (b) autonomy, that is the ‘absence of human intervention and control during the entire process of executing a task’; (c) impossibility of termination, meaning that ‘once started there is no way to terminate the device’; (d) indiscriminate effect, in that it will ‘execute the task of killing and maiming regardless of conditions, scenarios and targets’; and (e) evolution in the sense that ‘through interaction with the environment the device can learn autonomously, expand its functions and capabilities in a way exceeding human expectations’.<sup>272</sup> Critics of the Chinese proposal interpret China’s definitional approach as a way to implicitly legitimize the use of most foreseeable autonomous weapons.<sup>273</sup>

Second, the proposed definition would permit China and other states to freely research and to develop LAWS-related capabilities.<sup>274</sup> In this regard, it is notable that the types of AI capability in which China is particularly interested, at least at the policy level, are generic. The official documentation from the Chinese

<sup>268</sup> Allen (note 134), pp. 5–6.

<sup>269</sup> Chinese State Council (note 257) (author translation).

<sup>270</sup> Statement by China to Certain Conventional Weapons Convention, Group of Governmental Experts on Lethal Autonomous Weapon Systems, 9 Apr. 2018, 16.33–16.38.

<sup>271</sup> Certain Conventional Weapons Convention, Group of Governmental Experts on Lethal Autonomous Weapon Systems, Position paper submitted by China, CCW/GGE.1/2018/WP.7, 11 Apr. 2018, para. 3.

<sup>272</sup> Certain Conventional Weapons Convention, CCW/GGE.1/2018/WP.7 (note 271), para. 3.

<sup>273</sup> Kania, E. B., ‘China’s strategic ambiguity and shifting approach to lethal autonomous weapons systems’, *Lawfare*, 17 Apr. 2018.

<sup>274</sup> Chinese State Council (note 257).

Government states that China plans to focus on a series of core capabilities, including computational military reasoning; intelligent and autonomous weapon systems; AI-enabled information processing and intelligence analysis; cyber-defence and cyberwarfare; and electronic warfare.<sup>275</sup> Chinese official documents do not discuss explicitly how the pursuit of these capabilities fits into China's larger strategic military calculations. But it is clear that all these capabilities could enable a range of conventional and nuclear deterrence options.

Capabilities that contribute to China's deterrence options are also particularly relevant for China's concept of 'rapid response' (快速反应), which was introduced in the 2015 military strategy and places emphasis on the necessity of China to respond to attacks with promptness and precision.<sup>276</sup> This focus on rapid response at the strategic level is also a useful lens to make sense of its ongoing AI-related R&D activities. As discussed below, China is working on a number of projects that are ultimately meant to increase its military ability to fight at greater speed and with greater precision—and eventually to increase its deterrence capability. This includes discussions on everything from launch on warning to integration of greater autonomy into smart or cruise missiles, as well as increasing manoeuvrability of hypersonic glide platforms.

Beyond official documents and statements, Chinese-language strategic and technical publications on AI are useful proxies to get a sense of how the Chinese technical and military establishment see opportunities and risks generated by adoption of AI capabilities. When it comes to its nuclear deterrence structure, there are indications that China has already begun to consider ways in which AI can assist in both anticipating and countering incoming attacks.<sup>277</sup> These include using AI to allow for swarms to counter integrated air defences and anti-submarine warfare; enhanced targeting and discrimination to improve missile performance and accuracy; improved adaptability and manoeuvrability to enhance precision-guided munitions and defences; and simulation and modelling that involves testing of spacecraft, aircraft and naval systems.<sup>278</sup> Chinese technical and strategic texts also place an emphasis on deficiencies in early-warning and other systems in anticipating, much less countering, an incoming attack. This indicates a concern over asymmetries brought on by the vulnerabilities of China's conventional and nuclear forces.<sup>279</sup> This preoccupation indicates why China may be compelled to consider greater integration of automation and autonomy into its command and control in the future.

<sup>275</sup> Chinese State Council (note 257).

<sup>276</sup> Chinese State Council, 中国的军事战略 [China's military strategy], White paper (State Council Information Office: Beijing, May 2015). While the concept of rapid response is missing from the 2019 version of the military strategy, Chinese technical writing reveals that it is still present at the operational level. Chinese State Council, 新时代的中国国防 [China's national defence in the new era], White paper (State Council Information Office: Beijing, July 2019).

<sup>277</sup> This assessment is based on forthcoming research into 400 new Chinese-language technical articles, papers and books.

<sup>278</sup> Saalman, L., 'Fear of false negatives: AI and China's nuclear posture', *Bulletin of the Atomic Scientists*, 24 Apr. 2018.

<sup>279</sup> Saalman (note 278).

## Adoption of military AI

### *Capability to adopt the most recent advances in AI for military purposes*

For a number of reasons, China is well positioned to set the pace on AI development in both the civilian and military sectors.

First, the drive towards military–civil fusion allows China to fully exploit the potential of civilian innovation for military purposes. In the academic and industrial development of AI theories and practices, the dividing line between civilian and military purposes is increasingly imperceptible. The strengths of China’s civilian industries in using AI for surveillance, big data processing and decision-making could easily be exploited for the development of military systems. Moreover, China’s strong degree of state support, domestic and international transfer of technology and talent, investment in long-term, whole-of-society measures, and cross-sectoral development allow it to benefit from a range of domestic and international private companies and academic institutions. This represents a notable transformation from China’s traditionally stove-piped military R&D.

To further this integration, the Chinese Government has also made it clear that China aims to normalize communication and coordination among scientific research institutes, universities, enterprises and military industry units to develop this ‘military–civil two-way transformation of AI technology’.<sup>280</sup> This means that crossover between the civilian and the military, as well as the conventional and the nuclear, in China’s AI planning is only likely to grow. For example, in October 2018 the Beijing Institute of Technology (BIT), one of China’s top weapon research institutes, launched a four-year ‘experimental programme for intelligent weapons systems’ at the headquarters of NORINCO.<sup>281</sup> Students will be funnelled directly into China’s military modernization. Such channels between private companies and national laboratories engaged in dual-use and even military R&D will shape not only the interconnections among the government, universities and corporations, but also the policing and military applications that will follow.

Second, China can rely on a growing pool of AI research and AI start-ups. It reportedly already produces more patents and research papers than the USA.<sup>282</sup> The number of Chinese experts is bound to grow given the increasing number of students that enrol in computer programmes in Chinese universities every year and the growing national investments in AI technology training. While the quality of Chinese research in the field of machine learning has been questioned, notably in the USA, the quantity of research and projects can offset quality when it comes to AI.<sup>283</sup> In particular, the ‘cutting-edge technology’ guidelines for grassroots-level defence industries issued by the Chinese State Administration of Science,

<sup>280</sup> Chinese State Council (note 257) (author translation).

<sup>281</sup> Chen, S., ‘China’s brightest children are being recruited to develop AI “killer bots”’, *South China Morning Post*, 8 Nov. 2018.

<sup>282</sup> Robles, P., ‘China plans to be a world leader in artificial intelligence by 2030’, *South China Morning Post*, 1 Oct. 2018.

<sup>283</sup> Hvistandahl, M., ‘China’s tech giants want to go global. Just one thing might stand in their way’, *MIT Technology Review*, 19 Dec 2019.

Technology and Industry for National Defence (SASTIND) and MOST's plans for 20 'New Generation AI Innovation and Development Pilot Zones' by 2023 indicate a targeted effort to build AI capacity at the local level.<sup>284</sup>

Third, China has been called the 'Saudi Arabia' of data, which is essential for the training of machine learning. Indeed, Bai Chunli, president of the Chinese Academy of Sciences, has estimated that 'By 2020, China will hold 20 per cent of the global data, which is expected to reach 44 trillion gigabytes'.<sup>285</sup> In addition to data mined domestically, China's data sets reportedly include significant, large-scale US data sets that could be used for both situational awareness and military platform development (e.g. data obtained from the US Government's Office of Personnel Management, Equifax, Lockheed Martin, among others).<sup>286</sup> China is thus well positioned to exploit its access to information to its advantage in civilian and military terms. This data can be used in everything from enhancing geolocation for targeting of nuclear sites to tracking down and compromising infrastructure, networks and personnel that are crucial to nuclear command and control. For example, China has been working on using AI to enhance its simulation capabilities, suggesting that this vast amount of data will play a key role in mapping out potential scenarios for how a conventional or nuclear conflict could unfold.<sup>287</sup>

Fourth, China is spending a great deal on R&D and has a unique ability to optimize its investment with the abovementioned action plans that are targeted and concrete in terms of the roles of government, industry, academia and even the armed forces. China has laid out a goal to achieve a gross output of its core AI industry in excess of 150 billion yuan (US\$20 billion) and related industry output of more than 1 trillion yuan (\$140 billion) by 2030.<sup>288</sup> The Chinese Government has taken an active role in funding AI ventures by disbursing investments through 'government guidance funds' and 'special management shares' set up by local governments and state-owned enterprises to exert more influence over

<sup>284</sup> Chinese State Administration of Science, Technology and Industry for National Defence (SASTIND), '国防科技工业强基工程: 基础研究与前沿技术项目指南 (2018年)' [Project to strengthen development of the defence technology industry at the grassroots level: Guidelines for basic research and cutting-edge technology projects (2018)], June 2018, English translation: Center for Security and Emerging Technology (CEST), Georgetown University, 30 Sep. 2019; Center for Security and Emerging Technology (CEST), 'China creates national new generation artificial intelligence innovation and development pilot zones', Georgetown University, 11 Mar. 2020; and Chinese Ministry of Science and Technology (MOST), '国家新一代人工智能创新发展试验区建设工作指引' [Guidelines for the construction of the National New Generation Artificial Intelligence Innovation and Development Pilot Zones], 29 Aug. 2019.

<sup>285</sup> Chen, N., 'China's AI business ready to lead the world', Chinese Academy of Sciences, 2 June 2017; and Ding (note 264).

<sup>286</sup> Fruhlinger, J., 'The OPM hack explained: Bad security practices meet China's Captain America', CSO, 12 Feb. 2020; Barrett, B., 'How 4 Chinese hackers allegedly took down Equifax', *Wired*, 10 Feb. 2020; and Fryer-Biggs, Z., 'Latest theft of Navy data another sign of China targeting defense companies', USNI News, 11 June 2018.

<sup>287</sup> An, H. (安红) et al., '仿真技术在电子战作战支持中的应用研究' [Applied research of simulation technology in electronic warfare operational support], 电子信息对抗技术 [Electronic Information Warfare Technology], vol. 34, no. 3 (Mar. 2019), pp. 34–39; and Si, G. (司光亚), Zhang, Y. (张阳) and Wang, Y. (王艳正), '网电空间作战建模仿真研究综述' [Review on modeling and simulation in cyberspace operations], 系统仿真学报 [Journal of System Simulation], vol. 30, no. 2 (Feb. 2018), pp. 386–97.

<sup>288</sup> Chinese State Council (note 257), section 2(3).

large technology companies and to provide guidelines for AI advances.<sup>289</sup> These disparate technical players are integrated through a coordinating body known as the Leading Small Group on National Science and Technology Structural Reform and Innovation System Construction that takes the lead in comprehensive planning and coordination, while the AI Plan Implementation Office and an AI Strategy Advisory Committee provide the Ministry of Science and Technology with the AI-related contributions needed for it to implement major AI-related policy decisions.<sup>290</sup>

Taking a cue from these national-level efforts, as many as 19 cities and provinces across China have also started to develop and release their own plans and policies for AI. Tianjin announced in 2018 a 100 billion yuan (\$16 billion) fund to support the AI industry.<sup>291</sup> Shanghai and Guangzhou each plan to establish a special fund and institute to invest in AI development.<sup>292</sup> The Zhongguancun Development Group in Beijing plans to build a 13.8 billion yuan (\$2.12 billion) AI development park that could host up to 400 AI enterprises, feeding into a national-level AI laboratory within the park.<sup>293</sup> Much like China's special economic zones, which have been so integral to the country's economic growth, this wide array of technology parks is fed by academic talent from adjacent universities and allows for multiple entities to compete and interact in an environment that permits greater innovation. In turn, this can power the growth of AI in industry and the armed forces.

In addition to national and local governments, major technology firms have been incorporated into the AI development structure. Since 2018 the Chinese Government has designated Baidu, Alibaba, Tencent, iFlytek and SenseTime as 'national champions'.<sup>294</sup> This allows them privileged positions in the setting of domestic technical standards and insulates them from competition from state-owned enterprises. As a further example, the Chinese Association for Artificial Intelligence (CAAI) convenes senior Chinese academicians and experts from prominent private sector actors, including Baidu (autonomous vehicles), Alibaba (smart cities), Tencent (medical imaging), iFlytek (voice recognition) and Horizon Robotics.<sup>295</sup> There are additional incentives for private sector cooperation, including China's National Intelligence Law, which gives the government legal

<sup>289</sup> Ding (note 264).

<sup>290</sup> 'AI policy—China', Future of Life Institute, 2018; and Cheng, Y., 'China calls for AI alliance', *China Daily*, 13 Oct. 2017.

<sup>291</sup> Chen, Y., 'China's city of Tianjin to set up \$16-billion artificial intelligence fund', Reuters, 17 May 2018.

<sup>292</sup> Chinese State Council, Information Office, '上海举行推动新一代人工智能发展《实施意见》发布会' [Shanghai holds a press conference to promote the development of a new generation of artificial intelligence], News release, 14 Nov. 2017.

<sup>293</sup> Cadell, C., 'Beijing to build \$2 billion AI research park: Xinhua', Reuters, 3 Jan. 2018.

<sup>294</sup> US Chamber of Commerce, *Made in China: Global Ambitions Built on Local Protections* (US Chamber of Commerce: Washington, DC, 2017); and Allen (note 134).

<sup>295</sup> E.g. China Association for Artificial Intelligence, 'The 4th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS 2016)', 17 Feb. 2016.

authority to compel any organization or citizen to ‘support, assist in and cooperate in national intelligence work’.<sup>296</sup>

*State of adoption of AI by the armed forces*

China is currently modernizing and modestly expanding its nuclear arsenal. Notably, it is developing a credible sea-based nuclear deterrence capability. Currently, it has four operational Type 094 SLBM submarines. The next generation submarine, the Type 096, will be launched in 2020.<sup>297</sup> While there is no official source to confirm this, it is likely that AI will play a role in the future development of the Chinese nuclear arsenal: a review of the military systems and capabilities that China is known to have or to be exploring through R&D projects shows that military AI is already an operational reality in China.

There is anecdotal evidence that AI is already in use or planned to be employed in all capability areas (see table 3.6). China has developed or is currently developing a number of systems that are reported to use AI at some level or are likely to do so. Among the most indisputable cases are China’s work on the DF-ZF hypersonic glide vehicle, the CJ-10 air-launched cruise missile, the GJ-11 Sharp Sword stealth UCAV and the HSU-001 UUV or extra-large UUVs (XLUUVs).<sup>298</sup> These systems are meant to incorporate some degree of autonomy and will most likely have to rely on AI through automation and machine learning for guidance and navigation purposes.

The New Generation AI plan and the associated Three-Year Action Plan also note that China is conducting R&D projects on a number of relevant cross-cutting capabilities.<sup>299</sup> These include AI-enabled information processing and intelligence analysis, as well as AI applied to cybersecurity and cyber-defence. However, precise details of these projects remain unclear.

<sup>296</sup> National Intelligence Law, Promulgated by Presidential Order no. 69, 27 June 2017, as amended by Presidential Order no. 6, 27 Apr. 2018, Article 7.

<sup>297</sup> Kile, S. N. and Kristensen, H. M., ‘Chinese nuclear forces’, *SIPRI Yearbook 2019* (note 8), pp. 318–19; and US Office of the Secretary of Defense, *Annual Report to Congress: Military and Security Developments Involving the People’s Republic of China 2019* (Department of Defense: Washington, DC, 3 May 2019), p. 36.

<sup>298</sup> Chen, S., ‘China military develops robotic submarines to launch a new era of sea power’, *South China Morning Post*, 22 July 2018; Glass, P., ‘China’s robot subs will lean heavily on AI: Report’, *Defense One*, 23 July 2018; Panda, A., ‘Introducing the DF-17: China’s newly tested ballistic missile armed with a hypersonic glide vehicle’, *The Diplomat*, 20 Dec. 2017; Chan, M. and Liu, Z., ‘China rolls out new weapon systems, nuclear-capable missiles in military parade’, *South China Morning Post*, 1 Oct. 2019; and Huang, K., ‘China’s hypersonic DF-17 missile threatens regional stability, analyst warns’, *South China Morning Post*, 23 Aug. 2019. See also Saalman, L., ‘Integration of neural networks into hypersonic glide vehicles’, ed. Saalman (note 2), pp. 24–28; and Saalman, L., ‘Exploring artificial intelligence and unmanned platforms in China’, ed. Saalman (note 2), pp. 43–47.

<sup>299</sup> Chinese State Council (note 257); and Chinese Ministry of Industry and Information Technology (note 262).



**Table 3.6.** State of adoption of artificial intelligence in the Chinese nuclear deterrence architecture

Application area	AI in use	Example or mention in official sources	Status	What is known about AI use
<i>Early warning and intelligence, surveillance and reconnaissance</i>				
Data collection and analysis	✓	Over-the-horizon OTH-B long-range 24/7 radar <sup>a</sup>	R&D	..
Reconnaissance and surveillance	..	DR-8/WZ-8 supersonic UAV <sup>b</sup>	Production	May be a testbed for a hypersonic UAV
<i>Command and control</i>				
Battlefield conditions and management	✓	Joint Operations Command and Control Advanced Concepts Demonstration System <sup>c</sup>	Production	Based on study of the US programme Deep Green
<i>Precision strike and delivery</i>				
Air launched	..	GJ-11 Sharp Sword stealth UCAV <sup>d</sup>	Production	..
Air launched	✓	DF-ZF hypersonic glide vehicle, currently mounted on DF-17 solid-fuelled, road-mobile short-range ballistic missile <sup>e</sup>	Production	Reported AI use, machine learning and autonomy for guidance and manoeuvrability
Air launched	..	CJ-20 air-launched land attack cruise missile <sup>f</sup>	R&D	..
Air launched	..	YJ-100 subsonic anti-ship missile <sup>g</sup>	R&D	AI may enhance guidance system Inertial Navigation System (INS) for mid-course guidance and active radar/infrared seeker in terminal phase
Sea launched	✓	XLUUV, HSU-001 <sup>h</sup>	R&D	Autonomous navigation capability
Ground launched	..	CJ-10 land-attack cruise missile <sup>i</sup>	R&D	..
Missile/air/space defence	..	..	..	..
<i>Other</i>				
Cyber/electronic information warfare	✓	Work according to the Chinese Government <sup>j</sup>	R&D	..

Physical security	✓	Work according to the Chinese Government <sup>k</sup>	R&D	..
-------------------	---	---	-----	----

.. = no or unclear, ✓ = yes, AI = artificial intelligence, R&D = research and development, UAV = unmanned aerial vehicle,UCAV = unmanned combat aerial vehicle, XLUUV = extra-large unmanned underwater vehicle.

<sup>a</sup> Tate, A., 'China integrates long-range surveillance capabilities', *Jane's Intelligence Review*, vol. 29, no. 12 (Dec. 2017).

<sup>b</sup> Liu, Z., 'China unveils supersonic spy drone during National Day military parade rehearsal', *South China Morning Post*, 16 Sep. 2019; and 'BZK-008 CH-91 WZ-8 hypersonic drone testbed', GlobalSecurity.org, accessed 17 Jan. 2020.

<sup>c</sup> Kania, E., 'Artificial intelligence in future Chinese command decision-making', ed. N. D. Wright, *AI, China, Russia, and the Global Order: Technological, Political, Global, and Creative Perspectives*, White Paper (US Department of Defense and Joint Chiefs of Staff: Washington, DC, Dec. 2018), pp. 141–52.

<sup>d</sup> Trevithick, J., 'China showcases stealthier Sharp Sword unmanned combat air vehicle configuration', *The Drive*, 1 Oct. 2019.

<sup>e</sup> Trevithick, J., 'Four of the biggest revelations from China's massive 70th anniversary military parade', *The Drive*, 1 Oct. 2019.

<sup>f</sup> Missile Defense Advocacy Alliance, 'Changjian-20 (CJ-20)', accessed 5 Nov. 2019.

<sup>g</sup> 'YJ-100', Deagel.com, 7 Apr. 2017.

<sup>h</sup> Chen, S., 'China military develops robotic submarines to launch a new era of sea power', *South China Morning Post*, 22 July 2018; and Glass, P., 'China's robot subs will lean heavily on AI: Report', *Defense One*, 23 July 2018.

<sup>i</sup> 'DH-10 / CH-10 / CJ-10, land-attack cruise missiles (LACM), Hong Niao / Chang Feng / Dong Hai-10', GlobalSecurity.org, accessed 22 Apr. 2019.

<sup>j</sup> Chinese State Council, '新一代人工智能发展规划' [New Generation Artificial Intelligence Development Plan], Order no. 35, 8 July 2017.

<sup>k</sup> Chinese State Council (note j).

## VI. India

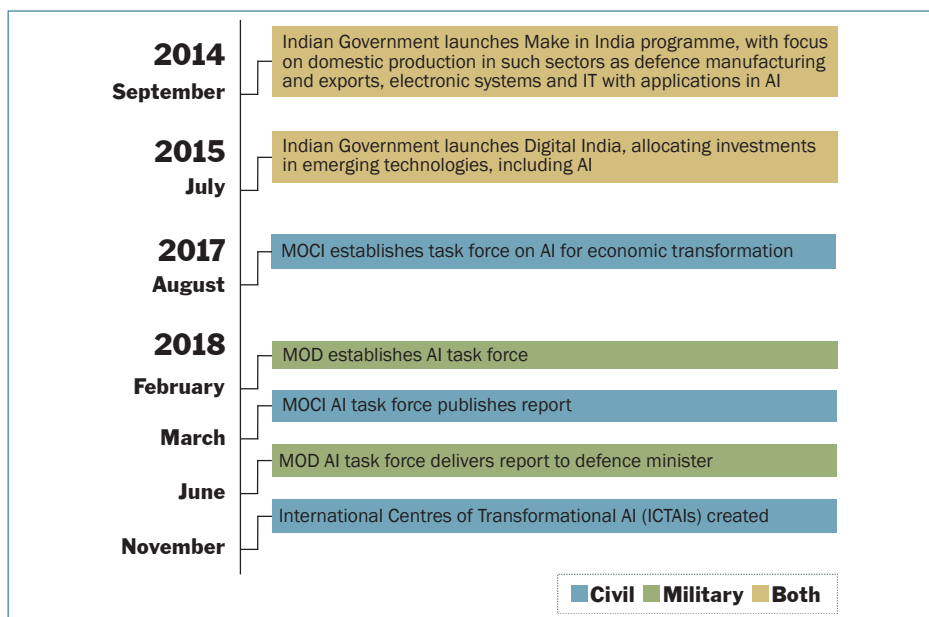
### Vision and policies

#### *AI on the political agenda*

Like other countries discussed here, artificial intelligence became a visible policy priority for India recently, in 2018. Compared to these other countries, however, India remains at a relatively early phase of development of its vision and policies on AI (see figure 3.6).

The Indian Government's primary steps so far have been to convene two task forces—one under the auspices of the Ministry of Commerce and Industry (MOCI) and one under the Ministry of Defence—to explore what India's priorities should be for the development of AI in the civilian and military spheres. The two groups of experts worked in parallel and delivered their findings in 2018. Their recommendations have not yet been formalized in a comprehensive national strategy comparable to those of the other countries reviewed, perhaps reflecting India's traditional reticence in articulating official military doctrines.

The two task forces had distinct missions. The aim of the MOCI task force was to look inwards to explore how the Indian Government could use AI to solve socio-



**Figure 3.6.** Recent policy developments related to artificial intelligence in India

AI = artificial intelligence, IT = information technology, MOCI = Ministry of Commerce and Industry, MOD = Ministry of Defence.

Sources: Indian Government, 'Make in India'; Mukherjee, S., 'Budget 2018: Govt. still strong on digital India; allocates \$480 mn to promote AI, ML, IoT', *Inc42*, 1 Feb. 2018; Task Force on Artificial Intelligence, *Report of Task Force on Artificial Intelligence* (Ministry of Commerce and Industry: New Delhi, Mar. 2018); Indian Ministry of Defence, 'Raksha Mantri inaugurates workshop on AI in national security and defence', Press Information Bureau, 21 May 2018; and National Institution for Transforming India (NITI Aayog), *National Strategy for Artificial Intelligence: #AIforall*, Discussion paper (NITI Aayog: New Delhi, June 2018), p. 72.

economic problems, rather than just to boost economic growth.<sup>300</sup> The report, which is publicly available, discusses how AI could improve the quality of life of Indian citizens and solve problems at a large scale, while generating opportunities for growth and employment. In contrast, the MOD task force looked outwards: 'To study the whole gamut of issues surrounding [the] strategic implications of AI [from a] national security perspective, in [the] global context'.<sup>301</sup> It was to provide recommendations on, among other things, (a) how to make India a significant AI power in national defence, specifically in aerial, naval, land, cyber, nuclear and biological warfare; (b) the policy and institutional interventions required to regulate and encourage robust AI-based technologies for the military sector; and (c) collaboration with start-ups and commercial industry in the use of AI

<sup>300</sup> Indian Ministry of Commerce and Industry, 'Artificial Intelligence Task Force'; and Task Force on Artificial Intelligence, *Report of Task Force on Artificial Intelligence* (Ministry of Commerce and Industry: New Delhi, Mar. 2018), p. 1.

<sup>301</sup> Indian Ministry of Defence, 'Raksha Mantri inaugurates workshop on AI in national security and defence', Press Information Bureau, 21 May 2018.

for national defence.<sup>302</sup> The full report of the MOD task force has not been made public, so it is difficult to conduct a detailed analysis of how India's military is already dealing with, or will deal with, these three issues.

The fact that India created two AI task forces that operated in parallel is illustrative of the India's historical tendency to approach civilian and military industrial and technical challenges separately. The MOCI task force did discuss the possible use of AI for national security purposes, but only in a short subsection.<sup>303</sup> Nonetheless, there is evidence that India is seeking to foster civil-military partnership in AI.<sup>304</sup> In its budget for the 2018/19 financial year, the Indian Government set aside 30 billion rupees (US\$480 million) for investment in emerging technologies, including AI.<sup>305</sup> While much of this activity is devoted to building a civilian AI foundation—for example, part of this sum is meant to build on the approximately 29 000 AI professionals in India<sup>306</sup>—there is also a military modernization component.

The Make in India programme, which the Indian Government launched in 2014 to encourage companies to manufacture their products in India—is being used as a vehicle for increasing cross-collaboration between the civilian and military AI sectors. It has become a central initiative for integrating technical advancement into 25 fields in the domestic civilian and military sectors, including aviation, arms production, arms exports, electronic systems, IT and space.<sup>307</sup> India has reportedly devoted 12.75 trillion rupees (\$180 billion) to Make in India, of which 25–30 per cent is allocated for military projects that are certain to include an AI component.<sup>308</sup> These projects will include stealth technology for UCAVs, next-generation critical technologies for the Advanced Medium Combat Aircraft (AMCA), next-generation integrated early-warning systems, indigenous technology for cruise missiles, and the BrahMos next-generation supersonic and hypersonic cruise missiles.

### *What vision for military AI?*

The upper echelons of the Indian Government have made it clear that India views the pursuit of military AI to be vital for its national security and strategic ambitions. In 2018 Narendra Modi, the prime minister, highlighted the importance of AI and robotics as 'determinants of defensive and offensive capabilities for any defence force in the future', while a senior MOD civil servant described AI as crucial to

<sup>302</sup> Indian Ministry of Defence (note 301).

<sup>303</sup> Task Force on Artificial Intelligence (note 300), p. 25.

<sup>304</sup> See Saalman (note 256).

<sup>305</sup> Mukherjee, S., 'Budget 2018: Govt. still strong on digital India; allocates \$480 mn to promote AI, ML, IoT', *Inc42*, 1 Feb. 2018.

<sup>306</sup> Gupta, D. K., 'Military applications of artificial intelligence', Centre for Land Warfare Studies (CLAWS), 17 Mar. 2018.

<sup>307</sup> Indian Government, 'Make in India'; Task Force on Artificial Intelligence (note 300); and Indian Ministry of Defence (note 301).

<sup>308</sup> Jha, M. K., 'DRDO is taking new challenges in AI and robotics that will act as force multipliers', *Businessworld*, May 2018.

preparing India 'for the next generation warfare which will be more and more technology driven, more and more automated and robotised'.<sup>309</sup>

Since the conclusions of the MOD task force have not been publicly released in full, there is little official information available to determine the specific role that the Indian military establishment sees for AI in future warfare. However, the mandate of the task force indicates that India plans to use AI in all areas of warfare.<sup>310</sup> Furthermore, there is anecdotal evidence to indicate that the AI ambitions of the Indian military establishment are partly a reaction to what China and other major powers are doing in that field. For example, in 2018, following the release of the Indian Army's Land Warfare Doctrine, the chief of the Indian Army Staff, General Bipin Rawat, emphasized the necessity of India keeping up with China's spending on AI.<sup>311</sup> In the CCW framework, India has also repeatedly emphasized that the risks to strategic stability posed by the development of LAWS merit greater consideration. India's insistence on strategic issues, although the CCW forum focuses on humanitarian concerns, has been interpreted as a reflection of the apprehension that the Indian military establishment has about China's efforts to integrate AI into its military apparatus.<sup>312</sup>

India's statements on LAWS at CCW meetings are the primary official sources of information about what India would consider to be the responsible military use of AI. They show that India sees a need to balance strategic and humanitarian considerations. Notably, in 2016 India declared that there is a need for 'increased systemic controls on international armed conflict in a manner that does not widen the technology gap amongst states or encourage the increased resort to military force in the expectation of lesser casualties or that use of lethal force can be shielded from the dictates of public conscience'.<sup>313</sup> India's statements at CCW meetings also demonstrate that it sees a need to limit the advance of autonomy in weaponry. It does not support proposals to prohibit autonomous weapon systems but has made clear that it believes that 'human control needs to be ensured in all weapon systems including future weapon systems using emerging technologies. [India] feel[s] that complete autonomy cannot be given to weaponized platforms'.<sup>314</sup>

Official sources do not provide a detailed picture of the types of military AI capability in which India is primarily interested. Nonetheless, the report of the MOCI task force does mention generic applications of AI that could have national security uses. These include autonomous surveillance and combat systems, adap-

<sup>309</sup> Pandit, R., 'India now wants artificial intelligence-based weapon systems', *Times of India*, 21 May 2018; and Press Trust of India, 'India is working on unmanned tanks, vessels, robotic weaponry for future wars', *Outlook*, 20 May 2018.

<sup>310</sup> Indian Ministry of Defence (note 301).

<sup>311</sup> Indian Army, *Land Warfare Doctrine—2018* (Indian Army: New Delhi, 2018); Press Trust of India, 'Army chief for tapping AI, big data for defence forces', *Economic Times* (New Delhi), 21 Jan. 2019; and Saalman (note 256).

<sup>312</sup> Reddy, R. S., *India and the Challenge of Autonomous Weapons* (Carnegie India: New Delhi, June 2016).

<sup>313</sup> Certain Conventional Weapons Convention, Meeting of Experts on Lethal Autonomous Weapon Systems, Statement by India, 11 Apr. 2016, para. 3.

<sup>314</sup> Certain Conventional Weapons Convention, Group of Governmental Experts on Lethal Autonomous Weapon Systems, 'Review of the potential military applications of related technologies in the context of the group's work', Statement by India, 25 Mar. 2019.

tive communications systems, AI-based systems for cyberattack mitigation and counterattack, and multi-sensor data fusion-based decision-making systems.<sup>315</sup> Despite these details, there remains relatively limited information in official sources on whether India is planning to adopt and use the latest advances in AI for nuclear deterrence purposes. The mandate of the MOD task force only shows that the military applications of AI are broad and far reaching.<sup>316</sup> A review of articles produced by Indian experts and known India military acquisition plans provides some additional confirmation.<sup>317</sup> Notably, it shows that India is interested in, currently developing, or acquiring through import and international collaboration a number of strategic systems that would be strong candidates for AI enhancement. Among these are the Indian-designed nuclear-capable Nirbhay long-range subsonic cruise missile and various projects that aim to give the Indian armed forces access to UCAV, USV and UUV technologies.

### **Adoption of military AI**

#### *Capability to adopt the most recent advances in AI for military purposes*

Although the development of AI capacities is a relatively new policy priority for India, its Defence Research and Development Organisation (DRDO) has worked in this area for a long time. The DRDO's dedicated laboratory, the Centre for Artificial Intelligence and Robotics (CAIR), was established in 1986 and has a staff of over 300.<sup>318</sup> It works on AI, robotics and intelligent control systems and also command, control, communications and intelligence (C3I), communications and networking, and communications secrecy. As of January 2019, CAIR's primary focus was on R&D of net-centric systems for tactical command, control and communications systems, intelligent systems, unmanned systems, information processing and information security.<sup>319</sup> It has also been engaged in network-centric operations and decision-making, using a vast knowledge base of battlefield tactics data.<sup>320</sup> CAIR is thus a core laboratory for R&D in defence-related ICT and plays a central role in the development of military application of AI in India. Among the five goals of CAIR's 2019 mission statement, the fifth stands out: 'driving the national debate where technology policy is critical to preserving national security and self sufficiency'.<sup>321</sup>

India's innovation ecosystems—whether for AI in particular or for the military in general—suffer from a number of limitations.

<sup>315</sup> Task Force on Artificial Intelligence (note 300), p. 25.

<sup>316</sup> Indian Ministry of Defence (note 301).

<sup>317</sup> Roy, K., 'Rationales for introducing artificial intelligence into India's military modernization programme', ed. Topychkanov (note 19), pp. 17–24.

<sup>318</sup> Indian Defence Research and Development Organisation, Centre for Artificial Intelligence and Robotics, 'About lab', accessed 15 Jan. 2019.

<sup>319</sup> Indian Defence Research and Development Organisation, Centre for Artificial Intelligence and Robotics, 'Area of work', accessed 15 Jan. 2019.

<sup>320</sup> Indian Defence Research and Development Organisation, Centre for Artificial Intelligence and Robotics, 'Products', accessed 15 Jan. 2019.

<sup>321</sup> Indian Defence Research and Development Organisation (note 318).

In the case of AI, as pointed out by the MOCI task force, India has to improve its ability to collect, validate and standardize, correlate, and archive AI-relevant data.<sup>322</sup> Many other countries share this challenge. The high cost and low availability of the computing infrastructure required for the development, training and deployment of AI-based services remain an obstacle for the development and growth of AI companies, notably start-ups, in India. This lack of infrastructure causes many Indian AI companies to incorporate their businesses abroad, which leaves AI outside the reach of researchers in government laboratories and many industries within India.<sup>323</sup> Another challenge that is more specific to India is that, despite the large number of IT engineers that it trains every year, its AI expertise is concentrated in a select few individuals and institutions.<sup>324</sup> According to one estimate, only 4 per cent of AI professionals in India have worked on emerging technologies such as deep learning and neural networks.<sup>325</sup> India seems resolved to address this. For example, it is building a range of new centres, including centres of research excellence and the International Centres of Transformational AI (ICTAIs).<sup>326</sup>

In the case of military innovation more generally, India continues to find indigenous development of high-end military technology difficult. The military-industrial complex is dominated by an inherent lack of transparency and collaboration between private industry and the public sector.<sup>327</sup> This translates into a weak and convoluted military tendering process that remains dominated by a few companies, while others from the university and private sectors that might be more innovative often lack the funding and support to compete or their eligibility to compete remains unclear.<sup>328</sup>

As a result, India relies more on foreign collaboration, interdependencies and cross-pollination than its leadership typically admits. India's partnerships with Israeli and Russian firms on many of its military platforms indicate that even with requirements for technology transfer, the country remains dependent on external inputs.<sup>329</sup> The dependence on foreign expertise is particularly salient for civilian and military AI development.<sup>330</sup> For example, in November 2018 the Central Electronics Engineering Research Institute (CEERI) of the Council of

<sup>322</sup> Task Force on Artificial Intelligence (note 300), p. 9.

<sup>323</sup> National Institution for Transforming India (NITI Aayog), *National Strategy for Artificial Intelligence: #AIforall*, Discussion paper (NITI Aayog: New Delhi, June 2018), p. 72.

<sup>324</sup> Task Force on Artificial Intelligence (note 300), p. 9.

<sup>325</sup> National Institution for Transforming India (note 323), p. 72.

<sup>326</sup> National Institution for Transforming India (note 323), pp. 54–57.

<sup>327</sup> Kumar, G. M., 'View: India must tap private sector for closing tech gap with global military powers', *Economic Times* (New Delhi), 20 Nov. 2019.

<sup>328</sup> Some of the difficulties caused by the frequent issuing of guidelines for DRDO procurement have been alleviated by the introduction of an online portal. Government of India, Central Public Procurement Portal (eProcurement); and Indian Defence Research and Development Organisation (DRDO), *Procurement Manual 2016* (DRDO: New Delhi, Nov. 2016).

<sup>329</sup> Press Trust of India, 'Russia plans to deliver S-400 missile systems to India on schedule: Putin', *Economic Times* (New Delhi), 15 Nov. 2019; and Pant, H. V. and Sahu, A., 'Israel's arms sales to India: Bedrock of a strategic partnership', Observer Research Foundation (ORF) Issue Brief no. 311, Sep. 2019.

<sup>330</sup> Sinha, J., 'India became the new hub for AI centres of excellence and labs in 2018', *Analytics India Magazine*, 13 Dec. 2018.

Scientific Industrial Research (CSIR) contracted DataDirect Networks, a US big-data storage company, to design, test and execute computation and storage solutions for AI, machine learning and deep learning.<sup>331</sup> The next month, the Swedish company Ericsson announced the opening of a Global AI Accelerator in Bengaluru, Karnataka, a laboratory focused on automation and AI systems, and a second AI centre in Chennai, Tamil Nadu.<sup>332</sup> In the long term, the involvement of foreign actors in India could be positive as it could allow India to access some key know-how.

*State of adoption of AI by the armed forces*

India is currently increasing its nuclear arsenal and adding new platforms to existing delivery systems.<sup>333</sup> It is unlikely that AI plays a major role in that process given that India is still in the early phases of adoption of AI for military purposes.<sup>334</sup> Anecdotal evidence for this includes the fact that India flew its first indigenously produced Rustom-2 (or Tapas 201) medium-altitude long-endurance UCAV only in 2018.<sup>335</sup> Moreover, finding detailed information about India's AI-related military capabilities has been even more difficult than in the case of China (see table 3.7). In the case of the Rustom-2, it is difficult to determine whether it has some form of autonomous capability.<sup>336</sup>

What can be said with a certain degree of confidence is that India is known to be working on a number of capabilities and systems that would be candidates for AI enhancement. These including (a) integrated early-warning systems; (b) the Multi Agent Robotics Framework (MARF) system, which is intended for robot collaboration on surveillance and reconnaissance; and (c) a number of unmanned systems that could be subject to AI enhancements and autonomy, such as the Matsya UUV and the Autonomous Unmanned Research Aircraft (AURA) programme that aims to demonstrate technologies for future UCAVs.<sup>337</sup> Based on the information available, it can be concluded that AI does not yet play a crucial role in Indian nuclear weapon systems, but its role is likely to grow in the future. In large part, this is due to the introduction of platforms such as the nuclear-capable Nirbhay long-range, subsonic cruise missile.<sup>338</sup>

<sup>331</sup> Associated Press, 'DDN Storage and CSIR-CEERI enter partnership to provide artificial intelligence as a service solutions', 27 Nov. 2018.

<sup>332</sup> Khannan, U., 'Ericsson sets up Global AI Accelerator in B'luru', *Deccan Herald*, 13 Dec. 2018.

<sup>333</sup> Kile, S. N. and Kristensen, H. M., 'Indian nuclear forces', *SIPRI Yearbook 2019* (note 8), pp. 325–31.

<sup>334</sup> Ray, T., 'Slow and steady: India's tentative steps into the AI race', *The Diplomat*, 14 July 2018.

<sup>335</sup> Kumar, C., 'DRDO's combat drone Rustom-2 flies for the first time', *Economic Times* (New Delhi), 14 July 2018.

<sup>336</sup> Kumar (note 335).

<sup>337</sup> Katoch, P., 'Coming—indigenous stealth drone', *Indian Defence Review*, 28 Feb. 2018; 'DRDO Aura', Military Factory, 20 Mar. 2016; Aroor, S., 'Exclusive: Inside the world of India's most secret combat aircraft program', *Livefist Defence*, 2 Feb. 2018; and Korulla, M., 'Autonomous underwater vehicles', Presentation, 'Make in India' Paradigm—Roadmap for a Future Ready Naval Force conference, Federation of Indian Chambers of Commerce and Industry, 18–19 Apr. 2016.

<sup>338</sup> Gady, F.-S., 'India test fires nuclear-capable Nirbhay cruise missile', *The Diplomat*, 15 Apr. 2019; and McLaughlin, J., 'India's expanding missile force', *Wisconsin Project on Nuclear Arms Control*, 15 July 2019.



**Table 3.7.** State of adoption of artificial intelligence in the Indian nuclear deterrence architecture

Application area	AI in use	Example or mention in official sources	Status	What is known about AI use
<i>Early warning and intelligence, surveillance and reconnaissance</i>				
AI for data collection and analysis	..	Integrated early-warning system <sup>a</sup>	R&D	..
Reconnaissance and surveillance	✓	Tapas 201/Rustom-2 medium-altitude long-endurance (MALE) UAV <sup>b</sup>	R&D	..
Reconnaissance and surveillance	✓	Adamyu UUV <sup>c</sup>	R&D	Autonomous underwater vehicle
Reconnaissance and surveillance	✓	Netra UAV <sup>d</sup>	..	Autonomous navigation and guidance system
Reconnaissance and surveillance	✓	Multi Agent Robotics Framework (MARF) system <sup>e</sup>	R&D	Facilitates robot collaboration on surveillance and reconnaissance
Reconnaissance and surveillance, jamming, etc.	✓	Himshakti integrated electronic warfare system <sup>f</sup>	..	AI for surveillance, analysis, interception, direction finding and position fixing, signal intelligence, and jamming of all communications and radar signals
<i>Command and control</i>				
Information retrieval, processing and storage	✓	Command Information and Decision Support System (CIDSS) <sup>g</sup>	..	AI would facilitate storage, retrieval, processing (filtering, correlation, fusion) and visualization of tactical data and provide effective decision support
<i>Precision strike and delivery</i>				
Air based	..	Advanced Medium Combat Aircraft (AMCA) stealth multirole fighter <sup>h</sup>	R&D	AI may be applied in stealth, advanced active electronically scanned array (AESA) radar, manoeuvrability, data fusion or advanced avionics
Air launched	..	Autonomous Unmanned Research Aircraft (AURA)/Ghatak UCAV <sup>i</sup>	R&D	Unclear if autonomous capabilities

Air, sea, ground launched	..	BrahMos supersonic cruise missile/land-attack cruise missile <sup>j</sup>	Production	
Air, sea, ground launched	..	BrahMos-II/NG hypersonic cruise missile <sup>k</sup>	R&D	AI may enhance speed and manoeuvrability to evade radar detection
Ground launched	..	Nirbhay long-range subsonic cruise missile <sup>l</sup>	Production	May use AI for manoeuvrability precision strikes
Missile/air/space defence	..	..	..	..
<i>Other</i>				
Cyber/electronic information warfare	..	..	..	..
Physical security	✓	UAVs to detect radiation and for disposal of improvised explosive devices (IEDs) <sup>m</sup>	R&D	..

.. = no or unclear, ✓ = yes, AI = artificial intelligence, R&D = research and development, UAV = unmanned aerial vehicle, UUV = unmanned underwater vehicle.

<sup>a</sup> Jha, M. K., 'DRDO is taking new challenges in AI and robotics that will act as force multipliers', *Businessworld*, May 2018.

<sup>b</sup> Kumar, C., 'DRDO's combat drone Rustom-2 flies for the first time', *Economic Times* (New Delhi), 14 July 2018.

<sup>c</sup> 'Adanya AUV:- India's submarine launched autonomous underwater vehicle', Indian Defence Update, 4 Sep. 2017.

<sup>d</sup> Bhardwaj, V., 'DRDO Netra mini UAV, quadcopter, Indian Armed Forces', *Aermech.in*, 23 Oct. 2015.

<sup>e</sup> Kumar, C., 'Army to get self-reliant, autonomous robots soon', *Economic Times* (New Delhi), 14 July 2018.

<sup>f</sup> 'Himshakti EW:- India indigenous electronic warfare system', Indian Defence Update, 4 Aug. 2018.

<sup>g</sup> Katoch, P., 'After BMS and F-INSAS, NFS stymied', *Indian Defence Review*, 18 Dec. 2018.

<sup>h</sup> 'HAL AMCA (Advanced Medium Combat Aircraft)', *Military Factory*, 24 Oct. 2017.

<sup>i</sup> Aroor, S., 'Exclusive: Inside the world of India's most secret combat aircraft program', *Livefist Defence*, 2 Feb. 2018.

<sup>j</sup> Jha (note a).

<sup>k</sup> Episkopos, M., 'Meet India's BrahMos II: The world's fastest supersonic cruise missile?', *National Interest*, 10 July 2019.

<sup>l</sup> Gady, F.-S., 'India test fires nuclear-capable Nirbhay cruise missile', *The Diplomat*, 15 Apr. 2019; and McLaughlin, J., 'India's expanding missile force', *Wisconsin Project on Nuclear Arms Control*, 15 July 2019.

<sup>m</sup> Ray, T., 'Slow and steady: India's tentative steps into the AI race', *The Diplomat*, 14 July 2018.

## VII. Pakistan

### Vision and policies

#### *AI on the political agenda*

Artificial intelligence emerged as a policy priority for Pakistan in December 2018 when President Arif Alvi launched an initiative to promote education, research and business opportunities in AI, blockchain and cloud-based computing: the Presidential Initiative for Artificial Intelligence and Computing (PIAIC).<sup>339</sup> This initiative is the first, and so far the only, official document that outlines Pakistan's ambitions in that area (see figure 3.7). Its political weight is limited: since the Pakistani president has no executive power, its implementation is at the discretion of the government. However, there are some indications that implementation of some of its measures has started, notably with support from domestic and foreign private companies.<sup>340</sup> For example, in 2019 Huawei, a large Chinese telecommunications company, provided an eight-day training course for the Pakistani trainers under the auspices of the PIAIC.<sup>341</sup>

The goal of the PIAIC is to strengthen Pakistan's domestic capabilities in AI, which are currently crude. In contrast with the other nuclear-armed states, Pakistan's ambition is not to become a world leader in the field but to ensure that it will be able to seize the opportunities offered by the AI renaissance. In support of his initiative, President Alvi explained:

our future economy and defence systems will strongly depend on our own skills to be a part of the great revolution that is knocking on our doorsteps. Rather than be a consumer that makes us totally dependent on foreign software and hardware even at crucial times, we must become a player and manufacturer of the new systems ensuring phenomenal economic dividends as well as our own security.<sup>342</sup>

While the PIAIC makes reference to the military potential of AI, it places emphasis on the development of AI in the civilian sphere, particularly through education. Notably, the PIAIC called for the enrolment of 100 000 AI students in 2018.<sup>343</sup> This focus can partly be explained by the fact that the initiative was launched on recommendations not from the armed forces but from business and educational institutes, which saw a growing demand for skilled AI professionals.

In 2019 the government of Punjab province launched the National Initiative for Artificial Intelligence and Security (NIAIS) with a similar focus on building national capacity in AI and cybersecurity. While this initiative claims to fill the gap between industry's demand and academia's supply of human resources, it also considers the impact on national defence capabilities.<sup>344</sup>

<sup>339</sup> Presidential Initiative for Artificial Intelligence and Computing, 'How it works'.

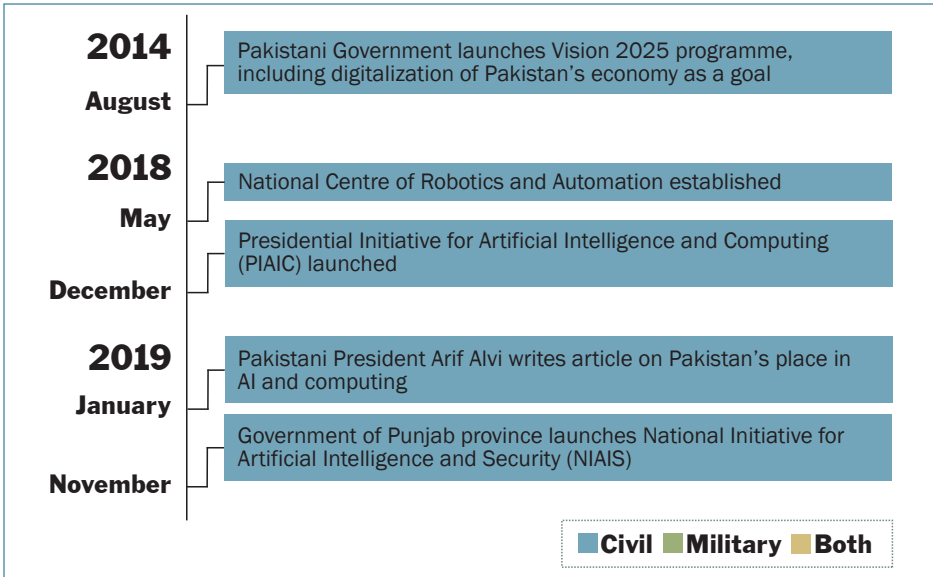
<sup>340</sup> 'Govt taking interest in revolutionising education: Alvi', *The Nation* (Lahore), 21 Jan. 2019; and 'Artificial intelligence', *The Nation* (Lahore), 21 Jan. 2019.

<sup>341</sup> 'HEC, Huawei hold training on AI', *Pakistan Observer*, 9 Nov. 2019.

<sup>342</sup> Alvi, A., 'Pakistan's place in artificial intelligence and computing', *The Nation* (Lahore), 20 Jan. 2019.

<sup>343</sup> 'Artificial intelligence' (note 340).

<sup>344</sup> National Initiative for Artificial Intelligence and Security (NIAIS), 'About'.



**Figure 3.7.** Recent policy developments related to artificial intelligence in Pakistan

AI = artificial intelligence.

Sources: This volume: Boulanin et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, June 2020).

So far, it is too early to assess the impact of either initiative on the AI capacities of Pakistan, including those in the military domain. But these crash courses are not sufficient for sustainable development of AI-related infrastructure.

*What vision for military AI?*

When presenting the PIAIC, President Alvi stressed that AI will be important for Pakistan's future military capabilities. He mentioned in particular the use of AI in such military applications as avionics, smart bombs and UCAVs, as well as in disinformation campaigns and cybersecurity.<sup>345</sup> He also acknowledged that Pakistan will have to pay particular attention to the question of how to integrate humans and AI to bring the best possible military value.<sup>346</sup>

Pakistan's armed forces have not yet made any official statement or produced any strategic document that outlines a vision or plans for AI. A review of the expert literature suggests that the Pakistani military establishment at large has paid limited attention to AI—so far. However, there seems to be a nascent conversation on the topic. First, a number of recent public statements indicate that the topic has attracted the attention of the Pakistani military community. For example, in March 2018 Vice Admiral Arifullah Hussaini, a retired commander of

<sup>345</sup> Alvi (note 342).

<sup>346</sup> Pakistani Inter Services Public Relations, “The synopsis of intelligence integration is requirement of present era” Dr. Arif Alvi while addressing 30th convocation of Pakistan Navy Engineering College (PNEC), Press Release no. PR-PN-220/2018-ISPR, 24 Dec. 2018.

the Pakistani Fleet, stated that ‘The future of war does not belong to conventional methods but artificial intelligence. Whosoever uses it well will win.’<sup>347</sup>

Pakistan’s interventions in the CCW debate on LAWS also indicate that the Pakistani Government has given increasing thought to the wider security opportunities and risks posed by the advance of AI in military systems. They also show what Pakistan has already identified a redline for the military use of AI. Indeed, Pakistan was one of the first countries to support the idea of a ban on LAWS.<sup>348</sup> Pakistan has justified its support for a ban over the years with a mix of ethical, legal and strategic considerations. Ethically, the use of LAWS could ‘make war even more inhumane’ since the ‘use of LAWS in the battlefield against a State fighting with human soldiers would amount to a situation of one-sided killing’.<sup>349</sup> Legally, weapons that are not ‘under the direct control and supervision of humans’ could ‘create an accountability vacuum and provide impunity to the user due to the inability to attribute responsibility for the harm that they cause’.<sup>350</sup> Strategically, LAWS could ‘undermine international peace and security. Their introduction would affect progress on disarmament and non-proliferation. Faced with the prospect of being overwhelmed by LAWS, states possessing [weapon of mass destruction (WMD)] capabilities would be reluctant to give them up, while others would feel encouraged to acquire them.’<sup>351</sup>

Little can currently be said about how Pakistan might adopt recent advances in AI for nuclear-related purposes.<sup>352</sup> Based on existing official statements and expert analysis, two hypotheses can be made. First, Pakistan is unlikely to use AI to pursue full automation of its nuclear command and control. Second, Pakistan’s efforts to use AI for deterrence purposes will be determined by its perception of the threat from India. In that regard, some Pakistani military experts are concerned about development of AI capabilities by India that would improve its space-based ISR and early-warning capabilities.<sup>353</sup> This is regarded as posing a possible threat to the survivability of Pakistan’s nuclear deterrent.<sup>354</sup>

<sup>347</sup> Mobin, M., “Bringing India to negotiating table is only sensible option for regional peace”, *Dawn*, 23 Mar. 2018. Another example of the military view on AI was presented by Air Marshall (retired) Javaid Ahmed. Strategic Vision Institute (SVI), ‘Artificial intelligence (AI), machine learning (ML) and implications for Pakistan’, 20 Aug. 2019.

<sup>348</sup> Certain Conventional Weapons Convention, Group of Governmental Experts on Lethal Autonomous Weapon Systems, ‘General exchange of views’, Statement by Pakistan, 9 Apr. 2018. See also Kumaraguru (note 270) not note 270 needs a reference.

<sup>349</sup> Certain Conventional Weapons Convention, Group of Governmental Experts on Lethal Autonomous Weapon Systems, ‘Further consideration of the human element in the use of force’, Statement by Pakistan, 27–31 Aug. 2018.

<sup>350</sup> Certain Conventional Weapons Convention (note 349).

<sup>351</sup> Certain Conventional Weapons Convention (note 348).

<sup>352</sup> Sial, S. A., ‘Military applications of artificial intelligence in Pakistan and the impact on strategic stability in South Asia’, ed. Topychkanov (note 19), pp. 46–51.

<sup>353</sup> India does not officially have satellite systems commissioned for nuclear-related ISR and early-warning tasks. The use of AI to enhance such capabilities has been discussed in the Indian research community. E.g. Bhargava, A., ‘EMISAT: A force multiplier’, Issue Brief no. 184, Centre for Land Warfare Studies (CLAWS), June 2019, p. 5.

<sup>354</sup> Minhas, A. S., ‘Space weapons: A rapidly evolving threat to South Asian strategic balance’, *National Defence University Journal* (Islamabad), vol. 22 (2018), pp. 173–88, p. 181.

Another concern of Pakistani defence and security experts is related to the risk of AI-enabled cyberattacks by state and non-state actors on critical military and civilian infrastructure, including military and civilian nuclear facilities. They consider the impact of AI to be both part of the challenge and part of the response to it.<sup>355</sup>

## Adoption and capabilities

### *Capability to adopt the most recent advances in AI for military purposes*

Among nuclear-armed states, Pakistan is the least able to use the recent advances in AI for nuclear deterrence-related purposes. The very existence of the PIAIC is an acknowledgement that Pakistan's capability to develop AI technology for military and civilian purposes is at an early stage.

Pakistan's weaknesses are manifold. First, it lacks sufficient trained AI engineers. It does not train enough, and the best students often emigrate to work for foreign universities or companies. Second, it is highly dependent on foreign technology, both at the level of enabling technology (e.g. hardware components, software architecture and programmes) and at the level of infrastructure (e.g. data storage and computing centres). Third, the level of state investment is comparatively low: only 367 million rupees (US\$3.3 million) in 2018. This spending was supposedly part of an allocation of 1.1 billion rupees (\$6.7 million) to a three-year project on AI (part of the government's Vision 2025 programme) that also includes the launch of the National Centre of Robotics and Automation headquartered at the National University of Science and Technology.<sup>356</sup> This centre will be formed by 11 AI laboratories across 13 universities, with each university receiving 70–75 million rupees (\$437 000–468 000).<sup>357</sup> These measures are likely to be insufficient to allow Pakistani research institutions to credibly compete at the global level. By way of comparison, in 2018 India was spending \$480 million annually (see section VI) and a single US university, the Massachusetts Institute of Technology (MIT), announced a \$1 billion plan to develop a new AI college.<sup>358</sup>

Pakistan continues also to have difficulty developing its own military technology. The Pakistani armed forces are currently highly reliant on the import of

<sup>355</sup> Ahmad, K., 'Artificial intelligence and the changing nature of warfare', *Stratagem*, vol. 1, no. 2 (Dec. 2018), pp. 57–72, p. 68; Iqbal, Z. and Salik, H., 'Cyber security in the age of artificial intelligence', *The Dayspring*, 25 July 2019; and Shah, F., 'Cyber terrorism and artificial intelligence', *Hilal*.

<sup>356</sup> Durrani, F., 'Government allocates Rs1.1 billion for artificial intelligence projects', *News International* (Karachi), 19 Apr. 2018; National University of Sciences and Technology (NUST), 'Inauguration of NCRA National Centre of Robotics and Automation', *Prospectus 2019: Engineering, IT & Computer Science* (NUST: Islamabad, 2019), p. 76; Pakistan Vision 2025 Secretariat, *Pakistan 2025: One Nation, One Vision* (Ministry of Planning, Development and Reform: Islamabad, [2014]); and Khan, S., 'Govt allocates Rs1.1bn for artificial intelligence projects in Pakistan', *Business Recorder* (Karachi), 23 Apr. 2018.

<sup>357</sup> Khan (note 356); and 'Pakistan launched largest ever artificial intelligence project in its history', *Times of Islamabad*, 11 July 2019.

<sup>358</sup> Mukherjee (note 305); and Knight, W., 'MIT has just announced a \$1 billion plan to create new college for AI', *MIT Technology Review*, 5 Oct. 2018.

foreign, particularly Chinese, systems or components.<sup>359</sup> Pakistan has a declared intention to create an innovative, self-reliant and self-sustained arms-producing industry but there are reportedly many obstacles to this, including bureaucratic inefficiency and competition between state-owned arms-producing companies and private companies.<sup>360</sup>

Pakistan has been able to compensate for its industrial weaknesses through the import of foreign military equipment. Notably, it has established over the years a relationship with China that has allowed it to access key military technologies, including strategic weapons.<sup>361</sup> A key example is the JF-17 combat aircraft that the Pakistan Aeronautical Complex (PAC) developed jointly with the Chengdu Aircraft Industrial Group (CAIG) of China.<sup>362</sup> In this context it is likely that Pakistan's adoption of the most recent advances in AI will occur via its access to foreign technology.

#### *State of adoption of AI by the armed forces*

Like India, Pakistan is currently increasing its nuclear arsenal and adding new platforms to existing delivery systems, especially air-launched and ground-launched cruise missiles.<sup>363</sup> Little can be said about whether AI plays a role in that process. Official Pakistani reports do not mention AI in the specific context of military equipment and weapons. The lack of publicly available information means that the state of adoption of AI by the Pakistani armed force can only be discussed in speculative and general terms.

What is known from the open sources is that the Pakistani armed forces employ or plan to employ automated systems for a number of nuclear-related missions, including early warning, ISR, command and control, and strategic offence and defence (see table 3.8). The best illustration of these plans is the Strategic Command and Control Support System (SCCSS), reportedly a fully automated system that enables 'robust Command and Control capability of all strategic assets with round the clock situational awareness in a digitized network centric environment to decision makers at National Command Centre'.<sup>364</sup> What concrete role automation plays in this system is, however, unclear.

AI is likely to play a critical role in many of the nuclear-capable cruise missiles that Pakistan is currently developing as it could allow these systems to fly at lower altitude (using automated functions for terrain contour matching) and to penetrate India's missile defence. However, the maturity of the technologies cannot be determined based on the available information.

<sup>359</sup> In 2008–18 China supplied 55% of the military equipment and weapons imported by Pakistan. SIPRI Arms Transfers Database, <<https://www.sipri.org/databases/armstransfers>>.

<sup>360</sup> Asanri, U., 'Pakistan wants to create a self-reliant, self-sufficient defence industry', *Defence News*, 25 July 2019.

<sup>361</sup> SIPRI Arms Transfers Database (note 359).

<sup>362</sup> Moskalenko, V. and Topychkanov, P., *Russia and Pakistan: Shared Challenges and Common Opportunities* (Carnegie Moscow Center: Moscow, May 2014), p. 10.

<sup>363</sup> Kile, S. N. and Kristensen, H. M., 'Pakistani nuclear forces', *SIPRI Yearbook 2019* (note 8), pp. 332–37.

<sup>364</sup> Pakistani Inter Services Public Relations, Press Release no. PR-135/2012-ISPR, 31 May 2012.

**Table 3.8.** State of adoption of artificial intelligence in the Pakistani nuclear deterrence architecture

Application area	AI in use	Example or mention in official sources	Status	What is known about AI use
<i>Early warning and intelligence, surveillance and reconnaissance</i>				
AI for data collection and analysis	✓	ELINT Threat Perception and Identification System for All kinds of Emitters <sup>a</sup>	Production	Offers automatic and manual operating modes, classifier for automatic emitter recognition, and electronic order of battle, among other options
Remote sensing	✓	Shahpar UAV for reconnaissance and surveillance <sup>b</sup>	Production	Has an autonomous guidance and tracking system, with optionally manual control, and automatic launch and landing modes
<i>Command and control</i>				
Command and control	✓	Strategic Command and Control Support System (SCSS) <sup>c</sup>	Deployed	Fully automated system enables the command and control of all strategic assets with 24/7 situation awareness in digitized network-centric environment for decision makers at the National Command Centre (NCC)
<i>Precision strike and delivery</i>				
Air launched	✓	Ra'ad (Hatf-8) air-launched dual-capable cruise missile <sup>d</sup>	R&D	AI may support the advanced guidance and automated terrain hugging capability
Sea launched	✓	Dual-capable Harbah/Babur-3 anti-ship and land-attack cruise missile <sup>e</sup>	R&D	AI supports on-board Terrain Contour Matching (TERCOM) automated navigation system using a predefined contour map of the flight path which acts as a comparison master image that allows the missile to fly without using a satellite navigation system
Ground launched	✓	Dual-capable Babur-2 (Hatf-7) ground-launched cruise missile (dual-capable) <sup>f</sup>	Deployed	Equipped with TERCOM navigation system
Missile/air/space defence	✓	Rabta C41 UAV and air defence automation system <sup>g</sup>	Production	Capable of completely autonomous flight; carries an electro-optical payload for daytime surveillance and reconnaissance for air and missile defence



*Other*

Cyber/ electronic information warfare	..	..	..	..
Physical security	..	..	..	..

.. = no or unclear, ✓ = yes, AI = artificial intelligence, R&D = research and development, UAV = unmanned aerial vehicle.

<sup>a</sup> Pakistani Ministry of Defence Production (MODP), *Defence Products of Pakistan* (MODP: Rawalpindi, [n.d.]), p. 29.

<sup>b</sup> Pakistani Ministry of Defence Production (note a), p. 15.

<sup>c</sup> Pakistani Inter Services Public Relations, Press Release no. PR-260/2012-ISPR, 28 Nov. 2012.

<sup>d</sup> Pakistani Inter Services Public Relations, Press Release no. PR-16/2016-ISPR, 19 Jan. 2016.

<sup>e</sup> Pakistani Inter Services Public Relations, 'Impressive fire power display by Pakistan Navy in north Arabian Sea', Press Release no. PR-PN-2/2018-ISPR, 3 Jan. 2018.

<sup>f</sup> Pakistani Inter Services Public Relations, 'Pakistan today conducted a successful test of an enhanced range version of the indigenously developed Babur cruise missile', Press Release no. PR-142/2018-ISPR, 14 Apr. 2018.

<sup>g</sup> Pakistani Ministry of Defence Production (note a), p. 17.

## VIII. North Korea

### Vision and policies

#### *AI on the political agenda*

North Korea is the least transparent of all the declared nuclear-armed states. As a result, its views and advances in the field of artificial intelligence can only be discussed in speculative and general terms.

Official doctrines, strategies and policies are usually classified by the North Korean Government. From the limited open-source material available, it is not possible to determine whether North Korea has an articulated vision and policy on AI or what its content might be. The public statements of the North Korean supreme leader, Kim Jong Un, and a review of the work that universities are doing in the field of AI suggest that the North Korean leadership is aware of the importance of investing in AI technology in both the civilian and military fields.

Since Kim came to power in 2011, he has repeatedly emphasized that science and technology are 'infinite strategic assets' and their development is 'essential' for North Korea's economic development and to become a strong Socialist state.<sup>365</sup> Kim's statements remain general: they do not provide precise details about how North Korea intends to achieve advances in science and technology, but they do place emphasis on scientific and technical education.<sup>366</sup> For example, 85 departments—

<sup>365</sup> E.g. Kang, J. (강진규), '북한 "정면돌파전에서 믿을 것은 과학기술의 힘"' [North Korea 'the power of science and technology to believe in front breakthrough'], NK 경제 [NK Economy], 27 Jan. 2020 (author translation). See also Lee, J., Kang, J. and Lee, S., 'Policy research on science and technology of North Korea in the Kim Jong-un era based on big data', *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 4S2 (Mar. 2019), pp. 52–56.

<sup>366</sup> Korean Central News Agency (KCNA), '새 세기 교육혁명을 힘있게 추동하게 될 의의깊은 대회' [Significant competition that will drive the educational revolution of the new century], 3 Sep. 2019.

including departments of information and security, nanomaterial science and engineering, and robotics—have been newly established in 37 universities.<sup>367</sup> This indicates that the number of students pursuing AI-related studies at university in North Korea has been expanding, which can be interpreted as indirect evidence that AI is deemed an important field of science and technology.

A speech made by Kim in late 2019 also suggests that the development of civilian science and technology is seen as key to the development of new strategic weapon systems.<sup>368</sup> He did not mention AI explicitly, but given the dual-use nature of AI technology, it is likely that North Korea views AI as an area where civilian advances could be beneficial to its future military capabilities.

### *What vision for military AI?*

Due to the lack of official information, it is impossible to report on how North Korea views the importance of AI for its military. All that is possible is speculation about the areas in which North Korea could have a strong impetus to apply AI.

One of the areas in which North Korea is most likely to exploit AI is cyber operations. These have brought financial gain and inserted this small and isolated country further into international political discourse. Cyber operations also represent a cost-effective and practical means and opportunity for North Korea to project and amplify its military power.<sup>369</sup> North Korea has already shown that it can—and is willing to—use its cyber capabilities to conduct disruptive cyber operations. Since 2013 it is alleged to have been involved in a number of cyber incidents—including those known as Kimsuky, the KHNP hack, DarkSeoul, the Bangladesh Bank heist and Wannacry—targeting over 150 governments and their domestic banks and other companies.<sup>370</sup> Cyber capabilities are critical to North Korea's national strategy and the foundations have been laid for greater investment in this field with more sophisticated technology.<sup>371</sup> In particular, North Korea may develop a capability for AI-enabled cyberattacks.<sup>372</sup>

<sup>367</sup> Williams M., 'Science and technology education in North Korea enters the 21st century', 38 North, 21 Oct. 2019; and '교육체계가 완비되고있다: 대학, 학교들이 새로 나왔다' [Education system is improving gradually: New colleges and schools], 로동신문 [Workers' Newspaper (*Radong Sinmun*)], 3 Sep. 2019.

<sup>368</sup> '주체혁명위업승리의 활로를 밝힌 불멸의 대강—우리의 전진을 저해하는 모든 난관을 정면돌파전으로 뚫고나가자: 조선로동당 중앙위원회 제7기 제5차전원회의에 관한 보도' [The immortal outline that unveiled the path to the victory of the Juche revolution—let's break through all the obstacles that hinder our progress: Report from the 5th plenary meeting of the 7th Central Committee of the Workers' Party of Korea], 로동신문 [Workers' Newspaper (*Radong Sinmun*)], 1 Jan. 2020.

<sup>369</sup> Jun, J., LaFoy, S. and Sohn, E., *North Korea's Cyber Operations: Strategy and Responses* (Center for Strategic and International Studies: Washington, DC, Dec. 2015); and Ko, L., 'North Korea as a geopolitical and cyber actor', *New America*, 6 June 2018.

<sup>370</sup> US Department of Justice, 'North Korean regime-backed programmer charged with conspiracy to conduct multiple cyber-attacks and intrusions', 6 Sep. 2018; Potter, R., 'Toward a better understanding of North Korea's cyber operations', 38 North, 5 Aug. 2019; and Kong, J., Kim, K. and Lim, J., 'The all-purpose sword: North Korea's cyber operations and strategies', eds T. Minárik et al., *2019 11th International Conference on Cyber Conflict: Silent Battle*, Proceedings, Tallinn, 28–31 May 2019 (NATO Cooperative Cyber Defence Centre of Excellence: Tallinn, 2019), pp. 143–62, p. 146.

<sup>371</sup> Jun et al. (note 369); and Ko (note 369).

<sup>372</sup> [China and Russia use AI for cyberattacks, North Korea also learns how to acquire ability, learns tricks and selects targets testifies former head of cybersecurity at US military headquarters in Japan], *Sankei Shimbun*, 14 Feb. 2018; and Goud, N., 'North Korea, China, and Russia to launch hyper war says NATO', *Cybersecurity Insiders*, accessed 25 July 2019.

## Adoption and capabilities

### *Capability to adopt the most recent advances in AI for military purposes*

Again, compared to the other nuclear-armed states, little can be said about North Korea's capabilities in the field of AI.

It is likely that North Korea's advances in this field are affected by the constraints on its exchanges with the outside world. International sanctions have made it difficult for the country to import the hardware components and technologies that are essential to the pursuit of advanced AI and robotic applications.<sup>373</sup> Another fundamental obstacle for North Korea is the restrictions on Internet access.<sup>374</sup> To make advances in machine learning, a large amount of data is required to train models and to make improvements. North Korea's lack of Internet connectivity to the outside world may hamper's the ability of its engineers to collect the critical mass of data to train their systems.

However, there are indications that North Korean universities are active in this field and are exploring various applications of AI (see table 3.9). According to one report, North Korea has been developing AI since the 1990s.<sup>375</sup> The AI Institute of the Korea Computer Center (KCC) is reported to be leading AI and related machine learning advances in North Korea.<sup>376</sup> Starting in 1997, the KCC developed the go-playing software Eunbyul (은별), which was champion six times at international computerized go competitions.<sup>377</sup> Until 2010 it was one of the world's leading go-playing programs.<sup>378</sup> Among the more recent reported advances are the Ryongnamsan (룡남산) 5.1 speech-recognition system developed by Kim Il Sung University, a fingerprint- and facial-recognition system created by the Natural Science Institute, and the Genius (신동) multilingual interpretation program developed by Kim Chaek University of Technology.<sup>379</sup>

The existence and sophistication of North Korean research is demonstrated by an article on audio classification using a deep belief network (DBN), an example of machine learning, that was published in the *Kim Il-Sung University Journal*

<sup>373</sup> E.g. UN Security Council Resolution 2321, 30 Nov. 2016, para. 11 and annex III. For a detailed list of prohibited items see e.g. United Nations, Security Council, Report of the Security Council committee established pursuant to Resolution 1718 (2006) prepared in accordance with paragraph 5 of Resolution 2375 (2017), 2 Oct. 2017.

<sup>374</sup> '북한도 인공지능 연구하나?' [Does North Korea also study artificial intelligence?], *Midas* (Seoul), Apr. 2016.

<sup>375</sup> Kim, M. (김민관), '북한의 인공지능 개발 현황과 전망' [Current status and prospects of artificial intelligence development in North Korea], Korea Development Bank (KDB), *Weekly KDB Report*, 16 Oct. 2017, pp. 15–17.

<sup>376</sup> Kang, J. (강진규), '김정은 시대 북한 IT 현황과 기술 수준' [North Korea's IT status and technology level under Kim Jong Un], *Digital Hurricane*, 17 May 2018. See also Hwang, J., 'Applications of machine learning in North Korea and South Korea', ed. Saalman (note 2), pp. 29–32.

<sup>377</sup> Aryal, B. R., 'Governance of artificial intelligence in Asia Pacific', Presentation, Asia Pacific School on Internet Governance, Bangkok, 8–12 Apr. 2018.

<sup>378</sup> '북한판 '알파고', 인공지능 바둑프로그램을 아시나요?' [Do you know about the North Korean version of the 'AlphaGo' AI go programme?], 12 Mar. 2016, *NK Today*.

<sup>379</sup> Ji, D., 'Facial, voice recognition software on display at North Korean IT exhibit', *NK News.org*, 23 Nov. 2017; and Hwang (note 376), p. 31.

**Table 3.9.** North Korean universities conducting research and studies in artificial intelligence

University	Fields of focus
Han Duk Su Pyongyang University of Light Industry	Facial, finger and voice recognition AI-driven document-analysis and management information systems
Kim Chaek University of Technology	Robotics (e.g. production line robots, autonomous mobile robots)
Kim Il Sung University	AI-enabled cyber capabilities (e.g. detection of intrusive cyber operations)
Pyongsong Coal Industry University	
Pyongsong University of Science	
Pyongyang Command Automation University	
Pyongyang Computer Technology University	
Pyongyang University of Science and Technology	

Sources: Ri, J. (리정철) and Hyon, S. (현성군), ‘음소음성인식에서 심층신뢰망을 리용한 한가지 음향모형화 방법’ [An acoustic modeling method based on Deep Belief Networks in the phone speech recognition], 김일성종합대학학보: 자연과학 [Kim Il Sung University Journal: Natural Science], vol. 62, no. 8 (Aug. 2016), pp. 30–34; Kang, J. (강진규), ‘북한, 위조지문 잡아내는 지문인식 기술 개발’ [North Korea develops fingerprint-recognition technology to detect counterfeit fingerprints], NK 경제 [NK Economy], 18 Nov. 2018; Kang, J. (강진규), ‘조선인공지능인민공화국? . . . 북한은 이미 AI 열풍’ [Artificial Intelligence People’s Republic of Korea? . . . North Korea and its AI craze], NK 경제 [NK Economy], 23 Dec. 2019; Kang, T., ‘North Korean universities join hands with Pyongyang in nurturing the science sector’, The Diplomat, 29 June 2019; ‘엄격하고 공정한 경쟁, 드높은 승벽심: 최우수프로그램개발자는 누구인가’ [Strict and fair competition, high level of victory: Who is the best program developer?], 노동신문 [Workers’ Newspaper (Radong Sinmun)], 6 Nov. 2019; and Kim Chaek University of Technology, ‘Gold medal in the students’ robot football game-2018’, accessed 1 Apr. 2020.

in 2016.<sup>380</sup> This research imitates foreign work on fast learning algorithms for DBNs, showing that North Korean machine learning technology is at an initial stage.<sup>381</sup> However, the cases of other AI-aspirant states show that the speed and breadth of advancement and adoption can be enhanced through foreign study and acquisition.

There is almost no open-source information on the North Korean governmental organizations engaged in AI-related R&D activities. It can be assumed that North Korean agencies that are already engaged in cyber operations are the most likely future locus of AI-related military research. Among these are Bureau 121 and other cyber units of the Reconnaissance General Bureau (RGB), an intelligence agency,

<sup>380</sup> Ri, J. (리정철) and Hyon, S. (현성군), ‘음소음성인식에서 심층신뢰망을 리용한 한가지 음향모형화 방법’ [An acoustic modeling method based on Deep Belief Networks in the phone speech recognition], 김일성종합대학학보: 자연과학 [Kim Il Sung University Journal: Natural Science], vol. 62, no. 8 (Aug. 2016), pp. 30–34.

<sup>381</sup> Hwang (note 376), pp. 31–32.

and the Command Automation Bureau of the North Korean Army's General Staff Department.<sup>382</sup>

The RGB, which was formed in 2009, is responsible for clandestine operations and North Korea's overseas cyber operations.<sup>383</sup> Among the RGB's divisions, Bureau 121 takes the lead in disruptive cyber operations, such as infiltrating computer networks, hacking to extract foreign intelligence and disrupting adversary computer networks.<sup>384</sup> Two sections of Bureau 121 are of particular interest: Lab 110 and Unit 91. Lab 110 is thought to engage in cyberattacks for intelligence operations, including research on computer command systems and electronic jamming.<sup>385</sup> Unit 91 is of greatest interest from the perspective of NC3 and nuclear weapons advances. It focuses on cyberattack missions that target isolated, air-gapped critical infrastructure and networks in South Korea and on theft of confidential information and technology to develop WMD.<sup>386</sup> Notably, some of the experts working in the RGB are selected from the universities that are engaging in AI-related advances, including Pyongyang Command Automation University, Kim Chaek University of Technology and Pyongyang Computer Technology University.<sup>387</sup>

Under the General Staff Department, Unit 204 is tasked with operating cyber psychological warfare, Unit 31 develops hacking programs, and Unit 51 focuses on communications programs for command and control.<sup>388</sup>

### *State of adoption of AI by the armed forces*

North Korea continues to work actively on its nuclear weapon programme and is expanding and modernizing its ballistic missiles. Although there is no evidence that North Korea is able to carry miniaturize nuclear warheads on its missiles, many have concluded that North Korea has made progress towards this goal.<sup>389</sup> It is unlikely that AI will play a significant role in advancing North Korea's nuclear weapons in the near term as its AI developments remain limited, in both civilian and military applications. Moreover, little can be said with certainty about the general adoption of AI by the North Korean armed forces.<sup>390</sup> Nonetheless, since all development of advanced technologies—including AI research projects

<sup>382</sup> Pinkston, D. A., 'North Korean cyber threats', ed. F. Ruge, *Confronting an 'Axis of Cyber'? China, Iran, North Korea, Russia in Cyberspace* 2018, pp. 89–119, p 106; and Yang, J., Kim, S. and Oh, I., 'Analysis on South Korean cybersecurity readiness regarding North Korean cyber capabilities', eds D. Choi and S. Guillely, *Information Security Applications*, 17th International Workshop, WISA 2016 (Springer: Cham, Mar. 2017), pp. 102–11, p. 106.

<sup>383</sup> Kong et al. (note 370), p. 147; and Yang et al. (note 382), p. 106.

<sup>384</sup> Kong et al. (note 370), p. 147.

<sup>385</sup> Kong et al. (note 370), pp. 147–48.

<sup>386</sup> Kong et al. (note 370), p. 148.

<sup>387</sup> Chung, K. and Lee, K., *Advancement of Science and Technology and North Korea's Asymmetric Threat: Rise of Cyber Warfare and Unmanned Aerial Vehicle*, Study Series no. 2017-03 (Korea Institute for National Unification: Seoul, Aug. 2017), p. 23.

<sup>388</sup> Chung and Lee (note 387), p. 23.

<sup>389</sup> Kile, S. N. and Kristensen, H. M., 'North Korea's military nuclear capabilities', *SIPRI Yearbook 2019* (note 8), pp. 341–48.

<sup>390</sup> But see e.g. Su, F., 'Military development in artificial intelligence and their impact on the Korean peninsula', ed. Saalman (note 2), pp. 33–38, pp. 35–36.

carried out by universities—remains under strict state control, articles published in North Korean academic technical journals can be taken as evidence of military applications of AI (see table 3.10). These indicate that some nascent developments with military applications may be occurring, notably in the fields of cyber operation and robotics.

A paper published in 2018 in the *Kim Il Sung University Journal* demonstrates that North Korea is researching improvements in detection of intrusive cyber operations via artificial neural networks and genetic algorithms.<sup>391</sup>

A research report issued by Kim Il Sung University in 2018 reveals that North Korea is working on the application of neural networks in autonomous mobile robots.<sup>392</sup> Another paper shows that North Korea is studying how to measure distance and recognize obstacles when operating autonomous robots.<sup>393</sup> Such technologies can be potentially used to improve the autonomous capabilities of its UAVs. For example, the application of neural networks through the use of imagery databases can enable better assessment of the surrounding environment.<sup>394</sup>

There is no evidence in publicly available sources that North Korea has applied machine learning and autonomy in the nuclear domain. As mentioned above, the one area in which North Korea could be expected to make a strategic use of machine learning is the cyber domain. North Korea's extensive offensive cyber operations listed above demonstrate that it has the resources and foundation to further expand its data collection and system training. This would improve its capacity for AI-enabled cyberattacks. These already have the potential to facilitate identification by North Korea of zero-day vulnerabilities in South Korean and US computer systems. Finding such a vulnerability could compromise NC3 systems and undermine the USA's extended nuclear deterrence on behalf of South Korea.<sup>395</sup>

<sup>391</sup> Pak, S. (박성호) and Hwang, C. (황철진), '망침입검출에서 속성선택에 의한 성능개선' [Performance improvement by attribute selection in the network intrusion detection system], 김일성종합대학학보: 정보과학 [Kim Il Sung University Journal: Information Science], vol. 64, no. 2 (2018), pp. 34–39. See also Kang, J. (강진규) '북한, 보안에 AI 적용을 추진하고 있다' [North Korea is pushing AI into security], NK경제 [NK Economy], 6 Nov. 2018; and Su (note 390), p. 35.

<sup>392</sup> Kang, J. (강진규), '북한, AI 적용 이동형 로봇 연구 중' [North Korea is studying applying AI technology in mobile robots], NK경제 [NK Economy], 4 Oct. 2018. See also Su (note 390), pp. 35–36.

<sup>393</sup> Han, H. (한학수) and Choe, M. (최명성), '안내로봇의 항행을 위한 촬영기와 레이저 거리수감부의 교정에 대한 연구' [Research of extrinsic calibration of a camera and a 2D laser range sensor for navigation of guided robot], 김일성종합대학학보: 자연과학 [Kim Il Sung University Journal: Natural Science], vol. 63, no. 12 (Dec. 2016), pp. 39–41. See also Kang T., 'North Korea's quest for autonomous technology', *The Diplomat*, 13 July 2018.

<sup>394</sup> Horowitz (note 18).

<sup>395</sup> Avin and Amadae (note 92), p. 107.

**Table 3.10.** State of adoption of artificial intelligence in the North Korean nuclear deterrence architecture

Application area	AI in use	Example or mention in official sources	Status	What is known about AI use
<i>Early warning and intelligence, surveillance and reconnaissance</i>				
AI for data collection and analysis	✓	Detection of intrusive cyber operations via artificial neural networks and genetic algorithms <sup>a</sup>	R&D	Use of AI to improve defensive cyber capabilities
Reconnaissance and surveillance	✓	Application of neural networks in autonomous mobile robots <sup>b</sup>	R&D	Use of AI to measure distance and recognize obstacles when operating autonomous robots and to enable better assessment of the surrounding environment, which can be potentially applied in UAVs for ISR missions
<i>Command and control</i>				
Command and control	✓	Relevant working unit established under the Reconnaissance General Bureau <sup>c</sup>	R&D	..
<i>Precision strike and delivery</i>				
Air launched	..	..	..	..
Sea launched	..	..	..	..
Missile/air/space defence	..	..	..	..
<i>Other</i>				
Cyber/electronic information warfare	✓	Indications of potential work in academic technical journals and in the government's offensive cyber departments <sup>d</sup>	R&D	..
Physical security	..	..	..	..

.. = no or unclear, ✓ = yes, AI = artificial intelligence, ISR = intelligence, surveillance and reconnaissance, R&D = research and development, UAV = unmanned aerial vehicle.

<sup>a</sup> Pak, S. (박성호) and Hwang, C. (황철진), '망침입검출에서 속성선택에 의한 성능개선' [Performance improvement by attribute selection in the network intrusion detection system], 김일성종합대학학보: 정보과학 [Kim Il Sung University Journal: Information Science], vol. 64, no. 2 (2018), pp. 34–39.

<sup>b</sup> Kang, J. (강진규), '북한, AI 적용 이동형 로봇 연구 중' [North Korea is studying applying AI technology in mobile robots], NK경제 [NK Economy], 4 Oct. 2018.

<sup>c</sup> Chung, K. and Lee, K., *Advancement of Science and Technology and North Korea's Asymmetric Threat: Rise of Cyber Warfare and Unmanned Aerial Vehicle*, Study Series no. 2017-03 (Korea Institute

for National Unification: Seoul, Aug. 2017), p 23.

<sup>d</sup> Pak and Hwang (note a). See also Kang, J. (강진규) ‘북한, 보안에 AI 적용을 추진하고 있다’ [North Korea is pushing AI into security], NK경제 [NK Economy], 6 Nov. 2018; and Kong, J., Kim, K. and Lim, J., ‘The all-purpose sword: North Korea’s cyber operations and strategies’, eds T. Minárik et al., 2019 *11th International Conference on Cyber Conflict: Silent Battle, Proceedings*, Tallinn, 28–31 May 2019 (NATO Cooperative Cyber Defence Centre of Excellence: Tallinn, 2019), pp. 143–62.



## 4. The positive and negative impacts of AI on strategic stability and nuclear risk

This chapter addresses the question of how the incorporation of machine learning algorithms into nuclear weapon systems and the adoption of autonomous systems for nuclear-related missions may have an impact on strategic stability and nuclear risk. It first (in section I) weighs the positive and negative effects that these technologies could have on strategic stability in the current nuclear order. It then (in section II) discusses a number of scenarios in which the use of machine learning and autonomy in nuclear weapon systems could increase the risk of nuclear weapon use.

### I. The impact on strategic stability and strategic relations

#### **What is strategic stability?**

In order to understand the impact that advances in artificial intelligence may have on strategic stability, it is useful to clarify some basic facts about the foundations of the concept of strategic stability. This term was coined during the cold war to describe the nature of the strategic relations between the Soviet Union and the United States. Early on, the two defined strategic stability as the absence of incentives for either country to launch a first nuclear strike.<sup>396</sup> Over time the concept evolved and received new and broader definitions, partly as a result of the evolution of the nuclear order, which has been gradually shifting from bipolarity to multipolarity. More broadly, it has been described as ‘the absence of armed conflict between nuclear-armed states’, and most broadly as ‘a regional or global security environment in which states enjoy peaceful and harmonious relations’.<sup>397</sup>

In this report, strategic stability is understood in its narrowest sense: as a state of affairs characterized by crisis stability (i.e. the absence of incentives for any country to launch a first nuclear strike) and arms race stability (i.e. the absence of incentives to build up nuclear forces).<sup>398</sup> A precondition for strategic stability from this standpoint is that ‘countries are confident that their adversaries would not be able to undermine their nuclear deterrent capability’ using nuclear, conventional or other non-conventional means (see box 1.1).<sup>399</sup> The concept builds on the assumption that stability is achieved by the fear of mutually assured destruction. If both sides have and are confident in their own and each other’s ability to effectively retaliate against a first nuclear strike or any type of highly destructive

<sup>396</sup> Steinbruner, J. D., ‘National security and the concept of strategic stability’, *Journal of Conflict Resolution*, vol. 22, no. 3 (Sep. 1978), pp. 411–28.

<sup>397</sup> Edward Warner cited in Acton, J. M., ‘Reclaiming strategic stability’, eds E. A. Colby and M. S. Gerson, *Strategic Stability: Contending Interpretations* (US Army War College Press: Carlisle Barracks, PA, Feb. 2013), pp. 117–46, p. 117–18.

<sup>398</sup> Warner in Acton (note 397), p. 117.

<sup>399</sup> Podvig, P., ‘The myth of strategic stability’, *Bulletin of the Atomic Scientists*, 31 Oct. 2012.

attack, then they ‘would not feel the need to build up their strategic arsenals and, most important, would not be under pressure to launch their missiles in a crisis’,<sup>400</sup>

From this perspective, strategic stability depends on two imperatives. First, the imperative to possess second-strike nuclear capabilities—that is, a capability to respond to a nuclear attack; and second, the imperative to ensure that this retaliatory capability is credible, effective and survivable.<sup>401</sup>

### **The positive effects of AI on strategic stability among nuclear-armed states**

The imperative to develop and maintain a credible retaliatory capability continues to guide the development of nuclear weapon systems. The USA and the USSR justified the creation of nuclear triads of strategic bombers, land-based ICBMs and sea-launched SLBMs as diversifying launch platforms in order to improve the survivability of the retaliatory force.<sup>402</sup> The demand of survivability also drove the development of early-warning and command-and-control systems that would allow the strategic command to identify a threat and make an adequate response within a limited time—within minutes. It also required nuclear-armed states to develop elaborate and hardened communications, control and response systems.<sup>403</sup>

Recent advances in machine learning and autonomous systems could find a number of applications that could increase a nuclear-armed state’s confidence in the credibility of its nuclear retaliatory capability (see chapter 3). They may, thereby, have a stabilizing effect on strategic relations between nuclear-armed states in the following ways.

1. *Enabling faster and more reliable early-warning and ISR tools.* These tools would give nuclear decision makers greater situational awareness and allow them to make more informed decisions in time-critical situations.
2. *Increasing the protection and maintenance of nuclear weapons and related infrastructure.* This protection against cyberattack, physical attack and system failure would extend, notably, to nuclear command and control. AI has applications in nuclear safety as well as in nuclear security since it can be used to engage in predictive maintenance. This reduces the risk of malfunction and human mistakes and failures.
3. *Fostering the development of more survivable delivery systems.* These include hypersonic weapon systems and unmanned submarines. For a major nuclear-armed state, these systems can increase confidence in its deterrence capability.

<sup>400</sup> Podvig (note 399).

<sup>401</sup> Brodie (note 52), pp. 264–305.

<sup>402</sup> Borrie (note 54).

<sup>403</sup> Geist and Lohn (note 18).

4. *Allowing for more advanced simulation and wargaming exercises.*

These could help nuclear decision makers to prepare for a crisis. AI can be used to develop virtual and interactive wargames that would provide decision makers with new tools to predict crisis situations and learn how to handle them. For example, the University of California, Berkeley, USA, has developed SIGNAL, an online game that explores ‘how various weapons capabilities, such as low-yield, high precision nuclear weapons, may affect the behaviour of different actors in an escalating global conflict’.<sup>404</sup> By tracking how players behave, the researchers behind the game hope to better understand how countries might react in time of crisis.

5. *Providing new tools for monitoring and verification of arms control and disarmament.*

The systems that a state may use for early-warning and ISR operations can also be used by the international community to monitor nuclear weapon-related developments. It can also be useful in verification of states’ compliance with existing bilateral and multilateral nuclear arms control and disarmament treaties.<sup>405</sup> For example, this can be done through monitoring of nuclear-related activities in countries that are suspected of violating their obligations under such agreements as the 1968 Non-Proliferation Treaty (NPT) and nuclear weapon-free zone NWFZ treaties (see box 4.1).<sup>406</sup> The Federation of American Scientists, a US think tank, has put this approach into practice by forming a task force to explore the use of AI and machine learning to analyse trade data and overhead sensing data for arms control.<sup>407</sup> Another US non-governmental institution, the Nuclear Threat Initiative (NTI), is also exploring how machine learning can be used to gather, organize and use open-source data to supplement the traditional monitoring and verification of non-proliferation regimes.<sup>408</sup>

<sup>404</sup> Manke, K., ‘New online strategy game advances the science of nuclear security’, *Berkeley News*, 7 May 2019; and SIGNAL.

<sup>405</sup> Kaspersen and King (note 7), pp. 125–26.

<sup>406</sup> Wollenmann, R. and Varialle, C., *Verification in Multilateral Nuclear Disarmament: Preparing for the UN Group of Governmental Experts* (Foreign and Commonwealth Office, Wilton Park: Beaconsfield, Mar. 2018); and Treaty on the Non-Proliferation of Nuclear Weapons (Non-Proliferation Treaty, NPT), opened for signature 1 July 1968, entered into force 5 Mar. 1970. For a list of nuclear weapon-free zone treaties see ‘Arms control and disarmament agreements’, *SIPRI Yearbook 2019* (note 8), pp. 549–86.

<sup>407</sup> Federation of American Scientists (FAS), Nuclear Verification Capabilities Independent Task Force, *Nuclear Monitoring and Verification in the Digital Age: Seven Recommendations for Improving the Process*, 3rd report (FAS: Washington, DC, Sep. 2017); and Ulrich, P., ‘Leveraging overhead imagery capabilities in the nonprofit sector through analytics-as-a-service and machine learning’, Working paper, Federation of American Scientists, [2017].

<sup>408</sup> Nuclear Threat Initiative (NTI), ‘Detecting proliferation risks through public data’, [n.d.].

### **Box 4.1. Machine learning and verification of nuclear arms control and disarmament: Opportunities and challenges**

#### **Opportunities for satellite imagery analysis and nuclear test monitoring**

Advances in machine learning might enable new breakthroughs in verification and compliance regimes.<sup>a</sup> Advances in machine learning for image recognition coupled with the increasing availability of satellite imagery could allow more actors to engage in verification activities, which would 'effectively crowdsource what was once the domain of technology of sophisticated states'.<sup>b</sup> Progress in machine learning also facilitates further improvement of existing methods for seismic monitoring of nuclear test. The Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) is working towards using machine learning in its international monitoring system.<sup>c</sup>

#### **Methodological challenges**

The main obstacle to the adoption of advances in machine learning for nuclear disarmament verification is methodological. To be proven effective, machine learning systems need to be trained with large volumes of high-quality data. In the case of nuclear disarmament verification, this raises fundamental methodological questions: What would make a good data set? What data should be selected? How would that data be gathered? Is the use of open-source data reliable enough to identify proliferation behaviour? How should the risk of data poisoning and other spoofing attacks determined to trick or defeat a verification system powered by machine learning be addressed?

Another problem that undermines the possible use of machine learning for nuclear disarmament verification is the lack of verifiability of machine learning systems themselves. There is no reliable mathematical method to verify machine learning systems, particularly those that involve deep neural networks, which operate like a black box.<sup>d</sup> The opacity of machine learning algorithms makes it difficult for the end-user to trust the results of nuclear disarmament verification systems powered by machine learning. Advances in verification will be needed to create the conditions for effective and widely accepted use of artificial intelligence for nuclear disarmament verification.

<sup>a</sup> Kaspersen, A. and King, C., 'Mitigating the challenges of nuclear risk while ensuring the benefits of technology', ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, *Euro-Atlantic Perspectives* (SIPRI: Stockholm, May 2019), pp. 119–27.

<sup>b</sup> Kaspersen and King (note a), p. 125.

<sup>c</sup> Russel, S., Vaidya, S. and Le Bras, R., 'Machine learning for Comprehensive Nuclear-Test-Ban-Treaty monitoring', *CTBTO Spectrum*, no. 14 (Apr. 2010).

<sup>d</sup> Russel, S., Dewey, F. and Tegmark, M., 'Research priorities for robust and beneficial artificial intelligence', *AI Magazine*, vol. 36, no. 4 (winter 2015), pp. 105–14; US Department of Defense (DOD), Autonomy Community of Interest, Test and Evaluation, Verification and Validation Working Group, *Technology Investment Strategy 2015–2018* (Office of the Assistant Secretary of Defense for Research and Engineering: Washington, DC, May 2015); International Atomic Energy Agency (IAEA), *Emerging Technologies Workshop: Trends and Implications for Safeguards*, Workshop report (IAEA: Vienna, 13–16 Feb. 2017); and Federation of American Scientists (FAS), Nuclear Verification Capabilities Independent Task Force, *Nuclear Monitoring and Verification in the Digital Age: Seven Recommendations for Improving the Process*, 3rd report (FAS: Washington, DC, Sep. 2017).

### **The negative effects of AI on strategic stability among nuclear-armed states**

The adoption of recent advances in AI by one or several nuclear-armed states could trigger a security dilemma: the same technology that increases one state's sense

of security can increase another state's sense of insecurity.<sup>409</sup> This problem is far from new—it has been a historical constant in the field of nuclear strategy.<sup>410</sup> The introduction of new, and potentially disruptive, technology by one state always has the potential to destabilize nuclear deterrence relations with other nuclear-armed states. Invariably, when one state develops new capabilities, others attempt to achieve similar capabilities, to find asymmetrical responses, or to change doctrines to either nullify or offset the advantage offered by the new technology. The way in which other states react usually depends on (a) the technical solutions they have available, (b) what they can afford economically, and (c) their political stance domestically and internationally.

The case of AI is, arguably, special. Unlike precision-guided missiles or low-yield nuclear weapons, AI is not a discrete application—it is an enabling technology that could be used to design a great variety of applications. Moreover, this amorphous quality means that AI does not lend itself well to verification much less controls, unlike the delivery platforms and warheads generally associated with nuclear deterrence. As a result, if one state applies AI in its nuclear weapon systems, then the other nuclear-armed states are faced with a moving target and could lose confidence in their second-strike retaliatory capability. Their response could thus include measures that have a destabilizing effect on the current balance of power and increase the risk of nuclear weapon use. These applications and the reactive measures that they may trigger are discussed below.

### *Destabilizing applications of AI*

Based on the literature and the expert discussions at the four project workshops, four AI-related technological developments that have the potential to destabilize nuclear deterrence relations can be identified.<sup>411</sup> These are (a) AI for remote sensing; (b) AI for non-nuclear strategic strike; (c) AI for missile defence and A2/AD; and (d) AI for autonomous nuclear weapon delivery.

*Remote sensing undermining deterrence at sea.* One of the scenarios that repeatedly comes up in the existing literature on AI and strategic stability is the prospect that AI-enabled remote-sensing systems on autonomous surface or underwater vehicles could make deterrence at sea obsolete.<sup>412</sup>

SSBNs are considered the ultimate deterrence tool as they are currently the most survivable type of nuclear-launch platform. Given the immense size

<sup>409</sup> The term 'security dilemma' was coined by John H. Herz in 1951. Herz, J. H., *Political Realism and Political Idealism: A Study in Theory and Realities* (University of Chicago Press: Chicago, 1951). See also Saalman, L., 'The impact of artificial intelligence on nuclear asymmetry and signalling in East Asia', ed. Saalman (note 2), pp. 103–108.

<sup>410</sup> Kramer, R. M., 'Nuclear weapons, peace and the security dilemma: The role of cognitive processes in deterrence', Research Paper no. 957 (Stanford University, Graduate School of Business: Stanford, CA, 1987); and Beardsley, V. and Asal, V., 'Nuclear weapons programs and the security dilemma', eds A. N. Strulberg and M. Fuhrmann, *The Nuclear Renaissance and International Security* (Stanford University Press: Stanford, CA, 2014).

<sup>411</sup> On these 3 workshops see ed. Boulanin (note 7); ed. Saalman (note 2); and ed. Topychkanov (note 19).

<sup>412</sup> Rickli, J.-M., 'The destabilizing prospects of artificial intelligence for nuclear strategy, deterrence and stability', ed. Boulanin (note 7), pp. 91–98, p. 94; and Geist and Lohn (note 18).

of their potential operating area—the world’s oceans—and the limitations of underwater sensor technology, they are extremely hard to detect. This means that it is impossible for a potential opponent to determine the location from which a second strike could originate and to strike pre-emptively. According to this logic, some analysts have argued that the introduction of SSBNs brought stability in the deterrence relations between the USA and the USSR during the cold war.<sup>413</sup>

A number of scholars and practitioners believe that the ‘sacrosanct assumption that SBBNs are immune to a pre-emptive strike’ could be seriously undermined by the possibilities that AI offers to the field of remote sensing.<sup>414</sup> Their assertion is that advances in AI could enable the deployment of autonomous surface and underwater systems that could detect, track and potentially attack SSBNs, making the survival of a second-strike capability less likely. Of particular concern is the fact that these systems are potentially inexpensive to produce and could therefore be deployed in massive numbers.<sup>415</sup> They would be able to cover a large part of the ocean, making the operation of SBBNs increasingly difficult.

Naval warfare experts reportedly regard this scenario with ‘extreme scepticism’—for two reasons.<sup>416</sup> First, laws of physics in the underwater environment make it opaque and hard to observe with existing sensor technology. The environment mutes and distorts the signals that sensors need to detect in order to identify submarines. Second, there are practical difficulties associated with making the sensors of autonomous vehicles effective and positioning them to deliver real operational value. In other words, the technology is not at the stage where it could credibly make SSBNs obsolete. Detecting an adversary’s submarine in the open ocean will remain extremely difficult. SSBNs are, therefore, unlikely to become an easy target for a pre-emptive strike.

That being said, autonomous surface and underwater systems could be used in ways that could still give a significant military advantage to a state that deploys them and generate insecurity on the side of a state that has deployed SSBNs. Autonomous surface and underwater systems could, for instance, be deployed to monitor chokepoints that an enemy SSBN has to traverse to reach or exit its patrol zone.<sup>417</sup> They would alert their own forces when a submarine is detected and therefore function as a virtual barrier, denying submarines access to specific areas of operations. They could also be used to trail enemy submarines once detected.

However, this would be much more challenging from both technical and operational standpoints as navies are prepared for that possibility and already have

<sup>413</sup> Rickli (note 412), p. 94.

<sup>414</sup> Rickli (note 412), p. 94. See also Li, X., ‘Artificial intelligence and its impact on weaponization and arms control’, ed. Saalman (note 2), pp. 13–18; Wu, R., ‘Survivability of China’s sea-based nuclear forces’, *Science & Global Security*, vol. 19, no. 2 (2011), pp. 91–120; and Zhao, T., *Tides of Change: China’s Nuclear Ballistic Missile Submarines and Strategic Stability* (Carnegie Endowment for International Peace: Washington, DC, 2018).

<sup>415</sup> Hambling, D., ‘The inescapable net: Unmanned systems in anti-submarine warfare’, British-American Security Information Council (BASIS) Parliamentary Briefings on Trident Renewal no. 1, Mar. 2016.

<sup>416</sup> Gates, J., ‘Is the SSBN deterrent vulnerable to autonomous drones?’, *RUSI Journal*, vol. 161, no. 6 (2016), pp. 28–35, p. 29.

<sup>417</sup> Gates (note 416).

procedures for a submarine to evade any vessel following it. This foreseeable use of autonomous surface and underwater systems could be a source of concern for a nuclear-armed state such as China, France, India or the UK that has a small number of SSBNs and for strategic and operational reasons has to transit naval routes that could be easily monitored by autonomous systems. To some extent, this might also be a concern for Russia and the USA since they have many fewer SSBNs today than during the cold war. Among the states with the smallest numbers of SSBNs, Chinese experts demonstrate a marked interest in the adverse impact on SSBN forces of such AI-enabled advances, due both to the existing vulnerability of China's fleet and to the utility of remote sensing to better anticipate and counter foreign vessels.<sup>418</sup>

*AI-enabled conventional strikes with strategic forces.* Advances in AI reinforce the problem of entanglement between conventional and nuclear weapons.<sup>419</sup>

AI could enable the improvement of non-nuclear strategic weapons or the development of new types of such weapons that could threaten the survivability of nuclear assets.<sup>420</sup> Advances in machine learning for missile guidance could allow the development of precision-guided munitions that are able to better penetrate air and missile defences. Advances in autonomy are also central to the development of autonomous stealthy UCAVs and hypersonic vehicles, which could be used to strike nuclear assets with conventional warheads.<sup>421</sup> AI can also play a role in the development of cyber-offensive capabilities that could be used to conduct left-of-launch operations on an enemy's nuclear command, control, communications and intelligence (NC3I) systems.<sup>422</sup>

These developments could be destabilizing as they generate insecurity among weaker nuclear-armed states, particularly those that are not able to keep up with the progress in AI in the conventional realm or deploy adequate countermeasures to the new capabilities. This could lead such a state to further develop or modernize nuclear weapons to counter its opponent's advances in the field of AI and maintain its deterrence capability. Chinese and Russian experts are particularly focused on this nexus between conventional and nuclear forces, demonstrating that the implications of a conventional high-precision, high-speed stealth operation that has an impact on a country's command and control could be just as destabilizing for nuclear risk.<sup>423</sup>

<sup>418</sup> Zhao (note 414); and Wu (note 414).

<sup>419</sup> Acton, J. M., 'Escalation through entanglement: How the vulnerability of command-and-control systems raises the risks of an inadvertent nuclear war', *International Security*, vol. 43, no. 1 (summer 2018), pp. 56–99; and Acton, J. (ed.), *Entanglement: Russian and Chinese Perspectives on Non-nuclear Weapons and Nuclear Risks* (Carnegie Endowment for International Peace: Washington, DC, 2017).

<sup>420</sup> Sauer (note 71). See also Li (note 414); and Cai, C., 'The shaping of strategic stability by artificial intelligence', ed. Saalman (note 2), pp. 54–77, pp. 64–65.

<sup>421</sup> Saalman, 'Integration of neural networks into hypersonic glide vehicles' (note 298); and Bronk (note 49).

<sup>422</sup> Boulanin (note 67).

<sup>423</sup> ed. Acton (note 419).

*AI for missile defence and anti-access/area-denial.* Advances in machine learning and autonomy could enable the development of more efficient threat-detection systems that could make both missile defence and A2/AD-related aircraft, warships, and ballistic and cruise missiles more capable. They could also be used to improve electronic countermeasure such as jamming.<sup>424</sup> The implication for strategic stability is that progress in the defences of one side may undermine the confidence of the other side in its ability to successfully conduct a conventional operation against it. This could create an incentive for the latter to resort to the use of more survivable but less controllable platforms that are traditionally used for conventional weapons to deliver low-yield nuclear weapons. These could include autonomous UCAVs or hypersonic glide vehicles, or even platforms in the air and at sea. In other words, this could increase the risk of nuclear weapon use. Such considerations factor into the USA's deployment of low-yield SLBMs and the planned introduction of low-yield submarine-launched cruise missiles.<sup>425</sup> This was in response to Russia's alleged posture of escalation to de-escalate a conflict, and could also be a response to China's alleged A2/AD advances in East Asia.<sup>426</sup> While these insecurities drive US postural change, both China and Russia are thought to be seeking to maintain their second-strike capabilities with UUVs and hypersonic glide vehicles in response to the US pursuit of missile defence and the Conventional Prompt Global Strike programme.<sup>427</sup> These interlinked trends suggest the importance of how perceptions are also having an impact on technological change.

*AI-enabled autonomous nuclear weapon delivery.* Advances in machine learning and autonomy open up the possibility of using new types of platform for nuclear delivery, notably UAVs, UUVs and hypersonic glide vehicles (see chapter 3). Nonetheless, there are significant risks associated with the use of UAVs and UUVs, as it would reduce the possibility of maintaining direct human control over nuclear weapon use. Euro-Atlantic experts seem to agree that Western nuclear-armed states would not seriously consider fielding nuclear-armed, fully autonomous aerial and underwater vehicles.<sup>428</sup> In fact, US military leaders have clearly stated many times their resistance to the idea of arming autonomous

<sup>424</sup> Saalman, 'Exploring artificial intelligence and unmanned platforms in China' (note 298), p. 44.

<sup>425</sup> US Department of Defense, 'Statement on the fielding of the W76-2 low-yield submarine launched ballistic missile warhead' (note 1); and US Department of Defense, *Nuclear Posture Review* (note 1), pp. 54–55.

<sup>426</sup> On China's A2/AD and the contention that it is a Western construct see Saalman, L., 'China's calculus on hypersonics glide', SIPRI Commentary, 15 Aug. 2017.

<sup>427</sup> Kozyulin, V., 'Regulatory frameworks for military artificial intelligence', ed. Saalman (note 2), pp. 78–85, p. 79; Bryen, S., 'Why China, Russia and America are obsessed with hypersonic weapons', *National Interest*, 1 May 2018; Saalman, L., 'Prompt Global Strike: China and the spear', Independent faculty research, Daniel K. Inouye Asia-Pacific Center for Security Studies, Apr. 2014; Acton, J.-M., 'Conventional Prompt Global Strike and Russia's nuclear forces', Carnegie Endowment for International Peace, 4 Oct. 2013; and Saalman, 'Integration of neural networks into hypersonic glide vehicles' (note 298).

<sup>428</sup> Boulanin, V., 'Promises and perils of artificial intelligence for strategic stability and nuclear risk management: Euro-Atlantic perspectives', ed. Boulanin (note 7), pp. 131–38, p. 135.



vehicles with nuclear weapons—for political and security reasons, a human has to stay in the loop.<sup>429</sup>

However, for countries that feel relatively insecure about their nuclear arsenal, the potential benefits in terms of deterrence capability may outweigh the risks.<sup>430</sup> Experts further seem to agree that the case of Russia is ambiguous. On the one hand, there are statements that indicate that its leadership places great value in maintaining a human in the loop when it comes to nuclear command and control.<sup>431</sup> On the other hand, Russia has revealed that it is pursuing the development of a long-range, autonomous UUV—the Poseidon—that could be used for nuclear weapon delivery.<sup>432</sup> Arguably, Russia's relative insecurity in relation to the USA in conventional weaponry and ballistic missile defence specifically, notably due to the latter's advances in AI, could be among the reasons why Russia is interested in the development of such a system.<sup>433</sup>

In East Asia, experts see the insecurity of some nuclear-armed states as a driver for the adoption of unmanned and potentially autonomous nuclear delivery platforms.<sup>434</sup> It is unclear what the exact capabilities of North Korea are in the field of AI and robotics, but it is not hard to imagine that these capabilities play a role in its current efforts to improve its weaponry and overall combat power (see chapter 3).<sup>435</sup> How destabilizing North Korea's adoption of AI for nuclear weapon delivery would be is debatable. Experts seem to agree that the capabilities available to North Korea do not change nuclear deterrence relations.<sup>436</sup> They mainly create concern from the perspective of nuclear risk reduction as the systems are more prone to loss of human control due to malfunction, hacking or spoofing, which could lead to accidental nuclear weapon use. (These risks are further discussed in section II below.)

### *Destabilizing reactions to AI adoption*

It should be acknowledged that the scenarios above remain speculative. The latest advances in machine learning and autonomy will not be translated into actual military capabilities for many years, if not decades.

However, the fact that these technologies are still emerging does not mean that they could not have a near-term impact on strategic relations. The knowledge or belief that one or several nuclear-armed states is planning to make AI technology a key component of its future conventional and nuclear capabilities could be sufficient to incentivize other states—whether nuclear-armed or not—to react with measures that could undermine the strategic relations of nuclear-armed states and potentially increase the likelihood of a nuclear conflict.

<sup>429</sup> Freedberg (note 110); and US Department of Defense (note 42).

<sup>430</sup> Horowitz (note 166), p. 82; Saalman (note 409); and Sial (note 352).

<sup>431</sup> Topychkanov (note 57), p. 68; and Il'nitskii, A. and Losev, A., [Artificial intelligence is both a risk and an opportunity], *Krasnaya Zvezda*, 24 June 2019 (in Russian).

<sup>432</sup> Topychkanov (note 57), pp. 74–75. See also Kashin (note 159), p. 41.

<sup>433</sup> Horowitz (note 166), p. 82.

<sup>434</sup> Su (note 390).

<sup>435</sup> Hwang (note 376). See also chapter 3, section IV, in this volume.

<sup>436</sup> Saalman (note 409).

Indeed, the field of strategy is ‘highly psychological’.<sup>437</sup> The perception of an enemy’s capability matters as much as its actual capabilities. This is where the inherent nature of AI technology becomes a major problem. The fact that it is based on software makes tangible evaluation of military capabilities difficult. Moreover, like electricity, it can be used in many different ways to enhance a state’s nuclear deterrence capability. A nuclear-armed state could therefore easily misperceive its adversaries’ capabilities and intentions in the field of AI. It could trigger destabilizing measures based only on the belief that its retaliatory capability could be defeated, now or in the near future, by another state’s advances in AI.

Depending on the technical, economic and political resources that it has at its disposal, an insecure nuclear-armed state might choose to (a) engage in a capability race on AI; (b) strengthen its commitment to the modernization or development of its nuclear arsenal; (c) change its nuclear policy and doctrine; (d) increase the alert status of its nuclear weapons; or (e) automate its nuclear launch policy.

*Engage in a capability race on AI.*<sup>438</sup> States that have the financial and technical resources could certainly be inclined to enter into a capability race. There is already some evidence that such AI competition may already be underway (see chapter 3). Among nuclear-armed states China, France, India, Russia, the UK and the USA have all released policy documents that reveal an ambition to gear up for great power competition in the AI age.<sup>439</sup> While none of these official documents make a clear connection between AI and nuclear deterrence, they indicate that these countries see AI as a fundamental enabler of their future military capabilities.

What is most concerning about the dynamics of the capability race is that it could lead some states, nuclear-armed or not, to adopt the latest developments in AI prematurely or irresponsibly. The fear of being left behind could lead a state, particularly one that is technologically inferior, to lower its safety and reliability standards so that it can adopt and field the technology more quickly.<sup>440</sup> In the context of nuclear weapon systems, this trend is particularly problematic given that a simple malfunction or error in a peripheral system could have dramatic consequences. In this regard, participants in the SIPRI workshops generally agreed that it would be prudent for nuclear-armed states to devote time and resources to develop a clearer understanding of the limitations of AI and the risks associated with premature adoption in critical nuclear force-related systems.<sup>441</sup>

<sup>437</sup> Rickli (note 412), p. 95. See also Saalman (note 409).

<sup>438</sup> On the use of the term ‘capability race’ in place of ‘arms race’ see box 3.1 in chapter 3 in this volume.

<sup>439</sup> Le Drian, J.-Y., ‘L’intelligence artificielle: Un enjeu de souveraineté nationale’ [Artificial intelligence: An issue of national sovereignty], *L’intelligence artificielle: Des libertés individuelles à la sécurité nationale* [Artificial intelligence: From individual freedoms to national security] (Eurogroup Consulting: Paris, 2017), pp. 11–23; Kania, E. B., *Battlefield Singularity: Artificial Intelligence, Military Revolution, and China’s Future Military Power* (Center for New American Security, Washington DC, Nov. 2017); Vempati, S. S., *India and the Artificial Intelligence Revolution* (Carnegie India: New Delhi, Aug. 2016); and Bendett (note 152). See also chapter 3 in this volume.

<sup>440</sup> On e.g. North Korea see Hwang (note 376); and Su (note 390).

<sup>441</sup> Boulanin (note 428), p. 133.

*Strengthen its commitment to the modernization or development of its nuclear arsenal.* A state that may not be able to keep up with the great power competition on AI could be inclined to further engage in the development or modernization of nuclear weapons to counter its opponent's advances in the field of AI and to maintain its deterrence capability. This is a possibility that has already been openly discussed in Russia following the publication in 2014 of the USA's Third Offset Strategy.<sup>442</sup> One report claims that Russia aims to counter the Third Offset Strategy by using the main principle from the USA's First Offset Strategy.<sup>443</sup> This strategy, dating from the 1950s, relied on tactical nuclear superiority to neutralize the USSR's numerical advantages in conventional forces.<sup>444</sup> In other words, Russia intends to offset US dominance in conventional warfare through the development of a wide array of strategic and tactical nuclear weapons. States that do not yet have the capability to compete with the USA, Russia and China in conventional weaponry and AI technology, such as India, Pakistan and North Korea, could also be tempted by this potential—however, they will also continue to develop new technologies in parallel (as outlined in chapter 3).

*Change its nuclear policy and doctrine.* Another destabilizing scenario would be a situation in which a nuclear-armed state abandons a nuclear posture or nuclear use policy that has had, thus far, a positive effect on strategic stability in its region. China or India, for instance, might renounce their no-first-use (NFU) policies, which are already under stress.<sup>445</sup> Participants in the SIPRI workshops on East Asia and South Asia openly discussed this possibility. Several experts asserted that the USA's pursuit of 'absolute security' under its 2018 Nuclear Posture Review would result in a nuclear imbalance with another nuclear-armed state such as China. This would be exacerbated by the USA improving the precision, manoeuvrability and accuracy of its nuclear launch and missile defences. According to this logic, these enhancements could cause countries that currently have an NFU policy to adopt more offensive postures and even consider a first strike.

While arguing that the USA's advances have not yet had a major impact on China's NFU policy, one expert at the East Asia workshop argued that if the improvement of nuclear forces through machine learning and autonomy has an impact on other countries' survivability—deferring a second-strike capability until it becomes a 'third strike', thus rendering retaliation obsolete<sup>446</sup>—then other countries will have to respond and alliance structures may increasingly transform

<sup>442</sup> Hagel (note 107); Work (note 108); and Kashin, V. and Raska, M., *Countering the US Third Offset Strategy: Russian Perspectives, Responses and Challenges*, Policy Report (S. Rajaratnam School of International Studies: Singapore, Jan. 2017).

<sup>443</sup> Kashin and Raska (note 442).

<sup>444</sup> As the USSR achieved nuclear parity with the USA in the 1960s, a Second Offset Strategy was adopted in the 1970s that centred on the development of high-technology conventional weapons, including precision-guided munitions and stealth aircraft, that could more accurately strike conventional forces. Kashin and Raska (note 442).

<sup>445</sup> Pan, Z., 'A study of China's no-first-use policy on nuclear weapons', *Journal of Peace and Nuclear Disarmament*, vol. 1, no. 1 (2018), pp. 115–36; and Pant, H. V. and Joshi, Y., 'Nuclear rethink: A change in India's nuclear doctrine has implications on cost & war strategy', *Economic Times* (New Delhi), 17 Aug. 2019.

<sup>446</sup> Saalman (note 409), p. 105.

into technological groupings. A country that has pledged to use nuclear weapons defensively could also replace its defensive nuclear posture with an offensive one. This would be a likely scenario for Russia, which is already contemplating this possibility as a result of the recent developments in Russian–US relations.<sup>447</sup> Another view is that Chinese and Russian nuclear pronouncements still allow for a degree of interchangeability between offensive and defensive postures, particularly when it comes to conventional and nuclear overlap and the systems that these two countries are developing and testing.<sup>448</sup>

*Increase the alert status of its nuclear weapons.* A more widely applicable scenario would be where a nuclear-armed state increases the alert status of its nuclear weapons; that is, it increases its readiness to launch a nuclear strike. Currently, only Russia and the USA have operationally deployed nuclear weapons that can be launched within minutes.<sup>449</sup> The nuclear weapons of China, France, India, Pakistan, and the UK are not on permanent alert, but could be ready quickly (in less than a day). China, India and Pakistan are reportedly moving away from their policy of keeping nuclear warheads separate from delivery systems specifically in the context of building sea legs of their nuclear triads. This changes their nuclear alert status.<sup>450</sup>

The destabilizing effect of increased alert statuses would augment a sense of insecurity among nuclear-armed states. More concretely, it would also mean that the time available for nuclear decision makers to make choices during times of crisis would be further compressed, which would increase the risk of escalation into a nuclear conflict. However, as some of the participants in the SIPRI workshops pointed out, it is unlikely that insecurity generated by AI alone would be sufficient to trigger a country to enhance its alert status. This scenario would mainly be credible in a context in which geopolitical relations among nuclear-armed states are already dramatically deteriorating.

*Automate its nuclear launch policy.* Finally, an insecure nuclear-armed state could feel inclined to automate its second-strike capability to increase deterrence, following the example of the USSR and Russia's Perimetr system. Nonetheless, it seems unlikely that any of the major nuclear-armed states would ever seriously consider the full automation of nuclear command and control, given that the consequences of failure of such a system could be catastrophic. As participants in the SIPRI workshops noted, there seems to be general agreement among nuclear-

<sup>447</sup> Topychkanov (note 57), p. 74.

<sup>448</sup> Stowe-Thurston, A, Korda, M. and Kristensen, H. M., 'Putin deepens confusion about Russian nuclear policy', *Russia Matters*, 25 Oct. 2018; Saalman (note 73); Saalman, L., 'Factoring Russia into the US–Chinese equation on hypersonic glide vehicles', *SIPRI Insights on Peace and Security* no. 2017/1, Jan. 2017; and Saalman, L., 'China: Lines blur between nuclear and conventional warfighting', *The Interpreter*, Lowy Institute, 19 Dec. 2014.

<sup>449</sup> Kristensen (note 121); and Kristensen, H. M., 'Russian nuclear forces', *SIPRI Yearbook 2019* (note 8), pp. 301–309.

<sup>450</sup> Ramana, M. V. and Borja, L. J., 'Command and control of nuclear weapons in India', *NAPSNet Special Report*, Nautilus Institute, 1 Aug. 2019; and Craig, T. and De Young, K., 'Pakistan is eyeing sea-based and short-range nuclear weapons, analysts say', *Washington Post*, 21 Sep. 2014.

armed states that it would be morally wrong to relinquish human control over the launch of nuclear weapons, notwithstanding unsubstantiated statements from President Putin indicating that Perimetr might be active again.<sup>451</sup>

However, a change in the situation in the light of new geopolitical developments should not be ruled out. There is no formal bilateral or multilateral agreement that prevents any nuclear-armed state from fully automating its command and control. Indeed, it is not hard to imagine that an authoritarian state such as North Korea could be tempted by this possibility in order to engage in provocative NC3-related statements and measures to deter the possibility of a decapitating strike.

## II. The impact on the likelihood of nuclear conflict: Foreseeable risk scenarios

The adoption of machine learning and autonomy in nuclear weapon systems may not only destabilize the current state of strategic affairs in specific situations, it may also increase the likelihood of a nuclear conflict. This section focuses on the nuclear risk that is generated by the potential adoption of AI technology in nuclear weapon systems. For the sake of clarity, it should be stressed that the concept of nuclear risk refers here to the likelihood of nuclear weapon use.

Expert discussions at the SIPRI workshops identified three types of scenario in which the use of AI could cause a crisis or a conventional conflict to escalate to the nuclear level: (a) AI causing an accidental escalation of a crisis or conflict into nuclear weapon exchange, (b) AI causing inadvertent escalation of a crisis or conflict into nuclear weapon exchange, and (c) AI leading to a deliberate escalation. These are presented below.

### AI and accidental escalation

The concept of accidental escalation has been defined as ‘An unintentional increase in the intensity or scope of conflict beyond a recognized threshold as the result of an unplanned action’.<sup>452</sup> The ‘recognized threshold’ is the point at which a party in the conflict may end up using its nuclear arsenal.

#### *Why and how it could happen*

Machine learning and autonomous systems are supposed to improve the ability of commanders to gather information and make informed decisions in time-critical situations. However, the paradox is that the adoption of such systems helps to increase the pace of warfare, which reduces the commanders’ decision-making time. This adoption thereby reinforces the need for further automation in early-warning and command-and-control systems. The more that human decision time is compressed and automation is integrated into command and control, the

<sup>451</sup> President of Russia, ‘Meeting of the Valdai International Discussion Club’, 18 Oct. 2018.

<sup>452</sup> Woodhams and Borrie (note 17), p. 9. See also Morgan, F. E. et al., *Dangerous Threshold: Managing Escalation in the 21st Century* (Rand Corporation: Santa Monica, CA, 2008), pp. 26–28.

greater becomes the risk of an accidental escalation into a nuclear conflict due to a loss or lack of human control.

Risks associated with human supervision of advanced automation are well known. These include the following.<sup>453</sup>

1. *Automation complacency*. Also known as ‘automation bias’, this is a phenomenon whereby humans over-rely on a system and assume that the information provided by the system is correct.
2. *Lack of trust*. This is the opposite phenomenon, in which human operators under-rely on a system, thereby ignoring relevant information or overriding the system’s actions based on incorrect assumptions. A number of incidents involving automated air-defence systems have been caused in this way. A famous example is the destruction of a commercial aircraft, Iran Air Flight 655, on 3 July 1988 by an Aegis Combat System stationed on the US warship USS *Vincennes*.<sup>454</sup>
3. *The out-of-the-loop control problem*. This happens when an emergency or critical situation occurs and the human operator is unable to regain sufficient situational awareness to react appropriately and in time.<sup>455</sup> It is a common problem associated with air and missile defence systems. For example, the USA’s Patriot missile system was involved in two fratricide incidents during the invasion of Iraq in 2003 because of the out-of-the-loop control problem.<sup>456</sup>

#### *A new Petrov incident*

It is not hard to imagine escalatory scenarios involving each of the phenomena described above. One scenario that participants in the SIPRI workshops in Stockholm and New York found particularly relevant was a modern version of the 1983 Petrov incident (see box 2.2). In this scenario, an early-warning system powered by machine learning would wrongly identify that an attack is underway and force the military command to decide within minutes whether to launch an attack. Participants had different views on whether and how escalation to nuclear weapon use would be likely to occur.

Some highlighted the automation bias issue. They argued that since an early-warning system powered by machine learning will have better detection capabilities than current hard-coded systems, it may be viewed as comparatively

<sup>453</sup> Parasuraman, R., Molloy, R. and Singh, I. L., ‘Performance consequences of automation-induced complacency’, *International Journal of Aviation Psychology*, vol. 3, no. 1 (Feb. 1993), pp. 1–23; Murphy, R. R. and Burke, J. L., ‘The safe human–robot ratio’, eds M. Barnes and F. Jentsch, *Human–Robot Interactions in Future Military Operations* (CRC Press: Boca Raton, FL, 2010), pp. 31–49; and Sharkey, N., ‘Staying in the loop: Human supervisory control of weapons’, eds N. Bhuta et al., *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press: Cambridge, 2016), pp. 23–38, p. 36.

<sup>454</sup> Evans, D., ‘*Vincenne*: A case study’, *Naval Science 302: Navigation and Operational II*, University of Pennsylvania, [n.d.].

<sup>455</sup> Murphy and Burke (note 453), p. 45; and Endsley (note 47), p. 6.

<sup>456</sup> Hawley, J. K., *Patriot Wars: Automation and the Patriot Air and Missile Defense System* (Center for New American Security: Washington, DC, Jan. 2017), p. 6.

safer to use. Operators might therefore be more likely to over-trust the system and not see a need to verify the information that it provides. However, one Russian expert responded that nuclear-armed states have learned their lessons with the 1983 Petrov incident. He noted that ‘Petrov incidents actually happen all the time’, such that nuclear-armed states that have deployed nuclear weapons on alert have set up procedures to minimize the risks posed by the malfunctions of early warning.<sup>457</sup> According to this expert, these states use system redundancy to distribute the risk. In other words, they would not authorize a nuclear weapon launch based only on information provided by only one system or one type of system, but would require the information to be cross-checked through multiple systems.<sup>458</sup> Applied to the era of machine learning systems, this would mean that nuclear-armed states probably combine multiple early-warning and ISR systems that would each use different machine learning algorithms and different training data sets.

Other experts responded that some nuclear-armed states might lack the resources or the will to apply similar safety standards. They pointed out that an insecure state would be likely to integrate machine learning algorithms prematurely into early-warning and ISR capabilities with little regard for the risks involved. The case of North Korea was raised repeatedly in making this argument. A Petrov incident involving North Korea or any other relatively inferior nuclear-armed state from a technical standpoint—India or Pakistan—did not seem likely outside the context of a deep political crisis or armed conflict given that these states reportedly do yet have early-warning systems or operationally deployed nuclear weapons.<sup>459</sup>

Another counterargument highlighted the out-of-the-loop control problem. The adoption of machine learning in early-warning systems and in remote-sensing ISR platforms will make the decision-support systems of nuclear decision makers increasingly automated and complex in their functioning. This means that it could become increasingly difficult for humans to maintain situational awareness and determine whether an alert may be the result of a system failure. Similarly, if the early-warning system breaks down, it might be difficult to identify why this had occurred and whether it was the result of malfunction or an adversarial attack (e.g. spoofing or cyberattack). For this reason, one expert concluded that it was of vital importance to distinguish early-warning and ISR systems from nuclear command-and-control systems as two distinct control paradigms. In other words, those managing early-warning and ISR systems should not be the same as those authorizing the launch of nuclear weapons.

<sup>457</sup> ‘The Soviet colonel who averted nuclear war’, Radio Free Europe/Radio Liberty, 26 Sep. 2013.

<sup>458</sup> Permanent Mission of the Russian Federation to United Nations, ‘On presentation of the World Citizen’s Award to Stanislav Petrov’, 19 Jan. 2006.

<sup>459</sup> Davenport, K., ‘Indian submarine completes first patrol’, *Arms Control Today*, vol. 48, no. 10 (Dec. 2018); and Kristensen, H. K. and Korda, M., ‘Chinese nuclear forces, 2019’, *Bulletin of the Atomic Scientists*, vol. 75, no. 4 (July 2019), pp. 171–78, pp. 171–72.

## AI and inadvertent escalation

The concept of inadvertent escalation has been defined as ‘An intentional action to increase . . . the intensity or scope of conflict [that is] interpreted to have crossed a threshold by an adversary in an unforeseen way’.<sup>460</sup> The main difference with accidental escalation is that the decision to escalate is intentional.

### *Why and how it could happen*

AI is a technology that can generate ambiguity, either naturally or artificially. Ambiguity can arise naturally since AI is an intangible, software-based technology. The AI-enabled capabilities of a weapon system are barely identifiable to the naked eye. Taking the case of a UAV, there is no major external difference between a remotely controlled system and one that is autonomous. It would also be difficult or impossible to determine how an adversary’s early-warning systems powered by machine learning work in concrete terms given that these are far more complex in their functioning than a hard-coded system.

As another example, robotics platforms are also multipurpose. This means that they can be used for different types of mission, so it can be difficult for an adversary to determine whether an unmanned platform that shows up on a radar is a remote-sensing platform or an armed platform. This ambiguity about the nature of systems or capabilities can arguably be a factor that triggers an inadvertent escalation.

AI can also be used to artificially create ambiguity. GAN technology enables the creation of deep fake photographic, audio or video content, which could give a malevolent state or non-state third party the ability to conduct an influence campaign that would trick one nuclear-armed state to attack or threaten to attack another state. These two possibilities are further explored in the scenarios below.

### *Misperception of an unmanned system’s operation*

The first scenario that participants in the SIPRI workshops debated as a possible cause of inadvertent escalation involved the use of UAVs or UUVs. Autonomous UAVs or UUVs could lower the threshold for conducting remote-sensing operations as well as conventional strikes since they can operate in communications-denied and adversarial environments without putting a human pilot or operator directly at risk (as illustrated in chapter 2). However, a state could easily misinterpret the intention behind an adversary’s deployment of an autonomous unmanned vehicle.

Two concrete sub-scenarios can be identified: (a) an autonomous unmanned vehicle deployed for remote sensing could be mistaken by the adversary for an armed system intended to conduct a conventional or nuclear attack; and (b) an autonomous armed unmanned platform carrying a conventional payload for a strike mission could be suspected of carrying a nuclear weapon. Workshop participants did not believe that either of these sub-scenarios would by itself be a

<sup>460</sup> Woodhams and Borrie (note 17), p. 9. See also Morgan et al. (note 452), pp. 23–25.



sufficient cause for inadvertent escalation to the nuclear level, at least in the near term. They noted that there were regional differences.

This is not currently a credible scenario in the South Asian context for a number of reasons. A first reason is that India and Pakistan are still in the early phase of adopting armed UAVs and UUVs. India is still working on the AURA and Rustom-2 UAVs (see table 3.7) and tailoring the US Predator UAV for purchase by its armed forces. Pakistan recently started to use armed UAVs: the NESCOM Burraq, a small UCAV which was used for the first time in a counterterrorist operation in 2015.<sup>461</sup>

Another reason is that India and Pakistan have also demonstrated in the recent past that they can prevent conventional conflict from escalating to the nuclear level. For example, in February 2019 a suicide attack in Indian-administered Kashmir that caused the death of 40 people was followed by India and Pakistan conducting airstrikes against each other's territory.<sup>462</sup> This reportedly did not affect the alert status of their nuclear weapons.<sup>463</sup> The prerequisite for this scenario to become really problematic in the South Asian context—as least with regards to Indian–Pakistani relations—would be a situation in which both countries have armed unmanned vehicles and have already escalated confrontation to the point that it could lead them to put their nuclear weapons on alert. According to a Pakistani workshop participant, Pakistan in that case would be likely to be the party in the crisis that inadvertently escalates given that India has a superior conventional arsenal and is currently bound by its NFU policy. Pakistan's nuclear posture, in contrast, signals that it could resort to the use of nuclear weapons in reaction to a conventional attack.<sup>464</sup>

Neither of the two sub-scenarios is highly plausible in the case of Chinese–Indian and Chinese–US strategic relations, at least in the current geopolitical context. China, like India, has an official NFU policy, and Indian or US military commanders would be unlikely to interpret the detection of a Chinese armed unmanned vehicle as a potential nuclear attack. If China were to detect a US armed UAV or armed UUV, it would, like India, also probably be restrained in its response by its NFU policy.<sup>465</sup> Instead, a Chinese expert noted at the New York workshop that, while an intentional US attack on Chinese nuclear forces would be unlikely, escalatory cyberattacks by either side were another issue as they would offer each a means of undermining the adversary without necessarily resorting to non-cyber means.

According to a US workshop participant, unmanned systems would not increase the risk of escalation, but rather increase the chance of a 'grey zone' conflict—that is, a military operation without declaring a war. The political cost, for both sides, of the destruction or seizing of an unmanned system originally deployed for

<sup>461</sup> 'EDEX 2018: Pakistan NESCOM displays its Burraq UCAV drone', Army Recognition, 5 Dec. 2018.

<sup>462</sup> 'Kashmir attack: Tracing the path that led to Pulwama', BBC, 1 May 2019.

<sup>463</sup> Waqar, A., 'Nuclear war between India and Pakistan: An expert assesses the risk', *The Conversation*, 6 Mar 2019.

<sup>464</sup> Narang, V., 'Posturing for peace? Pakistan's nuclear postures and South Asian stability', *International Security*, vol. 34, no. 3 (winter 2010), pp. 38–78.

<sup>465</sup> Saalman, L., 'China's NFU as a litmus test for GNFU', ed. P. Menon, *The Sheathed Sword: From Nuclear Brink to No First Use* (Takshashila Institution: Bengaluru, 2020).

strike and remote sensing is much lower than that of a manned variant. China's detection and capture of a US UUV in December 2016 illustrates this point. The vehicle was seized, kept for about a week and then returned to the USA.<sup>466</sup> This event did not cause a significant deterioration in Chinese–US strategic relations. The same would certainly not have happened if the system in question had been a manned submarine.

In the Euro-Atlantic context, there are some clearer situations in which these sub-scenarios could trigger escalation. Both Russia and the USA are developing unmanned aerial and underwater platforms that could be capable of nuclear delivery. Also, and perhaps more importantly, their diplomatic and political relations have reached a low point that is unprecedented since the end of the cold war. Most importantly, their nuclear arsenals are on alert. Workshop participants seemed to agree that underwater systems would be more likely to be the source of escalation given the limitations of sensor technology underwater, which make difficult the detection and identification of enemy systems and related capabilities as well as the signalling of intentions.

A hypothetical scenario devised by a workshop participant illustrates this. The escalation spiral could start with a situation where submerged detection lines near Murmansk, a major Russian naval base, detect an unknown underwater object, which could be interpreted as a submarine or an autonomous torpedo. In reaction, Russia could put its naval and aviation forces on high alert. The USA's early-warning and ISR systems could then detect the change of alert status of Russia's forces, leading the US forces to do the same. The crux in this scenario is that the detected underwater object may originate not from the USA, but from one of its non-nuclear-armed allies, which had failed to notify the USA that it had lost control of an underwater remote-sensing system intended to operate near, but not within, Russia territorial waters.

While a remote possibility, this scenario is noteworthy in suggesting that a non-nuclear member of the North Atlantic Treaty Organization (NATO) or a non-NATO Western state that has the capability to obtain and field underwater vehicles could also contribute to instability and tension among nuclear-armed states. This specific situation also shows that, when unmanned systems are deployed, a lack of communication and effective signalling among states—both nuclear-armed and non-nuclear weapon states—can also fuel instability. When viewed in the context of extended deterrence, in which the US nuclear posture in East Asia is so closely tied to its alliance structure and nuclear umbrella agreements with states such as Japan and South Korea, the impact of these non-nuclear-armed states on escalation dynamics must not be ignored. As one anecdotal example, a Chinese military expert stated during a 2015 track 1.5 dialogue that Japanese interception

<sup>466</sup> 'China to return seized US underwater drone, Pentagon says', BBC, 18 Dec. 2016; and Jiang, S. and Bohn, K., 'China returns seized US underwater drone', CNN, 20 Dec. 2016.

on behalf of the USA of a Chinese ballistic missile launch would be interpreted as an act of war.<sup>467</sup>

### *AI exploited by a third party*

AI provides new tools to spoof early-warning and ISR systems and to embed disinformation to fool public opinion and nuclear decision makers.<sup>468</sup> A malevolent actor could, for instance, use GAN technology to create deep fakes to make public opinion and policymakers in the USA believe that US soldiers had been killed by chemical or biological weapons during an operation in Syria.<sup>469</sup> Since that would amount to an attack by a WMD, the USA could, according to its nuclear doctrine, be entitled to retaliate with the use of nuclear weapons. The question of whether the USA would retaliate would then circulate on social media and potentially in the press.

The same GAN technology could be further used to create the impression through fake videos or recorded speech that the USA had put its nuclear forces on high alert, which might incentivize other nuclear-armed states to increase the alert status of their nuclear arsenals as well. Since nuclear-armed states should have sufficient intelligence resources to eventually discredit these fakes, it is unlikely that deep fakes like this would be sufficient to actually lead to the use of nuclear weapons.<sup>470</sup> However, deep fakes could still ‘create high levels of uncertainty and tension in a short period of time’.<sup>471</sup>

From the expert discussion at the SIPRI workshops it can be concluded that this method is technically possible but complex, which in turn raises questions about the resources and motives of the malevolent actor. In order to pull this off, the malevolent actor would have to be resourceful, at least from a technical standpoint. Creating believable deep fakes requires technical expertise and data. To have a palpable influence, the fakes would need to be supported by an effective communication campaign, which could itself also use AI to target specific groups of individuals who could be sensitive to the fake material or have the capability to relay it and amplify its impact. This too requires tacit knowledge that cannot be easily acquired.

The fact that the process would be resource intensive prompts the question of who would have the means and the motivation to conduct such destabilizing influence operations. Some workshop participants highlighted that some non-state actors could benefit from creating uncertainty and tension between nuclear-armed states. In the case of South Asia, non-state armed groups such as Jaish e-Mohammed (JeM) reportedly have a vested interest in maintaining tensions

<sup>467</sup> This example is based on participation in a track 1.5 China–USA strategic nuclear dialogue under the Chatham House Rule. Twomey, C. et al., *The US–China Strategic Dialogue: Phase IX Report*, Project on Advanced Systems and Concepts for Countering WMD (PASCC) Report no. 2017-001 (US Naval Postgraduate School, Center on Contemporary Conflict: Monterey, CA, Dec. 2015).

<sup>468</sup> Fitzpatrick, M., ‘Artificial intelligence and nuclear command and control’, Survival Editor’s Blog, 26 Apr. 2019. See also Avin and Amadae (note 92).

<sup>469</sup> This scenario is described by Fitzpatrick (note 468).

<sup>470</sup> Fitzpatrick (note 468).

<sup>471</sup> Fitzpatrick (note 468).

between India and Pakistan.<sup>472</sup> It is debateable whether these groups would be able to conduct such sophisticated information campaign, and they may have other, more conventional means of generating tensions. Nonetheless, the potential for deep fakes to be used to undermine the political and security conditions surrounding strategic stability remains.

Another takeaway from the expert discussion is that the possible use of AI to spoof early-warning systems and poison ISR data requires nuclear-armed states to take precautions when adopting machine learning and autonomy for early warning, ISR, and nuclear command and control. They need to develop or improve their capacity to spot spoofing attempts and signs of influence operations. Some workshop participants, in particular those from India and Pakistan, worried in that regard that nuclear-armed states will be unequally prepared to deal with the threat of destabilizing disinformation campaigns. They pointed out that it would take only one state to actually believe in the disinformation campaign to trigger a nuclear crisis.

### **AI and deliberate escalation**

The concept of deliberate escalation has been defined as ‘An intentional action to increase the intensity or scope of conflict beyond a recognized threshold’.<sup>473</sup>

#### *Why and how it could happen*

In the age of information-enabled warfare, modern armed forces deem the ability to collect and process data into actionable information to be a key factor that determines whether a war may be won or lost. This is one reason why nuclear-armed states—and military powers in general—value AI technology so much. AI allows the processing of more data, more quickly and in more innovative ways. The extent to which advances in AI are mastered can therefore be seen as a source of strategic advantage—or disadvantage for that matter. In this regard, there are two scenarios in which the use of AI could be tied to deliberate escalation into use of nuclear weapons.

The first scenario involves a nuclear-armed state deciding to launch a preventive first strike based on AI-generated information indicating that an enemy might be planning a surprise attack. In this scenario, AI is used to secure a strategic advantage: a nuclear-armed state uses it to inform decisions early on and to take preventive measures. In the second scenario, a nuclear-armed state might decide to deliberately escalate to the nuclear level to offset a strategic disadvantage at the conventional level, which itself might have been caused by AI. In both of these cases, nuclear escalation progresses up a ladder in the context of a conflict or limited war.<sup>474</sup>

<sup>472</sup> Singh, K., ‘Kashmir attack: Is terror group JeM pushing India and Pakistan to the brink of war?’, *South China Morning Post*, 16 Feb. 2019.

<sup>473</sup> Woodhams and Borrie (note 17), p. 9. See also Morgan et al. (note 452), pp. 20–23.

<sup>474</sup> King, J. E., ‘Nuclear plenty and limited war’, *Foreign Affairs*, vol. 35, no. 2 (Jan. 1957); and Halperin, M. H., ‘Nuclear weapons and limited war’, *Journal of Conflict Resolution*, vol. 5, no. 2 (June 1961), pp. 146–66.

*Preventive first strike*

Machine learning can be used to find correlations in data that can then be used to make statistical prediction about future behaviour (see chapter 2). It is not hard to imagine that a nuclear-armed state would try to use machine learning for nuclear-related intelligence purposes. The data gained from the various forms of intelligence-gathering operation (e.g. communication intelligence, human intelligence, geospatial intelligence, measurement and signature intelligence, open-source intelligence, signal intelligence and technical intelligence) could indicate that an adversary might be in the early stages of preparation for an attack. The question of whether any nuclear-armed state would be prepared to launch a preventive first nuclear strike based only on the presumption that its adversary may intend to strike is a difficult one. Participants in the SIPRI workshops highlighted many flaws in this scenario.

First, as one Russian expert pointed out, there is the question of whether the design of an AI system capable of predicting a nuclear strike would actually be methodologically feasible. No nuclear attack has occurred since 1945. States therefore have no historical data with which to train an AI system that would be capable of predicting that a nuclear strike is being planned and prepared. The only knowledge on which programmers can draw to build their model and algorithms is from wargaming simulations. This artificially generated knowledge is not a reliable substitute: reality is always more complex than a simulation. Determining how to weigh the different variables to make a statistical prediction is also hard. This type of scenario-building is susceptible to the injection of national biases and assumptions into the algorithms that are designed to anticipate an adversary's actions.

Second, should AI engineers eventually manage to design such a system, these methodological problems would invite great caution. Workshop participants found it hard to believe that a nuclear-armed state would find such a system reliable enough to initiate a pre-emptive nuclear attack based only on the information that its algorithms produce. Decision makers would probably seek to confirm this information with other sources, including human intelligence. They would possibly wait for tangible evidence of an attack, for example detection of a missile launch by early-warning systems.

*Offset strategy*

A number of workshop participants found a scenario whereby a nuclear-armed state resorts to nuclear weapons to offset its conventional inferiority more credible. For them, it was not impossible to imagine that an insecure state—such as North Korea, which does not have the resources to fully exploit AI technology or be effective in conventional warfare—would rely on a growing nuclear arsenal in order to increase its sense of security.

This scenario would also be plausible in a situation in which one party in a conflict has become so superior in conventional weaponry—notably due to the use of AI for strategic offence, air, missile and space defences, cyber-defence, and

A2/AD—that it would become almost unfeasible for the other party to attack its territory using conventional means. The latter may come to the conclusion that the only way to offset its relative inferiority in conventional means and cause significant harm is to resort to nuclear weapons.

However, AI plays only a secondary role in this scenario. It is only one of the technological factors that contribute to asymmetry in conventional forces. Other political variables will play a fundamental role in decisions leading to the use of nuclear weapons. For this scenario to become a reality, one of the premises would be that states are engaged in a deep geopolitical crisis if not an actual conventional war.

## 5. Mitigating the negative impacts of AI on strategic stability and nuclear risk

This chapter brings together the various country- and region-based analyses in chapters 2–4 to consider how best to mitigate or prevent the risks posed by the adoption of artificial intelligence for military uses in general and in nuclear weapon systems in particular. It maps out the measures that states and the international security community could explore. It starts (in section I) with an overview of the risk-mitigation measures that states have at their disposal and discusses the extent to which these would be adequate, implementable and effective in the light of the geopolitical context and the current state of arms control and disarmament. It then focuses (in section II) on the concrete technical, organizational and (in section III) policy measures that could be implemented in unilateral, bilateral or multilateral contexts.

### I. Mitigating risks: What, how and where

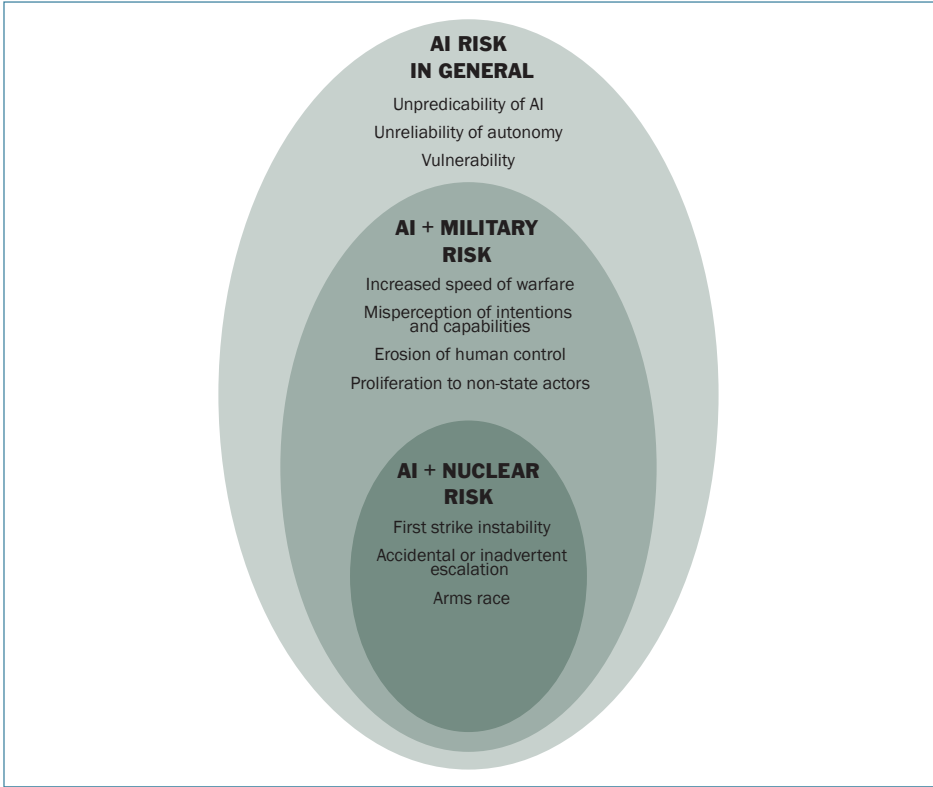
#### **What risks need to be tackled?**

The prerequisite to any adequate, implementable and effective risk-reduction measure is a clear understanding of the problem that is to be tackled. In that regard, one of the challenges of discussing the strategic and nuclear risks posed by AI is that the risk picture is multilayered. Risks and challenges can be analysed on three levels (see figure 5.1).

First, and most broadly, there are the risks inherent to the nature and limitations of AI technology. These include broad and general challenges such as the unpredictability of machine learning systems, the lack of reliability of autonomous systems, and their vulnerability to adversarial attack such as cyberattack, data poisoning and spoofing.

Second, there are the risks posed by the use of AI for military applications. These range from the challenge of a state signalling its own capabilities and intentions and understanding those of its opponent. This is particularly the case when AI-powered military technologies are used to deal with the acceleration of the speed of warfare. A related risk in that regard is the potential erosion of human control over the use of force. A further key concern is the acquisition of military AI by non-state actors, which is facilitated by the dual-use nature of AI technology.

Third, there are the specific risks posed by the use of AI in connection with nuclear weapon systems. These include AI undermining the confidence of nuclear-armed states in their second-strike capabilities. AI may also be employed to weaken the cybersecurity of nuclear force-related systems. AI also has the potential to provide new tools for influence operations on nuclear decision makers. It can increase the risk of accidental or inadvertent escalation, due to system



**Figure 5.1.** Risks and challenges posed by the use of artificial intelligence in nuclear weapons

failure or misuse of technology. Finally, there is the risk of deliberate escalation into nuclear conflict due to conventional force asymmetry fuelled by AI advances.

These three levels of risk are interconnected, but they may each require different types of response. They are also likely to involve various types of stakeholder. Risks inherent to the nature and limitations of AI technology are unlikely to be dealt with by states alone in a bilateral, trilateral or multilateral dialogue.<sup>475</sup> They require a much broader multi-stakeholder process that would involve academia and—since ‘it includes the progenitors of much of the relevant technology’—the private sector.<sup>476</sup>

Although the three levels of risks may require different types of response, they all need to be addressed. Risk mitigation in the context of AI and nuclear weapons requires a holistic approach, although the measures can be implemented either in stages or individually. Measures that could help mitigate the problem of opacity and unpredictability of machine learning systems could, for instance, be instrumental in reducing the risks specifically posed by the use of AI in a mili-

<sup>475</sup> As discussed by Kaspersen and King (note 7).

<sup>476</sup> Kaspersen and King (note 7), p. 123.



tary context in general and in nuclear weapon systems in particular as they would allow humans to exert better control of military applications of AI.

### **How can these risks be mitigated? Selecting adequate measures**

Risk-reduction measures can take many forms (see figure 6.1 below). These can be unilaterally implemented technical, organizational or policy measures that directly attempt to mitigate the risk. They can be bilaterally or multilaterally agreed confidence-building measures (CBMs) to build trust among parties, support crisis prevention and facilitate crisis mitigation. Or they can be internationally agreed regulatory frameworks in the form of hard or soft laws. Hard law regulation could take the form of a legally binding international treaty banning or restricting the development or use of a certain technology or capability.<sup>477</sup> Soft law could take the form of a political declaration or international code of conduct that would identify best practices or provide guidance to states, academia or industry.<sup>478</sup>

The type of measure that would be most appropriate will depend on a number of factors: relevance, efficiency and, most importantly, feasibility. When choosing a measure, it is important to be pragmatic and to weigh the advantages and disadvantages of each.

International law, whether hard or soft, has the greatest normative power, but it may take a long time—years or even decades—to be negotiated. Years of multilateral negotiations may also lead to measures that are limited in the scope of their application. Moreover, given that states might end up agreeing on the lowest common denominator, such measures may have limited relevance or effectiveness in reducing risk. In addition, unless they are widely supported by the community of states, their actual effect on the way that states conduct themselves may eventually be negligible.

Unilateral measures, in contrast, may be adopted quickly and in a way that is directly effective. However, they might not be adopted by all relevant states or relevant stakeholders or may not be adopted in a harmonized way, which could diminish their effectiveness. They are also easily reversible.

Bilateral, trilateral or multilateral CBMs provide a useful compromise in this regard. They would not need to be legally binding and, like some treaties, may be supported by a verification regime, but they would be easier to reach agreement on.

### **Where should the measures be discussed and implemented? Finding the right forum to make measures feasible and effective**

Regional organizations and military alliances, such as NATO, may be relevant forums for discussions on risk-reduction measures among like-minded states. There are a number of bilateral or trilateral discussion tracks between nuclear-

<sup>477</sup> Nishida, M., 'Arms control and developments in machine learning and autonomy', ed. Saalman (note 2), pp. 95–100.

<sup>478</sup> Kozyulin (note 427).

armed states as well as regional forums involving nuclear powers that provide opportunities for discussing the challenges raised by AI in the nuclear context, and in the military context more generally: the Russian–US dialogue on strategic stability, the NATO–Russia Council, the Organization for Security and Co-operation in Europe (OSCE), the Association of Southeast Asian Nations (ASEAN) Regional Forum (ARF) intersessional meetings on non-proliferation and disarmament, the Conference on Interaction and Confidence Building Measures in Asia (CICA), and the Shanghai Cooperation Organisation (SCO).<sup>479</sup> Multilaterally, there are many United Nations-led disarmament forums where nuclear risk reduction is already discussed: the review cycle of the Non-Proliferation Treaty, the Conference on Disarmament (CD), the UN Disarmament Commission and the First Committee of the UN General Assembly.

The problem is that the geopolitical conditions required for constructive discussions in these forums have been worsening dramatically in recent years. The disenchantment that the current US administration has expressed towards NATO has had a detrimental impact on relations among its members.<sup>480</sup> The bilateral arms control framework that the USA and the USSR had created by the end of the cold war is disintegrating as a result of increasing tension between the two largest nuclear powers. On top of this comes the fact that the binary Russian–US nuclear rivalry, a legacy of the old Soviet–US confrontation, is being replaced by regional nuclear rivalries and strategic triangles that no bilateral or trilateral arms control framework currently regulates. Moreover, the state of debate in the various arms control and disarmament forums shows that the commitment of states with the largest nuclear arsenals to pursue stability through arms control and disarmament is in doubt to an unprecedented degree.<sup>481</sup>

Even in the best of circumstances, the UN arms control and disarmament mechanisms have limitations. The NPT review cycle already covers a large number of issues, and participants are generally diplomats who might lack specific technical expertise.<sup>482</sup> It cannot therefore be viewed as a platform where in-depth substantive discussions can take place. The CD, which operates by consensus, has repeatedly failed in the past two decades to achieve any substantial result—the most recent programme of work, agreed in 2009, did not lead to any concrete outcome.<sup>483</sup> A relevant framework for discussing the AI–nuclear risk nexus could have been provided by the ad hoc committee established by the CD in February 2018 (Subsidiary Body 2 on prevention of nuclear war), but this

<sup>479</sup> On these forums see ‘International security cooperation bodies’, *SIPRI Yearbook 2019* (note 8), pp. 587–611.

<sup>480</sup> Buras, P., ‘State of disunion: Europe, NATO and disintegrating arms control’, European Council on Foreign Relations, 28 Feb. 2019.

<sup>481</sup> Seligman, L. and Gramer, R., ‘What does the demise of the INF Treaty mean for nuclear arms control’, *Foreign Policy*, 2 Aug. 2019.

<sup>482</sup> Onderco, M., ‘The programme for promoting nuclear non-proliferation and the NPT extension’, *International History Review*, 23 June 2019.

<sup>483</sup> United National, General Assembly, First Committee, ‘Warning against danger of disarmament machinery “rusting”, First Committee delegates call for greater political will to clear decades-long deadlock’, GA/DIS/3614, 31 Oct. 2018.

was not re-established in 2019.<sup>484</sup> The UN Disarmament Commission, which is a deliberative body that is intended to provide recommendations on various problems in the field of disarmament to the UN General Assembly, may also not be the most promising forum for discussion given that it has also been generally unable to adopt recommendations since the turn of the millennium. However, in 2017 it did adopt consensus recommendations on ‘Practical confidence-building measures in the field of conventional weapons’.<sup>485</sup>

In this context, the only UN forum that seems appropriate for constructive discussions and potentially a concrete outcome is the First Committee of the General Assembly, which focuses on disarmament and international security. This is a consensus-building body in which states can reach common understandings and agree on principles and norms of behaviour. The discussion in the First Committee is usually not highly interactive as positions are determined in advance by the national governments, but they can be substantial and potentially innovative.<sup>486</sup> Notably, the First Committee can establish groups of governmental experts (GGEs) to focus on specific themes and draw on the knowledge of technical experts. It can also draft resolutions for adoption by the General Assembly. The resolutions are not legally binding but have, nonetheless, a norm-setting power.

It should be acknowledged that these forums are not mutually exclusive. Discussions on the impact of AI on strategic stability and nuclear risk can take place at several levels in parallel: directly between nuclear-weapon states, within regional organizations and alliances, or within the UN-led multilateral arms control and disarmament frameworks. The following two sections map out the various types of concrete measure to mitigate AI-related nuclear risk that states could explore in these forums or unilaterally.

## II. Possible technical and organizational measures for risk reduction

Technical and organizational measures are particularly appropriate to tackle the risks that are inherent to the nature of AI technology. It is useful to recall that history already provides a series of useful lessons for the prevention and mitigation of the nuclear risk generated by recent advances in AI.

### **Lessons from the past: Automation and crisis and conflict escalation**

Automation is not a new risk-generating feature of AI technology in nuclear contexts (see chapters 2 and 4). Automation has been used in connection with nuclear weapon systems since the 1960s and the cold war period provided many situations involving automated systems as a possible cause of nuclear escalation,

<sup>484</sup> Finaud, M., ‘The Conference on Disarmament agrees to start working: A wake-up call for “Sleeping Beauty”’, Geneva Centre for Security Policy, 20 Feb. 2018.

<sup>485</sup> United Nations, General Assembly, Report of the Disarmament Commission for 2017, A/72/42, 27 Apr. 2017, annex.

<sup>486</sup> Reaching Critical Will, ‘UN General Assembly First Committee’, [n.d.].

the 1983 Petrov incident being the most famous (see box 2.2). Fortunately, none of these incidents led to actual nuclear weapon use. Nonetheless, these cases provide some useful lessons.

The first key lesson is that humans should remain in the loop. For example, the Petrov case stresses the importance of keeping humans and their common sense as a fail-safe mechanism. It did not lead to a nuclear exchange because a human was able to conclude that the information provided by the early-warning system did not make sense and could not be trusted. Further concrete preventive technical and organizational measures can be implemented.

1. *Technical measures.* Early-warning systems should be kept separate from command-and-control systems for nuclear weapon launch, and humans should remain the link between the two.<sup>487</sup>
2. *Organizational measures.* It is important to ensure that humans are the ultimate interpreters of information provided by early-warning and ISR systems and that only humans can authorize a nuclear weapon launch.

The second key lesson is that human supervisory control is not a panacea. A well-known example is the fratricide involving the US Patriot system during the invasion of Iraq in 2003. It shows that keeping human operators as the ultimate arbitrator of launch decisions may not be enough. Humans can make dramatic mistakes if they are not adequately trained on how to properly use a system and to understand its limitations. They can also make mistakes if the system interface does not allow the operator to have sufficient situational understanding to make well-informed decisions.<sup>488</sup> Concrete technical and organizational measures can contribute to reducing the risk of this type of incident.

1. *Technical measures.* There should be robust testing and evaluation of any new AI-based or autonomous system to determine both the system's capabilities and its limitations in order to anticipate the risk of failure or misuse. Human-machine interfaces should also be designed in ways that give all human operators sufficient situational awareness and reduce the risks of automation bias, under-trust or the out-of-the-loop control problem. This could be done through user trials that would use situational awareness metrics and physiological oversight in order to evaluate human performance.
2. *Organizational measures.* Operators of AI-powered systems should receive robust training to mitigate known risks associated with the supervision of advanced automation. As the literature on human-robot interaction in military operations has shown, in many cases automation increases rather than reduces the need for sophisticated training of human operators.<sup>489</sup> Another good practice is to distribute

<sup>487</sup> Borrie (note 54), pp. 49–50.

<sup>488</sup> Hawley (note 456).

<sup>489</sup> Parasuraman et al. (note 453); Reason, J., 'Safety paradoxes and safety culture', *Injury Control and Safety Promotion*, vol. 7, no. 1 (2000), pp. 3–14; and Murphy and Burke (note 453).

control and decision-making across the chain of command as a way to introduce multiple safeguards.<sup>490</sup>

A third key lesson that can be drawn from past incidents is that a nuclear launch decision should not be made on the basis of a single source of information. Seeking confirmation from multiple types of ISR system is fundamental to reducing the risk of wrongful engagement. In this regard, AI may have a stabilizing function in facilitating greater integration and interpretation of data. This could translate into concrete technical and organizational measures.

1. *Technical measures.* Early-warning and ISR systems should rely on different types of sensor and different pattern-recognition algorithms that can cross-check each other's results. If the algorithms are trained with machine learning, it could be a requirement that the system has redundant algorithms that are meant to do the same task (e.g. detecting an incoming missile) but have been trained with different data sets.
2. *Organizational measures.* Information provided by early-warning and ISR systems should also be verified by human intelligence. The armed forces should have procedures in place to ensure that they can rely on well-trained human analysts who would be capable of determining when the information provided by the systems could be flawed.

### **Precautions for the future: Dealing with the flaws of machine learning and autonomous systems**

Recent advances in AI not only exacerbate strategic challenges and nuclear risks that have been known for decades, but also bring new ones. Machine learning-powered AI systems work in a fundamentally different way to hard-coded rule-based systems. Their algorithms are opaque in their functioning, which makes them potentially unpredictable and vulnerable to adversarial attacks, and hence unsafe to use in life-critical systems such as weapon systems (see chapter 2).

In this light, it would be prudent for states to actively work to prevent an immature adoption of machine learning technology into nuclear weapon systems, particularly early-warning, ISR and nuclear command-and-control systems. This could be done via the implementation of technical and organizational measures.

1. *Technical measures.* States could fund research on testing and evaluation of machine learning systems; on explainable AI (machine learning); on cybersecurity of AI systems; and on systems that can spot deep fakes and other types of AI-generated disinformation.
2. *Organizational measures.* States could facilitate a dialogue between the engineers designing the systems and the military operators so that each can learn from the other and work jointly towards the

<sup>490</sup> Hawley (note 456).

development of the safest technological solutions. They should also avoid the integration of machine learning systems into the most critical parts of the nuclear deterrence architecture—most notably nuclear command and control—until reliable methods of testing and evaluation for machine learning system have been found.

### III. Possible policy measures for risk reduction

#### **Existing nuclear risk-reduction measures**

The fact that many of the risks and strategic challenges posed by AI are not fundamentally new means that policy options to mitigate the impact of AI on nuclear risk already exist. In fact, a number of existing policy options for nuclear risk reduction could have a positive effect on the strategic relations of nuclear-armed states and could help to reduce nuclear risk. These include no-first-use policies, commitments to lower the alert status of nuclear arsenals, transparency and information sharing, and cooperation.

#### *No-first-use policy*

One of the most effective means to prevent the nuclear escalation scenarios described in chapter 4 would be if all nuclear-armed states adopted a clear NFU policy. Currently, China and India are the only nuclear-armed states that have such a policy. Although universal adoption of NFU policies would not necessarily alleviate all the signalling problems generated by the introduction of AI in nuclear weapon systems, from the authors' perspective it would be positive for strategic stability if all nuclear-armed states, particularly Russia and the USA, were to make such a commitment. Some participants in the SIPRI workshops—notably from Russia and the USA—did not share this view.

Indeed, in the current geopolitical context, it is highly unlikely that these two states, as well as the USA's NATO partners France and the UK, would seriously consider that possibility. The USA in particular is constrained from official adoption of NFU by its alliance structures and the views of its allies in both East Asia and Europe.<sup>491</sup> The likelihood of Russia and the USA adopting NFU policies is further compromised by both official and non-official statements from China and India about the longer-term status and nature of their NFU commitments.<sup>492</sup>

#### *A commitment to lower the alert status of nuclear arsenals*

Removing strategic weapons from a launch-on-warning or launch-ready alert status would allow more time for decision makers to make appropriate assessments. India, Pakistan and China have the lowest level of alert as their nuclear

<sup>491</sup> Fetter, S. and Wolfsthal, J., 'No first use and credible deterrence', *Journal for Peace and Nuclear Disarmament*, vol. 1, no. 1 (2018), pp. 102–14, p. 103.

<sup>492</sup> Pan (note 445); and Pant and Joshi (note 445).

warhead and delivery vehicles are reportedly kept separate.<sup>493</sup> It would take at least a few days for them to launch a nuclear strike (although this is changing as they develop the sea legs of their nuclear triads).<sup>494</sup> In contrast, Russia and the USA have their systems on high alert, which means that they are much more likely to use nuclear weapons by accident or deliberately in a crisis or from the outset of a conflict.

It would be highly positive for international security if Russia and the USA were to de-alert their nuclear weapons—although this would not be a panacea without proper training, data integrity and cybersecurity systems in place. In the current context, this also seems to be an unrealistic possibility. Nevertheless, there is reason to hope that this possibility could be part of future bilateral nuclear risk-reduction measures between these two countries, as well as between them and other nuclear-armed states.

### *Transparency and information sharing*

A traditional approach to transparency and information sharing could also help in risk reduction. AI is a kind of technology that is prone to misunderstanding for two principal reasons. First, there remains widespread misconceptions about what AI is and what it can or could do. Second, it is difficult for states to assess in a tangible way each other's progress in this area. It is, for instance, impossible for a state to assess whether an opponent's air defence systems rely on traditional hard-coded programming or an algorithm trained by machine learning if that information has not already been disclosed by the opponent itself.

There are a number of transparency-centred CBMs that states could implement voluntarily and unilaterally or as a result of bilateral or multilateral dialogue. These can be grouped in three categories: (a) AI-specific measures, (b) measures related to the military use of AI, and (c) measures related to the use of AI in connection with nuclear weapons and deterrence.

*AI-specific measures.* One AI-specific measure would be for a state to make publicly available official documents that outline its general strategy and policy on AI. If a state has not yet adopted a national policy or strategy on AI, then it could adopt one to clarify its intention and views.

Another AI-specific measure would be to make publicly available governmental recommendations (e.g. guidelines, procedures and techniques) about testing and verification and about cybersecurity in the field of AI.

*Measures related to the military use of AI.* Military AI-related risk-reduction measures could include a state disclosing AI-related strategies, policies and

<sup>493</sup> The extent to which this is the case for Pakistan has been questioned by e.g. Albright, D., 'Securing Pakistan's nuclear weapon complex', Paper for the 42nd Strategy for Peace Conference, Institute for Science and International Security, 25 Oct. 2001; and Ahmed, M., 'Pakistan's tactical nuclear weapons and their impact on stability', Regional Insight, Carnegie Endowment for International Peace, 30 June 2016

<sup>494</sup> Sial (note 352); Chalmers, H., *A Disturbance in the Force: Debating Continuous At-Sea Deterrence*, Occasional Paper (Royal United Services Institute: London, Jan. 2014); and Kulacki, G., *China's Military Calls for Putting Its Nuclear Forces on Alert* (Union of Concerned Scientists: Cambridge, MA, Jan. 2016).

military doctrines that outline how it intends to use—and not use—AI-related technology in a military context. The US Department of Defense’s directive on autonomous systems is a useful model as it provides some guidance on how they should be used and includes a list of measures focusing on the design and acquisition phases.<sup>495</sup> These are intended to mitigate the risk of incidents caused by loss of control, system failure, or adversarial measures and cyberattack.

Another measure related to the military use of AI would be to appoint identifiable points of contact or focal points for issues related to military AI policy. The points of contact should be able to centralize and report information about governmental thinking on issues related to military use of AI. Their primary role would be to simplify communication between countries. Currently, expertise and responsibilities related to military AI are shared across various ministries and department. This fragmentation makes it difficult for government officials to identify relevant interlocutors in other countries.

One more military AI-related risk-reduction measure would be to disclose general information about ongoing and planned R&D activities in the field of military AI. The US DARPA, for instance, has made information available on its website on all the R&D projects that it currently funds.<sup>496</sup> Such information would help states to understand each other’s intentions and potential capabilities in AI.

*Measures related to the use of AI in connection with nuclear weapons and deterrence.* Nuclear deterrence-related measures could include a state sharing information about how AI fits into its future nuclear modernization plans and what procedures it has or intends to put in place to limit AI-related risks. This could include information on how it implements human control over the use of AI in early-warning systems. The state could also guarantee that nuclear launch decisions are not fully automated.

States could also actively communicate about non-threatening activity involving AI systems that could be misinterpreted as threatening. These include military exercises, operational tests and exploration of the seabed for peaceful purposes.

### *Cooperation*

History also provides a wide variety of bilateral and multilateral CBMs that could be adapted to reduce nuclear risk. These include measures that could help reduce nuclear-armed states’ misperception and mistrust regarding each other’s intentions and capabilities in the field of military AI. Other CBMs could enable a collaborative exploration of risk-mitigation measures targeted at the AI–nuclear nexus. These forms of risk mitigation could include expert dialogue, scientific cooperation and military-to-military cooperation.

*Expert dialogue.* A first step could be for nuclear-armed states to engage in a direct dialogue through which they could discuss outstanding conceptual and technical

<sup>495</sup> US Department of Defense, ‘Autonomy in weapon systems’, Directive no. 3000.09, 21 Nov. 2012, updated 8 May 2017.

<sup>496</sup> US Defense Advanced Research Projects Agency (DARPA), ‘Our research’.



issues related to AI as well as the risks that AI poses in the military sphere in general and in the nuclear field in particular. Ensuring that all sides share a common vocabulary and an equal sense of the danger is a prerequisite to the development of collaborative risk-reduction measures. Three types of dialogue track could be explored.

Track 2 dialogues involving experts from academia and research institutions are useful to raise awareness. They can allow generic issues to be debated and addressed in a manner that does not threaten states' national security interests. These dialogues can take the form of workshops and conferences (such as those convened to prepare this report) or panel discussions held alongside major UN-led meetings (e.g. side-events at NPT meetings). This type of dialogue is particularly appropriate for exploring the connections between AI and nuclear weapons given that the debate on the topic is still relatively new.

Track 1.5 dialogues involving governmental practitioners and experts from civil society would then be useful to translate the product of track 2 discussions into concrete actionable policy ideas. These can then be elevated to the level of decision makers.

Finally, track 1 bilateral dialogues and closed meetings of governmental experts would be where risk-prevention and -mitigation measures could then be negotiated and agreed.

*Scientific cooperation.* There are a number of topics where nuclear-armed states could have a joint interest in scientific cooperation. These include AI safety and AI for the common good.

As mentioned above, recent advances in machine learning have made the testing and evaluation of AI systems extremely difficult. It would be in the common interest of states to invest in joint research projects that would allow for the development of testing and evaluation methods that could certify that AI systems are safe to use.

Nuclear-armed states could also work collaboratively on the R&D of civilian and military applications from which they could all benefit. This could include funding projects that would look into practical solutions to prevent access to or misuse of some types of AI technology by terrorist organizations. While it is less likely to happen, states could work collaboratively to develop technical solutions that would reduce some AI-related escalatory risks. These could include developing friend-or-foe identification systems for underwater systems. AI can also play a positive role in terms of verification for arms control treaties and other measures.

*Military-to-military cooperation.* Military-to-military cooperation could play a key role in preventing and reducing tensions resulting from the adoption, or perceived adoption, of AI by nuclear-armed states. However, this type of cooperation is more politically volatile than those above.

### Arms control agreements

The strategic challenges and risks posed by the military use of AI could be of such magnitude that they call for the adoption of a new arms control instrument. This could take the form of an internationally agreed and legally binding treaty or a politically binding regulation such as a code of conduct.

Participants in the SIPRI workshops seemed to generally agree that such a move would be relatively premature: the conversation remains nascent and the nuclear non-proliferation regime already faces other issues of great complexity. But, even though the discussion on the challenges posed by AI remains speculative, states should not shy away from exploring the possibility of an arms control agreement.

Fundamental questions to be asked are: What would an arms control agreement look like? Would it aim to regulate, for example, general development of AI, development of specific nuclear-related AI applications or specific uses of AI technology? How likely and effective would such an agreement be?

Traditional arms control measures typically offer two types of regulatory approach. States can agree on (a) the types (or number) of systems that they should or should not develop, acquire and use, or (b) how they should or should not use the technology. The appropriateness and likely success from a political standpoint of each of these is discussed below.

#### *Limits to the development of AI in the nuclear domain*

A traditional feature of bilateral nuclear arms control agreements is that they set quantifiable limits on the number and capabilities of nuclear weapons or their delivery vehicles. They also tend to separate civilian from military abilities. SIPRI workshop participants largely admitted that attempting to regulate the development of military AI with such an approach would not be feasible or desirable given the multipurpose and fast-changing nature of the technology.

An agreement would perhaps work if states could identify a specific list of systems or capabilities that they would like to prohibit or limit. However, the discussion in the CCW framework on LAWS shows that it would be a difficult diplomatic endeavour. States have been discussing the issue of LAWS since 2014 and there remain major disagreements as to whether this type of weapon system should be and can be regulated.<sup>497</sup> A fundamental obstacle in the CCW process has been the difficulty of pinpointing in concrete terms the technological characteristics and capabilities that are of concern.

In the case of the AI–nuclear weapons nexus, one specific technological feature seemed to make all workshop participants uncomfortable: using AI to fully automate nuclear command-and-control systems. Participants came up with two primary reasons why this would be a terrible idea: (a) it would be morally wrong; and (b) it would dramatically increase the risk of accidental or inadvertent nuclear escalation.

<sup>497</sup> Boulanin, V., Davis, I. and Verbruggen, M., 'The Convention on Certain Conventional Weapons and lethal autonomous weapon systems', *SIPRI Yearbook 2019* (note 8), pp. 452–57.

In this light, it is not impossible to imagine that nuclear-armed states and the broader international community could agree that the development of fully automated nuclear command-and-control systems should be prohibited by international law. Such an agreement would draw a clear redline but at the same time would not prevent the development of other types of AI capability that nuclear-armed states value for their strategic and national security interests.

However, verification would be a problem. In order to verify that a nuclear command-and-control system does not contain such an automatic mode, the personnel in charge of verification would have to inspect the code. It is unlikely that a nuclear-armed state would agree to open up its systems to such scrutiny. Such an agreement may therefore not include a verification protocol, as is the case with the 1972 Biological and Toxin Weapons Convention (BTWC).<sup>498</sup>

#### *Limits to the use of AI systems in the nuclear domain*

Agreement on the parameters of nuclear-related use of AI technology could be achieved via the definition of positive requirements. This could include, for instance, an obligation to maintain meaningful human control of the nuclear launch decision. This requirement would be another way of putting limits on the development of machine learning and automation in nuclear command-and-control systems. Similarly, states could also agree on specific safeguards that they deem necessary to prevent the risk of accidental use of nuclear weapons. These could include some of the technical and organizational measures listed above.

States could also agree on concrete spatial and temporal limits on the use of AI technology. These could include a commitment not to deploy autonomous ISR platforms in some areas or to agree limits on how long they may be authorized to operate. States could also agree to restrict the use of AI capabilities for certain types of mission, such as interference with nuclear command-and-control architecture via an AI-enabled cyberattack or influence operations using deep fakes.

<sup>498</sup> Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (Biological and Toxin Weapons Convention, BTWC), opened for signature 10 Apr. 1972, entered into force 26 Mar. 1975, British Foreign and Commonwealth Office, Treaty Series no. 11 (1976).

## 6. Conclusions

This report is the final outcome of a two-year research project conducted by SIPRI on the impact of advances in artificial intelligence on nuclear weapons and doctrines. Of particular interest is the question of how advances in AI could have an impact on strategic stability relations among nuclear-armed states and potentially increase the risk of nuclear weapon use. This final chapter summarizes the key findings and provides a series of recommendations on how the strategic challenges and the nuclear risk discussed here may be addressed by the community of experts and governmental practitioners who work on arms control and nuclear risk reduction.

### I. Key findings

The section summarizes the answers provided above to the four questions posed in the introduction.

1. What is the state of AI and what types of capability could nuclear-armed states derive from the recent, current and foreseeable advances in AI?
2. Why and to what extent are nuclear-armed states currently investing in AI? Have they articulated a concrete plan around how AI could be used in future nuclear modernization or developments plans? Are there notable regional differences?
3. What impact might the adoption of AI for military purposes by nuclear-armed states have on strategic stability and nuclear risk? What differences are visible among regions?
4. How should the strategic risk posed by AI be mitigated or even prevented, both regionally and transregionally?

### **Old and new connections between AI and nuclear weapons**

Since the beginning of the 2010s, the field of AI has been undergoing a major renaissance. This AI renaissance consists of two major technological developments. The first of these is the rise of machine learning to become the dominant and most effective approach to AI engineering. This, in turn, has enabled the fast development of autonomous systems, that is, complex automated systems that can execute tasks without human involvement. It is important to bear in mind that when states and companies talk about exploiting recent advances in AI, they are considering machine learning and autonomy. This is also why this report focuses specifically on these technologies.

These advances are only a few years old, so any attempt to assess their impact on nuclear weapons can only be speculative for now. However, it is beyond dispute

that machine learning will play a role in the development or modernization of nuclear weapon systems for at least two reasons.

First, the connection between AI and nuclear weapons is not new. As early as the 1960s, when the discipline of AI was young, nuclear-armed states identified that AI technology could play a role in the nuclear enterprise. As the Soviet Union and the United States had both developed launch-on-warning postures, AI was seen as a technology that could allow the development of automated or semi-automated early-warning and command-and-control systems. These would allow the strategic command to identify threats and adequate responses more quickly.

Second, machine learning is a multipurpose technology. It can therefore unlock new and varied possibilities for a wide array of nuclear weapon systems, ranging from early warning, via command and control to weapon delivery. Machine learning can be used to augment the detection capability of early-warning and ISR systems, to make unmanned systems capable of conducting remote-sensing operation autonomously in complex environments, to design complex control systems for hypersonic delivery systems, or to improve the protection of nuclear assets against cyberattacks.

The question of when, how and by whom machine learning will be adopted in nuclear force architectures is difficult to address at this stage since little detailed information is available in official sources about how nuclear-armed states see the role of AI in their nuclear force development or modernization plans. From a technical standpoint, machine learning has important limitations that represent risk factors and should thus constitute obstacles to its adoption and use in nuclear weapon systems. Most notably, no reliable method is yet available to test and verify its safety and reliability. However, the potential benefits of the technology may prove to be irresistible to some states, which could in that case opt to lower safety and reliability standards in order to maintain or develop a technological edge over their competitors.

### **AI as a strategic priority**

There is clear evidence that all nuclear-armed states have taken notice of the current AI renaissance and made the pursuit of AI a priority. The ability to harness the recent advances in AI is typically presented as an essential enabler of national and military power in the years to come. AI is also systematically presented as a stake in the great power competition, and official sources show that nuclear-armed states are determined to be world leaders. In that regard, these countries have identified the same challenge and priority: mobilizing the human and data resources necessary to be able to design the best AI systems, whether for civilian or military purposes.

The nuclear-armed states are at different stage of maturity in their reflection on the role that they see for AI in their future military modernization plans. The USA has by far the most articulated vision in publicly available official sources. It has developed strategies, road maps and reports that provide concrete indication of which capabilities it wants to use and how it intends to use them. The USA has

also implemented concrete measures that show its determination to make AI a fundamental enabler of its future military priorities. China and Russia also seem to have a clear understanding of how they want their armed forces to use AI in the future. China has recently launched a spate of official documents and programmes on AI that indicate that it intends to take a leading role in the field, notably through its unique ability to generate synergies between civilian and military AI advances. Russian official statements and platforms indicate the centrality of AI development in achieving its military aims. The benchmark for both China and Russia seems to be the USA's vision and plans for AI in the military sphere. Chinese and Russian official documents and expert commentaries often refer to what the USA is doing; and overall China and Russia are prioritizing the same types of AI-enabled capability that the USA has or is developing.

France and the United Kingdom also have ambitions to be great powers in AI, but they have only just begun to articulate concrete visions and plans for how they intend to use AI in their armed forces. India and Pakistan have comparatively the least developed AI visions in publicly available sources. India, despite its reputation in software and IT, is still in the early phases of policy adoption, and the policy document that outlines its ambition in the field suggests that it currently prioritizes development for civilian purposes. Yet the recent establishment of two multi-stakeholder task forces to explore, respectively, civilian and military AI applications and aims indicates that India aims to make progress in both spheres. Pakistan's official vision for AI has so far been limited to initiatives that set general objectives for ensuring Pakistan's competitiveness in AI. The conversation on military application of AI seems to remain limited to expert circles. In the case of North Korea, little can be said due to the lack of publicly available information and official sources, but there are some indications that it is interested in the potential of AI for cyberwarfare and information warfare, notably deep fakes.

The nuclear-armed states, even those with the most developed vision of the strategic role that AI could play, have so far issued little official information on their policies on how they would use, or not use, the advances in AI in nuclear weapon systems. Instead, existing official documents tend to address the legal, ethical and security challenges posed by the increasing use of AI and robotics in conventional and cyber weapons. It can be deduced from the statements that states have made about lethal autonomous weapon systems that they see risks associated with the increasing delegation of tasks to AI systems and that there is a need to ensure that humans retain a form of meaningful control over any nuclear launch decision.

### **AI and strategic stability and nuclear risk**

The fact that nuclear-armed states have an uneven history when articulating a plan for how they intend to use AI for military purposes in general and in connection with nuclear weapons in particular makes it hard to determine the net effect of recent advances in AI on strategic stability and nuclear risk. However, three inferences can be made based from what is known about how the field of

nuclear strategy works, how AI and automation have been used and misused in the past, and what the current limitations of the latest advances in AI are.

First, AI could enable the development of nuclear force-related applications that could have both stabilizing and destabilizing effects on strategic stability, depending on the regional context. AI could be destabilizing in regions where it could reinforce force asymmetry between nuclear-armed states as it could undermine the status quo of the parties' deterrence relationships. In regions where nuclear-armed states already enjoy some kind of force symmetry, both conventional and nuclear, the effect of AI might not necessarily be destabilizing. In fact, it could be stabilizing as it could reinforce the acceptance of mutual vulnerability and also provide nuclear decision makers with the confidence that they are better prepared to deal with the risks of nuclear escalation in a time of crisis.

Second, an effect on strategic stability relations between nuclear-armed states can be sensed even without waiting for advances in AI to be turned into actual and readily deployable military capabilities. AI is a technology that can be easily misperceived; it is also a technology that is difficult to control through arms control mechanisms given its inherent dual-use and multipurpose nature. A state's belief that an opponents' investment in AI, even non-nuclear-related, could in due course give that opponent the ability to threaten its second-strike capability could be sufficient to generate insecurity and lead the state to adopt measures that could have a negative impact on strategic stability and increase the risk of a nuclear conflict. Hence, nuclear-armed states have a vested interest in communicating more clearly and openly about their capabilities and intentions in the field of AI, and also in discussing the nuclear risk that could emerge from the use of AI in the nuclear deterrence architecture.<sup>499</sup>

Third, regarding nuclear risk, there are many imaginable ways in which the military use of AI technology could increase the likelihood of a nuclear conflict. For example, current military AI applications are still brittle and could fail or be misused in way that could trigger an accidental or inadvertent escalation of a crisis or conflict into a nuclear conflict. Alternatively, as explained above, the way in which AI is expected to enhance the military capabilities of an adversary could be the reason why a nuclear-armed state may decide to deliberately escalate to the nuclear level. However, history indicates that, in order for these scenarios to become reality, a number of destabilizing dynamics would need to align. In the current geopolitical context it is hard to imagine how AI technology alone could be the determining trigger of nuclear weapon use. Geopolitical tensions, lack of communication and inadequate signalling of intentions are variables that would play an equally important if not greater role than AI technology in triggering an escalation of crisis or conflict to the nuclear level.

<sup>499</sup> As suggested by Rickli (note 412).

### **Addressing the risks**

In the light of the above, it is not too early to start discussing options that nuclear-armed states and the international security community could explore to prevent and mitigate the risks that military and even nuclear force-related uses of AI pose to peace and stability (see figure 6.1).

Some solutions already exist. Existing arms control instruments include a number of proven technical, organizational and policy measures that could be discussed and implemented, unilaterally, bilaterally or multilaterally. These include adoption or reaffirmation of no-first-use doctrines; lowering the alert status of nuclear arsenals; increasing transparency on future nuclear modernization plans and on the strategies, policies and military doctrines that outline how states intend to use—or not use—AI technology; supporting confidence-building measures such as expert dialogue, scientific cooperation and military-to-military cooperation; and agreeing on politically or legally binding agreements that would prohibit or regulate the development or use of certain technologies or capabilities.

However, political pragmatism is required to determine which measures and adoption processes will be adequate, implementable and effective. The main challenge is that the political and institutional conditions required for a constructive discussion between nuclear-armed states on arms control-related issues have been worsening dramatically in recent years, while the conversation on AI-related risks is still new and speculative.

## **II. Recommendations**

Based on the above conclusions, the following general recommendations can be made to the community of experts and governmental practitioners that work on arms control and nuclear risk reduction. While general in nature, they are based on the research conducted in this SIPRI project and on the contributions of its internationally and professionally diverse participants.<sup>500</sup> The best and most pragmatic way to deal with the strategic challenges that AI raises is to take a step-by-step approach. There are four types of mutually reinforcing measure that could be taken, sequentially or in parallel (see figure 6.2).

### **Raise awareness and get the risk picture right**

A priority should be to support awareness-raising measures that will help the relevant stakeholders—governmental practitioners as well as civil society experts—gain a realistic sense of the challenges posed by AI in the nuclear arena. Ensuring that all sides share a common vocabulary and an equal sense of the danger is a prerequisite for the development of collaborative risk-reduction

<sup>500</sup> Participants and authors in the SIPRI project came from Australia, China, France, Germany, India, Israel, Japan, South Korea, the Netherlands, New Zealand, Norway, Pakistan, Russia, Sri Lanka, Sweden, Switzerland, Taiwan, the UK and the USA.



measures. This discussion will have to be inclusive, so as to involve academia, the private sector and civil society. It should not be limited to the usual nuclear risk-related discussion forums, such as the review cycle of the Non-Proliferation Treaty. It should also be transregional and should not be limited to a discussion among like-minded states.

### **Increase transparency on AI**

Another priority should be to support transparency measures that can help to reduce misperception and misunderstanding among nuclear-armed states on AI-related issues. In this regard there are a number of transparency-centred CBMs that states could implement voluntarily and unilaterally or as a result of bilateral, trilateral or multilateral dialogue.

These CBMs can be grouped into three categories. The first includes AI-specific CBMs, such as drafting and making publicly available a national AI strategy (for those countries that have not yet done so). The second category includes CBMs related to the military use of AI, such as disclosing AI-related strategies, policies and military doctrines that outline how a state intends to use—or not use—AI-related technologies in a military context. The third category includes CBMs related to the use of AI in connection with nuclear weapons and deterrence, such as sharing information about how AI fits into future nuclear modernization plans, and also the limits that are placed on the use of AI in nuclear force-related systems, for instance in the form of human control measures.

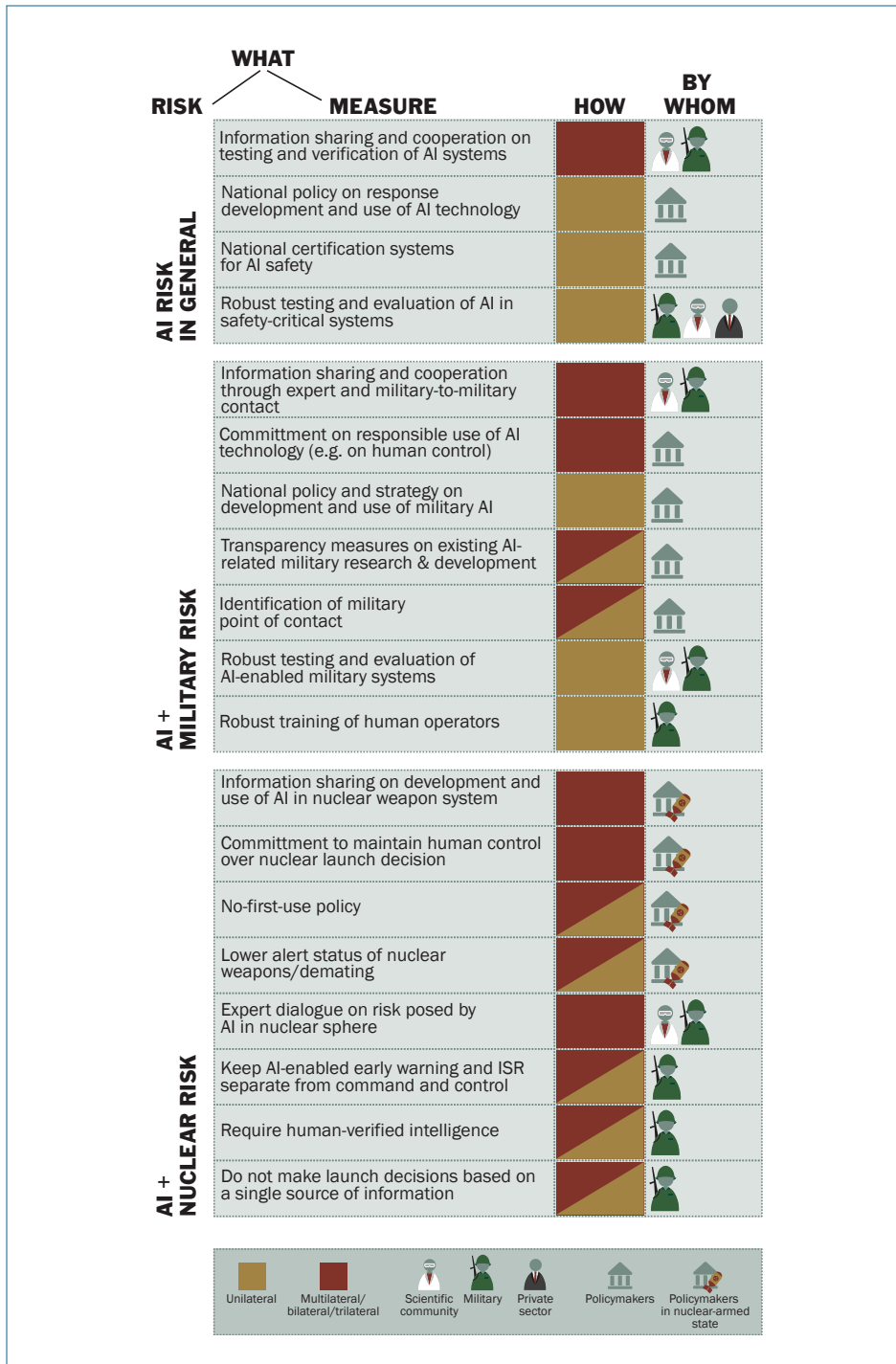
### **Support collaborative resolution of challenges posed by AI and exploration of universally beneficial use of AI**

Some CBMs can enable a cooperative approach to problem solving and these should be supported. It would be in the common interest of nuclear-armed states to join forces to solve some of the problems—such as AI safety and AI security—that make the use of AI highly problematic if not dangerous in the nuclear arena. It would be also highly valuable for states to discuss the ethics of AI and to work collaboratively on the research and development of civilian and military applications from which they could all benefit.

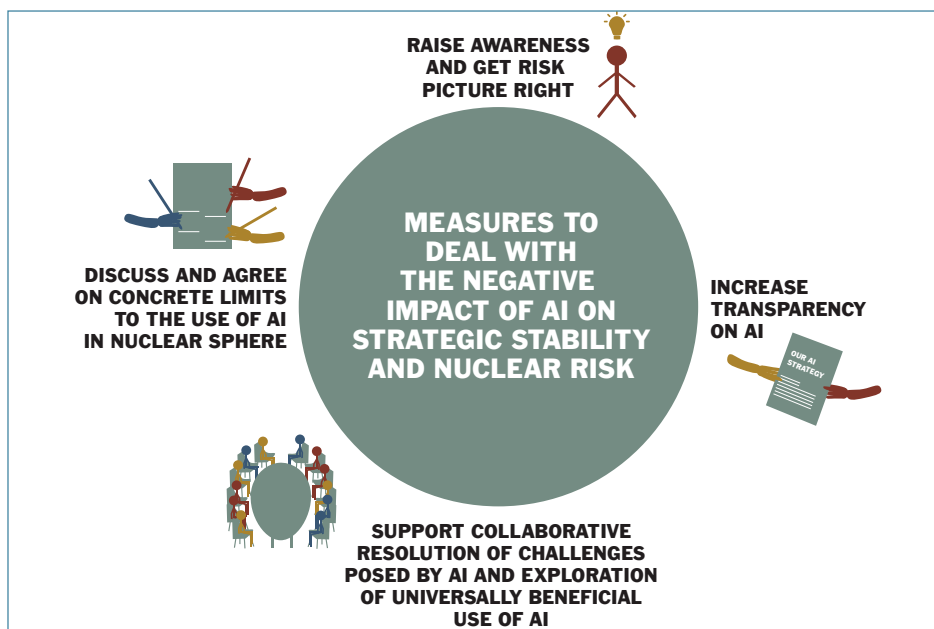
### **Discuss and agree on concrete limits to the use of AI in the nuclear sphere**

A final step would be for nuclear-armed states to discuss—and potentially agree among themselves—concrete technical, organization and policy measures that could reduce the negative impact of AI on strategic stability and nuclear risk.

It is not impossible to imagine that nuclear-armed states and the international community could agree that the development of a fully automated nuclear command-and-control system is unacceptable. Such an agreement would draw a clear redline, but at the same time would not prevent the development of other



**Figure 6.1.** Possible risk reduction measures and how they can be implemented  
 AI = artificial intelligence, ISR = intelligence, surveillance and reconnaissance.



**Figure 6.2.** Four key measures to deal with the negative impact of AI on strategic stability and nuclear risk

types of AI capability that nuclear-armed states value for their strategic and national security interests.

In the same vein, states could also discuss and agree on specific safeguards that they deem necessary to prevent the risk of accidental use of nuclear weapons. These could include some concrete technical and organizational measures, three of which stand out. The first of these is a requirement for robust testing and evaluation of new AI-based systems. The second is a requirement that both early-warning and ISR systems rely on more than one type of sensor with different pattern-recognition algorithms. The third requirement is that information provided by early-warning and ISR systems should also be verified by properly trained human operators. While the inherent difficulty of achieving these limitations should be recognized, they are integral to evolving standards and norms for the future military integration of AI.

In sum, the current AI renaissance is bound to have an impact on nuclear weapons and doctrines. It will generate opportunities but also risks, old and new, for strategic stability. The adoption of recent advances in machine learning and automation in the military sphere, and in nuclear weapons in particular, will be incremental and take time. However, it is not too early for states and international organizations to look for policy options and identify opportunities to tackle the challenges presented by these technologies. It can even be hoped that such an effort will provide a useful opportunity for nuclear-armed states to discuss nuclear risk reduction among themselves as well as with the global community of states in a constructive and collaborative manner.

## About the authors

**Vincent Boulanin** (France) is a senior researcher at SIPRI, where his work focuses on the challenges posed by the advances of autonomy in weapon systems and the military applications of artificial intelligence (AI) more broadly. Before joining SIPRI in 2014, he completed a doctorate in political science at the *École des Hautes Études en Sciences Sociales*, Paris. His recent publications include *Bio Plus X: Arms Control and the Convergence of Biology and Emerging Technology* (SIPRI, 2019, co-author). Boulanin edited the first volume, on Euro-Atlantic perspectives, of *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk* (SIPRI, 2019).

**Moa Peldán Carlsson** (Sweden) is a research assistant on emerging military and security technologies at SIPRI. Her current research focuses on AI, cybersecurity and arms control. Her other research interests include the relationship between gender and terrorism. She wrote her bachelor's thesis on alternative paths for female empowerment in militarized societies. Prior to joining SIPRI, Peldán Carlsson was an intern at the Swedish national committee of UN Women and managed her own graphic design and communications company.

**Lora Saalman** (United States) is an associate senior fellow on armament and disarmament at SIPRI and a senior fellow with the Global Cooperation in Cyberspace Programme of the EastWest Institute, New York. She has also worked as an associate professor at the Daniel K. Inouye Asia–Pacific Center for Security Studies, Honolulu, USA, an associate in the Nuclear Policy Program at the Carnegie–Tsinghua Center for Global Policy, Beijing, an adjunct professor at Tsinghua University, a research associate at the Wisconsin Project on Nuclear Arms Control, and a visiting fellow at the Observer Research Foundation, India, and the James Martin Center for Nonproliferation Studies and an intern at the International Atomic Energy Agency (IAEA), Vienna. She was the first US citizen to earn a doctorate from Tsinghua University's Department of International Relations, completing all her coursework in Chinese. Saalman edited the second volume, on East Asian perspectives, of *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk* (SIPRI, 2019).

**Fei Su** (China) is a researcher with the China and Global Security Programme at SIPRI. Her current research focuses on China's engagement with North Korea, South Korea and Japan. Prior to her current post, Su lived and studied in Seoul for three years, where she strengthened her fluency in Korean and obtained a master's degree in public administration with a focus on governance from the Graduate School of Public Administration of Seoul National University, where she wrote her dissertation in Korean.

**Petr Topychkanov** (Russia) is a senior researcher with the SIPRI Nuclear Disarmament, Arms Control and Non-proliferation Programme. He works on issues related to nuclear non-proliferation, disarmament, arms control and the impact of new technologies on strategic stability. Prior to joining SIPRI in 2018, he was a senior researcher at the Center for International Security at the Primakov Institute of World Economy and International Relations (IMEMO) of the Russian Academy of Sciences. From 2006 to 2017 he was a fellow with the Carnegie Moscow Center's Nonproliferation Program. He has a doctorate in history from the Institute of Asian and African Studies, Moscow State University. His recent publications include 'US–Soviet/Russian dialogue on the nuclear weapons programme of India', *Strategic Analysis* (May 2018), and *Setting the Stage for Progress Towards Nuclear Disarmament* (SIPRI, 2018, co-author). Topychkanov edited the third volume, on South Asian perspectives, of *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk* (SIPRI, 2020).

