

2023-05-18

Review of James Johnson, Artificial intelligence and the future of warfare: the USA, China and strategic stability. Manchester University Press, 2021 xi+221 pp.

Cladi, Lorenzo

<https://pearl.plymouth.ac.uk/handle/10026.1/20922>

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

James Johnson, *Artificial intelligence and the future of warfare: the USA, China and strategic stability*. Manchester University Press, 2021xi+221pp.

This is the accepted version of a book review which appears on *Defence Studies*

Among the many potential sources of instability and great power competition, there is a mysterious and largely unknown area to the political science and international relations community: artificial intelligence. James Johnson's book is a welcome addition to the literature on strategic studies. It really compels the reader to engage with the many and unanswered questions about the role that artificial intelligence could play in international politics. Whilst the book ultimately generates more questions than answers, the analysis is worth engaging with and it will be intelligible to students with previous knowledge of strategic studies and international relations. The book is divided into three parts. Part I provides a theoretical framework to understand the likely consequences of artificial intelligence (AI) for nuclear stability. Part II delves into the competition between the US and China and how this increasingly occurs in artificial intelligence. Part III looks at a number of case studies to explain to the reader what the escalation risks associated with AI are.

The field of AI encompasses machines that reproduce capabilities that are normally in the realm of humans. These typically include language, reasoning, learning and observation. The advent of AI is good news for human society: in many cases, it can reduce the time needed to perform a given task. However, the impact of AI on international politics is largely unexplored. Also unknown remains how, in what ways, and if, great powers might use AI in a potential military confrontation with one another. If that were the case, then the future of warfare could be very different and

this book succeeds in getting the reader to think about what could come. Johnson's area of inquiry lies in the field of strategic studies and he seems particularly interested in AI's impact on strategic stability. There is a myriad of potential applications and analyses of the impact of artificial intelligence. Johnson's choice to focus on the potential military implications of AI for nuclear risk ties the book to relatively well known (at least to scholars of international relations) debates on strategic stability, deterrence and escalations management between rival nuclear-armed forces.

Strategic stability, broadly speaking, entails the absence of armed conflict between states that possess nuclear weapons. Strategic stability can also describe a situation in which states lack incentives to use nuclear weapons. In a situation of strategic stability, states enjoy a harmonious relationship because they also lack incentives to build nuclear weapons. In today's international political system, the US is concerned with maintaining 'first mover' advantage. The most likely challenger to the US 'first mover' advantage, so one of the book's underlying assumptions goes, seeking to both capture and to then obtain that first mover advantage, is China. Combining this alleged Chinese desire with the proliferation of AI leaves us with the deep uncertainty over the future of strategic stability, with potential repercussions for nuclear stability.

How the advent of AI risks affecting strategic stability between the US and China is the biggest, most ambitious, question that Johnson seeks to tackle. Johnson deserves considerable credit for seeking to tackle this question. Admittedly, he plays it relatively safe up front by arguing that AI's impact upon strategic stability will likely

be 'more prosaic and theoretical, than transformational' (p.4). Yet, as one delves into the book, several alarm bells ring, bringing the reader to seek to navigate through the uncertainty that permeates the use of artificial intelligence. In several places throughout the book, the reader comes to terms with very disturbing scenarios such as the potential use of drone swarms and hypersonic weapons. These, Johnson argues, can bedevil missile defences, undermine states' nuclear-deterrent forces, and increase the risk of escalation. It is possible that, under competitive geopolitical pressures, states could decide to authorise unknown military algorithms to take decisions on military issues. Similarly, whilst Johnson argues that for the moment nuclear-armed states have agreed not to pre-delegate nuclear command and control to machines, we are not immune from a potential decision to automate nuclear forces and launch postures.

In Johnson's account, this risk is especially high for authoritarian states because, as Johnson puts it, 'authoritarian states may perceive an adversary's intentions differently from a democratic one' (p.172). Furthermore, a regime that sees its second-strike capabilities as exposed could be more inclined to automate its nuclear forces and launch postures. Democracies are not as likely to reach the decision to automate nuclear weapons because of political processes, accountability, nuclear-launch protocols, more mature civil-military relations and strategic values between allies. Here comes what to me looked like one of the key lessons from Johnson's book: the perceptions that adversaries have of each other's capabilities matter somewhat more than what a particular AI-enabled application is technically capable of doing. Whilst the theoretical foundations of this lesson are not entirely novel, the reader still has to come to terms with how AI might upset strategic stability and how

long it will be before it does so. On this matter, Johnson's analysis does not quite reach far enough because no theory is offered as to how AI might ultimately constitute a brand new level playing field among the great powers in the 21st century. The reader is left to wonder over that point. Moreover, the controversies over AI arguably pose a risk to accountability and transparency in democracies too. Without enough information on AI-related technology, criticism abounds over its use, and potential misuse. It also casts a doubt over whether democracies can effectively play a crucial role in shaping the norms over AI. Balancing the need to achieve strategic aims with the need for accountability is a challenge for democracies. How democracies are going to be able to strike this balance is, again, left hanging by the analysis. Johnson's theoretical foundations cannot lead to suggest ways in which AI could be the 'feared game changer'. Neither do we have enough elements to substantiate the claim that adversaries might think about AI to surprise other great powers. Nor do we convincingly understand why every single great power, in this day and age, might be interested in capturing 'first mover' advantage. Arguably, the US is very concerned about losing its 'first mover' advantage and therefore about the rise of challengers such as Russia and China. However, to really decipher what the latter makes of AI we arguably need to examine in a deeper and more sophisticated way to understand what China thinks. Subsequently, we could put Johnson's assumptions to the test in a more clear-cut manner.

In terms of where to go from here, it is always difficult to predict what to do when tackling a topic surrounded by so much uncertainty. Put differently, the devil is in the details. Johnson still seeks to offer us some ways forward. Firstly, he argues that great powers should establish an international framework for governance, norms,

regulations, and transparency in the development of AI-augmented military capabilities. Secondly, decision-makers must carefully consider the nuanced trade-offs between increasing degrees of complexity, interdependency, and the vulnerabilities that military AI could engender. Thirdly, the think-tank community, and AI experts and academics, should pool their resources to investigate the implications of military AI for a range of potential security scenarios.

Johnson's policy recommendations should be taken seriously to make sure that artificial intelligence does not become a fundamentally destabilizing force, thereby increasing the risk of greater geopolitical instability, possibly involving the use of nuclear weapons. Geopolitical tensions between great powers to retain (in the case of the US) or to capture (in the case of China) the first-mover advantage in the pursuit of AI will likely create incentives not cooperate. What this means for the future of nuclear deterrence is something that readers should think about and reading this book will undoubtedly set you up for it.

Lorenzo Cladi
University of Plymouth