Should Strategists Worry About the Philosophy of Artificial Intelligence?

Article · September 2023

1 author:

Vincent J Carchidi
Middle East Institute
8 PUBLICATIONS **7** CITATIONS

Should Strategists Worry About the Philosophy of Artificial Intelligence?

Vincent J. Carchidi -
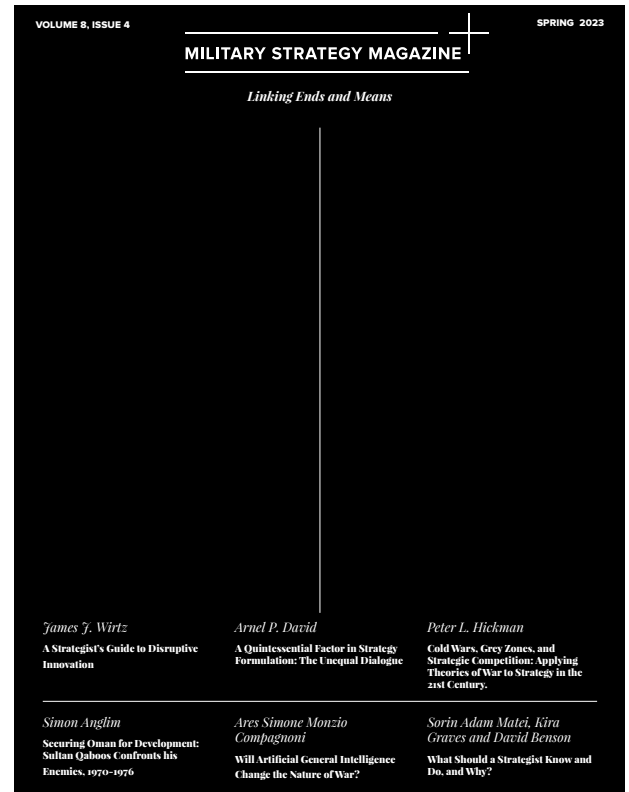*Middle East Institute - Strategic Technologies and Cyber Security Program*

This article explores an implicit philosophical assumption in strategy formation: that biological intelligence can be replicated, granting capabilities sufficient to justify a medium- and long-term strategic focus on artificial intelligence. It explains how an implicit and unstudied philosophy of AI often makes its way into force structure planning and national policy objectives across states.

Should Strategists Worry About the Philosophy of Artificial Intelligence?

Vincent J. Carchidi – Middle East Institute – Strategic Technologies and Cyber Security Program

Photo 171270756 | Artificial Intelligence © BiancoBlue | Dreamstime.com

About the author

Vincent J. Carchidi is a Non–Resident Scholar at the Middle East Institute's Strategic Technologies and Cyber Security Program, where his work focuses on the intersection of emerging technologies, defense, and international affairs.

Introduction: The Unspoken Assumption

When individuals talk (and write) about technologies like artificial intelligence (AI), they do so in different personalities: in some instances, as people recreationally interested in the technology, in others as participants in large-scale social experiments with new technological deployments. In either case, individuals take on the assumptions of these personalities. In forums such as this, individuals take on a rather specific personality: that

of strategists. In doing so, they confront these underlying beliefs, going to great lengths to uncover what Francis J. Gavin calls our "unspoken assumptions"[i] about how the world works. Strategists aim to assess and create strategies by pinpointing their "prevailing assumptions" with a "future-leaning"[ii] bias.

Here, I explore an "unspoken assumption" about AI with which strategists have insufficiently grappled: that artificial intelligence is *possible*. There is an unspoken assumption that the wet, fleshy stuff within human (or animal) skulls is unique but replicable, reproducible on silicon substrates. But what if it is not possible to replicate intelligence via artificial means? What if today's "narrow" AI is merely a series of engineering-based workarounds that function as band-aids on the fundamental problems of reproducing intelligence?

Much of the strategic interest in AI is derived from an implicit chain of reasoning. It begins with the assumption that biological intelligence can be reproduced via artificial means. With this reproduction comes capabilities once exclusive to biological organisms. The reasoning ends by linking such artificially intelligent systems with capabilities relevant to, say, new force structures (e.g., the long-term development and adoption of semi- or fully autonomous "AI-piloted" jets[iii]). The "unspoken" part of this reasoning concerns the assumption that biological intelligence can, in fact, be reproduced via artificial means; that the upper limit on such technological innovation is comparable, equivalent to, or higher than that of biological organisms' capabilities. This is *not* about the distinction between "narrow" and "general" AI which, although useful, can incorporate this unspoken assumption in the former in nearly as pernicious ways as the latter.

Strategy's orientation to the *future* is what makes this unspoken assumption problematic for strategists. The assumption that biological intelligence can be replicated informs a medium- and long-term developmental trajectory for AI. For our purposes, what is meant by both "strategy" and "strategists" is not just anything and anyone. For the former, the unspoken assumption about AI directly implicates the three legs of strategy—policy ends, strategic ways, and military means—identified by this journal, albeit in varying degrees. For the latter, this article is for policy analysts, military personnel, academics, wargamers, and interested individuals across nations who see medium- and long-term potential for AI's impact on force structures, doctrine formation, and national policy objectives.

This article begins with a breakdown of how strategists employ an implicit philosophy of AI in their dealings with the technology. This allows for a clearer understanding of how pernicious this "unspoken assumption" can be in strategic thought, allowing us to then pinpoint its origin. This origin story is told in lively detail, illustrating how comparisons made between biological brains and artificial neural networks have thoroughly shrouded the assumption that biological intelligence can be replicated via artificial means. The relevant strategist, it is explained, cannot assume an ever-improving "narrow" AI, as the developmental potential of the technology is sharply limited. The article closes with insights into the relationship between strategists, strategy, and AI.

The Strategist's Philosophy of AI

Strategy in the "fourth industrial revolution"[iv] is decidedly interdisciplinary.[v] With the breadth of scientific endeavors that accompany it and its intersection with defense and international affairs come a litany of assumptions about science and technology. Yet there is "no such thing," as Daniel Dennett observed, "as philosophy-free science."[vi] The unspoken assumption about the possibility of AI often reflects an implicit and unstudied philosophy of AI.

Indeed, the unspoken assumption about AI is operative in multinational government statements, documents, and initiatives.

In May 2023, U.S. Air Force Col. Tucker Hamilton's hypothetical misstatements about a rogue autonomous drone[vii] highlighted a broader effort within the U.S. military to develop and adopt AI-enabled autonomy technologies with a long-term focus.[viii] Such efforts are supported by figures including U.S. Chairman of the Joint Chiefs of Staff Gen. Mark Milley who predicted that roughly one-third of "the advanced industrial militaries of the world likely will be robotic" in the next 10 to 15 years.[ix] Milley's comment reflects a view now instantiated in the National Security Strategy of the United States that AI, alongside other emerging technologies, promises to "transform warfare"[x] and serve as one of the "foundational technologies of the 21st century."[xi]

The United States is not alone in this long-term AI focus. In April 2023, Germany's Bundeswehr released its "2035 and beyond" objectives for German naval forces, laying out a need for "comprehensive" integration of unmanned systems alongside AI for surface and underwater warfare as well as enhanced maritime domain awareness.[xii] In February 2023, Japan's Self-Defense Ministry announced plans to abolish its "obsolete" attack and observation helicopters with a reduction of 1,000 required human personnel as it adopts new uncrewed systems.[xiii] Finally, in late-2020, the newly-minted state-backed Beijing Institute for General Artificial Intelligence took up the goal of creating AI systems trained on "small data" while emulating human cognitive abilities,[xiv] with Director Zhu Song-Chun calling Artificial General Intelligence "the global strategic high ground of technology and industrial development."[xv]

Underlying each of these examples is a philosophy of AI. The critical feature common to all is that each lays claim to

a specific *developmental potential* for AI; each assumes that AI can be sufficiently developed to help fulfill goals such as the reliable reduction of human personnel, the robust execution of uncrewed maritime warfare or surveillance in adversarial conditions, or that AI can one day reproduce or emulate higher-order cognitive functions.[xvi]

What if these plans are riding a wave of AI enthusiasm that is destined to fail?

These questions may sound overthetop, especially as generative AI systems drive commercial and defense engagement. But the unspoken assumption that what we do as humans (or what animals do), in principle, can be replicated, is a serious, underappreciated factor in strategy formation in the fourth industrial revolution. If the assumption is incorrect, it will directly impact the utility of medium- and long-term force structure planning, the formation of military doctrines concerning the use of AI-enabled weapons and platforms, and even deliberation on national policy objectives pertaining to AI.

This unspoken assumption is *fundamental*—it considers the distinction between "narrow" and "general" AI a useful but ultimately insufficient conceptualization of the technology's capabilities and applications, casting doubt that the former is, in fact, replicating intelligence and barring the latter from ever coming to fruition.

The titular "worry" is *not* that strategists should be worried about AI becoming "general" and upending the fundamental nature of warfare, as Ares Simone Monzio Compagnoni[xvii] argues. Rather, the worry is that strategists who have witnessed the increasing capabilities of "narrow" AI and now seek to apply it more widely across domains with an eye towards its incremental improvement may be working towards a partial impossibility. It is an *assumption* that narrow AI can sufficiently outgrow its propensity to real-world failures sufficient to justify its medium-and long-term strategic focus. The assumption may be incorrect.

For the strategist who sees the transformative potential of AI, the possibility that the biological stuff is the only game in town is worrying indeed.

Origin of the Unspoken Assumption

In 2021, AI expert J. Mark Bishop made a prominent case against the possibility of AI in an article bluntly titled, "Artificial Intelligence Is Stupid and Casual Reasoning Will Not Fix It."[xviii] Some field-specific history is embedded in this title that we should briefly review. (I promise it will not be boring—we are talking about our brains, after all).

Artificial neural networks were originally inspired by the composition of neurons within biological brains. The idea, recently emphasized by figures like Geoffrey Hinton,[xix]

goes something like this: the human brain is composed of billions of neurons bound together by trillions of connections. Activity between these neurons takes the form of signals sent between them (through "synaptic connections"). The final result of these signals can be expressed through *arithmetic*. Simply put, the neuronal activity of the human brain can be characterized in *computational* terms.[xx]

Over time, this conception of the brain inspired the idea that intelligence can be *replicated* via computational means. We see the analogy between brains and AI in the structure of neural networks today, the most basic component of which is an artificial "neuron." Artificial neurons are arranged in layers, with a simple network consisting of an input layer, which feeds into "hidden" layers, and then results in an output layer. A positive or negative number assigned to connections between neurons in successive layers determines how impactful the output from one neuron will be to the next. That is, the strength or weakness of the connection is determined by this number (positive is stronger, negative is weaker). As a model is trained, these weights change to yield the appropriate output.[xxi]

The "deep" in "deep neural networks" refers to the hundreds of layers of neurons they possess. These networks, in contrast to older, shallower artificial neural networks, are dependent on *enormous* amounts of data to properly train. More than this, their recent successes owe as much to increases in the available computing power needed to process data as to the amount and quality of the data themselves. And while many AI success stories of the past decade use more than just deep learning, this technique underpins most examples: the Go-playing systems AlphaGo and AlphaGo Zero, software underpinning Tesla's and Waymo's self-driving vehicles, large language models like GPT-3 and GPT-4, text-to-image generators like DALL-E and DALL-E 2, and text-to-video generators like Meta's Make-A-Video.

While it would be an exaggeration to say this is all "just math,"[xxii] computation underwrites all of deep learning.

Burying the Lede

Just as interesting as deep learning's successes are its *failures*—these systems tend to be surprisingly stupid. Deep neural networks are so data-centric that they are confined to the data on which they are trained. Popular systems like ChatGPT—which is designed to simply predict reasonable continuations of text[xxiii]—sometimes appear to be doing something more "general," but this is because of natural language's open-ended uses and our predilection to anthropomorphize its human-like outputs. ChatGPT suffers from serious, unintelligent problems including hallucinations, unreliability, and an inability to distinguish possible from impossible.

ChatGPT is not alone. OpenAI's computer vision system CLIP incorrectly classified a granny smith apple as an iPod simply because somebody stuck a label with the word "iPod" on it.[xxiv] KataGo, a state-of-the-art open-source Go-playing agent, was beaten by an *amateur* human Go player by employing a fairly simple technique (the creation of a large "loop" of stones while distracting the agent in the corner of the board).[xxv] A DARPA object recognition system tasked with detecting human movement was fooled by marines doing somersaults and hiding under cardboard boxes as they approached it without being detected.[xxvi]

These problems are not dissimilar. They are the result of deep learning systems' detachment from any understanding of the world and an ability to reason over data, even though they can often perform certain tasks far better than humans. They are well-documented categories of problems, triggering contentious debates about how to best resolve them. Some, like Judea Pearl and Dana Mackenzie,[xxvii] argue that these systems need *causal reasoning* abilities: the ability to not only *associate* raw data (as deep learning systems do) but also to infer outcomes from active changes in the environment and to imagine counterfactual scenarios.[xxviii] Whatever the proposed cure one prescribes for machine learning systems, talk of an AI "Winter" or "Summer" refers, by proxy, to how well these problems are perceived to be dealt with.

The unspoken assumption is that these problems can, in fact, be resolved.

Bishop's claim is that they never will be. "No matter how sophisticated the computation is, how fast the CPU is, or how great the storage of the computing machine is, there remains an unbridgeable gap (a "humanity gap") between the engineered problem solving ability of machine and the general problem solving ability of man."[xxix] The reason, he argues, is that computation *alone* can never realize human understanding.

He draws from interdisciplinary arguments to reach this conclusion, the most prominent of which is John Searle's famous "Chinese Room Argument."[xxx] In a nutshell: the mechanistic use of rules to execute a method (i.e., an algorithm) can *never* lead to an understanding of the program's target output. Sure, a machine can translate a language, complete a sentence, or generate new sentences altogether, but all it is doing is executing a method—it is *mindless*, having no idea of what language is, the world that sentences describe, or why a joke in the target language is funny. The machine does nothing *except* execute software— that's it. Modern AI is fundamentally dependent on computational methods operating in exactly this fashion.

Critically, Bishop takes the distinction between *engineering* solutions for automated behavior and *intelligence* via computation seriously: "While causal cognition will undoubtedly be helpful in engineering specific solutions

to particular human specified tasks, lacking human understanding, the dream of creating an [Artificial General Intelligence] remains as far away as ever. Without genuine understanding, the ability to seamlessly transfer relevant knowledge from one domain to another will remain allusive."[xxxi] The idea is that AI systems *will* continue to improve, but they will "remain prey to egregious behavior" while forever "lacking genuine understanding of the bits they so adroitly manipulate."[xxxii] The trajectory of AI, in this view, is fundamentally limited without the possibility of resolution.

Should Strategists Start Worrying?

Strategy takes us to unexpected places, and the philosophy of mind is not the most comfortable landing point for a discipline with much to worry about already. But it might be time to start worrying given the integration of AI with medium- and long-term strategic thought.

The implication of Bishop's argument is that, while AI-enabled systems will see improvements in areas including automated target recognition, human-machine teaming and interaction, and semi- and fully-autonomous tasks, among others, they will always be prone to stupid, potentially catastrophic mistakes—it is just a matter of how likely they are to make them. This implicates strategists who see an urgent need to refine, adopt, and deploy narrow AI-enabled systems, as their real-world deployment will never match the medium- and long-term ambitions humans set for them.

AI systems will never, furthermore, dynamically transfer knowledge from one domain to another, meaning they will remain "narrow." Conceptions of future warfare like the "singularity"[xxxiii] or "hyperwar"[xxxiv] that appear to rely on AI-enabled machines moving with a remarkable speed and seamlessness across domains and between one another is sci-fi now and forever, in this view.

To be sure, Bishop's arguments are by no means a consensus view. He observes that Searle's Chinese Room Argument against the possibility of intelligent computation is one of the most divisive philosophical problems of the twentieth century.[xxxv] Whether it ultimately holds up to scrutiny is not a matter we will resolve here.

Perhaps the ambiguity gives the strategist some comfort, tempted to pin the hopes for AI's strategic advantages less on the intelligence of the technology but the novel engineering workarounds it has afforded—permitted by Bishop's argument. Any potentially "disruptive" technology requires a fortification of individuals', organizations,' and governments' *willingness* to capture the benefits of innovations, as James J. Wirtz argues,[xxxvi] and the engineering aspects of AI may instead be inflated at the expense of its alleged intelligence. This organizational

effort, indeed, appears to be General Milley's aim. It is also the aim of venture capitalists who are practically 'begging' U.S. Secretary of Defense Lloyd Austin to streamline the innovative technology adoption process by the Department of Defense.[xxxvii]

This is an evasion of the problem. Because strategy is "future-learning" and faces an "unavoidability of assumptions"[xxxviii] that never achieve empirical certainty, any strategy involving AI must confront the possibility that biological intelligence is not reproducible via computational means. Otherwise, investments in basic research and adoption of AI may continue to see improvements in engineering but will never escape the problems that plague them today.

Or, maybe, the setbacks that AI has faced over decades really do boil down to the fact that reproducing and inventing intelligence is possible but extraordinarily difficult. Strategists should still not get too comfortable. While I remain agnostic on Bishop's argument, my own work argues that certain aspects of human behavior—but not intelligence wholesale—are unlikely to ever be replicated by machines for separate reasons. On such arguments, no organizational change or investment in basic research will ever yield the technical trajectory for AI that some strategists may desire.

Now is the time to confront the possibility that the crown jewel of the fourth industrial revolution's "commanding heights"[xxxix] is an impossibility. Strategists should have zero illusions about their individual abilities to decisively conclude the debate, as this challenge is premised on philosophical work stretching back centuries, recently instantiated in overlapping fields like the cognitive and neurosciences. Strategists may, nonetheless, be forced to worry about the philosophy of AI eventually, and they would be wise to do so sooner than later.

Conclusion

The unspoken assumption—that biological intelligence can be reproduced via artificial, computational means—directly supports strategic thought incorporating AI today. While some defense analysts[xl] recognize the poor track record in predicting AI's future capabilities, grasping what this technology's ups and downs over the years might mean for strategy formation remains essential. Because Bishop's argument directly implicates the three legs of strategy's triad,[xli] a diverse range of strategists should confront the uncomfortable possibility that what we do can never fully be reproduced by our creations.

References

[i] Francis J. Gavin, "Unspoken Assumptions," Texas National Security Review, 6, no. 2 (Spring 2023): 3-6. DOI: http://dx.doi.org/10.26153/tsw/46147.

[ii] "What is Strategy?", Military Strategy Magazine, MSM Brief, March 2013.

[iii] Tom Ward, "The US Air Force Is Moving Fast on AI–Piloted Fighter Jets," Wired, March 8, 2023, https://www.wired.com/story/us-air-force-skyborg-vista-ai-fighter-jets/.

[iv] Schwab, Klaus. The Fourth Industrial Revolution (New York, NY: Crown Business, 2016).

[v] Baptiste Alloui-Cros, "On Strategy," Baptiste's Substack, July 6, 2023, https://baptisteallouicros.substack.com/p/on-strategy. Alloui-Cros makes a careful distinction here between strategy in practice and the study of strategy, the latter of which he deems "interdisciplinary."

[vi] Daniel C. Dennett, Darwin's Dangerous Idea (New York, NY: Simon & Schuster Paperbacks, 1995), 21.

[vii] Stephen Losey and Colin Demarest, "Air Force Official's Musings on Rogue Drone Targeting Humans Go Viral," Air Force Times, June 2, 2023, https://www.airforcetimes.com/unmanned/uas/2023/06/02/air-force-officials-musings-on-rogue-drone-targeting-humans-go-viral/.

[viii] Joseph Trevithick, "Future of Artificial Intelligence Dominated Air Combat Showcased in New Air Force Video," The Drive, July 5, 2023, https://www.thedrive.com/the-war-zone/future-of-artificial-intelligence-dominated-air-combat-showcased-in-new-air-force-video; on U.S. Air Force planning for autonomous "collaborative combat aircraft," see Stephen Losey, "US Air Force Eyes Fleet of 1,000 Drone Wingmen as Planning Accelerates," Defense News, March 8, 2023, https://www.defensenews.com/air/2023/03/08/us-air-force-eyes-fleet-of-1000-drone-wingmen-as-planning-accelerates/.

[ix] Jim Garamone, "Milley Makes Case for Rules-Based Order, Deterrence in New Era." U.S. Department of Defense, June 30, 2023, https://www.defense.gov/News/News-Stories/Article/Article/3446709/milley-makes-case-for-rules-based-order-deterrence-in-new-era/.

[x] The White House, National Security Strategy (Washington, D.C.: The White House, 2022), 21.

[xi] Ibid., 33.

[xii] Bundeswehr, German Navy Objectives for 2035 and Beyond (Bundeswehr: Rostock, 2023), 1-12.

[xiii] Mike Yeo, "Japan to Replace Attack, Observation Helicopters with Drone Fleet," C4ISRNET, February 9, 2023, https://www.c4isrnet.com/smr/defending-the-pacific/2023/02/09/japan-to-replace-attack-observation-helos-with-drone-fleet/.

[xiv] Huey-Mei Chang and William Hannas, Spotlight on Beijing Institute for General Artificial Intelligence (Washington, D.C.: Center for Security and Emerging Technology, 2023).

[xv] Irene Zhang, "AI Proposals at 'Two Sessions': AGI As 'Two Bombs, One Satellite'?," ChinaTalk, March 8, 2023, https://www.chinatalk.media/p/ai-proposals-at-two-sessions-agi.

[xvi] It may be useful to keep in mind that the same reasoning applies to biological organisms like humans who are also assumed to have developmental potentials. Individuals routinely assume humans will develop linguistic, visual, auditory, musical, and social capabilities of very specific kinds—but the only reason why fields like neuroscience and cognitive science exist is that such capabilities are realized to be poorly understood once our unspoken assumptions are brought to light.

[xvii] Ares Simone Monzio Compagnoni, "Will Artificial General Intelligence Change the Nature of War?," Military Strategy Magazine, Volume 8, no. 4 (Spring 2023): 32-37. https://www.militarystrategymagazine.com/article/will-artificial-general-intelligence-change-the-nature-of-war/.

[xviii] J. Mark Bishop, "Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It," Frontiers in Psychology, 11, no. 1 (2021): 1-18.

[xix] Will Douglas Heaven, "Geoffrey Hinton Tells Us Why He's Now Scared of the Tech He Helped Build," MIT Technology Review, May 2, 2023, https://www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/.

[xx] Bishop, "Artificial Intelligence," 2-4.

[xxi] For a somewhat technical but general overview of artificial neural networks, see, Janik Tinz, "Understand the Fundamentals of an Artificial Neural Network," Towards AI, February 4, 2023, https://towardsai.net/p/l/understand-the-fundamentals-of-an-artificial-neural-network#:~:text=Artificial%20Neural%20Networks%20in%20general&text=A%20neural%20network%20has%20one,a%20Feed%20Forward%20Neural%20Network.

[xxii] Human design and architectural choices matter immensely for any AI system to function successfully.

[xxiii] Stephen Wolfram, What Is ChatGPT Doing and Why Does It Work? (New York, NY: Kiligry, 2023).

[xxiv] James Vincent, "OpenAI's State-of-the-Art Machine Vision AI is Fooled by Handwritten Notes," The Verge, March 8, 2021, https://www.theverge.com/2021/3/8/22319173/openai-machine-vision-adversarial-typographic-attacka-clip-multimodal-neuron.

[xxv] Richard Waters, "Man Beats Machine at Go in Human Victory Over AI," Financial Times, February 17, 2023, https://www.ft.com/content/175e5314-a7f7-4741-a786-273219f433a1; see also, Tony T. Wang, et. al., "Adversarial Policies Beat Superhuman Go AIs," ArXiv (July 13, 2023)" 1-87. DOI: https://doi.org/10.48550/arXiv.2211.00241.

[xxvi] Paul Scharre, Four Battlegrounds: Power in the Age of Artificial Intelligence (New York, NY: W.W. Norton & Company, 2023), 231.

[xxvii] Judea Pearl and Dana Mackenzie, The Book of Why: The New Science of Cause and Effect (New York, NY: Basic books, 2018).

[xxviii] See also, Bishop, "Artificial Intelligence," 10.

[xxix] Ibid., 17.

[xxx] John R. Searle, "Minds, Brains, and Programs," Behavioral and Brain Sciences, 3, no. 3 (1980): 417-424. DOI: https://doi.org/10.1017/S0140525X00005756.

[xxxi] Bishop, "Artificial Intelligence," 17.

[xxxii] Ibid.

[xxxiii] Elsa B. Kania, "Battlefield Singularity," Center for a New American Security, November 28, 2017, https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power.

[xxxiv] John R. Allen and Amir Husain, "On Hyperwar," USNI Proceedings, 143, no. 7 (2017): https://www.usni.org/magazines/proceedings/2017/july/hyperwar.

[xxxv] Bishop, "Artificial Intelligence," 11.

[xxxvi] James J. Wirtz, "A Strategist's Guide to Disruptive Innovation," Military Strategy Magazine, 8, no. 4, (Spring 2023): 4-9. https://www.militarystrategymagazine.com/article/a-strategists-guide-to-disruptive-innovation/.

[xxxvii] Sydney J. Freedberg, Jr., "Venture Capitalists, Tech Firms Beg Defense Secretary to Speed Up Innovation," Breaking Defense, June 26, 2023, https://breakingdefense.com/2023/06/venture-capitalists-tech-firms-beg-defense-secretary-to-speed-up-innovation/.

[xxxviii] "What is Strategy?", Military Strategy Magazine, MSM Brief, March 2013.

[xxxix] Rush Doshi, The Long Game: China's Grand Strategy to Displace American Order (New York, NY: Oxford University Press, 2021), 5.

[xl] Scharre, Four Battlegrounds, 284-285.

[xli] "What is Strategy?", Military Strategy Magazine, MSM Brief, March 2013.