CSET CENTER for SECURITY and EMERGING TECHNOLOGY

AUGUST 2020

The AI Triad and What It Means for National Security Strategy

AUTHOR
Ben Buchanan

CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

Established in January 2019, the Center for Security and Emerging Technology (CSET) at Georgetown's Walsh School of Foreign Service is a research organization focused on studying the security impacts of emerging technologies, supporting academic work in security and technology studies, and delivering nonpartisan analysis to the policy community. CSET aims to prepare a generation of policymakers, analysts, and diplomats to address the challenges and opportunities of emerging technologies. During its first two years, CSET will focus on the effects of progress in artificial intelligence and advanced computing.

CSET.GEORGETOWN.EDU | CSET@GEORGETOWN.EDU

AUGUST 2020

The AI Triad and What It Means for National Security Strategy

AUTHOR
Ben Buchanan

ACKNOWLEDGMENTS

The author would like to thank John Bansemer, Beba Cibralic, Tarun Chhabra, Teddy Collins, Andrew Imbrie, Igor Mikolic-Torreira, Michael Sulmeyer, Danielle Tarraf, Lynne Weil, and Alexandra Vreeman for their comments on an earlier draft of this paper.

PRINT AND ELECTRONIC DISTRIBUTION RIGHTS

Cover photo: paul/AdobeStock

Contents

Executive Summary

A single sentence can summarize the complexities of modern artificial intelligence: Machine learning systems use computing power to execute algorithms that learn from data. Everything national security policymakers truly need to know about a technology that seems simultaneously trendy, powerful, and mysterious is captured in those 13 words. They specify a paradigm for modern AI—machine learning—in which machines draw their own insights from data, unlike the human-driven expert systems of the past.

The same sentence also introduces the AI triad of algorithms, data, and computing power. Each element is vital to the power of machine learning systems, though their relative priority changes based on technological developments. Algorithms govern how machine learning systems process information and make decisions. Three main classes of algorithms are common today: supervised learning, which draws insights from structured data sets; unsupervised learning, which excels at finding structure or clusters in unorganized data sets; and reinforcement learning, which builds up a machine learning system's capability through trial and error. Usually, these algorithms run on neural networks, a type of computer program architecture. Neural networks provide enormous flexibility and power, but come with their own tradeoffs—chiefly, the lack of transparency behind their reasoning.

Data is often, though not always, how machine learning systems learn about the world. If a fact is not in the data provided to the machine, the system will never learn it, especially in the case of supervised learning. Acquiring larger and more representative datasets can further empower

machine learning systems. Without such datasets, bias can creep into systems, causing them to perform poorly in hard-to-detect ways.

Often overlooked, computing power is increasingly essential as algorithms grow more complex and datasets larger. The past eight years have seen a revolution in the amount of computing power applied to cutting-edge machine learning, expanding by a factor of several hundred thousand. Computing power increasingly affects the performance of machine learning systems and presents a significant cost during system development.

Each part of the triad offers its own policy levers. Algorithmic progress depends on a nation acquiring and developing talented machine learning researchers. Larger and better datasets require tricky policy choices involving bias, privacy, and cybersecurity. Computing power can provide a point of leverage for export controls in foreign policy, as well as a bottleneck for AI research at home. In order to judiciously wield the levers available in AI policy, policymakers must first understand the technology itself and how it will reshape national security. The concept of the AI triad is one framework for doing so.

Introduction

A single sentence can summarize the complexities of modern artificial intelligence: Machine learning systems use computing power to execute algorithms that learn from data. Everything policymakers need to know about a technology that seems simultaneously trendy, powerful, and mysterious is captured in those 13 words.

AI matters. At home, it is already fundamental to everyday life, voiced by Alexa and Siri and tucked inside smartwatches and phones. In science, it contributes to major breakthroughs, from diagnosing disease to aiding drug discovery to modeling the climate. In business, it shapes the economic competitiveness of nations and alters how trillions of dollars pulse through global markets. In national security, it bolsters logistics and intelligence analysis and—with visions of lethal autonomous weapons, drone warfare, and self-guided cyberattacks—seems poised to do much more.

The breadth of the technology is why the single-sentence articulation of AI is so important, and why the concepts alluded to within it matter so much. If policymakers do not understand AI, they will be a passive audience to technological pioneers charging onward, slow to recognize what AI can do for the issues they care about. Maybe worse, policymakers who do not understand AI will fail to recognize what the technology *cannot* yet do, despite its hype. They will ignore AI's current structural flaws, such as a lack of transparency and the potential for insidious bias—challenges that must be mitigated with both technical and policy solutions.

The concise definition of AI first specifies a paradigm for modern AI: machine learning. Machine learning stands in direct contrast to the previous era's dominant paradigm, expert systems, which focused on formally

encoding human knowledge in a way a computer could process. For example, IBM's computer program DeepBlue drew heavily on inputs provided in advance by chess grandmasters to beat world chess champion Garry Kasparov in 1997. In machine learning, AI developers reject direct instructions in favor of a system that can learn on its own. This paper focuses on the dominant paradigm of machine learning known as deep learning, explained in more detail in the next section.

Three components make deep learning happen: data, algorithms, and computing power. Together, I call these components the AI triad. Computer scientists have long used this tripartite division to study machine learning, and I argue that it offers a framework for understanding the power of machine learning and what it means for policy.

The elements of the AI triad work in combination to achieve stunning results. For example, OpenAI, a leading AI research lab, trained a text generation system, known as GPT-3, to write whole paragraphs in response to a given prompt and to perform other linguistic tasks. The engineers assembled almost a trillion words, which they filtered down to around 540GB of human writing. They then devised a clever algorithm with around 175 billion learned parameters that could predict which word would come next in a sentence based on patterns in the collected data; in essence, the algorithm learned to imitate the writing it saw. The engineers set GPT-3 loose on high-performance computers for days, performing quadrillions of calculations as it refined its own capacity for mimicking human language.

GPT-3's combination of data, algorithms, and computing power produced a breakthrough. The system wrote some text that 88 percent of readers thought was convincingly human. Perhaps most impressive was GPT-3's ability to mimic its prompt, from continuing a news report to writing the next stanza of a poem in the style of a particular poet.[1]

GPT-3 is one in a long line of machine learning breakthroughs. The drumbeat of advancement and machine learning's relevance to national security show no signs of diminishing. Policymakers need a deeper understanding of machine learning's three components and why they matter so much.

1 The AI Triad

ALGORITHMS

An algorithm is a series of instructions for processing information. Schoolchildren, for example, are taught the algorithm PEMDAS for the order of operations in solving math problems: parentheses, exponents, multiplication/division, addition/subtraction. When given information—such as the problem 7+5(1+3)—the algorithm tells them explicitly how to process it: first add one and three in the parentheses, then take the resulting four and multiply it by the five next to it, and finally add the resulting 20 to the seven.

Another algorithm, one that works from left to right and ignores order of operations, processes the information differently—adding seven to five, multiplying by one, and then adding three. It yields a different (and incorrect) answer.

In the same way, computer programmers issue direct instructions—algorithms—to their systems on how to process information. These algorithms often contain conditional logic—if this, then that—specifying that a program should do one thing with one set of information but a different thing with a different set of information. For decades, engineers built AI programs with a similar kind of design. As mentioned, DeepBlue's defeat of chess champion Garry Kasparov was enabled by a detailed series of direct instructions culled from grandmasters that guided that program's play in the tournament match.

Machine learning is different. Machine learning algorithms do not apply explicit strategies or directions provided by humans. Instead, these algorithms derive their own insights from datasets, and use these insights as a basis for operation. To do this, machine learning systems often deploy

deep learning or neural networks, the technical details of which are not important for this paper.[2] Using these networks, a machine learning algorithm might discover the PEMDAS system from studying reams of solved equations and working backward to find the rules.[3]

Machine learning can do more than just reverse-engineer rules. It can solve problems much less structured and well-defined. Consider an analogy of teaching a robot to play with Legos. In one approach to training the robot, a human gives it the exact Legos required to build a certain structure and step-by-step assembly instructions. A robot with enough physical dexterity (no small feat, but irrelevant in this context) will be able to assemble the pieces; so, too, can traditional software programs execute certain instructions.

In this approach, the human is the architect and the robot the builder. Asked to build something new with limited pieces and no instructions, the robot will likely underwhelm, just as many traditional kinds of AI failed when asked to perform complex tasks or adapt to unforeseen situations. The robot has been programmed to follow human-given guidance and cannot succeed without it.

In a different approach to Lego construction, the robot is both the architect and the builder. The human, meanwhile, offers only evaluation and feedback, providing some examples of previous successful and failed structures but very little current direction. Through trial and error, the robot will attempt various approaches, getting feedback from the human each time and learning iteratively which sorts of structures earn praise. The robot will eventually learn to build new structures without instructions or preset designs. This second approach to developing skill with Legos will likely take longer than one predicated on following directions, but will also yield a robot far more capable and versatile—and Lego creations that are much more creative.

However, this creativity comes at a price. In the first approach, the robot will act in a way entirely comprehensible to humans, executing human commands. In the second approach, the robot gains the freedom to design as well as to build, but cannot explain why it makes the design choices it does. The robot might consistently make beautiful Lego creations, but will not have the capacity to explain its own reasoning in a step-by-step way. This inability to explain is one of the most challenging aspects of current machine learning systems. While perhaps inconsequential for Lego structures, lack of explainability raises legal and ethical concerns in other areas.

This example gives some intuition for how machine learning algorithms gain power from data and feedback instead of through explicit commands, as well as some of the limitations of that approach. Looking at the broad classes of machine

learning algorithms can enable us to go deeper still. These three approaches are known as supervised learning, unsupervised learning, and reinforcement learning; neural networks are often used to implement all three types of algorithms.

A supervised learning algorithm derives patterns from well-organized data usually provided by humans. Once developed, the system can then deploy these pattern-recognition capabilities in new situations. For example, in a commercial application, machine learning engineers might give a pattern-recognition algorithm data on thousands of car sales in the United States, including the make, model, year, and condition of each car sold, plus information on additional features. This data serves as the foundation for building pattern-recognition capabilities.

In a supervised learning system, the algorithm is also given the completed sale price for all of these sales. This price reflects how much humans value cars with different features. The algorithm then determines what it thinks the relationship is between various features of the car and the amount the customer ultimately paid. In doing so, it derives insights both large and small, such as that newer cars with fewer miles are likely to sell for more, and that extras like sunroofs increase a car's value; this is known as "training" the system.

With these insights assembled, the supervised learning algorithm is ready to attempt a task called "inference." In this case, inference involves making a prediction about the future sale price of a car still on the market, given the information known about the car. The system is likely to be very good at inference if well trained; if it had poor training data or was poorly calibrated, it is likely to fail. In general, well-trained supervised learning algorithms prove adept at forecasting a wide variety of outcomes, from spam filters to housing markets, and even including predictions like the resale value of fine wine.[4]

What could supervised learning do for national security? Perhaps the same kinds of algorithms that can predict car sales or identify lung cancers in CT scans can predict terrorist attacks or help sort through large numbers of satellite intelligence photos. The United States military made a substantial investment in this area in 2017 when it announced the creation of the Algorithmic Warfare Cross-Functional Team (better known as Project Maven). One of the team's first tasks was to apply supervised learning to photos and videos collected by U.S. drones around the world; another task was using supervised learning to better predict equipment failures in special operations helicopters. Then-Deputy Secretary of Defense Robert Work made clear that he saw these efforts as the first of many military forays into machine learning, writing that the Defense Department "must integrate artificial intelligence and machine learning more effectively across operations to maintain

advantages over increasingly capable adversaries and competitors."[5]

For all their power, supervised learning algorithms still have limits. They depend on having the "right answer" for the data provided to the algorithm, such as prices for past car sales, examples of human-analyzed intelligence photos, and helicopter maintenance records. Without these right answers from which to learn, supervised learning systems cannot derive the patterns needed to make predictions for new data. If absent from the training data, a supervised learning system will never know it.

This is where unsupervised learning algorithms come in. Unsupervised learning thrives when there isn't a neat, well-organized set of data with the right answer provided for each data point. These algorithms can help disentangle complex webs of data and provide some structure.

For example, one of the most common tasks for advertisers and politicians is to know their market. While they can gather data, their customers or constituents represent a vast group, rife with patterns but also contradictions and complexity. Making sense of the market might mean segmenting or clustering this large group into a series of more meaningful smaller groups. With these smaller groups identified, advertisers and politicians can target and tailor their messages more effectively.

Some ways of clustering this data are obvious, such as filtering based on age, gender, or ethnicity. Other clusters are less apparent, but unsupervised learning can help identify them. Much of the online advertising space is based on more nuanced algorithmic clustering than age, gender, or ethnicity, or even a combination of all three. Rather, advertisers seek information including online history, purchases, and expressed interests to feed into unsupervised learning algorithms. The systems produce smaller and more accurate clusters, often grouping users based on combinations of interests and personal characteristics. Advertisers can then tailor their marketing to the audiences most receptive to it.

What works for advertisers could also work for propagandists. The business of generating and distributing disinformation is, unfortunately, an intrinsic part of modern geopolitics, and AI prompts significant concerns in this regard. The 2016 Russian election interference campaign demonstrated what happens when hackers and propagandists work together. In some respects, clustering algorithms shaped the terrain upon which their information operations unfolded. The Facebook ad-targeting algorithm, used by the Russians to deploy the ads they purchased, depends in part on clustering users, as do the algorithms helping posts go viral.[6] It remains to be seen how unsupervised learning will enable the targeting of future propaganda efforts.

Neither supervised nor unsupervised learning excels at the long-term strategic assessment and planning integral to national security. A third type of algorithm, reinforcement learning, can help. These algorithms are especially powerful in areas with

a clearly defined environment, such as in board or video games. Through trial and error, they make decisions and receive feedback from the environment. A reinforcement learning algorithm, for example, may be awarded points for finding moves leading to victory in a game, but docked points for moves resulting in defeat. As they seek to maximize rewards, the algorithms improve over time at navigating the environment and performing tasks, sometimes even surpassing human capacity.

DeepMind, a Google-owned leading research lab, deployed reinforcement learning to great success in its program AlphaZero, which could beat all humans and computers at the board games Go, chess, and shogi.[7] Its success at Go, an ancient board game originating in Asia, is notable because of the game's complexity: there are more possible positions on the Go board than atoms in the universe. In fact, there are more possible positions on the Go board than total atoms if every atom in the universe contained a universe of atoms within it.[8] With this many possibilities, calculating a path to victory is impossible, even for a computer; players must use intuition to win. The adaptive nature of reinforcement learning has proven better at this kind of strategic intuition on the Go board than supervised or unsupervised learning—or, for that matter, human intelligence.

Reinforcement learning algorithms also have applications in national security. After creating AlphaZero, DeepMind used reinforcement learning as a key part of a program called AlphaStar, attaining grandmaster status at the strategy video game Starcraft II. The algorithm's success was remarkable as Starcraft requires more decisions than Go, unfolds in real time rather than in discrete turns, and offers only imperfect information—players do not see all of their opponents' moves, and opponents can actively deceive one another. For these reasons, Starcraft is a much closer proxy for military strategy than Go. AlphaStar's success signals the increasing strategic relevance of reinforcement learning.[9]

Another notable area of reinforcement learning success is robotics: the algorithm makes a decision, the robot carries it out, and the robot's sensors detect how the environment responds, and whether that response was good or bad. For this reason, reinforcement learning appears in some self-driving car technology. It could also power autonomous military vehicles or aircraft capable of swarming targets at high speed or using complex tactics in real time. In battlefield environments with degraded or non-existent command and control capabilities, reinforcement learning algorithms capable of making decisions on their own could be essential.

DATA

"Data is the new oil"—or so we're told. The phrase has become a cliché, mentioned everywhere from corporate marketing to presidential debates. It is easy to

see why observers make this comparison. As the previous overview of algorithms shows, data—information about the area of focus of the machine learning system, such as car sales or drone photos—is crucial, especially for supervised learning. Without it, there would be no patterns to recognize, and many algorithms would be much less effective, if they worked at all. One foundational study in 2001 suggested that the amount of training data available mattered more than the algorithm used.[10]

The amount of training data has a large role in determining machine learning system effectiveness. For example, in the case of car sales, if the data on future car sale prices given to the supervised learning algorithm was minimal, it might be dominated by outliers; a car in the dataset might have sold for a lower price because it was a transaction between friends or because the car had a flaw obvious to the buyer but not captured in the data. These outliers average out in larger datasets and diminish in relevance.

All else being equal, larger and more representative datasets better capture the real world. A small dataset of car sales might not include any sales from niche but important car manufacturers like Ferrari, since many more Toyotas and Chevrolets are sold. In this case, when the machine learning system predicts the sale price of a luxury car, it will more likely fail given the lack of relevant training data. Thus, in machine learning as in other kinds of statistics, a fuller and broader sample size is usually better.

Yet data collection can itself present a challenge. For this reason, among others, companies with direct access to large amounts of consumer data, such as Facebook, Google, and Amazon, are market leaders. Once gathered, data must be organized, stored, and made accessible, all of which are technically and organizationally challenging. Legal and regulatory hurdles, especially around privacy, also constrain what organizations can do throughout the process of assembling a large dataset.

Crucially, not just the amount of data matters, but also the salience of that data to the problem at hand. An infinite amount of data on bike sales, for example, is unlikely to provide much value in deriving patterns about car prices. Those designing and building machine learning systems need a well-defined problem and data relevant to that problem. This demand for specific data makes it is difficult to judge the value of data in the abstract, especially across sectors and countries. Chinese companies, for example, might have granular insights on app-based food deliveries in their country, but that data is unlikely to improve Chinese military competitiveness.

One final component of data merits discussion, especially for supervised learning systems: the accuracy of the right answer in the training data. It is common for

training data to exhibit some kind of bias. When this occurs, the machine learning system can absorb this bias in the same way it learns other patterns in the training data. For example, a planned Amazon system for scanning resumes was scrapped for discriminating against women, likely because the training data given to the system—past hiring decisions—exhibited an anti-woman bias.[11]

Biased data could amplify the aforementioned problem of explainability: when machine learning systems do inherit biases from their training data, they can make biased decisions without explaining why. A machine learning system can thus take the biased data and present it as impartial. One researcher and entrepreneur called machine learning "money laundering for bias," as flawed algorithmic outputs tend to be viewed as fair and objective.[12] This problem can cause machine learning systems to fail in particularly insidious, if unintentional, ways. For example, if a machine learning system designed to help spot terrorists developed a bias against a certain ethnicity, it might consistently recommend more scrutiny for innocent people of that ethnicity while ignoring suspects of other ethnicities—all without ever telling its human operators that ethnicity was a key factor in its recommendations.

COMPUTING POWER

In many parts of the modern world, computers are commodities. Whereas desktop and laptop computer manufacturers once competed on the speed and power of their processors and graphics cards, most modern computer users—save perhaps for cryptocurrency miners and hardcore video gamers—have no idea what kind of computer chips they use every day.

In the context of AI, however, to ignore computing power (or "compute" to the machine learning community) is to make an enormous mistake. Largely overlooked by news media and in other popular narratives, compute has underpinned a great deal of modern AI progress. Rich Sutton, widely considered one of the founders of modern AI, called the centrality of compute "the bitter lesson" of machine learning and one that researchers have been slow to learn.[13] He contended that the preeminence of compute is uncomfortable because it reduces the role of humans in building AI. Sutton argues that a great deal of AI progress has been enabled not by making systems more similar to humans, or by imparting more human knowledge to computers, but rather by giving machine learning systems greater processing power to learn on their own. In this view, the architecture of an algorithm and the data of human knowledge are simply less significant than the computer hardware enabling machine learning. If Sutton is right, then compute may well be the most important part of the triad.

Indeed, ample evidence shows that compute correlates strongly with AI ad-

vancements. OpenAI studied how compute drove AI progress from 2012 to 2018. What they found was remarkable: during that period, which was filled with tremendous AI achievements, the amount of compute applied to the training of top AI projects increased by a factor of 300,000.[14] To put that into context, if a cell phone battery lasted one day in 2012 and increased at the same rate, in 2018 that battery would last more than 800 years.[15]

This increasingly potent computer power leads to breakthroughs otherwise inaccessible. For example, XD Huang, a leading Microsoft AI researcher, believes the transition to graphics processing units to better execute machine learning calculations was "the real weapon" that enabled a great deal of Microsoft's advances. Some projects, he said, would have taken as much as five more years to complete without the increased compute.[16]

In other cases, the amount of compute given to a machine learning system helps shape the power of that system. For example, OpenAI developed four versions of GPT-2, the precursor to the GPT-3 system mentioned in the introduction, which takes a prompt from a user and generates text in response. These four versions differed in their number of parameters, with more parameters requiring more compute to train; the training data and the overall algorithmic design remained the same for each. The difference in parameters significantly changed performance. The biggest system emerged so powerful that OpenAI delayed its public release for months because of potential national security concerns, such as automated propaganda operations. In OpenAI's view, the difference in number of parameters meant the difference between a fun, impressive, and harmless system and one so powerful that it had to be kept locked away until its dangers could be studied—a controversial and unusual decision in the machine learning research community, which has a strong preference for openness.[17]

Three factors have driven this tremendous increase in compute. First, there is the continued drumbeat of Moore's Law (famously, Intel CEO and cofounder Gordon Moore suggested that computing power would double every 24 months as a result of improved processor engineering).[18] But even Moore's Law would predict a much smaller increase in compute applied to AI systems than what has occurred in recent years.

The second factor is the increased application of parallelized computing in machine learning chips. Parallelization enables many computer chips to train a machine learning system at the exact same time. Like an orchestra playing a symphony, the tasks involved in training are divided into many parts and managed all at once. While the idea of parallelization has been around for years, modern systems take it

to an extreme degree, with hundreds of processors working simultaneously.

The third factor is the increased efficiency of machine learning computer chips. Running machine learning algorithms is different from running an Excel spreadsheet or a web browser; as discussed, the former uses neural networks to learn from data while the latter execute direct human instructions. As a result, the kinds of optimizations present within typical computer chips and operating systems do not yield the same gains in efficiency in machine learning calculations. Yet specialized chips can be built and tailored to run machine learning algorithms much more efficiently. From 2012 to today, several paradigm shifts in machine learning compute have occurred, transitioning the industry from regular compute processors to graphics processing units to dedicated chips built for efficiency in machine learning.

None of these three factors comes cheap. Even though Moore's Law seems poised to continue for a few more years, the production of new chip factories increases in cost and complexity as semiconductor engineering problems get harder. The growing parallelization of machines is a boon, but purchasing more machines adds expense. The increased efficiency from custom-built chips specialized for machine learning has enabled significant advances, but requires large investments to design and build new hardware. As attaining compute continues to grow more expensive and complex, it increasingly becomes a bottleneck for machine learning researchers and relevant for national security policymakers.

2 What the AI Triad Means for Policymakers

The AI triad is useful for demystifying and understanding modern AI, especially given the rapid pace of progress and complexity of cutting-edge systems. It can help categorize advances in machine learning, differentiating algorithmic genius from computationally intense success.

Perhaps even more useful are the ways in which the AI triad can frame and inform decisions in national security policy. Beyond just making sense of what is happening, it can help in developing intuitions about what to do about it. Each part of the triad lends itself to distinct policy levers, challenges, and opportunities. The comparative policy importance of each part has implications for the national security strategy; when one part of the triad is of a higher priority than others, different policy prescriptions follow.

In a world in which algorithms reign, the research talent and resources to develop those algorithms become preeminent. Current supply of this talent cannot meet global demand. As a result, policymakers at the national level must find ways to attract foreign talent to their country, to retain the talent that does come, and to develop new talent.[19] The resulting policy levers are things like visa controls, industrial strategies, worker retraining and certification frameworks for AI skills, and educational investments to meet AI faculty and teacher shortages. Given the centrality of AI talent for algorithmic advances, these routine government functions can take on significant national security and economic implications. Though seemingly mundane, this ground is the terrain on which geopolitical competition in the age of AI is first fought.

Yet if data is of highest priority, different policy levers emerge. Under a machine learning paradigm driven by human-curated data, the prospect of biased systems resulting from biased datasets substantially grows. Tracking and measuring the risk of bias thus becomes more important with data at the center of AI; subject-matter experts should be consulted to understand potential sources of unintentional bias. Some concrete ideas, such as a system akin to nutritional labeling for machine learning, can also help provide clarity about the underlying data on which algorithms were trained.[20] Even if algorithms remain opaque and unable to explain their decisions, transparency about training data used and information within that data set could increase confidence in the systems.

Privacy issues rise in importance the more data matters for AI. Insofar as tension exists between the privacy rights of users and the value of their data in training machine learning systems, governments must manage the balance. They will have to craft privacy laws and regulations that protect the civil liberties and rights of individuals without unduly constraining the innovation that using their data for training might enable. It is not a zero-sum equation, as additional technical research into privacy-preserving machine learning systems can help algorithms learn from data without revealing information about individuals. While promising, these kinds of algorithms comprise a comparatively small fraction of current machine learning research and merit additional government funding. Governments could also require that high-consequence algorithms—such as those relating to parole decisions, credit risk, and healthcare—undergo thorough vetted by technically informed regulators before they are deployed.

Other policy questions will emerge if data is central to machine learning progress. The increased importance of data could prompt acquisition and storage of ever-larger data sets, creating second-order considerations of cybersecurity and data breach liability policies, as this data must be protected. The role of government as a data collector and provider emerges most fully in this data-centric world, too. For example, how should the government assemble datasets to solve its problems and what government procedures need to be changed to collect and organize this data? More generally, which of the government's vast stores of data should be made available, how, and to whom? All of these questions will need careful policy-making to address.

The final scenario is that compute is of highest priority. If so, it is vital to manage the flow of powerful computer chips optimized for machine learning calculations. Export controls thus emerge as significant policy levers, especially for the United States and its allies, who currently enjoy an edge in advanced computer chip manufacturing. China depends on access to Western companies for advanced

photolithography and other semiconductor manufacturing equipment.[21] For China, a strategy for further developing its domestic computer chip industry becomes essential for preserving economic and national security flexibility in the age of AI; for this reason, among others, Beijing has aggressively sought alternatives to Western chips.[22] Export control effectiveness depends on the technology denied to adversary nations surpassing that which those nations can produce or obtain. While open for now, the window for using such controls against China is likely beginning to close.

More generally, the cost of compute looms large. If compute becomes too expensive for academic researchers to employ, then research will shift to the private sector, with potential negative effects on long-term innovation. Government could play a role in making compute accessible to academic researchers so they can continue to train new experts and contribute to AI progress.[23]

So which part of the triad should policymakers prioritize? It depends a great deal on what happens behind closed doors in research labs. On balance, though, data seems somewhat overvalued and overhyped in the modern era, especially with the emergence of specific technological innovations, such as generating representative data from artificial sources or developing algorithms that do not rely on human-curated training data. While privacy and data aggregation concerns are real, such concerns are likely independent of truly cutting-edge machine learning research. Whereas a decade ago, data seemed central—Google's chief scientist Peter Norvig famously said, "We don't have better algorithms than anyone else; we just have more data"—it has somewhat diminished in comparative importance, as the power of algorithms and compute has become more apparent.[24]

Algorithms seem more fairly assessed, if only in theory. Policymakers increasingly recognize the importance of innovations in this area, but in the United States, attracting the talent necessary to develop algorithms hasn't become enough of a national priority. In contrast, allies such as the United Kingdom, Canada, and France attempt to nurture their domestic AI industries and attract new researchers from overseas.[25] China, too, has aggressively developed its Thousand Talents program to recruit top AI minds and researchers in other fields.[26] The United States, which educates a great deal of the world's AI talent in its universities, could do much more to build on this home-field advantage before it slips away.

While data appears overhyped and algorithms get lip service but no major policy action in the United States, compute seems undervalued and underhyped nearly everywhere. Computational advances are hard to explain and harder still to visualize, perhaps explaining the oversight. That said, as OpenAI's research shows, the exponential growth in compute applied to machine learning systems in the last few years has driven a tremendous amount of the observed progress. Sutton's larger

observation about this pattern is striking as well.[27] The boom of Silicon Valley start-ups working on advanced computing for AI suggests that even more progress in this area is to come, with potentially significant effects for the future of machine learning, national security systems that rely on it, and the choices available to policymakers.

Conclusion

In practice, determining the most important component of the AI triad is an academic question rather than a policy one. Nations will compete in all three areas, though the relative priority will shift as different parts of the triad advance at different rates. Policymakers thus must devise a cross-cutting AI strategy that addresses data, algorithms, and compute, while simultaneously assessing which part of the triad—and thus which policy levers—are most significant.

Policymakers will have to make this judgment in a forward-looking way, cognizant that their choices carry both short- and long-term consequences. For example, a decision to place export controls on computer chips might provide a benefit for a number of years, but risks China developing its own chip manufacturing industry uninhibited by Western competition—probably a net negative in the long run for U.S. policymakers. Conversely, a decision to try to attract and develop algorithmic talent might take a great deal of effort right away and not pay dividends for more than a decade.

Sorting through these various ramifications will not be easy. It will depend on marrying geopolitical imperatives with the present and the future of algorithms, data, and compute. As with so much else, making good AI policy starts by demystifying the underlying technology.

Endnotes

1. Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al., "Language Models are Few-Shot Learners," *arXiv preprint arXiv:2005.14165* (2020), https://arxiv.org/abs/2005.14165.

2. For more on neural networks, see Ben Buchanan and Taylor Miller, *Machine Learning for Policymakers: What It Is and Why It Matters* (Belfer Center for Science and International Affairs, 2017), https://www.belfercenter.org/sites/default/files/files/publication/MachineLearningforPolicymakers.pdf.

3. Silviu-Marian Udrescu and Max Tegmark, "AI Feynman: A Physics-Inspired Method for Symbolic Regression," *arXiv [physics.comp-Ph]* (May 27, 2019), http://arxiv.org/abs/1905.11481.

4. Michelle Yeo, Tristan Fletcher, and John Shawe-Taylor, "Machine Learning in Fine Wine Price Prediction," Journal of Wine Economics 10, no. 2 (2015): 151–72.

5. Cheryl Pellerin, "Project Maven to Deploy Computer Algorithms to War Zone by Year's End," U.S. Department of Defense, July 21, 2017, https://www.defense.gov/Explore/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/.

6. For the definitive discussion of the Russian operation, see Robert S. Mueller III, "Report On The Investigation Into Russian Interference In The 2016 Presidential Election" (Department of Justice, March 2019), https://www.justice.gov/storage/report.pdf. For more on Facebook's algorithms generally, see Jeff Horwitz and Deepa Seetharaman, "Facebook Knows It Encourages Division. Top Executives Nixed Solutions," *Wall Street Journal*, May 26, 2020, https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499.

7. David Silver et al., "A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go through Self-Play," *Science* 362, no. 6419 (December 7, 2018): 1140–44.

8. John Tromp and Gunnar Farnebäck, "Combinatorics of Go," in *Computers and Games* (Springer Berlin Heidelberg, 2007), 84–99.

9. Oriol Vinyals et al., "Grandmaster Level in StarCraft II Using Multi-Agent Reinforcement Learning," *Nature 575*, no. 7782 (November 2019): 350–54.

10. Michele Banko and Eric Brill, *Scaling to Very Very Large Corpora for Natural Language Disambiguation, Proceedings of the 39th Annual Meeting on Association for Computational Linguistics* (Association for Computational Linguistics, 2001).

11. Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women," *Reuters*, October 10, 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

12. Maciej Cegłowski, "The Moral Economy of Tech" (Society for the Advancement of Socio-Economics, June 26, 2016), http://idlewords.com/talks/sase_panel.htm.

13. Rich Sutton, "The Bitter Lesson," Incomplete Ideas, March 13, 2019, http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

14. The amount spent on compute also increased during this time, raising the question of whether, due to the importance and price of cutting-edge compute, future machine learning advances are likely to occur only in well-funded labs. One academic researcher wrote, "Access to compute is abysmal even at good places." Eric Jonas, Twitter, May 27, 2020, https://twitter.com/stochastician/status/1265711669050839041.

15. Dario Amodei and Danny Hernandez, "AI and Compute," May 16, 2018, https://openai.com/blog/ai-and-compute/.

16. Cade Metz, "How AI Is Shaking Up the Chip Market," *Wired*, October 28, 2016, https://www.wired.com/2016/10/ai-changing-market-computer-chips/. For more on different kinds of AI chips, see Saif M. Khan and Alexander Mann, "AI Chips: What They Are and Why They Matter," Georgetown University Center for Security and Emerging Technology, April 2020, https://cset.georgetown.edu/wp-content/uploads/AI-Chips%E2%80%94What-They-Are-and-Why-They-Matter.pdf.

17. Irene Solaiman et al., "Release Strategies and the Social Impacts of Language Models," *arXiv [cs.CL]* (August 24, 2019), http://arxiv.org/abs/1908.09203.

18. R. R. Schaller, "Moore's Law: Past, Present and Future," *IEEE Spectrum* 34, no. 6 (June 1997): 52–59.

19. Remco Zwetsloot et al., "Keeping Top AI Talent in the United States," Georgetown University Center for Security and Emerging Technology, December 2019, https://cset.georgetown.edu/wp-content/uploads/Keeping-Top-AI-Talent-in-the-United-States.pdf; Remco Zwetsloot, "China's Approach to Tech Talent Competition: Policies, Results, and the Developing Global Response," Georgetown University Center for Security and Emerging Technology, April 2020, https://cset.georgetown.edu/wp-content/uploads/Zwetsloot%E2%80%94Chinas-Approach-to-Tech-Talent.pdf.

20. Margaret Mitchell et al., "Model Cards for Model Reporting," *arXiv [cs.LG]* (October 5, 2018), http://arxiv.org/abs/1810.03993.

21. Saif M. Khan, "Maintaining the AI Chip Competitive Advantage of the United States and Its Allies," Georgetown University Center for Security and Emerging Technology, December 2019, https://cset.georgetown.edu/wp-content/uploads/CSET-Maintaining-the-AI-Chip-Competitive-Advantage-of-the-United-States-and-its-Allies-20191206.pdf.

22. Lorand Laskai, forthcoming paper. More generally, China has sponsored research in different AI architectures, such as those that are more directly inspired by the human brain. William Hannas and Huey-Meei Chang, "China's Access to Foreign AI Technology: An Assessment," Georgetown University Center for Security and Emerging Technology, September 2019, https://cset.georgetown.edu/wp-content/uploads/CSET-Chinas-Access-to-Foreign-AI-Technology-2.pdf.

23. For discussion of this idea, see Amodei and Hernandez, "AI and Compute." Danny Hernandez, "AI and Efficiency," OpenAI, https://openai.com/blog/ai-and-efficiency/.

24. Scott Cleland, "Google's 'Infringenovation' Secrets," *Forbes*, October 3, 2011, https://www.forbes.com/sites/scottcleland/2011/10/03/googles-infringenovation-secrets/#7bff9df030a6.

25. Roxanne Heston and Zachary Arnold, "Strengthening the US AI Workforce," Georgetown University Center for Security and Emerging Technology, September 2019, https://cset.georgetown.edu/wp-content/uploads/CSET_US_AI_Workforce.pdf.

26. Zwetsloot, "China's Approach to Tech Talent Competition: Policies, Results, and the Developing Global Response."

27. See the addendum to Amodei and Hernandez, "AI and Compute."

CSET CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

CSET.GEORGETOWN.EDU | CSET@GEORGETOWN.EDU