TECHNOLOGY FOR GLOBAL SECURITY

AI and Human Decision-Making: AI and the Battlefield

Technology for Global Security and Center for Global Security Research

November 28, 2018

Recommended Citation

T4GS, "AI and Human Decision-Making: AI and the Battlefield", T4GS Reports, November 28, 2018, http://www.tech4gs.org/ai-and-human-decision-making.html

AI and Human Decision-Making: Considering AI on the Battlefield

Introduction

As the 21st century geopolitical balance shifts in uncertain ways, there is an increasing eagerness to deploy AI technologies into both the physical and digital battlefields to gain both tactical and strategic advantage over adversaries. However, the nature of increasingly powerful and unpredictable AI demands a measured and balanced approach to deploying these tools before the limitations, risks, and vulnerabilities are fully understood and addressed. Indeed, these technologies may not currently be "ready for primetime", on a number of levels. We begin this discussion - meant to be a series of posts on this domain of issues - focused on the following. This initial paper is based off numerous small-group workshops and ongoing engagement with the AI research community in the San Francisco Bay Area:

- Current AI capabilities remain limited to narrow, well-defined domains
- The "black box" nature of state-of-the-art AI/ML algorithms gives limited insight as to their decision-making processes - and conclusions
- Deploying AI's could preempt ethical considerations that have yet to be fully understood, identified, or agreed upon, and is in the potential context of an industry-driven race to the bottom

Narrow Remains Narrow…for Now

Increased AI performance in a specific task does not translate to other, unrelated tasks: the current generation of AI remains limited to constrained environments—which warzones are not—making the deployment of current AI technologies in a military context highly unpredictable. Moreover, AI-assisted systems are disconnected from one another and unable to collaborate outside of their specific design parameters. Some of the limitations of AI for military purposes are rooted in the sources of data used within the training process of the systems. More specifically, in correctly identifying the input and output of the system, as well as the context of its application(s). Additional limitations for data sets in the military context include AI's lack of ability to capture and comprehend complex geo-political concepts. Indeed, certain concepts are unable to be learned by AI systems because they cannot be quantified/represented. Ethics, values and social norms are highly contextual and dependent on circumstances. Within a military context, much of the high-level decision processes involve abstract ideas that cannot be broken down into correlated pairs. What is needed is not just data—of which there may be an abundance—but correlated pairs. Thus, another main concern with the deployment of AI is data validation, because models of complex real-world situations require enormous amounts of human-tagged correlated pairs of data points, which direct the computer to its goal through human involvement and supervision. This remains a significant barrier in AI. For AI to learn something, it has to be "representable" quantitatively. If it cannot be represented, the system cannot learn it. In looking at the development of AI towards human-level performance, there is a narrow context for AI systems in their specific problem domains, meaning that machine performance may degrade dramatically if the original task is modified even slightly. This means it would take a very different set of AI systems to perform a set of diverse tasks that would require only the intelligence of a single human. Additionally, if the data set used for training systems has been polluted, significant questions remain concerning how well that can be

detected. While there is great potential in AGI research, the exponential gains in "easy problems" with narrow scope in today's research do not scale easily to the "hard problems" that are necessary to provide a holistic view of a highly complex, open-ended problem space such that a battlefield will demand (digital/cyber or physical).

The "Black Box" and Decision-Making

The oft-referenced 'black-box' problem makes it difficult for the user to understand the exact process by which the AI technology comes to its conclusions/decisions. The speed with which decisions must be made—especially in wartime—means there may be limited human interference/participation. This necessitates a predictability and interpretability of system behaviors well beyond current capabilities. For example, consider a situation in which a military commander relying upon an AI system for information on enemy whereabouts and logistical planning for military positioning. If there is not a transparent understanding of the machine's decision-making process, then the veracity of the information being provided for real-time human decision making results in almost impossible decision points—not to mention that the after action review of the situation is rendered unbelievably complex not only due to the opacity of machine-learning decision making, but also due to the dynamic between the human and machine coming to a decision at all. As these systems are increasingly relied upon, humans will likely no longer be able to participate in large portions of the decision-making process, as battlefield applications become either too complicated and/or fast, and humans cannot keep pace to decide quickly enough which courses of action to choose. This could present a situation where it is advantageous to attack first with AI rather than defend, which results in a bias towards striking first while one still can. Machine learning is also advancing the ability to quickly reconstruct alternate scenarios as part of the planning phase of operations. Such applications could be applicable to wargaming. In terms of deployment, the current limitations of AI can be traced to the system design of these technologies. It is the system design that determines if the developments in AI can become suitable for warfare or not. Using AI for planning and analysis is different than using it to locate and attack enemy forces.

The AI "Fog of War": Ethical & The Race to the Bottom

Decisions made by autonomous weapons systems would need to be calculated to account for ethics such as shooting a child versus an adult, and the consideration of collateral damage. The incentives built into the systems manifest different behaviors, and if the functions of a system are not clearly understood, there is a risk of a lack of control and direct responsibility of the technology that is created as decision-making process is delegated to machines. Without a grasp of the above-mentioned limitations, executives, policymakers and pundits tend to focus on the upside of AI at the expense of adversarial ability to leverage more mundane technologies to disrupt. The uptick in spending in AI by commercial applications, and increasingly by the U.S. Department of Defense and other governments globally, risks the likelihood of a race to the bottom—both in industry but also more dangerously between nation states to leverage technological developments to deter, gain, and hold advantage. Overpromising is a problem within the community, and technology is a field that is driven by profit. As a result, AI tends to be over emphasized without a clear understanding of what is possible and what is lacking with these technologies ("we do not have the capabilities, but people think we do"). Additionally, current policymakers are not anticipating adversaries dangerously repurposing current cheap consumer-level AI. The speed with which such technology is being developed/deployed is

worrisome to those concerned with establishing ethical guidelines to support principles and norms.

Recommendations

At the heart of the breathless pursuit of AI primacy is gap between researchers and the policy makers who want to deploy the technology. Essentially, each side tends to have an oversimplified view of the other. The policy community is not familiar with using AI as a strategic lever; its strengths and weaknesses are largely unknown beyond superficiality. In turn, AI researchers are not familiar with the nuanced, visceral world of global security. Collaboration between technical and policy experts can help bridge this gap – collaboration based on regular interaction, robust debate, and a lot of listening. Such collaborations could begin to consider the following:

- What types of problems can be solved with our current data and methods?
- What possible unanticipated behaviors can result from a system that interleaves human and AI processes?
- What are the ethical concerns for an AI with a great deal of autonomy as we consider the economic drivers for most of the technologies coming online?
- What are the non-kinetic threats that AI could pose, in terms of decision making, for example?
- While Artificial General Intelligence (AGI) may not be on the immediate horizon, what plans/frameworks need to be developed in anticipation of its potential development?

Conclusion

In reality, the incentives for deploying powerful AI technology onto the battlefield and beyond will likely outweigh any ethical apprehensions, as both money and political power are at stake when competing for technological dominance. Discussions of both the needs of the government on a shifting, violent geopolitical landscape, as well as the current capabilities to serve those needs from a technical perspective should help calibrate AI technologies with military requirements. All AI is not equal, and the military missions for which it is being considered are similarly diverse. Without understanding the risks and benefits of the technology, AI applied to military purposes could produce very unpredictable and undesirable results. The time is right for broad engagements involving the technical AI community, military officials, and policy experts to plan for the near term applications of AI and to discuss the implications of if and when AI becomes truly powerful.