Computer-Centered Humans: Why Human-AI Interaction Research Will Be Critical to Successful AI Integration in the DoD

E. S. Vorm
U.S. Naval Research Laboratory

***Abstract*—The technology development of Artificial Intelligence (AI) in the U.S. Department of Defense (DoD) is proceeding at an unprecedented pace, with record breaking levels of funding to support unprecedented breakthroughs. As these technologies progress through technology readiness levels and make their way into the hands of human beings, however, the need for human-centered design practices will become more evident. This article briefly illustrates the emerging need for more human-AI interaction research in the Department of Defense to ensure an appropriate and cohesive integration strategy of AI in warfighting and defense sectors.**

Funding for research and development in the U.S. Department of Defense (DoD) today is at an all-time high. Of the total $134.1 billion proposed for fiscal year 2020, nearly half (44.3%) was devoted to the defense sector.[1] If there is one theme that permeates all research and development in the DoD today, it is probably the promise of artificial intelligence (AI). It is difficult to overstate just how prevalent discussions

Digital Object Identifier 10.1109/MIS.2020.3013133
Date of current version 3 September 2020.

involving AI have become across the entire spectrum of warfighting and defense capabilities. But while the speed of research in the technology of AI is clearly increasing, research in human-AI interaction does not appear to be keeping pace. Issues such as usability, interaction modalities, visualization, and knowledge representation techniques are vital parts of a coherent technology integration strategy. While it is critical that the United States keep pace with its near-peer allies and adversaries by developing advanced technologies, it is also vital that these technologies are developed in ways that humans can understand and use appropriately. This article discusses two principal reasons why research in the psychology of human-AI interaction and human-centered design work needs to keep pace with technology development in the DoD: to ensure AI systems that are safe and reliable for humans to use, and to ensure these systems are integrated in ways that do not inject new kinds of error and risk to existing systems.

HOW TO ASSURE SAFE AND EFFECTIVE HUMAN-CENTERED AI

Conflicts between humans and advanced technologies always have the potential to arise in complex socio-technical systems. Examples of these conflicts can range from errors in human perception that originate from poorly designed interfaces, to errors in human judgement and decision making that are introduced when human interactions with said systems confuse, distract, or disrupt normal information processing. Decades of research has taught us that the potential for these conflicts increases exponentially in relation to the complexity of the system, and the degree to which subsystems are tightly coupled to one another.[2] Technologies being developed by the DoD today, such as many autonomous weapons systems, and systems that provide advanced data analytics and decision support are amongst the most sophisticated and technologically complex systems ever brought to bare. Compounded by the black-box nature of many of these systems, whose intricate dimensionality renders a quick explanation of system behavior often impossible, the potential for significant human factors conflicts to occur as these systems begin to come online in the DoD is tremendous. Contrary to popular opinion, however, these conflicts are seldom the result of technological failures. Instead, these conflicts in human-AI interaction tend to originate in a far more variable and lesser understood system the user's brain. How humans use and react to complex technologies is a delicate dance between perception, sense making, decision making, and acting, with sticky ingredients such as trust thrown in for extra measure. Preventing these conflicts takes deliberate and careful design work to ensure systems that are safe for humans to operate and reliable enough to employ in high-risk domains. This is why the Explainable Artificial Intelligence (XAI) program at U.S. Defense Advanced Research Agency (DARPA) is so vital to the development and integration of AI into the DoD.

The XAI program is funding research that seeks to develop methods to make machine learning models more understandable to human beings. Creating explainable AI is especially important for DoD applications such as autonomous systems, whose stochastic nature can sometimes produce baffling and confusing behavior that is difficult to understand. Building tools that help humans to understand, appropriately trust, and effectively use AI is an example of the type of research that will need to keep pace with technology development, especially as autonomy becomes a larger and more prevalent component of the U.S. National Defense Strategy.[3]

DEVELOPING AN INTELLIGENT AI INTEGRATION STRATEGY

Another reason that research in human-AI interaction needs to keep pace with technology development is because failing to do so essentially guarantees that we will integrate AI into the DoD through a strategy of assimilation, where we try to shoehorn new technologies into existing processes and workflows. The injection of advanced technologies into the workplace always transforms those processes while creating new forms of error.[4] At the same time, new technology often renders older tasks considerably easier or even obsolete. A classic example is the QWERTY keyboard, a design that was originally developed to limit the speed of typists because the hammers in early typewriters could not cope with excessively fast keystrokes.[5] Today there are several more efficient keyboard layouts, and given that computer keyboards do not have the same physical limitations of their late-19th century cousins, the use of the QWERTY layout today is mostly a testament to habit.

AI has the potential to change nearly every aspect of how the United States defends itself and projects power to deter aggression. Rather than asking "what else can we automate?" and proceed by trying to fit new technology into old, clumsy processes, we need instead to adopt a strategy of accommodation. Accommodation, in this case, would mean developing new processes that carefully consider how to best leverage the advantages of AI in ways that support and blend with the strengths of human beings, while also compensating for the weaknesses of both.[6] This is the cornerstone of the concept of human–machine teaming, and is the focus of research at the Navy Center for Applied Research in Artificial Intelligence at the U.S. Naval Research Laboratory (NRL). The principal goal of human–machine teaming is to identify the right balance of task and function allocation between advanced technologies and human beings in order to synergize the strengths of both. Efforts in human–machine teaming at NRL are exploring how humans interact with technology, especially in high-risk domains where decision speed and accuracy are critical. Without this modeling, systems engineers are forced to make broad assumptions about how humans will respond to and interact with autonomous systems. History has unfortunately demonstrated that assumptions about user interactions can often be drastically incorrect, which can lead to deadly consequences.[7] Using cognitive modeling techniques, we are working to model interactions between humans and advanced intelligent systems so that we can build systems that are safe and assured, and are easier for humans to use, understand, and appropriately trust.

OPEN CHALLENGES FOR HUMAN-AI INTERACTION

There are a number of ongoing, open challenges unique to the DoD that will need to be addressed to ensure the appropriate and effective implementation of AI. One is the use of automated planning algorithms to aid commanders in making complicated and time-critical decisions. Planning in any domain is a time-consuming and labor-intense process, largely dictated by the specific domain and operational level in which the process occurs (e.g., ship routing through dangerous waters; tactical orders of battle; satellite scheduling, etc.). The DoD has invested a considerable amount in the recent development of a large variety of computer-automated planners with the goal of making the traditional planning process faster and less human-resource intensive. These planners, however, have proven less useful than hoped, largely because computers are currently unable to incorporate higher-level reasoning and constraint consideration into their planning process. Developing techniques that expand the ability to represent these higher order dependencies and context-specific intents in automated planning are needed to help realize the promise of intelligent automated planning under real-world complex constraints. Alongside the algorithmic development, there is also need for efforts that enable human users to provide feedback to the algorithms in ways that are natural, flexible, and that result in appropriate updating and future learning of the algorithm. This concept is often referred to as "closing the loop." In this context, the loop begins with an intelligent algorithm having learned from training data, and providing an output in the form of a prediction or recommendation to the user. Currently, the user is unable to provide meaningful feedback or coaching to the algorithm regarding its output; the only way to update the model is to go back and retrain it which can be a very time consuming process, and requires specific expertise. New generations of intelligent algorithms will need to possess the ability to receive feedback from human users, and update their models to reflect the needs and expectations of those users. This will greatly increase the utility of automated planning algorithms and other intelligent decision support systems. Multiple research programs by various DoD funding agencies are currently in development and will provide a coherent research strategy to support the study and acquisition of these functionalities.

Another open challenge for AI in the DoD today is the use of predictive data dashboards for unit readiness. Dashboard analytics have garnered immense attention from the highest levels of military strategy, policy, and planning in recent years. The DoD is currently using dashboards to provide predictive analytics for a variety of activities, including acquisition program planning, manning, and recruiting. Recently there has been much discussion around the use of dashboards for making behavioral predictions to support unit readiness. The vision of these dashboards is to provide unit commanders with a snapshot look of unit readiness in a manner that could encourage early interventions to prevent unplanned losses from destructive behaviors like suicide, domestic abuse, and sexual assault. Using data from service members' various military and health records, algorithms could predict in aggregate unit-level risk for these behaviors, and alert staffs who could respond appropriately before these behaviors occur. There are a number of sensitive issues that need to be carefully considered in conjunction with the building of these dashboards. These concerns include addressing the ethical use of behavioral predictions, methods to detect and deter bias in both the algorithm, as well as the human decision making component to ensure that individuals cannot be unfairly targeted or discriminated against based solely on algorithmic predictions of their behavior. The U.S. NRL and its partners, through funding from the Office of Naval Research, are currently leading an effort to address these sensitive concerns in order that unit readiness dashboards can be developed using techniques that ensure they are fair, accountable, and transparent.

CONCLUSION

The DoD has made a strong commitment to always having humans in the loop of autonomous and lethal weapon systems.[8] Studying how human beings perceive, comprehend, and make decisions while interacting with AI, therefore, remains an absolutely necessary component of the DoD's AI integration strategy. Research in human-AI interaction and human–machine teaming is vital if we are to consider seriously how best to capitalize on AI in the DoD in ways that improve mission success. Failing to do so only perpetuates a strategy that asks humans to fit into uncomfortable and ill-fitting technology rather than designing it for their principal benefits strategy that increases the likelihood that these future "disruptive technologies" may end up disrupting only ourselves.

ACKNOWLEDGMENTS

This work was supported in part by the U.S. Defense Advanced Research Project Agency (DARPA) under the Explainable Artificial Intelligence (XAI) program. I would like to thank Dr. M. Turek, program manager in the DARPA Information Innovation Office for his support. I would also like to thank Dr. B. Clancy and Dr. R. Hoffman of the Florida Institute for Human–Machine Cognition, and G. Trafton of the U.S. Naval Research Laboratory for their thoughtful remarks and suggestions in preparing this article.

REFERENCES

1. Congressional Research Service, "Federal research and development funding: FY2020," Tech. Rep. R45715, Mar. 2020.
2. C. Perrow, *Normal Accidents*. Princeton, NJ, USA: Princeton Univ. Press, 1984.
3. Department of Defense, National Defense Strategy, USA, Jan. 2018.
4. A. M. Pendleton-Jullian, J. S. Brown, P. N. Edwards, and G. C. Bowker, *Design Unbound*. Cambridge, MA, USA: MIT Press, 2018.
5. R. L. Dorit, "Marginalia: Keyboards, codes and the search for optimality," *Amer. Sci.*, vol. 97, no. 5, pp. 376–379, Sep. 2009.
6. P. Neuhaus, A. Raj, and W. J. Clancey, "Human-centered cognitive orthoses: Artificial intelligence for, rather than instead of, the people," *AI Mag.*, vol. 36, no. 4, pp. 9–11, Dec. 2015.
7. J. K. Hawley, "Not by widgets alone," *Armed Forces J.*, vol. 51, no. 2, pp. 13–19, Feb. 2011.
8. Department of Defense, Autonomy in Weapons Systems, DoD directive 3000.09, 2012, pp. 1–15.

E. S. Vorm is currently an Aerospace Experimental Psychologist with the U.S. Naval Research Laboratory in Washington, D.C., USA. He conducts research in areas of human–machine teaming, human-autonomy interaction, and human-systems integration. As an Active Duty Officer in the U.S. Navy Medical Service Corps, he holds a Ph.D. degree in human–computer interaction from Indiana University, Bloomington, IN, USA. He served on the front-lines of combat in the middle-east with the US Marine Corps during Operation Iraqi Freedom, and is currently a member of the IEEE Robotics and Automation and Systems, Man, and Cybernetics Societies. Contact him at eric.vorm@nrl.navy.mil.