DEVELOPING ETHICAL, LEGAL, AND POLICY ANALYSES RELEVANT TO THE USE OF MACHINE LEARNING ALGORITHMS IN NATIONAL SECURITY

Andrew H. Peterson, Jesse L. Kirkpatrick

Institute for Philosophy and Public Policy

and

Deborah A. Boehm-Davis

Department of Psychology

George Mason University

e-mail: apeter31@gmu.edu; jkirkpat@gmu.edu; dbdavis@gmu.edu

This white paper is in response to the call for white papers designed to provide input to the U.S. National Academies Social and Behavioral Sciences Decadal Survey, which is being conducted with the goal of helping guide the development of a research agenda in national security for the next ten years. This paper focuses on policy issues related to the use of machine learning algorithms in national security. The paper addresses the following areas taken from the call for papers:

- Modeling and/or improving understanding of behaviors of relevance to national security (including those with multiple actors)
- Strategies and techniques for avoiding errors and biases in decision making
- Decision support for national security initiatives
- Support systems in the workplace (e.g., for managing workloads, volumes of data/information, and stressful events)

A machine learning algorithm is a data analysis method that uses a computer program to sort, interpret, and classify data without human intervention (Goodfellow, Bengio & Courville, 2016). A program first is trained on a data set drawn from a larger data set. The program develops an algorithm that predicts a specified outcome using those data, and is then set loose on the larger data set for analysis. Machine learning algorithms are used broadly within the social and behavioral sciences. Some argue that machine learning—deep learning, in particular—is a major technical step toward general artificial intelligence (LeCun, Bengio & Hinton, 2015). The use of machine learning in national security contexts raises important policy issues now, and likely will continue to raise more within the next 10 to 20 years. There are at least 3 areas where machine learning is currently being used for national security purposes (although there likely are many more):

- **Autonomous weapons systems**: An autonomous weapons system is a weapon or weapons platform that can select and engage targets without human intervention. Machine learning algorithms play a role in the target selection process. For example, machine learning can be used to interpret satellite images and select targets (e.g. weapons instillations or military vehicles).

- **Biometrics**: Biometrics is the process of gathering biological information on subjects to track, and potentially predict, their movement and behavior. A contemporary example of biometrics is facial recognition software.

- **Cyber warfare**: The growing cyber domain raises the possibility of engaging in virtual warfare in the next 10 years. Machine learning algorithms will play a crucial role in cyber weapons and cyber defense. Machine learning algorithms can respond faster than humans, and may provide superior defense for cyber-attacks.

The above applications of machine learning raise important policy issues. Among these issues are:

- **How should we account for false positives and false negatives generated by machine learning?** Often machine learning provides a better assessment of data than humans. However, there are cases in which machine learning outputs are wrong. A target selection system that uses machine learning could generate false positives, and misidentify a friendly unit as a threat. For example, in the 2003 Iraq War, a U.S. Patriot Missile guidance system misidentified a coalition aircraft as a threat. This was the first friendly fire incident of Operation Iraqi Freedom (Piller, 2003). Biometric systems using machine learning might also misidentify individuals. In fact, facial recognition software has been shown to be subject to error in people with dark skin (Magnet, 2011; Introna & Wood, 2004).
- **Who is legally responsible if a machine learning algorithm is wrong?**
  Should a system using machine learning make an error (e.g. select the incorrect target), who is legally responsible when those who were injured by the incorrect action sue for damages? Is it the programmer who developed the machine learning algorithm? Is it the company that she works for? Or is it the organization who contracted for the system to be built?
- **Do biometric machine learning algorithms violate a right to privacy?**
  The use of machine learning algorithms in biometrics raises important concerns regarding the 4th amendment. Under what conditions should U.S. citizens have a reasonable expectation of privacy if there are sensors in public spaces that gather and analyze data regarding their biological signature?
- **How do we regulate the development of machine learning algorithms?**
  U.S. federal agencies (e.g. IARPA and DARPA) may no longer be the unchallenged leaders in technical and scientific innovation. Rather, private industries are currently leading the charge, particularly in the development of machine learning (consider Google image recognition or IBM's Watson). This could be problematic for imposing regulations; the goals of industry might not align with U.S. national security priorities and the U.S. government is not able to regulate the development of new technologies by simply turning off funding.

There are also issues that arise when humans interact with, and make decisions based on the output of, machine learning. Among these issues are:

- **What might happen if machine learning is used in command and control scenarios?**
  Machine learning algorithms may soon be used in command and control. Machine learning can analyze the movement of ground units, and provide command recommendations to improve mission outcomes. However, the use of machine learning in command and control raises difficult questions. What happens when an operator disagrees with the recommendations derived from machine learning? Will an operator follow the recommendations or will she rely on her best judgement? There may be situations in which a human operator is aware of more contextual features of a command and control scenario than an algorithm can encode. This may provide a "sanity" check on the system, yet it is difficult to define when these checks ought to occur.
- **Who is responsible if machine learning is used in command and control scenarios?**
  If machine learning is integrated into command and control, who is ultimately responsible for the human-machine output, and who is responsible for the final decision? Does this change if the human operator is "in-

the-loop"? If so, would the human operator be responsible for the final decisions, rather than the machine or the machine's programmer?

- **What happens to human operators in the context of human-machine interaction?** Research in psychology and human factors, as well as the first-hand experience of pilots flying aircraft equipped with automated systems, suggests that when automated systems are in use, the human operator can lose vigilance of the current situation. This is referred to as being "out-of-the-loop" (Endsley & Kiris, 1995; Boehm-Davis, Curry, Wiener, & Harrison, 1983; and Wiener & Curry, 1980). Operators may also not be clear about how a machine learning system is drawing its conclusions. This can lead operators to question—or not question—machine learning outputs appropriately. Developing methods that ensure operators understand what a system is doing, and how it is doing it, could facilitate active monitoring of machine learning and improve outcomes of human-machine interaction. This issue has broader policy implications regarding the training of operators, and the national security operations that should—or should not—rely on machine learning.

Recommendation: The issues raised in this white paper suggest that there is a need for researchers working in fields including computer science, philosophy, cognitive psychology, and human factors/ergonomics to come together to conduct ethical, legal, and policy analyses that are integrated in the development of new technologies—whether it is machine learning or other innovations in the behavioral sciences. A successful model that supports these multidisciplinary interactions is the funding design of the NIH Human Genome Project and BRAIN Initiative. Both programs have funding streams that support ethical, legal, and policy analyses of basic science. A similar endeavor in the domain of national security and behavioral sciences would likely generate beneficial scholarly and policy outcomes.

References

Boehm Davis, D. A., Curry, R. E., Wiener, E. L., & Harrison, R. L. (1983). Human factors of flight deck automation: Report on the NASA-Industry workshop. *Ergonomics*, 26(10): 953-961.

Endsley, M. R., and Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2): 381-394.

Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep learning*. Cambridge: MIT Press.

Introna, L., Wood D. (2004). Picturing algorithmic surveillance: The politics of facial recognition systems. *Surveillance & Society*, 2(2/3): 177-98.

LeCun, Y., Bengio, Y., Hinton, G. (2015) Deep learning. *Nature*, 521(7553): 436-44.

Magnet, S. (2011) *When biometrics fail: Gender, race, and the technology of identity*. Durham: Duke University Press.

Piller, C. (2003). "Vaunted Patriot Missile Has a 'Friendly Fire' Failing." *Los Angeles Times*, April 21. http://articles.latimes.com/2003/apr/21/news/war-patriot21.

Wiener, E. L., and Curry, R. E. (1980). Flight deck automation: Promises and problems. Ergonomics, 23, 995–1011.