Challenges on value-laden research of AI for military applications

Yunji Woo

1. Introduction

Could science and technology be value-free in the considerations of military applications? The legitimate role of values in science is one of the central topics in the contemporary philosophy of science. According to Heather Douglas(2000), for science that has clear non-epistemic impacts, being “value-free” is not an achievable goal. Holman and Wilholt(2022) argue that value necessarily plays a role in core areas of scientific inquiry, and have proposed the ‘New demarcation problem’ as a way to distinguish between legitimate and illegitimate influences of values in science. If the technology is developed under consideration of societal applications, the research is more engaged with non-epistemic values and harder to be value-free.

AI has a wide range of potential applications, including automating routine tasks and processes, improving decision-making by providing insights and recommendations based on data analysis, and enabling new forms of human-computer interaction. Industrial sectors widely adopted the concept of AI with great interest to process complex information, collecting multi-modal information, and learning algorithms based on the collected information to adjust new information. The technology has a widespread interest within the industry, which funds research on its applications. However, governments are still major funders, especially for military applications.

DARPA(Defense Advanced Research Projects Agency) has achieved huge attention on AI implications on defense systems. After showing off the possibility of usage of AI by AlphaDogfight trials, the defense sector has a high interest in AI as a human-machine association. DARPA’s Air Combat Evolution(ACE) program seeks to automate air-to-air combat and build human trust in AI as a step toward improved human-machine learning. The Ministry of Defense(MOS) in the USA published Defense Artificial Intelligence Strategy on June 2022,  which aims to modernize the armed forces based on AI.

The potential of AI to improve efficiency and accuracy, reduce human labor, as well as enhance cyber security and facilitate logistics, had led to high expectations for its use in the military field, particularly for surveillance, decision-making, and action on the battlefield. The circumstance was accelerated as ‘International ARMS races’ between China and the USA as ongoing competitions.

Initially, AI was thought of as a value-free idea, free from the biases and prejudices of humans. However, this belief has been disputed in recent years as evidence has emerged showing that AI systems can replicate and even amplify the biases present in the data they are trained on. This has led to a reevaluation of the idea that AI can be completely value-free and has sparked discussions about the need to consider the implications of AI and to develop methods for mitigating bias in AI systems.

There has been criticism of the research on AI applications for not adequately considering philosophical perspectives, which some belief are essential to the responsible development of AI. As a rapidly advancing technology, AI raises several significant philosophical issues, such as the inductive risk and the non-epistemic value of its applications. The technology has also been criticized for its lack of trustworthiness and the presence of biases in the data it uses.

The military use of AI enhances the applied non-epistemic value that influences the inductive risk. Some key inductive risks associated with AI in military systems include the potential for biased or unfair decision-making and the possibility for unethical or illegal use. There are also concerns about accountability, responsibility, and the protection of human rights when AI is used in military contexts. The inductive risks of AI in military systems will vary depending on how the technology is used and the specific context in which it is deployed.

In this research, I would describe the epistemic and non-epistemic values applied to AI, regarding the national innovation expectations, inductive risks of the technology, uncertainty, and explainability, and especially the operational risk of military applications. I argue that these philosophical difficulties are aggregated in the consideration of military applications, because the AI is trained with value-laden datasets, with value-purposed training.

1. Public-funded emerging science and technology research in view of national innovation system

The study of military science is considered to be value-neutral. It is the study of the principles and techniques of military operations, including strategy, tactics, logistics, and other related subjects. The technical aspects of warfare do not necessarily consider the moral or ethical implications of using military force. As such, it does not typically involve the implementation of non-epistemic values, which are values that are not related to knowledge or belief. But, when it is applied in practice, non-epistemic values might include issues such as national security, the protection of human rights, and the promotion of economic and political stability. These values may influence how military science is developed, applied, and evaluated, and also be considered in deciding how research funding should be allocated. The specific non-epistemic values have legitimate roles depending on the goals and priorities of the stakeholders.

The mission-oriented or project-oriented science that is focusing on achieving objectives that are determined by non-scientific considerations is becoming more common, challenging the ideal of value-free science. This approach to science often involves close collaboration between scientists and stakeholders from other fields, such as policymakers, industry leaders, and community organizations, to ensure that the research is focused on addressing real-world needs and has the potential to generate tangible benefits for society.

National states have a high interest in emerging science and technology because of national innovation systems. From the concept of national innovation systems, science and technology is playing a high-priority role in modern economics after WWII. (Damon, 2017) Although the opportunity of developing national economics was equalized by globalization which leads to the development of transportation and communications, the national disparities in innovation are described for various explanations. 1) Country’s population or economy will determine its national innovation rate. 2) Military spending determines national innovation rates. 3) First-mover advantages in some countries while others are locked in competition. 4) Late industrialization explains national innovation rates. 5) National culture matters.

I would like to focus on military spending and defense-related innovations that claims military science has a role in the development of the national science and technology industry. There is a strong belief that military investment in R&D leads to technologies that also benefit commerce and welfare in the form of defense-related innovation that consistently spills over into the civilian economy. Military spending and weapon production correlate with national innovation rates, and this comes with geopolitical concepts of national S&T innovators to accelerate the ARMS race with its competitors.

Military funding for AI research is often focused on specific goals or applications, rather than on general AI research. This can limit the ability of researchers to pursue more fundamental or exploratory research, which may be important for advancing the field as a whole. Holman and Elliot(2018) argued that industry-funded science has raised philosophical discourse promoting practical concerns about industry bias. In the considerations of military application, the stakeholder ecosystem could augment the role of non-epistemic value on AI.

Moreover, mission-oriented research could shift the assessment of research products and potential from scientists to external factors, such as governmental non-epistemic values. This can lead to what Elzinga (1997) refers to as "epistemic drift," which is a result of the clash between different value systems and cultures. Governmental non-epistemic values, such as national security or economic interests, may influence the direction and priorities of research funding, leading to a focus on certain areas of study or approaches to problem-solving at the expense of others. This can have implications for the quality and trustworthiness of the research, as well as its potential impact and applications. It is important to carefully consider the potential impacts of mission-oriented research and the role of non-epistemic values in shaping the direction and priorities of scientific research.

1. Inductive Risk of AI

Values influence scientists directly when values are treated as if they are a form of evidence. The concept of inductive risk is a consequence of direct and indirect ways that value could influence scientific reasoning, which affects accepting or rejecting scientific hypotheses. How much evidence science should demand before concluding socially relevant topics? Hempel(1965) explains how inductive risk fits with other works on the legitimate uses of non-epistemic values in science, and how consideration of inductive risk can require the use of non-epistemic values. Heather Douglas(2000) claims that the public questioning of science generates much heat but little on the values of science questions.

The values that are embedded in the design and use of a machine learning model can also influence the decisions made by the model. For example, if a model is designed to prioritize certain values, such as transparency or fairness, this can affect the decisions made by the model. The values that are held by the people who design, train, and use a machine learning model can also influence the decisions made by the model. Values are closely associated with data in machine learning, and understanding and addressing these values is important for ensuring that machine learning models make fair and accurate decisions.

Regarding military applications, the non-epistemic values affecting the system are significant. The data used to train a machine learning model can reflect if it contains information that embodies military values.

Inductive risks come with the question, what if we are using the wrong standard? According to Douglas(Douglas, 2000), inductive risk plays an ineliminable role in the evaluation of scientific hypotheses. The statistical significance levels and decision of how to weigh evidence are responsible, for instance, false positive and false negative errors. Both errors could have a significant influence on military-applied AI systems, where the consequences of such errors can be serious. To mitigate this risk, military AI systems should be carefully designed and tested to ensure that they can accurately and reliably identify potential threats.

A false positive error would refer to a situation in which the AI incorrectly identifies a threat or hostile action when there is none. This could lead to a military response that is unwarranted and potentially dangerous. False positive errors can occur when the AI system is not trained on a diverse enough dataset, or when it is not able to accurately interpret the data it is given.

A false negative error would refer to a situation in which the AI system fails to identify a threat or hostile action when it is present. This could lead to a failure to respond to a potentially dangerous situation, and put soldiers and civilians at risk. False negative errors can occur when the AI system is not trained on a diverse enough dataset, or when it is not able to accurately interpret the data it is given.

In the following paragraphs, I will argue that AI involves incorporating values into the training process, which can introduce inductive risk. The uncertainty of data and lack of explainability in AI are major concerns, and the military applications of AI can either worsen or mitigate these risks.

1. Uncertainty of data

The nature of AI is an explanation of optimization to a dataset. Values encoded in the data itself can influence the decisions made by a machine learning model. If a dataset contains biased or discriminatory information, a machine learning model trained on that data may learn to make biased or discriminatory decisions.

Inductive risk, as a generalization error, is the risk that a machine learning model will make incorrect predictions on new data that it has not seen before. This risk arises from the fact that machine learning models are trained on a limited dataset, and they may not be able to accurately generalize their predictions to new, unseen data.

From the training process, two types of uncertainty arise Aleatory uncertainty and Epistemic uncertainty. Aleatory uncertainty is the uncertainty caused by the inherent noise contained in the data. It cannot conclude that the data is clean when acquiring data from a sensor, etc. The inherent noise of the data is included from the measurement stage, and it does not disappear just by acquiring a large amount of data.

The types of aleatory uncertainty can be separated into homoscedastic uncertainty, which assumes that noise is constant for each type of input data, and heteroscedastic uncertainty, in which noise is different for each input. The practical samples are important to model heteroscedastic uncertainty.

Apart from the data, epistemic uncertainty is caused by the model parameter, stating how confident the model is. In machine learning, the more data, the less uncertain the model is. However, the military samples have problems with data.

To address the uncertainty in data, it is important to choose and carefully select the training data and use appropriate techniques for evaluating and selecting the model. However, military applications often involve sensitive or classified data, which may not be readily available to researchers outside of the military. This can limit the ability of researchers to develop and test new AI models. Security concerns that restrict the openness and transparency of AI development could also limit the ability of researchers to collaborate and share knowledge, which is important for driving progress in the field. The acquisition and quality control of data can be difficult in these circumstances, leading to epistemic uncertainty.

There are challenges associated with generating synthetic data using GAN algorithms and modeling simulation to address security concerns in military AI applications. Synthetic data generation is a technique that allows researchers to create artificial data that is similar to real-world data but does not contain any sensitive or classified information. By using synthetic data, researchers can still train and evaluate machine learning models without compromising the security of sensitive data. This can be particularly useful for military AI applications, as it enables researchers to develop and test models that can be used in real-world scenarios without exposing sensitive data to unauthorized individuals. The synthetic data generated can be used to augment real-world data, creating larger and more diverse datasets that can improve the performance and generalizability of the models, making them more effective in real-world scenarios.

1. Lack of explainability

AI is often referred to as a  ‘black box’ due to the lack of transparency in its decision-making processes. This lack of explainability is a significant challenge in many forms of AI, including machine learning algorithms that rely on complex mathematical models to analyze and interpret data. The construction of multi-layered machine learning models involves optimizing initial values through successive layers of non-linear calculations. Deep learning, which refers to models with more than 10 layers, can have hundreds of thousands of optimized parameters. While multiple iterations of training can help optimize these models, it does not provide insight into how the results were achieved. A major concern with machine learning models is their lack of reproducibility, as they may be influenced by undetermined model parameters or unique aspects of the training process.

The lack of transparency in the inner workings of deep learning algorithms, known as algorithmic opacity, can also lead to epistemic opacity or a lack of understanding about how the algorithm arrives at its predictions or decisions. This lack of interpretability can be a problem when it comes to accountability and transparency.

In the case of simple classifier AI, the probability of different categories is calculated for samples that pass the classifier, with the category with the highest probability being the predicted outcome. However, the confidence of this prediction is calculated based on the probability, which can make the result uncertain. In contrast, the traditional decision tree models are logic-based and can explain the results they produce.

There is no strict epistemic standard for determining how much evidence is necessary to accept a theory, and the decision depends on both the risks involved and non-epistemic values. Logical relevance to the proposed hypothesis is important, as evidence that does not contribute to the support or disconfirmation of the theory is not useful. In highly regulated or critical environments, such as the military, it is especially important to consider inductive risk in decision-making, to identify how much evidence is sufficient to confirm or refute hypotheses that involve non-epistemic values. Machine learning models, which are based on empirical inference, are subject to error as indicated by the presence of training error and test errors. Therefore, it is important to carefully consider the potential risks and limitations of these models when applying them in real-world situations.

The lack of explainability in AI systems can be a concern in the military context, as it can make it difficult for humans to understand how the system arrived at a particular decision or to identify and correct any biases or errors in the system’s decision-making process. To address these issues, the Defense Advanced Research Projects Agency's (DARPA) XAI (Explainable Artificial Intelligence) research project focuses on developing techniques for improving the explainability of AI systems. Explainability refers to the ability of an AI system to provide clear and understandable reasons for its decisions and actions. The XAI project aims to improve explainability through the development of techniques for generating human-readable explanations of AI system behavior, as well as by incorporating explainability into the design of AI systems from the beginning. By improving the explainability of AI systems, the XAI project hopes to mitigate some of the inductive risks related to the use of these systems in military contexts and improve trust in their decisions and actions.

1. Conclusion

The demarcation problem refers to the challenge of defining clear criteria for distinguishing between scientific and non-scientific theories. Holman and Wilholt (2022) have proposed five types of demarcation strategies, including approaches that define science in terms of its methods, goals, or subject matter. The mission-oriented research proposed by Holman and Wilholt (2022) suggests that non-epistemic values can play a legitimate role in science, as long as they are well-coordinated with expectations within and outside the research community and considered within a broader context. This approach to the demarcation problem, known as the "new demarcation problem," acknowledges that social values can play a necessary role in determining the quality and trustworthiness of science.

Values also play a role in practical engineering, influencing the choices and decisions made by engineers when designing and implementing technological systems and solutions. Values such as safety, reliability, efficiency, and sustainability can guide engineers in selecting materials and components, determining the appropriate level of testing and validation, and making trade-offs between competing design objectives. Additionally, values such as fairness, transparency, and accountability can influence how engineers engage with stakeholders and the public and can help ensure that the impacts of technological systems are considered and addressed responsibly and ethically. The specific values that are relevant to practical engineering will depend on the context and goals of the engineering project, as well as the broader societal and ethical considerations at play.

The use of AI in military contexts brings with it many risks and challenges, including the potential for AI systems to make decisions or take actions that result in unintended harm or negative consequences. There is also the risk that AI systems could be used in ways that violate ethical norms or international laws, and the potential for AI systems to be used in situations where they may not fully understand or predict the consequences of their actions. The risk of an arms race or other forms of destabilization is also a concern. These risks are not unique to military AI but are particularly relevant due to the high stakes and potential for harm involved in its use for military purposes.

The inductive risk of AI applied in the military refers to the potential for AI systems to make incorrect inferences or conclusions based on the data they have been trained on, leading to faulty decision-making and potentially harmful actions on the battlefield. To address these risks and challenges, it is important to consider inductive risk during the training process and design AI systems that are more transparent, interpretable, and fair. By considering the values that are important to society, it may also be possible to develop AI systems that are more ethically and socially responsible, which could help to build trust and confidence in these technologies.

Technological developments have the potential to dissolve some of the ethical disputes surrounding the use of AI in military contexts. The development of robust and reliable AI systems that are capable of operating reliably and safely could help to mitigate concerns about the potential risks and unintended consequences of using AI in military contexts. Additionally, the development of AI ethics frameworks and guidelines could help to address ethical concerns and establish clear principles for the use of military AI. Finally, by considering the values that are important to society, it may be possible to develop AI systems that are more ethically and socially responsible, which could help to build trust and confidence in these technologies. Scientific research and philosophy research could both contribute to the development of AI technology that is more attuned to the values of society and helps to address the risks and challenges associated with its use in military contexts.

1. Reference
2. Elliott, Kevin C. (2017), A Tapestry of Values: An Introduction to Values in Science, New York: Oxford University Press,
3. Douglas, Heather. “Inductive Risk and Values in Science.” Philosophy of Science, vol. 67, no. 4, 2000, pp. 559–79.
4. Coletta, Damon V. “The Politics of Innovation: Why Some Countries Are Better Than Others at Science and Technology. By Mark Zachary Taylor. Oxford: Oxford University Press, 2016. 444p. 27.95 Paper.” *Perspectives on Politics*, vol. 15, no. 3, Sept. 2017, pp. 807–09. *Cambridge University Press*, <https://doi.org/10.1017/S153759271700113X>.
5. Holman, Bennett, and Kevin C. Elliott. “The Promise and Perils of Industry-Funded Science.” *Philosophy Compass*, vol. 13, no. 11, 2018, p. e12544. *Wiley Online Library*,<https://doi.org/10.1111/phc3.12544>.
6. Elzinga, Aant. “Large-Scale Military Funding Induces Culture Clash.” *Space Policy*, vol. 6, no. 3, Aug. 1990, pp. 187–94. *ScienceDirect*, <https://doi.org/10.1016/0265-9646(90)90062-3>.
7. Elzinga, Aant. “The Science-Society Contract in Historical Transformation : With Special Reference to “Epistemic Drift” Social Science Information, vol. 36, no. 3, 1997, pp. 411–45.
8. Hilligardt, Hannah. “Looking beyond Values: The Legitimacy of Social Perspectives, Opinions and Interests in Science.” *European Journal for Philosophy of Science*, vol. 12, no. 4, Oct. 2022, p. 58. *Springer Link*, <https://doi.org/10.1007/s13194-022-00490-w>.
9. Schmid, Stefka, et al. “Dual-Use and Trustworthy? A Mixed Methods Analysis of AI Diffusion Between Civilian and Defense R&D.” *Science and Engineering Ethics*, vol. 28, no. 2, Mar. 2022, p. 12. *Springer Link*, <https://doi.org/10.1007/s11948-022-00364-7>.
10. Kendall, Alex, and Yarin Gal. *What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?* arXiv:1703.04977, arXiv, 5 Oct. 2017. *arXiv.org*, <https://doi.org/10.48550/arXiv.1703.04977>.
11. Karaca, Koray. “Values and Inductive Risk in Machine Learning Modelling: The Case of Binary Classification Models.” *European Journal for Philosophy of Science*, vol. 11, no. 4, Oct. 2021, p. 102. *Springer Link*, <https://doi.org/10.1007/s13194-021-00405-1>.
12. Buckner, Cameron. “Deep Learning: A Philosophical Introduction.” *Philosophy Compass*, vol. 14, no. 10, 2019, p. e12625. *Wiley Online Library*, <https://doi.org/10.1111/phc3.12625>.
13. Biddle, Justin B. “On Predicting Recidivism: Epistemic Risk, Tradeoffs, and Values in Machine Learning.” *Canadian Journal of Philosophy*, vol. 52, no. 3, Apr. 2022, pp. 321–41. *Cambridge University Press*, <https://doi.org/10.1017/can.2020.27>.