Value-laden guidelines on the Implication of Artificial intelligence in Lethal autonomous weapon systems (LAWS)

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 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 play 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. As governments increasingly prioritize the development of science and technology, the mission-oriented or project-oriented science is becoming more common, challenging the ideal of value-free science.

One of the most recent and highly criticized cases is the implication of AI in the military field. Industrial sectors widely adopted the concept of AI with great interest to process complex information, collect multi-modal information, and learn algorithms based on the collected information to adjust new information.

In fact, DARPA(Defense Advanced Research Projects Agency) has achieved huge attention on AI implications on defense systems. After showing off the possibility of usage on AI by AlphaDogfight trials[[1]](#footnote-1), 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 USA has published Defense Artificial Intelligence Strategy on June 2022[[2]](#footnote-2), that aims to modernize the armed forces based on AI.

The potential of AI to improve efficiency, accuracy, and 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 in the battlefield. The circumstance was accelerated as ‘International ARMS races’ between China and the USA as ongoing competitions.

However, the AI technology itself has been criticized for its lack of trustworthiness and availability of data biases. The inductive risk of AI applied military refers to the potential for the AI to make incorrect inferences or conclusions based on the data it has been trained on. This can lead to faulty decision making and potentially harmful actions in the battlefield.

The potential for AI systems to make decisions or take actions that result in unintended harm or negative consequences, from a variety of factors including the limitations of current AI technology. The system could be used in ways that violate ethical norms or international laws, and the potential of AI systems to be used in situations where they may not be able to fully understand or predict the consequences of their actions. The risk that the use of AI in military contexts could lead to an arms race or other forms of destabilization. These risks are not unique to military AI, but they are particularly relevant in this context due to the high stakes and potential for harm involved in the use of AI for military purposes.

In this research, I would describe the epistemic and non-epistemic values applied on AI, regarding the national innovation expectations, and inductive risks of the technology, uncertainty and explainability, and especially the operational risk of military applications.

1. Power risks in emerging science and technology in view of national innovation system, Roles of value in choosing of defense in National Innovation Systems

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.

However, the values have legitimate roles in deciding how the government should influence research funding, and military science is highly prioritized. National states have high interest in emerging science and technology in view 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 that 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. The idea of value in defense-related innovation consistently spills over into the civilian economy, prioritizing the mission-based research. Military spending and weapon production correlates with national innovation rates, and this comes with geopolitical concepts of national S&T innovators to accelerate ARMS race with its competitors.

1. Inductive Risk of ARMS 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, that affects on accepting or rejecting scientific hypothesis. How much evidence science should demand before drawing conclusions about 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 in science questions.

Inductive risks comes with the question, what if we are using the wrong standard? According to Douglas(Douglas, 2000), The inductive risk plays an ineliminable role in the evaluation of scientific hypotheses. The statistical significance levels and decision of how to weigh evidence is responsible, for instance of false positive and false negative error. Both errors could have significant influence on military applied AI system.

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.

Both errors can be a significant risk in military applications, 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.

However, mission-oriented research could shifts the assessment of research products and potential from scientists to external factors. (Elzinga, 1997) Mission funding induces ‘epistemic drift’ which is part of a more general clash between different value systems and cultures.

The inductive risks of AI in military implications would refer to the potential negative consequences that could arise from the use of AI in military systems, as a result of the limitation of the technology itself. These risks could arise from the fact that AI systems are based on mathematical models and algorithms that are designed to make predictions and decisions based on data, and are therefore subject to certain inherent limitations. Some of the key inductive risks of AI in military systems include the potential for AI to make decisions that are biased or unfair, the possibility of AI systems being unable to cope with novel or unexpected situations, and the potential for AI to be used in ways that are unethical or even illegal. Additionally, the use of AI in military systems could raise concerns about accountability, responsibility, and the protection of human rights. The inductive risks of AI in military systems will depend on how the technology is used and the specific context in which it is deployed.

* 1. Uncertainty of data

The nature of AI is an explanation of optimization to a dataset. The data is composed three-fold as a rule: train, valid, and test data. The designed network model is trained to optimize train data, verified through valid data, and confirm model quality with test data.

From this 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 from data. The concerns of defense security make acquisition and quality control of the data hard, which leads to epistemic uncertainty.

* 1. Lack of explainability

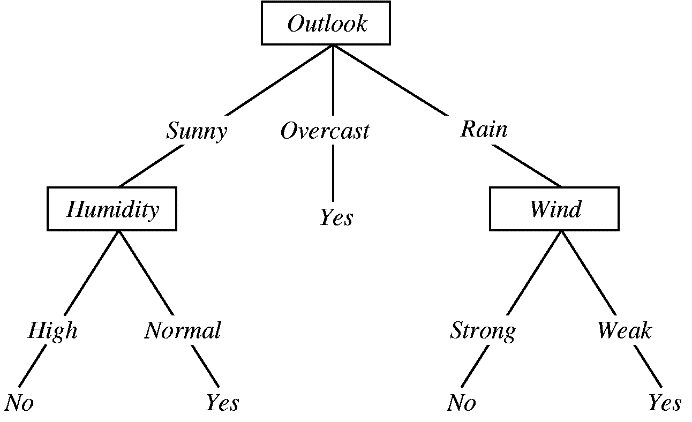
AI is often called a ‘black box’, the abstract concept in which input and output occur although the internal structure is unknown. The construction of multi-layered machine learning model evolves optimization procedures, as randomly applied initial values are gradually optimized through each layer of non-linear calculations. Deep learning, which often refers to models with more than 10 layers, can have more than hundred-thousands of optimized parameters. However, a major concern is that the results of machine learning models are often not easily reproducible and may be influenced by underreported model parameters or unique aspects of the training process.

The algorithmic opacity entail epistemic opacity in the context of deep ML models?

Multiple iterations of train help optimization, but does not explain its result.

In case of simple classifier AI, if the samples pass the classifier, the probability of the categories are calculated to determine which categories have higher probability. The confidence of the result is calculated from the probability, which makes the result uncertain. Decision tree model, on the contrary to machine learning, is logic-based model to explain the result.

all logical relevance to the proposed hypothesis since they can contribute neither to its support nor to its disconfirmation. No strict epistemic standard for how much evidence is enough to accept a theory, and the decision depends on the risks involved.



Logic based Decision tree model, contrast to Machine Learning

This would open it up for risk in its application in highly regulated or critical environments. Inductive risk is called for decisions, to identify how much evidence is sufficient to confirm or refute hypotheses that require non-epistemic values.

Because the construction of machine learning models is an estimation process based on empirical inference, the results of these models are always subject to error, as indicated by the presence of training error and test error.

* 1. Operational risks

Although current artificial intelligence technology has several internal philosophical problems, the operational risks rise from the reliability, fragility, and security of AI systems in the applications of military fields. Some of the key intentional risks of AI in military systems include the potential for AI to be used in ways that are unethical or even illegal, the possibility of AI systems malfunctioning or being hacked, and the potential for AI to be used as a tool of war by hostile actors.

Militaries are likely to use AI to assist with decision making. The widespread adoption of AI raises concerns about shifting decision-making away from humans, as AI cooperates military programs, humans play a lesser role. However, this concern does not seem to take place since AI is taking its role to converge information to assist commanders. In reality, human factors become more important in decision-making when using the technology. The members of the military can trust the technologies that they're given.

The problem rather arises from the internal technology as described, the transparency of the system is ambiguous with matters of data, and models. The increasing complexity provokes automation biases, and inability to understand the determination process.

Also, militaries are concerned about the attack on training machine learning such as a designated dataset that makes models trained on errors to give false results. Thus, the requirements of more evidence-based decisions are requested to verify the reliability.

As the weapon systems control lethal areas, reliability is excessively important in the field of defense. Military standards are one of the most highly reliable qualification systems. Test and evaluation processes developing for internal use, they'll be able to advocate for and socialize and normalize among other militaries around the world. This non-deterministic, non-linear, high-dimensional, and probabilistic method requires high cost of implementation, the traditional validation techniques are insufficient. The development of legislation and standardization would help to resolve the ambiguous issues on AI, not only in the military field, but also in academia.

1. Power risk

Technological losers created by emerging science and technology. The cultural losers are often called by indicating the religious or other cultural groups that have conflicts with ethical or normative values as science and technology progresses. Innovation will alter their ability to access or negatively affect the costs, risks, or benefit of an existing technology. This will also affect minorities.

* The role of Think tanks
* Case of compromises of UN

1. Military science as a demarcation problem

The demarcation problem refers to the challenge of defining a clear set of criteria that can be used to distinguish between scientific theories and non-scientific theories. Holman and Wilholt(2022) has proposed five types of demarcation strategies Some common approaches to the demarcation problem include defying science in terms of its methods, goals, or subject matter. The mission-oriented research

The New demarcation problem is certain non-epistemic values are well coordinated with expectations within and without the research community, and take place within a broader context.

Non-epistemic values can play a legitimate role in science has largely come to a close, “New demarcation problem”[[3]](#footnote-3), If we agree that social values play a necessary role in science, epistemic considerations are no longer sufficient to determine when science is good. That is: authoritative and trustworthy.

1. Reference ( Not completed & not in required format yet. Will edit soon!)

1. Elliott, Kevin C. (2017), A Tapestry of Values: An Introduction to Values in Science, New York: Oxford University Press,
2. Douglas, Heather. “Inductive Risk and Values in Science.” Philosophy of Science, vol. 67, no. 4, 2000, pp. 559–79.
3. 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>.
4. 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>.
5. 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.
6. 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>.
7. 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>.
8. 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>.
9. 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>.
10. 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>.
11. 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>.

1. <https://www.darpa.mil/news-events/2020-08-26> [↑](#footnote-ref-1)
2. <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf> [↑](#footnote-ref-2)
3. The new demarcation problem, Bennett Holman, Torsten Wilholt [↑](#footnote-ref-3)