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) has proposed ‘New demarcation problem’ as certain non-epistemic values are well coordinated with expectations within and without the research community, and take place within a broader context. As governments configure their role in national innovation systems, the development of science and technology has become a state goal. The mission-oriented or project-oriented science is taking place as a rebuttal of the value-free ideal of 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, 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 Defence Artificial Intelligence Strategy on June 2022,  that aims to modernize the armed forces based on AI.

As AI has potential to improve efficiency and accuracy, reduce human labor, improve sensing, enhance cyber security, and facilitate logistics, 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. Also, the implication on industries and military could raise ethical issues. 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. There are several cases of compromises on the usage of AI in warfare, so I am planning to look into the specific cases and its results.

1. Power risks in emerging science and technology in view of national innovation system
   1. Role 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 does not necessarily take into account 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.

Nationals 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. 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 spendings and defense-related innovations. The idea of value in defense-related innovation consistently spills over into the civilian economy, prioritizing the 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.

Military science has a role in the development of the national science and technology industry. Government lead initiatives and Firm lead initiatives. International ARMS races in China and the USA are ongoing competitions.

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

Non-epistemic values can play a legitimate role in science has largely come to a close, “New demarcation problem”, 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.

Power risk emerges from this value-laden developments

Mission-oriented research shifts the assessment of research products and potential from scientists to external factors. Mission funding induces ‘epistemic drift’ which is part of a more general clash between different value systems and cultures. The culture clash between the military and industry has led to a shift in a debate from spin-off to dual-use technologies.

For industry, the objective is economic gain, and in academic science it is the pursuit of knowledge.

This makes military related science and technology hard to be value-free. AI is no longer treated as an R&D project. Many industries are developing AI technology with great interest because it is easy toIn particular, military fields are interested in AI because it can play a big role in the development of unmanned weapon systems.

The case of AI could be different from mission-oriented fundings, because the technology itself flows from private firms to federal organizations.

How to frame?

1. Value free ideal in ARMS

Explanation of Hempel’s idea of inductive risk in emerging technology

Hempel 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.

Global nations have developed and deployed lethal autonomous weapon systems, artificial intelligence technologies to be applied to military weapons, joining the global competition to develop autonomous arms. In the field of defense, AI is playing a role in target identification and autonomous response, and decision making.

The acceleration of technological race has also made us reconsider norms and ethics to the implications of AI in warfare. The military race is being held with a geopolitical perspective based on national security, to consolidate power and technological advance of nations. Moreover, the widespread adoption of AI raises concerns several risks on the technology and its implications.

1. Inductive Risk of ARMS AI

Statistical problems (Douglas, 2000)

False positives

Inductive risks: what if we are using the wrong standard?

risk plays an ineliminable role in the evaluation of scientific hypotheses.

1. Inductive risk of AI technology

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.

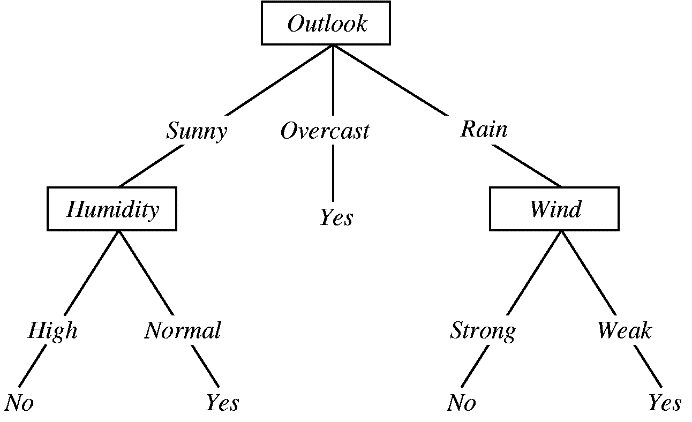
Apart from the data, epistemic uncertainty is caused by the model parameter, stating how confident the model is. In machine learning, the more data, the less uncertain the model is. However, the military samples have problems from data. The concerns of defense security make acquisition and quality control 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. AI lacks explainability. 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.

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

Forcing scientists to respond to a seemingly endless wave of unnecessary and unhelpful objections and demands.

Creating an atmosphere in which scientists fear to address certain topics and/or to defend hypotheses as forcefully as they believe is appropriate. This makes scientists fear to draw strong conclusions, and also makes them waste resources.

Inductive risks and the evaluation of dissents

Dissents in science be epistemically detrimental?

1. Operational risk of AI warfare

How can we resolve

1. Cases

1. Case of Project Maven (2018)

In 2018, Google employees protested the company’s involvement in Project Maven. The Pentagon contracted Project Maven that used Google’s artificial intelligence technology for military effort to develop drone surveillance footage.

The company argued that they could contribute to international knowledge development on AI beyond the geopolitical perspective. Also, they insisted the defense industry was still able to use this technology even if they did not participate in the project, because the project uses open source software released by Google.

The protests ended with around a dozen employee resignations, but Google did not renew the contract and announced international guidelines for future AI projects that forbid work on weapons and surveillance projects violating internationally accepted norms.

1. Compromise of UN ODA Office for Disarmament Affairs(August 2019)

As the AI ARMS race accelerates, the UN advocated against the deployment of technologies in the field. They found it is unlikely that a complete ban can be enforced, the best way forward is to define a set of broad guidelines for its deployment to secure the world. The UN stated the guidelines as follows.

“AI alone should never be allowed to make judgment calls in matters of arms. There should be human surveillance of its decisions before they are executed in the field. Persons entrusted with deploying AI must have a thorough knowledge of this tech. Humans should have sufficient oversight and the ability to disengage a malfunctioning system immediately.”

Risk of inductive error means that values must play a role in science. The diverse communities of stakeholders are starting to conduct robust dialogues as they endeavor to maximize the benefit of AI while mitigating the misapplication of this important technology. In the field of defense science, the funding source is the government. The groups of governments or associations of academia should take roles of ethics and philosophy.

1. Reference
2. Large-scale military funding induces culture clash

<https://www.sciencedirect.com/science/article/abs/pii/0265964690900623>

# The science-society contract in historical transformation: with special reference to “epistemic drift”

<https://journals.sagepub.com/doi/abs/10.1177/053901897036003002?casa_token=op4BTunIg9sAAAAA:KAvsd9V2RPnQdcqoomEdKodpRISF1Gk5G2zlskQPi96QfpYO_mP7dqJgEPOPurq_BZ7xSW7aZY4bVw>

# Looking beyond values: The legitimacy of social perspectives, opinions and interests in science

<https://link.springer.com/article/10.1007/s13194-022-00490-w>

1. Schmid, S., Riebe et al, Dual-Use and Trustworthy? A Mixed Methods Analysis of AI Diffusion Between Civilian and Defense R&D. Sci Eng Ethics 28, 12 (2022)
2. Sisson, M. et al, The Militarization of Artificial Intelligence, UN Report (2020)
3. Kendall, A. et al, What Uncertainties Do We Need in Bayesian Learning for Computer Vision?, 31st Converence on Neural Information Processing Systems (2017)