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

Yunji Woo

1. Introduction

AI is no longer treated as R&D project. Many industries are developing AI technology with great interest because it is easy to process complex information, collects multi-modal information, and learns algorithms based on the collected information to adjust new information. In particular, military fields are interested in AI because it can play a big role in the development of unmanned weapon systems. AI has potential to improve efficiency and accuracy, reduce human labor, improve sensing, enhance cyber security, and facilitate logistics.

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

The acceleration of technological race has also made reconsider norms and ethics to implications of AI in the warfare. The military race is being held with 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.

In this research, I would describe the internal risk of AI itself, regarding uncertainty and explainability, and the inductive risk of military applications. There are several cases of compromises on the usage of AI on warfare, so I am planning to look into the specific cases and its results.

1. Internal risk of AI
   1. Uncertainty of data

The nature of AI is an explanation of optimization to 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 large amount of data.

The types of aleatory uncertainty can be separated into homoscedatastic 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.

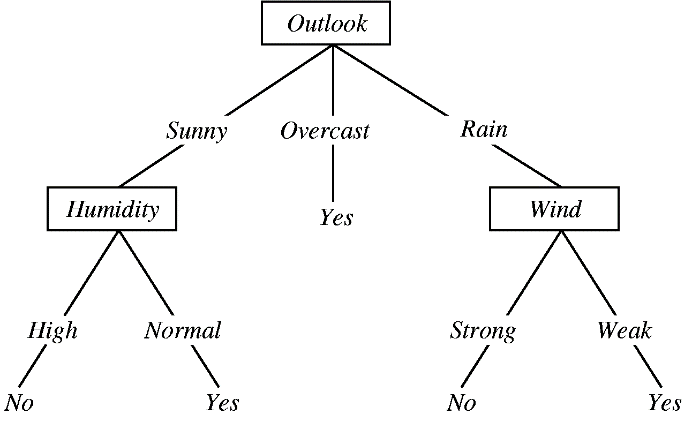
Apart from the data, epistemic uncertainty is caused by model parameter, stating how confident the model is. In machine learning, the more data, the less uncertain the model is. However, the military samples have problem from data. The concerns of defense security make acquisition and quality control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

* 1. Lack of explainability

AI is often called as ‘black box’, the abstract concept in which input and output occur although the internal structure is unknown. AI is lack of 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 has higher probability. The confident 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 enough to sufficient to confirm or refute hypotheses require non-epistemic values.

1. Operational risk of AI warfare

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.

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, human play a lesser role. However, this concern does not seem to take places since AI is taking its role to converge information to assist commander. 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 system is ambiguous with matter of data, and models. The increasing complexity provoke automation biases, and unable to understand the determination process.

Also, militaries are concerned about the attack on training machine learning such as designated dataset that makes model 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 controls lethal area, the 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 military field, but also in academia.

1. Conclusion
   1. Case of Project Maven (2018)

In 2018, Google employees protested on the company’s involvement in Project Maven. 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, 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. UN stated the guidelines as follows.

“AI alone should never be allowed to make judgement calls in matter 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. Human should have sufficient oversight and the ability to disengage a malfunctioning system immediately.”

Risk of inductive error meant 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. 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)
3. Sisson, M. et al, The Militarization of Artificial Intelligence, UN Report (2020)
4. Kendall, A. et al, What Uncertainties Do We Need in Bayesian Learning for Computer Vision?, 31st Converence on Neural Information Processing Systems (2017)