Theresan Military-Academic Forum

Perspectives on AI Security in Defense-critical Contexts

Author:

- **Clemens Kerschbaum** has a background in business law. During his PhD he specialized on Knowledge Management, focusing on different types of knowledge with particular emphasis on non-rational knowledge and its influence on strategic decisions. He published in different academic journals and presented his work at various international conferences.
- **Raphael Dachs** has a background in Applied Machine Learning and organizational AI design. Over the past ten years he has mainly been working together with governmental agencies across Europe in establishing AI expert teams and enabling them to produce AI products. His interest is in AI security and Adversarial AI for high-risk applications and in the organizational knowledge creation through the use of AI.

Abstract: With the release of Open AI's Chat GPT, AI has made its way into the mainstream. Yet, already before that, the technology showed game-changing potential for various fields. This includes, as to be expected, the defense industry and the use of AI in a military context. AI is not only transforming the functioning and interaction of weapon systems but changes the way battles are fought. These developments pose several important questions for military organizations: First, in order to understand the full potential of AI on the battlefield (beyond isolated applications in specific weapon systems), both decision makers and operative personnel requires a basic understanding of the functioning of AI. Second, as we already see in modern-day battlefields, we must ask how to defend against AI driven weapon systems. And third, if we use AI driven systems ourselves, we must ask how to develop systems that are robust enough to accomplish their goal under all operational circumstances, which raises the question of AI-security. Thus, with this contribution we aim to give an overview of the functioning of AI technology from a general perspective and shed light on different approaches to machine learning (supervised, unsupervised, and reinforcement learning) and the current state of AI technology (artificial narrow or general intelligence). After the general explanation we progress to give examples of the use of AI technology in military scenarios and how these systems already have been, or could be manipulated. We conclude by giving an outlook on our view of AI security and some considerations on how to create more safe, secure and robust AI systems, takin into consideration technology, its use case, and human factors.

Bottom-line-up-front: Just as most IT systems, Artificial Intelligence can be manipulated or attacked. Yet, AI-Security might differ from 'traditional' Cyber-Security due to the systems complexity and potential use. Hence, in order to develop robust AI systems, we ought to consider not only technology but especially the use case and the user. This articles explains in detail what we mean by that.

Problem statement: Through our ongoing work in the fields of AI, Machine Learning, and ML Ops, questions regularly arose about securing AI systems. With our contribution we aim to answer the question, how AI systems can be manipulated and what measures could be taken to counter such attacks. To this end we also briefly explain basic concepts with regards to the functioning of AI.

So what?: We argue that robust systems can only be possible when safety and security are thought in the context of the use case and the users. Thus, the better we understand AI, the safer we can use it. With this article we hope to give readers a basic understanding of AI alongside some examples of how such systems can be and have been manipulated in the military context.