

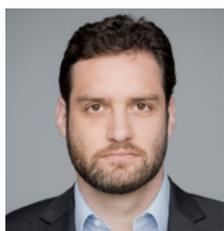


Killer Robots - A Psychologist's Personal Reflections on AI in Warfare



● **Marisa Tschopp**
scip ag · Research | Master of Arts

She has studied Organizational Psychology at the Ludwig-Maximilians-University in Munich. She is conducting research about Artificial Intelligence from a humanities perspective, focusing on psychological and ethical aspects.



● **Marc Ruef**
scip ag · Research

He is active in cybersecurity since the late 1990s and well-known for his many publications and books. The last one called «The Art of Penetration Testing» is discussing security testing in detail.

1. Preface

This paper was written in 2019 as part of a research project at scip AG, Switzerland. It was initially published online at <https://www.scip.ch/en/?!-abs.20191003> and is available in English and German. Providing our clients with innovative research for the information technology of the future is an essential part of our company culture.

2. Introduction

A few weeks ago, I met with a former colleague of mine, who had worked with me as a makeup artist in a beautiful hotel in Germany over 15 years ago. We were doing hand or facial treatments, where I even had the chance to paint the nails of no other than Alanis Morissette, who was giving a concert that day. The hotel closed many years ago and we talked about what we are doing now. She told me proudly that she has opened her own

beauty salon in her house, in the basement with great products, candles, flowers, and all that beauty stuff. She loves it. And you? She asked curiously. I said, I am doing research on Killer Robots. *End of conversation.*

Many people I tell about the topic of how AI and robotics are used in warfare, express reactions that fall somewhere between surprised, curious, scared or even irritated. I find these reactions come especially from moms, psychologists, business owners, or anyone else that is not working in the fields of computer sciences or the military.

Who am I to talk about killer robots? I am neither a computer expert nor an expert in politics, military or international law. I left the beauty industry to study business and psychology. I was working at a University when I attended a presentation by IBM, where I witnessed an AI, the Watson, defeating the world champion in the TV quiz show Jeopardy. I was fascinated, not only by the skills of the Watson, but the fear and resistance I saw in the eyes of other people.

That day I decided to take a deeper look into this bizarre fascination that surrounding artificial intelligence and robotics. Moreover, as a mother of two, I wanted to know how this technology will impact the future of my children. I envision a future where they are safe and protected. Where technology can advance alongside human values. When I first heard about AI in weapons, I was deeply disturbed, as this seemed to inflict on the deepest level with my need for peace and harmony. These so-called killer robots are machines that can detect, select and destroy a target (human or any other object) without any human oversight. But there was also this bizarre attraction and curiosity, which was proof enough, that this is not only worth exploring, but further, where I get the chance to take concrete action on the legacy I want to leave behind for future generations.

***It is not just about caring.
In this case it is rather daring to care.***

But I probably not the only one to ask: Who am I to talk about this? The classic imposter syndrome, the fear of failure holds back to learn something new. Women, especially, are extremely reluctant to step up, if they do not feel like experts, because this is way outside the comfortzone. But apparently, learning begins outside the comfort zone. So why not dare to try and see if it is worth it?

Once I followed a Twitter conversation and one autonomous weapons expert said, that the topic is not very complicated – everyone can discuss this! Then proceeded to cats. At that time, I felt quite ashamed that for me, it was hard to understand. Later I understood, the intention was to lower the threshold for non-experts to enter the discussion.

However, downplay can backlash in a great way, even if we intend to help. The same accounts for threatening statements like shocking videos proclaiming a dark future. It creates reactance, a psychological phenomenon, which occurs if cognitive freedom of choice is threatened. It creates turbulence in the mind, which leads to lethargy and denial. The opposite of what is needed. We must find additional ways to communicate. There are more effective, authentic ways, integrating empathy for self (I am unsure how to talk about this) and others (I want to understand, how you feel about this). Instead of downplaying or threat, one could try saying: I am already so used to this, that I forgot to acknowledge, that this is not as easy to understand, and one can be easily overwhelmed. How can we find a common ground that we both feel comfortable discussing this?



Many people feel very uncomfortable with the moral finger topics. Topics like war, climate crisis, vaccination, and more, you can barely mention without pointing fingers. Who does not feel guilty when you would rather spend your day talking about cats and croissants? Activists and politics are good at pointing fingers (this is an inherent part of their task!), but isn't there a point that we are missing?

Pointing fingers, is missing the point.

One of the most frequently asked questions I get is: Don't you get depressed when seeing this war stuff? Yes, I do. However, I understood that everybody serves with his own gifts and talents, whether it is cats, or politics, or both. Or neither of them.

To answer the question a bit deeper, yes, at times I feel helpless in this situation. However, it is not the power of weapons that frighten me. Rather it is the cruelty of war, the human-made horrors, that leave me in despair. Take for example the Abu Ghraib Scandal, where US Soldiers have cruelly tortured the inmates. But now, I understand, why people would like to replace humans with robots. One cannot avoid asking: Could these horrors have been prevented?

Machines do not see the world as humans do. They don't know stress or fear. They do not have the desire to dominate or thirst for revenge. What if a machine would take over in critical situations, untouched by human hysteria? Although, for me, something feels inherently wrong about weapons, and moreover robots with weapons that can kill, I could not stop myself from thinking: What if? What if, that would be a better way to solve armed conflicts? What if, it would help to kill only the bad guys? What if, less soldiers must be sent to war to die and leave their friends and families behind?

What if, AI would make war more humane?

To answer this question, one must understand how Artificial Intelligence, and war works. How could you approach such complex and sensitive topics? To understand the hows and whys AI in warfare, I found it very useful to look at the OODA-Loop, which is a military concept of how decisions are being made. It helps because it operates on the level of the individual, which makes it easier to take another person's perspective, rather than trying to understand a system. Bear in mind, it is about defense, where the military aims to defend their country and the civilian population. This idea is worthy of support, because there are soldiers in war zones who must be protected, right? Different technologies can help protect these soldiers to act more accurately, more effectively: Technological enhancement in warfare is about being faster than an attacking opponent. Military strategist, John Boyd, developed the OODA Loop named above. The beauty of this OODA loop lies in its simplicity and intuitiveness. It helps to understand (1)

human behavior and decision making, as well as (2) the concepts of Artificial Intelligence along the autonomy continuum, and last but not least, the different possibilities for how humans and machines interact (3), the human in, on or off the loop.

OODA stands for Observe (There is something, a stimulus), Orient (What is it?), Decide (How to react?) and Act (Do something, a response). From a psychological perspective, it relates to the classic stimulus-response pattern.

The concept of information processing in a computer is quite similar to this stimulus-response pattern. A computer reacts to a certain stimulus through an input and reacts in a certain way. The most exciting difference of both lies in the black box, the processes that happen between stimulus and response. Whereas the brain has always been extraordinary, the computers' processing units have become increasingly extraordinary within the past 30 to 50 years. Simply observing does not make a computer intelligent, it is just a camera. But if a machine has a processing unit in addition to the sensor, where objects can be detected based on machine learning, then we can talk about intelligence.

However, it gets interesting, even dangerous, when even the experts do not really understand what is happening in this computer black box. Some world-renowned experts claim that there is a realistic chance that humanity might lose control over these intelligent systems. Especially, if they can execute decisions without any human oversight. This can mean to kill another human being, theoretically a foe, not a friend.

To go a level deeper, the OODA Loop not only gives you an understanding on human decision making and information processing, it also illustrates the dimensions of autonomy and human-machine interaction.

Autonomy, what it is, and when it is achieved lacks universal definitions. The scale of autonomy can be seen on a continuum from automatic (not intelligent), to automated to autonomous (intelligent). The further it moves towards autonomous the more intelligent is a machine. For us nontech people it gets more scary, because we less and less understand what's happening in this computer. This is one reason, why experts are working hard on transparent, explainable AI to increase trust and presumably, likelihood of use.³

What is simply algebra and probability to others, is kind of magic to many.

The magic fades, if we understand AI fully, which is unlikely to happen and very impractical. We use normal cars, but most do not know how they work. But what we must know is, how to use a car safely. The more technology takes over, the more we are becoming supervisors. Are we underestimating how much it takes to be in this supervisor role?

Nowadays, when it comes to war, and autonomy in weapons, there is a human in the loop somewhere to supervise what the program is doing. A security net, so to speak, so responsibility is not blindly delegated away. But, it is still unclear where exactly this trustworthy human in the loop is, we need so badly?

Augmented reality glasses for example, can enhance human sight and support decisions making processes, based on more data than the human can possibly process or simply cannot see (like data from drones or people behind a thick wall). The stimulus and information processes are enhanced (presumably, considering the error rate), but the execution is within the human.

Maybe AI could make war not only more efficient but also more humane, because computers do not have emotions, like fear or hysteria, all steady companions in warfare. Machines can do the dull work without getting tired and can be sent out unscrupulously to do the dirty and dangerous work. The vision is strong and emotional, namely, to replace soldiers on the battlefield in order to save lives. For the same noble reason, Richard Gatling, invented the automatic gun: Create stronger, faster weapons to send less soldiers to war to die for, but it resulted in more blood spilt than ever before. Landmines were built to replace soldiers on the battlefields, but it resulted in over 20000 killings per year, often innocent women and children, before landmines were banned in 1997. Chemical, biological and nuclear weapons are banned. Blinding lasers were banned, before they even ever saw a battlefield. This same sort of preemptive ban of lethal autonomous weapons has been called out by experts for many years. However, no such treaty has been established so far. Among others, the US, UK, and China are investing vast amounts of money in research and development of AI in weapons. AI seems to offer the long-expected panacea to eradicate human weaknesses. We are outsourcing responsibility, even if we pretended to be in the loop; however, these machines may act so fast, that humans are not even close to supervising a decision.

The idea that AI makes war more humane is an illusion.

It makes conflict even more messy. We are overestimating not only the machines abilities to make a decision, but also our own abilities to build or supervise such a machine. Let alone the fact, that no machine should be the one taking a live away, a line we must never cross. But let's be fair: The desire to use AI in warfare may not be born out of the desire to kill, it is still the idea of inventing something to save human lives. In the end, every human being (excluding severe mental illness), wants peace and security, we just choose different strategies for how to get there. But lethal autonomous weapons is a very dangerous strategy to achieve goals.

If we want to make conflicts more humane, we must turn to humans, not to machines.

We live in interesting times, change is what we need on an systems level, but also on a personal level. These days of deciding, how we want to use technology for the benefit of humankind are ours and the time is now. All of us, but especially us psychologists, have the unique skills to help solving dilemmas of disharmony, before they escalate. No matter if you are a make-up artist in your basement or the President of the United States, you can be the human in the loop.

2.1. References

- Arkin, R. (2010) The Case for Ethical Autonomy in Unmanned Systems (gatech)
- Mohr, T. (2014) Why women Don't Apply for Jobs Unless They're 100% Qualified (hbr)
- Noone, G., & Noone, D. (2015). The Debate Over Autonomous Weapons Systems. (CWR)
- Scharre, P. (2018). Army of None: Autonomous Weapons and the Future of War. (WWN)
- Sparrow, R. (2016). Robots and Respect: Assessing the Case Against Autonomous Weapons (Cam)
- Zimbardo, P., Gerrig, R., & Graf, R. (2008). Psychologie. (Pea)

