Good AI, Bad AI - Psychological Aspects of a Dual-Use Technology

Article · October 2019

Marisa Tschopp
scip ag

Marc Ruef
scip AG

Good AI, Bad AI - Psychological Aspects of a Dual-Use Technology

Marisa Tschopp
Research Department, scip AG
mats@scip.ch
https://www.scip.ch

Marc Ruef (Editor)
Research Department, scip AG
maru@scip.ch
https://www.scip.ch

Abstract: AI is a dual-use technology that can be used for both civilian and military purposes or alternatively with good or bad intentions. Tech giants are in a precarious position where their technology is being used more and more for military purposes, against the values of users and employees. Responsibility grows on all sides, behavioral alternatives become more and more complex. Anxiety and feeling overwhelmed lead to lethargy and paralysis within the affected people and make communication and motivation for action more difficult. Communication about AI and its implications must be moved away from threats and fear mongering, to facts, visions and clear reasons why.

Keywords: AD, AI, Amazon, Artificial Intelligence, Detect, Drone, EPFL, Fake News, Gender, Google

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/?labs.20190411* 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

With *Hololens for Good, not for War* [1], several hundred Microsoft employees are raising their voices against use of the *Hololens* for military purposes, with no effect so far. Thousands of Google employees protested against *Project Maven* [2] in 2018, using the same open letter strategy: The employees do not want to be profiteers of war. The consequence of the Google protest was the dissolution of the contract with the US Department of Defense. The dual-use dilemma arouses public interest: Can we ensure that research and technology will be used exclusively for the benefit of humanity?

The publicized cases of the *Hololens* or *Project Maven*, show impressively on what thin ice tech companies are moving, especially those giants who are in the limelight, such as Microsoft or Google. While it is not enough to become a big scandal, these two cases are again causing people to think about the implications of *dual-use technologies* [3] . These are used not only commercially but also in the military, or are used with good or bad intentions (Brundage et al, 2018). These problems have not existed only since AI but have a long tradition and are firmly anchored in the chemical or engineering sciences. Rocket technology (good intent: Moon landing, evil intent: Offensive weapon) or nuclear technology (mostly good intent: Energy, evil intent: Nuclear bombs) are prime examples.

> *Right now: Technology, computing power, programming languages, computer code, encryption, information, big data, algorithms, investment, they all have dual use. Forbes, 2019 [4]*

3. Main aspects of ethical debates

AI and its related technologies and sub-technologies together form a volatile mix that positively transforms almost every sector from medicine to urban planning but also brings questionable or even dangerous implications with it. From super-precise hacking of data platforms to the surveillance state and loss of privacy without opportunities for public consent. The biggest *ethical concerns in the technological context* [5] according to Vallor et al. of the Santa Clara Universiy are (2018):

- Declining transparency of technological processes
- Algorithmic bias and injustice (such as discrimination)
- Reduced control over data
- Manipulation of the human psyche (e.g. in political elections)
- Creating a tech monoculture (e.g. humanity without high technology is not possible)
- Monopolies of tech companies
- Surveillance society (e.g. facial or motion detection)
- Attention Economy (e.g. media addiction)
- Watering down human responsibility and control
- Digital Taylorism (e.g. workplace control)
- Declining social trust (e.g. because of fake news)
- Environmental concerns (e.g. e-waste)
- Democracy vs. Technocracy (e.g. better technology is always the solution)

4. A Code of Ethics is only part of the solution

Unfortunately, there is no such thing as a patent solution for dealing with AI properly. As a rule, a case-by-case analysis of the ethical dilemma is necessary. *Dual-Use* [6] always poses the question of whether research, technology or the final product poses a specific threat or a high enough risk to be used for malicious purposes. This raises the question of whether they can endanger health or safety of humans, even if malicious outcomes were never wanted nor the primary goal. Not only does this dilemma confront developers with the choice of developing a product that they might help others with but which definitely also enables malicious people to do evil. Does that make them accomplices? Or the institution, the investors or the company? Does this justify the project? Does the potential benefit outweigh the foreseeable danger? (Brazzetti et al, 2018)

Often, potential negative consequences arise only in the course of research and development, and were not previously anticipated, which is to say, there was at no time the intention to develop something that causes harm to humans. Expressed in the jargon of *Google Code of Ethics* [7] there two important distinctions: *Don't be evil* (intentional negative consequences) and *Don't provide means to do evil* (unintended negative consequences). As an example one can look at the decision of *OpenAI to not publish their program* [8] because it carried too much risk of abuse, for example for generating fake news.

5. Technology on the battlefield – tech giants are war profiteers

In the military context, the dilemma is more difficult when it comes to research and development for technologies and weapons. For offensive military research, the case is relatively clear. This is ethically unacceptable because the primary purpose is to harm people. The bigger dilemma is defensive military research, the development of defensive weapon systems, which basically serve the public good. Lastly, it is questionable when technologies are developed that have civilian, commercial as well as military applications, since now economic conflicts of interest also come into play (Brazzetti et al. 2018).

Collaboration with the US Department of Defense is lucrative. They see a need for sophisticated technologies to better defend the country, its borders, its soldiers and the civilian population. This is largely an idea that is worthy of support because there are currently soldiers in war zones who must be protected. Looking at decision-making processes, different technologies can help act more accurately, faster and more effectively. Talking about defense, it is about being faster than an attacking opponent. Military strategist John Boyd was the inventor of the *OODA Loop*, an information and decision-making concept. The following graphic illustrates this process and serves as a basis for thinking about which phases various technologies should be used in to increase the speed.

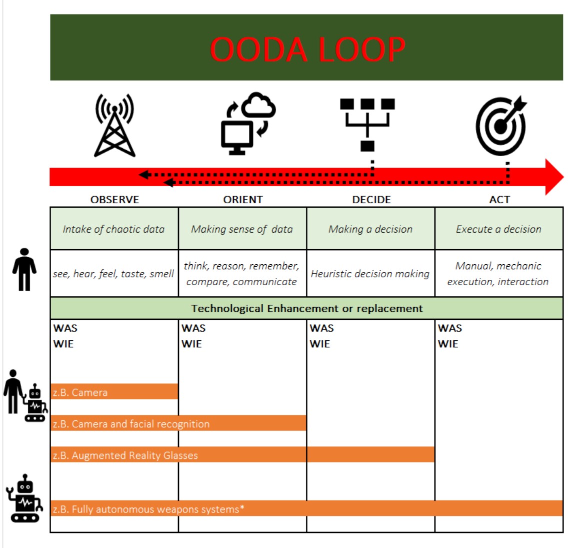
OODA LOOP
OBSERVE
ORIENT
DECIDE
ACT
Intake of chaotic data
Making sense of data
Making a decision
Execute a decision
see, hear, feel, taste, smell
think, reason, remember, compare, communicate
Heuristic decision making
Manual, mechanic execution, interaction
Technological Enhancement or replacement
WAS
WIE
WAS
WIE
WAS
WIE
WAS
WIE
z.B. Camera
z.B. Camera and facial recognition
z.B. Augmented Reality Glasses
z.B. Fully autonomous weapons systems*

Figure: The OODA Loop: Decision-making cycle first described by military strategist J. Boyd

The whole human information processing system, but also the final execution of a decision, can be supported or even completely replaced by technology (through software and hardware). The separation into the respective areas is not quite as strict as illustrated in the picture, especially since the components of course depend on each other and cannot be considered by themselves. Nevertheless, this division certainly makes sense to shed more light on the critical functions. The crucial part at the level of international humanitarian law is the last one that represents the actual *act* of executing a decision. Once this is no longer performed by a human, the system can be considered a lethal autonomous weapon. Although all tools used currently are theoretically legitimate, the situation is still somewhat blurred. Either because the decision-making process is so fast that no more meaningful human control can be maintained or because it is unclear how good the data is that has led to the human decision (especially considering potential defects and manipulation). Thus, the seemingly legitimate hololens, with its individually built algorithms might move into a gray area.

5.1. The Microsoft Case: What is the Hololens Dilemma about? (2018/19)

Microsoft's second-generation *Hololens* is a augmented reality device that can visualize human anatomy or parts of a motorcycle using *augmented reality* (AR). Among other things, artificial intelligence in form of neural networks can be used to process sensor material. The fact that Microsoft is using this technology to help soldiers kill more effectively upsets some employees and ethicists. One reason, because there was no common agreement upon how the technology will be used.

Microsoft employees demand to stop the 480 Million Dollar project with an *open letter to the CEOs* [9]:

> *We are a global coalition of Microsoft workers, and we refuse to create technology for warfare and oppression. [...] We did not sign up to develop weapons, and we demand a say in how our work is used. [...] The contract's stated objective is to rapidly develop, test, and manufacture a single platform that Soldiers can use to Fight, Rehearse, and Train that provides increased lethality, mobility, and situational awareness necessary to achieve overmatch against [...] our adversaries. [...] (it) works by turning warfare into a simulated video game [...]. Brad Smith's (CEO of Microsoft) (suggests that employees) would be allowed to move to other work within the company [...]. (This) ignores the problem that workers are not properly informed of the use of their work.[...] As employees and shareholders we do not want to become war profiteers.[...]*

The key points of their problems are:

- Lack of transparency
- Insufficient alternatives
- No possibilities for a compromise
- Contradictory, different, arbitrary alignment of corporate culture and values
- Clear demand: No cooperation with the American military

The company clearly backs the American government and sticks to its commitment, as Microsoft does not want to deny its homeland, with its democratic values, access to the best possible technology. The ability to work on another project is entirely pointless as that does not address the needs of the employees at all. The only concession is the endorsement of an open dialogue – but a dialogue with hardened fronts will be a tough conversation. Involvement of the public, NGOs and/or mediators is absolutely necessary in order to find solutions together.

In this video one thing is clear: It is about *speed*. Second, it is about expanding the context, for example using data from drones to make better decisions based on more data. This is reflected in the aforementioned OODA loop: Whoever comes to a decision faster *wins*. What *better* decisions are, is too large a discussion for this space. A visual system, such as augmented reality glasses, is only brought to life by the provided algorithms. The Machine Learning (here used as a generic term) provider is freely selectable. Thus, the OODA loop can be influenced accordingly in the phase *Observe*, *Orient* and *Decide*. Person in charge is still the human, who is influenced in his decision, but ultimately still gives the command to execute. The responsibility within this *human in the loop* process is pretty clear from a *humanitarian* [10] point of view.

The more and more complicated technology used in war, the more fuzzy responsibility becomes, which causes worries worldwide. The call for clear rules is loud, but politics are stuck arguing about definitions for years. The general public, even the employees of the involved companies, do not get much information because the whole topic is understandably extremely sensitive. War in general, but in particular in connection with AI and killer robots is very sensitive and therefore Google in *Project Maven* did not wish to talk about it openly.

5.2. The Google Case: What is the Maven Dilemma about? (2018)

Thousands of Google employees demand and end of the 7 – 30 (numbers vary) million dollar Pentagon contract with an *open letter* [11]. 10 – 20 employees left their job because of this affair:

> *We believe that Google should not be in the business of war. Therefore we ask that Project Maven be cancelled [...]. Google is implementing Project Maven, a customized AI surveillance engine that uses Wide Area Motion Imagery data captured by US Government drones to detect vehicles and other objects, track their motions, and provide results to the Department of Defense.[...] By entering into this contract, Google will join the ranks of companies like Palantir, Raytheon, and General Dynamics. The argument that other firms, like Microsoft and Amazon, are also participating doesn't make this any less risky for Google. Google's unique history, its motto Don't Be Evil, and its direct reach into the lives of billions of users set it apart.[...] Building this technology to assist the US Government in military surveillance – and potentially lethal outcomes – is not acceptable.*

The key points here clearly emphasize the distinctive and famous corporate culture of Google, which, despite the cases of discrimination, is a great asset for current as well as potential employees. As with Microsoft, the employees involved are fundamentally liberal in their ideas and do not want to be profiteers of the war. The whole story gets worse as Google was trying to not publish anything, yet emails found their ways into the public.

Both cases are of course very different and difficult to compare if you don't know the internal details. This is not to compare right from wrong rather it serves as a basis to think about how we want to use technology. According to Dr. *Frank Sauer* [12], Senior Researcher at the *Bundeswehr University* in Munich, the discussion about killer robots can seem strange and exotic. Yet he sees it as part of the bigger picture. It is one aspect of the big question about the future relationship between humans and machines, prompting humanity to answer not if but how it wants to deal with science and technology.

6. Understanding, Listening, Discussing: What is it really all about?

Microsoft employees demand a clear exit like Google: But does this exit strategy really solve the problem? According to *Paul Scharre* [13], Senior Fellow and Director of the Technology and National Security Program in the Center for a New American Security, the provider is simply replaced and someone else will do the job. In worst case, maybe someone with less know-how and skills than the big players. Thus, the ethical dilemma is perfect.

On a philosophical note, *Megan Welle* [14], a philosopher in a machine learning company, adds the following idea, which again, takes this whole issue further:

> *Microsoft decided to work with the US Military was because it supports the defense of the US and wants the people who defend it to have access to the nation's best technology, including from Microsoft. The underlying ethical stance is that companies have a moral obligation to maintain loyalty to the defense of the country of their origin. It's important to note that Microsoft is HQ'ed in the US, although the CEO is originally from India. Microsoft supported their position by indirectly implying that if US companies don't develop this technology for the US, it will be developed and used elsewhere. This unspoken fear perpetuates throughout the military and business world. In the light of Google's resignation from Project Maven, Google has been accused of indirectly helping China's military.*

An extensive stakeholder analysis is needed, in which theoretically not only the needs of all involved have to be included, but also those who are affected, and that is basically the whole civilization:

- What is important to employees? e.g. Participation, genuine decision-making options, transparency, authenticity in dealing with the company values …
- What is important to the CEO? e.g. Ensuring financial security through innovation, return on investment, business viability, patriotism, military security, support for democratic values worldwide …
- What is important to the state / military? e.g. Physical security through progress and innovation, life support, recognition from the people, differentiation from other ideologies …
- What is important to the population? e.g. Freedom, life support, identity, community, security …

So what is this about? Money? Power? Transparency? Democracy? Enlightenment? World peace? Almost no one knows this from the other, often not even from himself. There is a path that is characterized by empathy, cognitive understanding and authentic communication.

- The *authentic* willingness and openness to be free of judgment to deal with oneself and others
- The willingness through *education and facts* to make a picture of the overall situation that is as accurate as possible
- The willingness to talk to each other and to seriously understand conflicting opinions

The big problem today is that they are difficult issues that are difficult to communicate and difficult to hear without feeling patronized. There are thousands of topics, such as climate change, famine, gender equality, civil wars, mass migration, rainforest deforestation and animals going extinct to cope with. The general public is overwhelmed and annoyed, possibly scared and feels powerless. Nationalism and othering, lethargy and paralysis are the logical consequences. These defensive behaviors were described as early as 1966 by J.W. Brehm. As he puts it, once people perceive a restriction of their will and agency, they shut down and may choose behaviors that are completely irrational even to themselves, only to restore the cognitive freedom of choice and the resulting empowerment of the self.

The moral of the story is clear: There must be a change in how critical issues are communicated e.g. AI or war, or AI at war. Away from threats and hype to facts and visions. But how? Important findings come from consumer and advertising psychology. A study on donation behavior has shown that people were more willing to donate when the ad shows a happy child on the way to school than when the child was hospitalized (Zimbardo & Gerrig, 2004, Kroeber-Riel et al., 2009). Shock pictures and videos attract attention but do not allow dialogue. However, exactly this is so urgently needed.

7. Conclusion

These two cases show in a great way how technological progress individualizes and complicates life and that there is no one true solution. From a psychological perspective, war lies at the intersection of aggression and prejudice, triggered by religious, ethical or cultural differences. It is important to understand that evil deeds are not necessarily committed by evil people. This banality of evil has repeatedly been confirmed by psychological research: Every human being is capable of doing evil. The following situational factors are much more likely to lead to forms of highly organized aggression: Difficult living conditions, political upheaval, scarce resources, charismatic leaders. Research has shown that verbal resistance is still expressed but what is lacking are easier behavioral exit strategies. Verbal resistance without change on the behavioral level often does not bring the desired effect (cf. Zimbardo & Gerrig, 2004). It is important to follow these two cases carefully, as important insights can be drawn from them. In the Google case, the actual resignations probably caused the exit decision. Whether this was a better decision than supporting Microsoft's decision to support the military remains to be seen.

Of course, it can never be guaranteed that research and technology will not fall into the wrong hands. Violence and aggression will always be part of humanity. Yet, the capacity for highly organized violence in the form of war is as much in the hands of humans as is the capacity for highly organized conflict resolution in the form of peace.

7.1. References

- Barazzetti, G., Diezi, J. & Benaroyo, L. (2018). A brief introduction to dual-use in engingeering sciences. EPFL Research Office
- Brundage, M. Avin, S., Clark, et al., (2018). The malicious use of artifcial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228
- Kroeber-Riel, W., Weinberg, P., & Göppel-Klein, A. (2009). Konsumentenverhalten (9th ed.). München: Vahlen.
- Scharre, P. (2018). Army of None: Autonomous Weapons and the Future of War. W. W. Norton & Company

- Vallor, S., Green, B. & Raicu, I. (2018). Ethics in Technology Practice: An Overview. The Markkula Center for Applied Ethics at Santa Clara University
- Zimbardo, P., Gerrig, R., & Graf, R. (2008). Psychologie. München: Pearson Education.

8. External Links

[1] https://twitter.com/MsWorkers4/status/1099066343523930112

[2] https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/

[3] https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

[4] https://www.forbes.com/sites/cognitiveworld/2019/01/07/the-dual-use-dilemma-of-artificial-intelligence/#7560ecff6cf0

[5] https://www.scu.edu/ethics-in-technology-practice/overview-of-ethics-in-tech-practice/

[6] https://tto.epfl.ch/wp-content/uploads/2018/09/Web-page-dual-use-version-3.1.2018-and-briel-introduction-.pdf

[7] https://abc.xyz/investor/other/google-code-of-conduct/

[8] https://www.forbes.com/sites/nicolemartin1/2019/02/19/new-ai-development-so-advanced-its-too-dangerous-to-release-says-scientists/#7d77en994a80

[9] https://twitter.com/MsWorkers4/status/1099066343523930112

[10] https://blogs.icrc.org/law-and-policy/2018/04/03/autonomous-weapon-systems-ethical-basis-human-control/

[11] https://static01.nyt.com/files/2018/technology/googleletter.pdf

[12] https://www.unibw.de/politikwissenschaft/professuren/lehrstuhl-ip/sauer/dr-frank-sauer

[13] https://www.cnas.org/people/paul-scharre

[14] https://www.craftinity.com/