Conference Papers as in the Abstract Booklet
Responsible Innovation: Values and Valorisation (21 – 22 May 2014)

Designing for Responsibility - Five Desiderata of Military Robots

Maaike Harbers, Thomas Baar, Marieke Peeters

Introduction

Recently, the use of military robots – which may be drones, unmanned aerial vehicles (UAVs), remotely piloted systems, autonomous weapon systems or 'killer robots' – has been debated in the media, politics and academia. Military robots are increasingly automated, which means that they can perform tasks with decreased human involvement. On the one hand, this may lead to faster and better outcomes (Arkin, 2009), but on the other hand, it raises the concern 'Who is responsible for the (failed) actions of military robots?' (Matthias, 2004; Sparrow, 2007; Lucas, 2011).The issue becomes particularly stringent in the prospect of a future in which armies may deploy military robots that apply lethal force without human interference.

In this abstract, we approach the responsibility question from an engineering perspective, and suggest a solution that lies in the design of military robots. First, we would like to make a distinction between legal and moral responsibility (Noorman & Johnson, 2014). Legally, the person or organization deploying military robots, i.e. the army here, is responsible for their behavior, rather than the designer, programmer, manufacturer or the robot itself. The army's legal responsibility, however, does not imply that it is in the position to take moral responsibility. In accordance with the Value Sensitive Design approach (Friedman et al., 2006; Van den Hoven, 2007), we argue that the way technology is designed affects moral responsibility. For instance, most people will agree that in principle the person firing a gun, and not the manufacturer or the gun itself, should be held responsible for the consequences of a shot. In this case, the gun's design supports moral responsibility. Acting responsible is harder, however, when you rely on a decision support system that is incomprehensible, or when you have to use a weapon that may fire accidentally. In these examples, the system's design hinders moral responsibility. A gap between moral and legal responsibility is undesired. We therefore argue that military robots should be designed such that the army is in the position to take moral responsibility for the behavior of military robots. In other words, we have to design for responsibility.

Five Desiderata of Military Robots

Humans can perform a wide range of tasks, but they have limited cognitive capacities. Automation, instead, computes faster and does not get tired, but can usually only perform one or a few tasks, often under very specific conditions. Humans and robots that work together in a team can strengthen each other's capabilities (Rathje, 2013), e.g. robots performing simple, repetitious or computationally complex tasks allow humans to focus on tasks requiring creativity or moral decision making. Thus, we believe that human-robot teamwork can support responsible human behavior, and that to design for responsibility, we have to design robots that are good 'team players'. There are multiple overviews with characteristics of good team players (Klein et al., 2004; Bradshaw et al., 2011). Here, we will present five of them and explain how these support responsible human behavior. We use the term 'operator' to refer to the responsible representative of the army. In reality, the decision of an operator may be his/her own or those of a superior, depending on the army's line of command.

- Directable. The operator should be able to direct the robot's behavior at any time, where behavior is not only constituted by the robot's actions, but also by its sensing and reasoning activities. This makes it possible for the operator to initiate new tasks and actions, and to adjust or change previous instructions if circumstances require. For example, the operator can command the robot to explore a new area. The characteristic of directability makes it possible for humans to enact control over the robot, this allows them to take responsibility for the team's actions.
- Predictable. The robot should be predictable in the sense that it is dependable and can be counted on to work, and that it follows instructions and performs tasks in a

predictable way. For instance, if the operator instructs the robot to go to a certain location, he/she should be able to rely on the robot to go there in an effective way. If the robot's behavior would involve a lot of randomness, a human has less control over the robot and it is harder to take responsibility for its actions. Thus, predictability is important for responsibility.

- Transparent. The robot should be transparent with respect to its state, which includes its current physical and informational state, the status of the tasks it is performing, and its physical and cognitive capabilities and limits. The robot's transparency contributes to its predictability, and facilitates direction of the robot. For instance, if you know that a robot needs power and has the intention to return to its base, your can direct the robot to record its surrounding on the way back. Thus, the better an operator knows what is going on, the better his/her position to take responsibility.
- Understandable. Besides being transparent, the robot should communicate and provide information in such a way that it is understandable for humans. A robot's internal representation of its state, knowledge and capabilities is not always accessible for humans. For instance, an operator gains more insight in a robot's functioning when it shows knowledge-based reasoning rules rather than a neural network. Thus, sometimes a translation of the robot's internal representations to information understandable for human is needed.
- Selective. Most robots collect and produce more data than a human can process in real time. To avoid cognitive overload and keep the data stream manageable for the human operator, the robot should be selective. Robots can be selective, for instance, by aggregating data and focus the operator's attention on important information. While the robot supports the operator by proactively selecting information, the operator should be able to direct the robot regarding the information it provides and thus access any information that is available.

Conclusion

The five characteristics presented above aim to support the moral responsibility of humans in a human-robot team. The characteristics are rather general, and would be beneficial to any human-robot team. A more specific way to support responsibility of the army would be for example to equip military robots with moral decision-making support (see e.g., De Greef & Leveringhaus, 2013). From a broader perspective, parties other than the army bear responsibility for the behavior of military robots as well. For instance, during the design process, designers have the responsibility to consider all possible consequences of the military robots they design, both for direct and indirect stakeholders (Friedman et al., 2006).

References

Arkin, R. (2009). Governing lethal behavior in autonomous robots. CRC Press.

Bradshaw, J. M., Feltovich, P., & Johnson, M. (2011). Human-agent interaction.Handbook of Human-Machine Interaction, 283-302.

De Greef, T., & Leveringhaus, A. (2013). An ethical boundary agent to prevent the abdication of responsibility in combat systems. In Proceedings of the 31st European Conference on Cognitive Ergonomics (p. 18). ACM.

Friedman, B., Kahn Jr, P. H., & Borning, A. (2006). Value sensitive design and information systems. Human-computer interaction in management information systems: Foundations, 5, 348-372.

Klein, G., Hoffman, R. R., Feltovich, P. J., Woods, D. D., & Bradshaw, J. M. (2004). Ten challenges for making automation a" team player" in joint human-agent activity. IEEE Intelligent Systems, 19(6), 91-95.

Lucas Jr, G. R. (2011). Industrial Challenges of Military Robotics. Journal of Military Ethics, 10(4), 274-295.

M. Aaronson & A. Johnson (eds.), Hitting the Target: How New Capabilities are shaping international intervention (London: Royal United Services Institute, 2013).

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information Technology, 6(3), 175–183.

Noorman, M., & Johnson, D. G. (2014). Negotiating autonomy and responsibility in military robots. Ethics and Information Technology, 16(1), 51-62.

Rathje, J. M., Spence, L. B., & Cummings, M. L. (2013). Human-Automation Collaboration in Occluded Trajectory Smoothing. Human-Machine Systems, IEEE Transactions on, 43(2), 137-148.

Sparrow, R. (2007). Killer robots. Journal of Applied Philosophy, 24(1), 62-77.

Van den Hoven, J. (2007). ICT and value sensitive design. In The information society: Innovation, legitimacy, ethics and democracy in honor of Professor Jacques Berleur SJ (pp. 67-72). Springer US.