Malevolent Cyborgization

Nadisha-Marie Aliman(✉)

University of Stuttgart, Stuttgart, Germany
nadishamarie.aliman@gmail.com

Abstract. While significant progresses in AI research are expanding the presumed limits of feasibility, the dangers of future AI agents with human level intelligence or beyond exhibiting a hostile behavior towards humans have been increasingly discussed. A lot of ethical concerns have been expressed in this context, whereby AI Safety research was classically focused on how to create safe and ethical AI systems. By contrast, Pistono and Yampolskiy (2016) proposed a new important approach inspired by the cybersecurity paradigm and analyzing the unethical development of an AI with malice in design. In this paper, we connect the ethical concerns raised by a Malevolent Artificial Intelligence (MAI) as characterized in their work, to those raised by a possible maliciously crafted human-machine intelligence merger. We elaborate on how both concepts could be related or even intertwined, but would also exhibit specific differences. Our analysis reveals a wide array of alarming potential risks and suggests integrating considerations concerning the safety of AI systems as well as such affecting the safety of cyborgian systems into a joint interdisciplinary framework covering various developments towards Superintelligence.

Keywords: Cyborgization · Malevolent Artificial Intelligence · Superintelligence

1 Introduction

In their paper titled "Unethical Research: How to Create a Malevolent Artificial Intelligence", Pistono and Yampolskiy (2016) described possible developments towards a future unethical AI. The authors argue that, unlike in the domain of cybersecurity, where a certain balance is ensured by a research concept covering both potential malicious exploits and measures to maintain safety, AI Safety researchers so far only focused on the general conditions of implementing safe AI systems, while possible malicious exploits on such remained disregarded. According to them, the lack of information resulting out of it should be resolved, since the consequences of an intentional malicious exploit on superintelligent AI systems in the future could be devastating for humanity. In our opinion these claims are accurate for the following reasons: first the previous publications in the field of AI safety before, predominantly contained considerations on how to design safe AIs and a deeper differentiated analysis was missing, although there is no

T. Everitt et al. (Eds.): AGI 2017, LNAI 10414, pp. 188–197, 2017.
DOI: 10.1007/978-3-319-63703-7_18

reason why malicious exploits should not be performed intentionally alike on AI systems, since the same principle of taking advantage of security holes in cybersecurity can be transferred to AIs as being software/hardware entities. Secondly, the level of intelligence of AIs steadily increases, it is to be expected that superintelligence will be reached in the not so far future (Bostrom 2014; Chalmers 2010) and an intentionally crafted attack especially using a superintelligent system would imply unforeseen and unintelligible effects for human experts whose minds are going to be overcharged or too slow to counteract. A type of *"Hazardous Intelligent Software"* may even stay undetected a long time because of the gap of intelligence – just like monkeys cannot comprehend complex human behavior patterns. It will therefore offer exceptionally much power to the attacker to harm humans and as stated in the paper, it is known in history that "absolute power corrupts absolutely".

The authors described a variety of reasons why several stakeholders like military, governments, corporations, psychopaths or even AI Safety researchers with unethical intentions could intend to implement a MAI ranking from acquiring control and dominance, gaining financial benefits, to initiating the extinction of mankind among other things. In any case, there is a kind of cooperation between a human entity and an artificial one to achieve an unethical objective, whereby mostly the human entity initiates the cooperation with a malicious intent. Our view is that the intensity of such an alarming human-machine cooperation could be much higher in the future, since – according to the foreseeable scientific progress in fields like Bionics, Nanorobotics or Brain-Computer Interfaces (BCI) research – it could be possible to merge human and intelligent artificial entities to obtain a hybrid system with an enhanced cognitive performance, which could be used to follow similar unethical objectives as mentioned earlier and would concern the same stakeholders. For instance, psychopaths could as well maliciously intend to merge with an AI entity to become more intelligent or get greater knowledge than their fellow men and in doing so, to be able to manipulate and control others on a large scale or the military could encourage cyborgization techniques to be able to deploy cyborg armies in wars wiping out opponents through intellectual, strategical or/and possibly physical superiority. In this paper, we analyze the concept of a human-machine intelligence merger with intentional malice in design which we call *Malevolent Cyborgization*, and relate it to the MAI concept introduced by Pistono and Yampolskiy.

Outline: In the next Sect. 2, a brief explanation concerning present trends towards cyborgization from both a technical and a societal point of view is provided, followed by a short general overview briefly introducing different approaches to a definition of the term cyborg as a concept of human-machine merger. In Sect. 3, we discuss possible societal impacts and ethical concerns in connection with Malevolent Cyborgization and highlight common features with MAI scenarios, but also specific differences. Thereafter, in Sect. 4, we argue about a possible cyborgian path to Superhuman Intelligence, which could be linked to (superintelligent) AI and indicate potential impacts on society. Finally, the last Sect. 5 concludes.

2 Cyborgization

Already today, first technological efforts to make cyborgization possible can be noticed and are considered by AI researchers as well as scientists from different fields. The Defense Advanced Research Projects Agency (DARPA) is working on a new project concerning brain implants allowing a *"channel between the human brain and modern electronics"* (DARPA 2016) and the company Kernel with the goal to build *"the world's first neural prosthetic for human intelligence enhancement"* was founded this year (Mednitzer 2016). Committed persons like Elon Musk are emphasizing the need of an injectable neural lace bypassing a surgical intervention to *"achieve symbiosis with machines"* (Bhavsar 2016) which is not that utopian as it might seem, since Liu et al. (2015) successfully tested such an engine they called "syringe-injectable electronics" on mice. Furthermore, Musk recently founded the BCI company Neuralink with the long-term goal to achieve human enhancement. Initial steps towards wireless BCIs have been taken in the form of in-animal trials of what the researches entitled "neural dust" (Seo et al. 2013) – miniature wirelessly working sensors to monitor brain activity. A first brain-to-text system which performs "automatic speech recognition from neural signals" has recently been implemented by Herff and Schultz (2016). This system represents first steps towards an automated transcription of imagined sentences to text. Moreover, cyborgization has already been perceived by the general public and is seriously thematized as a phenomenon of the near future by some researchers. For example, many people encountered the topic of cyborgs in a broader sense through the first Cybathlon hold in Zürich including a competition with Brain-Computer Interfaces. During this event, ethical discussions referring to the topic of enhancement through cyborgization amongst others, were televised. Furthermore, Kurzweil (2006) prognoses the concept of wirelessly connecting the neocortex to a synthetic one in the cloud, which could be feasible in the mid-century according to him. In this context, he postulated future developments denoted as the "human body version 2.0 scenario" and explained: *"Computers started out as large, remote machines in airconditioned rooms tended by white-coated technicians. They moved onto our desks, then under our arms, and now into our pockets. Soon, we'll routinely put them inside our bodies and brains. By the 2030s we will become more nonbiological than biological."* All these developments show that the path towards cyborgization is actually considered in the digital age and that cutting edge research already started. In our opinion, Brain-Computer Interfaces and Computer-Brain interfaces, which were already used for several proof-of-concept Brain-to-Brain Communication scenarios, will play a decisive role in this development by providing a new quality of intimacy between human brains and machines (and also between different human brains).

In the literature, there are different types of definitions for the notion of cyborg. Etymologically speaking, the word cyborg comes from "cybernetic organism" and was first introduced in an article by Clynes and Kline (1960) dealing with the adaptation of humans under the conditions of outer space. Some researchers argue, that humans are already cyborgs today due to the omnipresence of technical devices used to facilitate the daily life and that it is only a matter of time till

the devices will be located under the human skin. Spreen (2010) describes the transition between low tech bodies and high tech bodies as a spectrum with variable proportions illustrated by a slider, whereby the middle of the spectrum represents the "skin border" (originally "Hautgrenze"). According to him, a human becomes a cyborg as soon as the skin border is exceeded and likewise, a human ceases to be a cyborg and can therefore put the slider back e.g. if he removes the technical device(s) under his skin. Another definition is provided by Haraway (1987) which views cyborgs in a feminist context and sees the concept as possibility to break out of traditional patterns. In "Cyborg Morals, cyborg values, cyborg ethics", Warwick (2003), which performed self-experimentation and experienced a chip implant at first hand, narrows the usage of the term cyborg in his paper to the cases where the cyborg *"is formed by a human, machine brain/nervous system coupling"*, hence cases where a human is *directly* linked to technology via his brain/nervous system and excluding more superficial variants like intelligent glasses or smartphones. In order to provide clarity in the following, when we refer to the term cyborg, we specifically mean (unless otherwise stated) a human whose brain is *directly* linked to technology able to enhance his cognition/intelligence. We accordingly refer to the underlying process to become a cyborg or to "develop" a cyborg system as cyborgization.

From the perspective of an individual, they are a lot of reasons why cyborgization is worth striving for. Warwick (2003) mentions *"use the computer part for rapid maths"*, *"call on an internet knowledge base, quickly"*, *"understand multi dimensionality"*, *"communicate in parallel, by thought signals alone, i.e., brain to brain"* as possible motivations amongst others. Further possible advantages could be: position oneself in the labor market by exhibiting above-average analytic abilities, extend the limits of perception and remember countless details leading to a photographic memory, achieve unforeseen ingenuity in research fields, earn a lot of money and so forth. But equally, malicious motivations with the aim to harm other people such as being able to manipulate and subjugate other people or exploiting the ignorance and vulnerability of non-enhanced humans, could emerge.

3 Ethical Concerns of Malevolent Cyborgization

In the light of the above, it becomes clear that Malevolent Cyborgization (MC) could be desirable for a wide range of stakeholders with a heterogeneous set of goals. There is even an overlapping between possible entities, which could be interested in MC and those eligible for MAI as described by Pistono and Yampolskiy. In the following, we first take up the exemplary stakeholders for MAI mentioned in their paper and indicate which motives could justify them likewise as stakeholders for MC showing the parallels between those two phenomena. Thereafter, we introduce additional global effects specific to MC and differing from the MAI scenario.

- *Military:* As already mentioned in the introduction, the military could maliciously employ cyborg soldiers similarly as MAIs *"to achieve dominance"* through intellectual, strategical or/and possibly physical superiority.
- *Governments:* Through cyborgization, governments could acquire intellectual superiority with the same intentions as for the MAI case: *"to establish hegemony, control people, or take down other governments"*. Note that these goals can also be reached through a forced cyborgization of inferior quality carried out by governments on people with the aim to subjugate them (e.g. a kind of digital lobotomy suppressing the functionality of the frontal cortex, body hijacking or an automatic red-out of personal information using BCI data could be possible).
- *Corporations:* The authors state the following motives for the MAI case: *"trying to achieve monopoly, destroying the competition through illegal means."* As cyborgs with enhanced intelligence could be able to process considerably more information than non-enhanced humans, the transparency on the market may suffer of it allowing them to take over and build monopolies in different fields, which is again similar to the MAI goals.
- *Villains:* Following Pistono and Yampolskiy, possible goals why a MAI could be desirable for villains are: *"trying to take over the world and using AI as a dominance tool"*. In the case of MC, the same goals would be valid except that Cyborgization will be used primarily as means rather than AI. (But obviously, cyborgs of all the eligible entities could also merge with AI, which would represent an extremely risky scenario. We will analyze this matter later in the next section).
- *Black Hats:* Through their enhanced cognition/intelligence, cyborgs could have an enhanced ability to detect security holes at their disposal and could for instance develop better heuristics for password-guessing. They might therefore, likewise black hats with a MAI, secretly attempt *"to steal information, resources or destroy cyberinfrastructure targets"*.
- *Doomsday Cults:* The goal of *"attempting to bring the end of the world by any means"* using a MAI can obviously also build a basis for cyborgs involved in doomsday cults.
- *Depressed:* Depressed cyborgs could hand over the liability for their live or death to their artificial part e.g. by setting a self-destruction mode stopping vital functions in the brain. They could thereby reach the goal to commit suicide such as depressed people using MAI to be able to *"commit suicide by AI"*.
- *Psychopaths:* As described in the introduction, psychopaths could be interested in cyborgization to be able to manipulate and control others. Moreover, psychopaths could wish to historically gain notoriety with regard to their wrongdoings. The aim to *"trying to add their name to history books in any way possible"* seems to not only be a possible motivation for MAI, but also for MC.
- *Criminals:* According to the authors, criminals could attempt *"to develop proxy systems to avoid risk and responsibility"*. The same is possible in the MC scenario. A malevolent cyborg could for instance wirelessly establish a

connection to proxy systems to commit crimes at other places. He could conceal the fact he is a cyborg so that nobody would suspect him of being involved in crimes.
- *AI Risk Deniers:* For the case of MC, it would be appropriate to instead address "Cyborgization Risk Deniers". This stakeholder could let people believe that cyborgization is not more than a Science Fiction scenario and leave non-enhanced humans in ignorance yielding an even greater disparity between cyborgs and the regular humans.
- *AI Safety Researchers:* For the MAI scenario, the authors state "*AI Safety Researchers, if unethical, might attempt to justify funding and secure jobs by purposefully developing problematic AI.*" If in the future a discipline like "Cyborg Safety" existed, malicious people working in this field could deliberately develop unsafe cyborg systems e.g. such that than can easily be exploited, so that they can ensure their occupation over and over again.

After having pointed out the similarities between the entities which could be interested in the usage of MAI on the one hand and MC on the other hand, as well as having clarified the conformity of the achievable unethical objectives in both scenarios, we will now allude to some additional societal impacts that can be specifically caused by MC (and not necessarily by MAI) through the phenomenon of the human mind transcending its biological boundaries:

- *New hierarchy in mankind:* Cyborgization could lead to a hierarchy of enhancement forming an open-ended continuum ranging from completely non-enhanced humans to cyborg versions 1.0, 2.0 and so forth, even if performed with positive or neutral intentions toward humans. This development follows from the common practice of software updates and hardware tuning. Over time, the biological part of the cyborgs is furthermore going to be surpassed by the non-biological part getting faster with exponential pace. Like in many other cases, the quality of the "products" people can afford would depend on their financial status and a lot of people might irreversibly stay behind. This circumstances could lead to social unrest and conflicts. This background provides a strategical basis for every conceivable kind of MC.
- *Global identity crisis:* Cyborgization could initiate an unforeseen social transformation shaking the notion of "human being", "identity" and "self" for the questions could be: "At what time does someone stop to be a human?", "Does the self include the machine part?", "What happens if the non-biological part starts to prevail – does the cyborg become a machine?". Psychological studies actually demonstrated, that human self-perception is extremely flexible (Clark 2004). Likewise, the first officially recognized cyborg (the expression cyborg is here used in a broader sense) Neil Harbisson, which is equipped with an eyeborg stated (Jeffries 2014): *"I don't feel like I'm using technology, or wearing technology. I feel like I am technology. I don't think of my antenna as a device - it's a body part."* The additional perception through the eyeborg fully integrated the functionality of his brain leading to a seamless unity. This gives an indication that future cyborgs might extend the limits of "identity"

and "self" in addition to the higher level of intelligence. This could lead to a strong sense of alienation between non-enhanced humans and cyborgs raising tensions and providing a fertile ground for MC.

– *Evolutionary upheaval:* A world of work with extremely productive and superintelligent cyborgs could piece by piece make less enhanced humans superfluous. MC could at a certain point introduce the extinction of those people leading to a disaster for humanity. Equally, cyborgization could lead to an evolutionary advantage and some could consider it as the next step in evolution. In this case cyborgs would supersede non-enhanced humans in the long run and this process might be accelerated by means of MC. The historian Yuval Noah Harari claimed: *"I think it is likely in the next 200 years or so homo sapiens will upgrade themselves into some idea of a divine being, either through biological manipulation or genetic engineering of by the creation of cyborgs, part organic part non-organic. [..] It will be the greatest evolution in biology since the appearance of life. Nothing really has changed in four billion years biologically speaking. But we will be as different from today's humans as chimps are now from us"* (Knapton 2015). But he also addressed the increasing gap between poor and rich in this future, which could lead to a dying out of the poor, while the rich could live forever.

4 Cyborgization, AI and Superhuman Intelligence

Warwick (2013) postulates: *"We must be clear that with extra memory, high-powered mathematical capabilities, including the ability to conceive in many dimensions, the ability to sense the world in many different ways and communication by thought signals alone, such cyborgs will be far more powerful, intellectually, than humans"*. He furthermore describes proof-of-concept experiments he performed concerning human-machine merger. He comes to the conclusion that human-machine merger is going to be feasible from a technological point of view and that *"[..]connecting a human brain, by means of an implant, with a computer network could in the long term open up the distinct advantages of machine intelligence, communication and sensing abilities to the implanted individual"* and warns that this development will also raise fundamental ethical questions. We support this view relating to communication, sensing abilities and memory for reasons already mentioned in the last sections; for the matter of advantages through machine intelligence (or generally speaking AI) in the context of cyborgization, we will hereinafter shed some light on some possible outcomes and distinguish different associated scenarios.

With the joint aim to produce a higher intelligence, cyborgization and AI are not necessarily disjunctive developments. Progresses in AI can even provide an ideal ground for cyborgization efforts. Some view the dangers of superintelligent AI as a motive for cyborgization in order to forestall a future domination of AI over mankind. An example for this is the statement of Stephen Hawking 15 years ago (Highfield 2001): *"There is a real danger that computers will develop intelligence and take over. We urgently need to develop direct connections to*

the brain so that computers can add to human intelligence rather than be in opposition." illustrating this kind of consideration.

Note that cyborg-systems could already be implemented without real AI components and could nevertheless reach Superhuman Intelligence. A superintelligent cyborg would obviously be able to implement more intelligent AIs than regular humans could and moreover, the development of such cyborg-systems could even be the first step towards an intelligence explosion (Chalmers 2010) making a cyborg able to develop AIs or cyborg-systems more intelligent than itself. A restriction to this scenario could be that the biological part of the cyborg limits the speed of self-improvement, but this could be compensated for instance by increasing the percentage of the non-biological part or/and by enhancing the intelligence of the non-biological part e.g. by using narrow AI components.

Of course one could argue about a possible boundary for when a cyborg stops to be a cyborg and becomes a machine/an AI entity. In our view, there will be no such clearly ascertainable boundary due to conceivable designs of cyborg-systems in the future. For instance could concepts inspired by ensemble learning be deployed. In the case of a cyborg-system, it could be beneficial to combine the strengths of the biological part with those of the non biological one by means of ensemble learning on a meta level e.g. to improve the intelligence of the overall system. From an abstract point of view, a cyborg-system could then act like a self-optimizing dynamical ensemble with adjustable weights where the most intelligent entity controls and adjusts the weights given actions in certain contexts. Initially, the biological part could be in control, but at a certain point, the non biological part would be able to inhibit the biological choices for actions if they happen to be disadvantageous for the goal setting and thus reduce the weights of the biological one up to its vanishing. The transition from cyborg to machine could therefore rather be fluent and occur gradually without a precise boundary between cyborgs and machines (Kurzweil 2006).

Imagine a cyborg which – in addition to the access to a huge memory and knowledge base, superfast calculation capacity, parallel communication to other brains and so forth all by thoughts – could permanently delegate the organization of his thoughts, his perception and rational thinking to numerous specialized AI agents in parallel. The cyborg could be able to understand big data and extract comprehensive information out of it. For instance, he could be computing the statistics of the current situation on the financial markets, while walking on the street and taking note of the biography of a totally unknown person he retrieved by face recognition and search on the internet, having a phone (or brain-to-brain) call and at the same time running multiple simulations for different variants of AIs he developed. Analogously, a cyborg could decide to merge with a superintelligent AI. In this case, issues related to control might raise. Depending on the level of intelligence, it could stay a human-machine cooperation, maybe on equal terms by means of control at the beginning or also depending on the goal sets, leading to a quite "dissociative" construct exhibiting certain symptoms for which the dissociative personality disorder could give a premonition. But the superintelligent AI would presumably have a considerable advantage in the

long run. If the superintelligent AI happens to be more intelligent than the cyborg, it might result in a scenario where the superintelligent AI uses the cyborg as "delegate" in the same way described, where a cyborg could use narrow AIs. In this context, issues related to the notion of "free will" might arise additionally.

The just described scenarios leave much space for the actual emergence of MC. Equally, malicious stakeholders could strive for cyborgization to be able to develop a MAI or conversely plan to merge with a MAI for malicious purposes. It becomes additionally clear, that cyborgization is not a guarantee to prevent the creation of MAI, since MC is possible and could also directly lead to MAI.

5 Conclusion and Future Prospects

Cyborgization is only one possible path towards superintelligent enhanced humans. It could in principle also be reached e.g. by means of genetic engineering/breeding and biotechnology. Yampolskiy and Spellchecker (2016) conclude that *"augmented humans with IQ beyond 250 would be superintelligent with respect to our current position on the intelligence curve but would be just as dangerous to us, unaugmented humans, as any sort of artificial superintelligence."* We come to the same conclusion with regard to Malevolent Cyborgization, because it could serve similar stakeholders to accomplish the same unethical goals representing existential risks for humanity as in the MAI case.

In the future, cyborg-systems could become a daily reality offering a variety of promising perspectives regarding human enhancement, but their development and deployment will then need to be regulated. Besides a legal obligation for open source cyborg-sytems, a possible approach for a society willing to prevent MC and related risks could for instance be measures inducing an obligation for all stakeholders developing cyborg-systems to adhere to "Cyborg Safety" guidelines, which could be defined by an ethical board for superintelligence. Nowadays, there is yet no explicit binding international interdisciplinary ethical board for superintelligence containing e.g. AI, AI Safety, Cybersecurity, Neuroscience, Biotechnology, Nanotechnology, Law experts (just to name a few) at the same time. Such a collaboration would though be of great value to maintain an overview of all critical developments with the aim to reach superintelligence. However, forward-thinking interdisciplinary frameworks similar to the Asilomar AI Principles (FLI 2017) could serve as a basis and should be extended, since there are always security holes that remain undetected and characteristically, only one specially selected successful MAI or MC attack trial could be enough to drastically change the world in a negative sense.

Acknowledgements. We would like to especially thank Roman Yampolskiy and further Kevin Baum for the helpful feedbacks on an earlier draft of this paper. We would also like to thank Martin Bevandic for his support.

References

Bhavsar, N.: Elon Musk: To Survive The Age of Artificial Intelligence, Humanity Must Achieve Symbiosis With Machines (2016). http://futurism.com/neural-lace-what-humanity-needs-to-survive-the-age-of-artificial-intelligence/

Bostrom, N.: Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Oxford (2014)

Chalmers, D.J.: The singularity: a philosophical analysis. J. Conscious. Stud. **17** (2010)

Clark, A.: Natural-Born Cyborgs. Oxford University Press, Oxford (2004)

Clynes, M., Kline, N.: Cyborgs and space. Astronautics, 74–75 (1960)

DARPA: Bridging the Bio-Electronic Divide (2016). http://www.darpa.mil/news-events/2015-01-19. Accessed 11 Nov 2016

Future of Life Institute: Asilomar Ai Principles (2017). https://futureoflife.org/ai-principles/

Haraway, D.: A manifesto for Cyborgs: science, technology, and socialist feminism in the 1980s. Aust. Feminist Stud. **2**(4), 1–42 (1987)

Herff, C., Schultz, T.: Automatic speech recognition from neural signals: a focused review. Front. Neurosci. **10**, 429 (2016)

Highfield, R.: Interview with Stephen Hawking (2001). http://www.telegraph.co.uk/news/science/science-news/4766816/Interview-with-Stephen-Hawking.html. Accessed 11 Nov 2016

Jeffries, S.: Neil Harbisson: the world's first cyborg artist (2014). https://www.theguardian.com/artanddesign/2014/may/06/neil-harbisson-worlds-first-cyborg-artist. Accessed 11 Nov 2016

Knapton, S.: Humans 'will become God-like cyborgs within 200 years' (2015). http://www.telegraph.co.uk/culture/hay-festival/11627386/Humans-will-become-God-like-cyborgs-within-200-years.html. Accessed 11 Nov 2016

Kurzweil, R.: The Singularity is Near: When Humans Transcend Biology. Penguin (Non-Classics) (2006)

Liu, J., et al.: Syringe injectable electronics. Nat. Nanotech. **10**(7), 629–636 (2015)

Mednitzer, R.: Kernel's Quest to Enhance Human Intelligence (2016). https://lifeboat.com/blog/2016/10/kernels-quest-to-enhance-human-intelligence/

Pistono, F., Yampolskiy, R.V.: Unethical Research: How to Create a Malevolent Artificial Intelligence. In: 25th International Joint Conference on Artificial Intelligence (IJCAI-16). Ethics for Artificial Intelligence Workshop, AI-Ethics-2016, 9–15 July 2016, New York, NY (2016)

Seo, D., Carmena, J.M., Rabaey, J., Alan, E., Maharbiz, M.M.: Neural Dust: An Ultrasonic, Low Power Solution for Chronic Brain-Machine Interfaces (2013). https://arxiv.org/abs/1307.2196

Spreen, D.: Der Cyborg: Diskurse zwischen Körper und Technik. Suhrkamp **1971**, 166–179 (2010)

Warwick, K.: Cyborg morals, cyborg values, cyborg ethics. Ethics Inf. Technol. **5**(3), 131–1137 (2003)

Warwick, K.: The Future of Artificial Intelligence and Cybernetics (2013). https://www.bbvaopenmind.com/en/article/the-future-of-artificial-intelligence-and-cybernetics/

Yampolskiy, R.V., Spellchecker, M.S.: Artificial Intelligence Safety and Cybersecurity: A Timeline of AI Failures (2016). https://arxiv.org/abs/1610.07997