Emerald Insight

Journal of Information, Communication and Ethics in Society

IT-ethical issues in sci-fi film within the timeline of the Ethicomp conference series
Anne Gerdes

Article information:

Users who downloaded this article also downloaded:

(2015),"Computer ethics beyond mere compliance", Journal of Information, Communication and Ethics in Society, Vol. 13 Iss 3/4 pp. 176-189 http://dx.doi.org/10.1108/JICES-10-2014-0043

(2015),"Just say "no!" to lethal autonomous robotic weapons", Journal of Information, Communication and Ethics in Society, Vol. 13 Iss 3/4 pp. 299-313 http://dx.doi.org/10.1108/JICES-12-2014-0065

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

IT-ethical issues in sci-fi film within the timeline of the Ethicomp conference series

Anne Gerdes

Department of Design and Communication,
University of Southern Denmark, Kolding, Denmark

Abstract

Purpose – This paper aims to explore human technology relations through the lens of sci-fi movies within the life cycle of the ETHICOMP conference series. Here, different perspectives on artificial intelligent agents, primarily in the shape of robots, but also including other kinds of intelligent systems, are explored. Hence, IT-ethical issues related to humans interactions with social robots and artificial intelligent agents are illustrated with reference to: Alex Proyas' I, Robot; James Cameron's Terminator; and the Wachowski brothers' Matrix. All three movies present robots cast in the roles of moral agents capable of doing good or evil. Steven Spielberg's Artificial Intelligence, A.I. gives rise to a discussion of the robot seen as a moral patient and furthermore reflects on possibilities for care and trust relations between robots and humans. Andrew Stanton's Wall-E shapes a discussion of robots as altruistic machines in the role as facilitators of a flourishing society. Steven Spielberg's Minority Report allows for a discussion of knowledge-discovering technology and the possibility for balancing data utility and data privacy.

Design/methodology/approach – Observations of themes in sci-fi movies within the life span of the ETHICOMP conference series are discussed with the purpose of illustrating ways in which science fiction reflects (science) faction. In that sense, science fiction does not express our worries for a distant future, but rather casts light over questions, which is of concern in the present time.

Findings – Human technology interactions are addressed and it is shown how sci-fi films highlight philosophical questions that puzzle us today, such as which kind of relationships can and ought to be formed with robots, and whether the roles they play as social actors demand that one ought to assign moral standing to them. The paper does not present firm answers but instead pays attention to the selection and framing of questions that deserve attention.

Originality/value – To relate sci-fi movies to topics raised during the past 20 years of the ETHICOMP conference series, seemed to be an appropriate way of celebrating the 20-year anniversary of the ETHICOMP conference series.

Keywords Artificial intelligence, Philosophy, Surveillance, Autonomy, Trust, Artificial agents

Paper type Research paper

1. Introduction

Terrell Ward Bynum and Simon Rogerson fostered the ETHICOMP conference back in 1995, and, over the years, they managed to cultivate an inspiring and open-minded community. They stepped down from the throne after ETHICOMP 2013, which took place in Kolding at University of Southern Denmark. In his closing remarks to this

conference, Terrell Ward Bynum noticed that (freely quoted by memory): "at the first ETHICOMP, back in 1995 there were around thirty delegates. By now, we are hundred humans and one robot". Hereby, he pinpointed the impact of ETHICOMP, namely, the fact that the ETHICOMP community strives to explore and reflect pro-actively in responding to philosophical, social and moral questions raised by technology.

In the same sense, it is well-known that art often captures emerging scientific advances or even, to some extent, functions as a source of enlightenment and inspiration to scientific discovery. But probably most of all, the expressive power of art provides us with ways of understanding ourselves and our being in the world. Here, particularly sci-fi movies stand out in their effort to grasp what is puzzling today seen through the lens of the future. Typically, the storylines in sci-fi movies reveal important philosophical questions regarding moral agency and patiency, consciousness, identity, social relations and privacy to mention just a few. In this paper, I shall be outlining an analysis of selected IT-ethical perspectives in sci-fi movies covering the timeline of the ETHICOMP conference series, which was launched in 1995 by the Centre for Computing and Social Responsibility (CCSR), Terrell Ward Bynum and Simon Rogerson.

2. Technology as a social actor

Traditionally, we have conceived machines as mere tools or extensions of human activity. But the development of relation-based technology, such as different types of artificial companions, implies a change in perspective to account for the role these technologies play as social actors:

> ACs [artificial companions] are a technological success because they are not the outcome of some unforeseeable breakthrough in strong AI, but the social equivalent of Deep Blue: they can deal well with their interactive tasks, even if they have the intelligence of an alarm clock. And they are philosophically significant precisely because they are neither Asimov's robots nor Hal's children. Out of the realm of thought experiments and unrestrained speculations, they posit concrete, philosophical questions. (Floridi, 2014, p. 153)

Sci-fi movies also display philosophical and moral issues and highlight ways in which our formation of identity and self-understanding are influenced by robots, which have developed to become social actors. Likewise, within the field of human–robot interaction studies, several empirical studies illustrate ways in which we enter into relationships with social robots (Turkle, 2011; Dautenhahn, 2007; Schärfe *et al.*, 2011; Carpenter, 2013; Bartneck *et al.*, 2007). For instance, when it comes to the case of inflicting pain on a robot, Bartneck *et al.*'s experiment, in which research subjects were asked to switch off a robot (which had assisted them during a test session), illustrates that research subjects were emotionally affected by the pleas of the robot and even entered into long dialogues with the robot before turning it off (Bartneck, 2007).

Although, we are aware that robots are nothing but machines, the fact that we are highly meaning-seeking and interpretative beings causes us to anthropomorphize robots and intuitively interpret them as fellows, as illustrated by Schärfe *et al.* (2011), who refer that relations of trust can be observed in studying physical interaction between a person and a geminoid. Through empirical studies, they find evidence of initial trust formation, which springs from a hand placed on the shoulder of the geminoid, likewise eye contact with the geminoid also promotes intimacy. Similarly, at the ETHICOMP 2013 conference, Henrik Schärfe and his alter ego, Geminoid DK, were present as keynotes (www.geminoid.dk), and some among the female audience felt like

they were able to make eye contact with the geminoid. Consequently, as summarized in a phrase by Sherry Turkle, robots press our "Darwinian buttons" (Turkle, 2011, p. 8) and we respond with care even toward very simplistic robot design (Benford and Malartre, 2007, p. 165). Moreover, Turkle (2011, p. 49-ff.) warns us that the *robotic moment* is lurking in the horizon, which refers to the moment at which we prefer risk-free relationships with robots over challenging engagements with humans:

> As I have said, my story of this moment is not so much about advances in technology, impressive though these have been. Rather, I call attention to our strong response to the relatively little that sociable robots offer – fueled it would seem by our fond hope that they will offer more. With each new robot, there is a ramp-up in our expectations. I find us vulnerable – a vulnerability, I believe, not without risk. (Turkle, 2011, p. 52)

On the other side, the World Robot Declaration was issued in 2004 in Japan, implying that maybe the time has come to discuss if there is a need for granting rights to robots? So far, we have dealt with whether we can assign responsibilities to artificial agents or not (Floridi and Sanders, 2004; Verbeek, 2011). But, according to David Gunkel, we have not yet seriously started to consider whether we ought to have moral duties toward machines (Gunkel, 2012, p. 99). Hence, in *The Machine Question – Critical Perspectives on AI, Robots, and Ethics* (2012) – Gunkel argues that already by now, the term "person" has opened up to also include animals and corporations, and as we have started to grant rights to non-human agents, it might be a matter of time before we include machines into the category of persons (Gunkel, 2012, p. 39-ff.). If we do so, the question arises whether the kind of responsibilities we would have toward robots would be on par with the kind of responsibilities we have toward animals, corporations or other human beings?

In a similar vein, sci-fi movies often highlight what is morally at stake in *the relation, per se*, rather than framing the plot around a distinction between personhood versus machine. The first mentioned perspective is illustrated in *Artificial Intelligence, A.I.* (Section 4), but, in what follows, we shall be paying attention to the second perspective, i.e. ways in which artificial moral agency constitute a threat to human autonomy.

3. Robots as (in)moral agents
The exploration of the impact of technology on civilization and especially the theme of loss of human control and autonomy due to robot interference, is a classic theme in sci-fi movies, as spelled out in *A Space Odyssey*, by Kubrick (1968), in which the intelligent computer H.A.L 9000 sets the mission of the spaceship above the crew. A similar theme is raised in the first Alien movie – here an android researcher risks the crew to preserve the alien for the sake of science. Within the timeline of ETHICOMP, the robot cast in the role as a moral agent with the ability to do (mostly) evil and good represents a strong narrative.

This is shown in the movie *I, Robot*, by Alex Proyas (2004), which takes place in 2035 in a world in which social robots interact with humans as polite and caring servants. The overall storyline highlights challenges in the relationship between imperfect humans and perfect moral robots, which (or who?) by deduction from Asimov's laws of robot ethics, turn against humanity in illustrating the paternalistic point that robots find it necessary to protect humans against themselves – with reference to Asimov's zeroth law: "A robot may not harm humanity, or, by inaction, allow humanity to come to harm". Fortunately, this logical inappropriate implication is short-circuited by the detective, Del Spooner. In this way, paternalism evaporates in favor of human autonomy, which

includes our capabilities for imperfection and failures, thereby characterizing what makes us moral beings in the first place:

> There is nothing particularly ethical about adding up utilities or weighing pros and cons, until the decision maker feels the urge *not to follow* the ethical course of action it arrives at. (McDermott, 2008, p. 6)

Next, and bearing the ETHICOMP timeline in mind, we shall be looking at *Terminator 3 – the raise of the Machines*, By James Cameron (2003). The plot develops around the latest version of a killing robot T-X, which is sent from the future by Skynet, a consciousness AI, whose main purpose is to get rid of humanity. The Terminator's outstanding information processing abilities include infrared sight and context-sensitive action-guiding scripts. Moreover, the Terminator's deliberations are not distracted by emotions but purely guided by rational calculations. These advantages enable it to always have a lead over humans. And, although we can never know what it is like to be a Terminator (Nagel, 1974), we are still allowed a first-person's perspective on its sense impressions. Thus, we get an inside glimpse of the makeup of its perceptual apparatus and its mind. This camera technique gives rise to anxiety because it reveals a dedicated action minded species, which acts in accordance with efficient calculations without emotional distractions – as summarized by one of the leading actors: "It can't be bargained with, it can't be reasoned with, it doesn't feel pity or remorse or fear, and it absolutely will not stop, ever, until you're dead!"

Hence, the Terminator travels back to our time to hunt down and kill Sarah Connor to bypass a future in which she gives birth to a son, John, who will grow up to become the future leader of a revolt against the machine rulers. In this particular movie, from 2003, Skynet also enters into cyber warfare and infects US army computer systems, thereby leaving the country open to attacks. As a minor curiosity in relation to the Terminator films, one might note that traces of our ambivalence toward humanoid machines are observable in the *100 heroes and Villains' list* of Hollywood films, on which *Terminator* was ranked the 22nd villain of all time, while the Terminator in *Terminator 2: Judgment Day* was ranked the 48th hero of all times. It has never happened before that the same character appears on these two opposing lists (Wikipedia: Terminator).

Finally, in The Wachowski brothers' Matrix movies, of which only the first, *The Matrix* (1999), is outstanding, we are introduced to a future-simulated reality named "the Matrix", i.e. a computer-generated world run by machines to control humans and use their bodies as bio-electric raw material as sources for energy consumption. But a small group of hackers discover the virtual reality fraud. The rebellions grow supernatural powers and their knowledge of the true nature of the computer-simulated reality makes them capable of bending the natural laws. Hence, they try to overthrow the Matrix and win back human autonomy.

Both films encapsulate an archetypical AI-narrative, which describes intelligence with reference to logic governed behavior:

> Some, at least, of what makes minds rational is their ability to perform computations on thoughts; when thoughts […] are assumed to be syntactically structured, and where "computation" means formal operations in the manner of Turing. (Fodor, 1998, p. 205)

This narrative carries over to the present motivations for introducing lethal autonomous weapon systems (LARS) in warfare. Here, a main argument for assigning moral agency to LARS mainly rests upon the assumption that eventually, with the right kind of

implementation of a formal moral framework, such robots shall be able to outperform human beings and lead to more moral warfare (Arkin, 2007; Lin *et al.*, 2008). Hence, moral robots are assumed to be better at moral decision-making and subsequently acting in accordance with rules of engagement under stressful conditions. Moreover, future combat robots presumably also hold promises of progressively more bloodless and less harmful warfare (Arkin, 2007, p. 6; Lin *et al.*, 2008, p. 50-ff.), which implies that war might turn into a risk-free business – "When a robot dies, you don't have to send a letter to its mother" (Singer, 2009, p. 31). But, this route down the future warfare lane seems to be paradoxically paved, as in line with Just War Theory (Walzer, 1977), war ought to be the last resort and not something that we turn to too easily, otherwise the global conflict level might increase dramatically.

So far, our present discussion has concerned whether it is plausible to assign moral agency and responsibility to an AI agent, a discussion which is increasingly raised in philosophy (Floridi and Sanders, 2004; Sparrow, 2004, 2007; Verbeek, 2011) in tandem with the discussion of whether this implies that we ought to be responsible toward a robot, i.e. consider a robot as a proper receiver of moral considerations (Gunkel, 2012; Floridi, 2014; Benso, 2000). Consequently, in what follows, we shall be addressing issues of moral patiency in human–robot interactions.

4. Robots as moral patients

Should their [artificial companions] non-biological nature make us discriminate against them? The question casts an interesting light on our understanding of what kind of persons we would like to be. (Floridi, 2014, p. 154)

Artificial Intelligence, AI., by Steven Spielberg (2001), tells the story of the first robot child, David, programmed to love and brought into an ordinary family, in which he functions as a substitute for a beloved son, who is in coma:

Programmed to give and elicit love and affection, David is soon calling Monica "Mommy" and, bereft of her only natural child, she accepts and warms the artificial boy. (Clocksin, 2003, p. 1738)

Later on, the son recovers and the robot is abandoned. David then sets out on a journey during which he undergoes a process of development. Throughout the film, David wishes to be real to be able to evoke his mother's love, and the movie reflects upon issues of authenticity but at the same time also addresses what we become from entering into relationships with robots, which (whom) we choose to treat as inferior or slaves.

This line of argument is reflected in Gunkel's (2012) book *The Machine Question*, in which he establishes a framework for a thorough philosophical investigation of questions of whether machines are going to be held responsible to us for their actions and whether we are going to be responsible toward the machines' rights. Gunkel emphasizes that a demand for robot rights may arise from their social interactivity (Gunkel, 2012). Consequently, while we may argue about the status of intelligence, people have already started to form relationships with artificial companions. The fact that they look like us, combined with their growing ability to behave as social actors, will cause new forms of friending and bonding in connection with human-robot interaction. If we treat these artifacts as mere machines without questioning whether robots ought to be worthy or deserving of ethical considerations, we risk facilitating a

master–slave society. On this background, Gunkel advices us to (re)consider what kind of questions we ought to pay attention to in analyzing human–robot interaction.

"When god-like Odysseus", Leopold (1966, p. 237) recalls, "returned from the wars in Troy, he hanged all on one rope a dozen of slave-girls of his household whom he suspected of misbehavior during his absence. This hanging involved no question of propriety. The girls were property. The disposal of property was then, as now, a matter of expediency, not of right and wrong". As long as others – whether human, animal, machine or otherwise – are defined as mere instruments or the property of a ruling group, they can justifiably be used, exploited and dispensed with in a way that is purely expedient and beyond any moral considerations whatsoever (Gunkel, 2012, p. 108).

A contradicting viewpoint is reflected in revisiting Turkle's earlier mentioned concept of *the robotic moment* (Turkle, 2011, p. 49-ff.), which reflects her worry that when we turn to robots we might gradually lose sight of that trust – as opposed to the kind of supposedly risk-free relationships we may enter into with robots – is essential for the formation of flourishing human relationships. Consequently, Turkle (2011, p. 6) argues for human superiority in claiming the existence of fundamental obstacles for forming trustful relationships with robots. According to Turkle, our interactions with robots are biased by our tendencies to anthropomorphize technology, which lacks any kind of authenticity and represents nothing more than *a clever collection of* "as if" performances. In this sense, it does not make sense to ascribe moral agency or patiency to robots:

> At the robotic moment, we have to be concerned that the simplification and reduction of relationship is no longer something we complain about. It may become what we expect, even desire. In this book I have referred to our vulnerabilities rather than our needs. Needs imply that we must have something. The idea of being vulnerable leaves a lot of room for choice. There is always room to be less vulnerable, more evolved. We are not stuck. To move forward together – as generations together – we are called upon to embrace the complexity of our situation. We have invented inspiring and enhancing technologies, and yet we have allowed them to diminish us. The prospect of loving, or being loved by, a machine changes what love can be. We know that the young are tempted. They have been brought up to be. Those who have known lifetimes of love can surely offer them more. When we are at our best, thinking about technology brings us back to questions about what really matters. (Turkle, 2011, p. 295)

To elaborate on Turkle's observations, it might be valuable to discuss the concept of trustful relations as contrary to "risk-free" relations. Hence, it is generally acknowledged that trust is vital for the flourishing of human life and a precondition of any cultural ordering (Løgstrup, 1997; Rawls, 1999, p. 433); Fukuyama, 2003, p. 126). According to the moral philosophy of Løgstrup (1997), our fundamental human condition is rooted in the fact that life is interpersonal; we are mutually dependent on each other. Consequently, openness, in the sense of trusting, i.e. daring to risk ourselves in coming forward to meet the other, is a definitive feature of human co-existence and inherent in all communication. When we place trust in others, it involves genuine risk-taking, as we surrender ourselves to the other. Thus, the mutual dependence, from which the ethical in human life springs, can only be comprehended in understanding what is at stake in a given context between two given persons.

The structure of all human encounters, or any kind of approaching, presupposes a fundamental trust that the other will meet your request and carries within it a risk that you will be met with rejection or carelessness. Hence, we are obliged to act out of

consideration to the best of the other. Løgstrup is not addressing the kind of universality we find in Kant's (1785, p. A54) categorical imperative, rather Løgstrup points to that we find ourselves situated in a given context with an particular person in a specific situation – what is ethically important is, so to speak, consideration for the other in an abstract, which allows us to put our moral dispositions to test (Fink, 2005, p. 99). But, in entering into relationships with robots, we are not faced with having to surrender ourselves to the social robot. Even though the robot acts in a human-like way and displays emotions, there is nothing at stake besides "as-if" behavior caused by programmed internal states and executed so as to ensure adequate adaption to the surroundings. We know that this is the case about our relationship – the robot simulates, and we invest without cost.

By now, we have two strong lines of arguments regarding what interacting with social robots might mean and do to us. On the one side, Turkle worries that we will lose something important to our humanity and discover it too late – as the title of her book indicates: "Alone together – why we expect more from technology and less from each other" (Turkle, 2011). On the other side, Gunkel emphasizes that we ought to start to think about rights beyond living things, i.e. give moral standing to robots to frame possible ways in which we perceive of and respond to robots. Hence, it does not matter whether Paro, the seal, is "authentic", what matters is the fact that we will be affected by its social interactivity; grandfather will get emotionally attached to Paro and it will become an important part of his life. From this perspective, it seems inhumane to treat Paro with cruelty and, for instance, tear out its batteries in front of grandfather. A demand for moral rights may arise precisely from such social relationships.

In summarizing, *Artificial Intelligence, A.I.* encapsulates both perspectives; on the one hand, the robot boy, David, is obviously non-authentic, without emotions and even to some extend not able to imitate appropriate responses toward the surroundings (in one scene, he laughs loudly without sense of situational awareness). In this manner, the film expresses that there might be something morally disturbing about the whole idea of the kind of non-authentic substitute relationship the robot boy offers. On this background, we should go with Turkle and her concerns for "as if" relationships. But despite this, we feel sorry for David and find him worthy or deserving of ethical considerations when kids are bullying David and when his mother leaves him behind in a forest – once her real son has recovered and the conflict level between him and David is building up.

Both Gunkel and Turkle address what kind of persons we would like to be, but they stay in separate camps: in conclusion, one could say that Turkle emphasizes that personhood is what links moral agency to responsibility and moral patiency to a claim for rights. Gunkel uses the social relationships we enter into with robots as a springboard for asking what happens when robots grow into our lives and if our attachment to social robots implies that we ought to assign moral standing to machines? *Artificial Intelligence, A.I.*, manages to show us both perspectives.

5. Altruistic robots

As a research field, AI, was coined at the Dartmouth Conference in 1956 and Turing wrote his famous article *Computing Machinery and Intelligence* six years earlier (Turing, 1950), since the fifties, the classical mind–body problem has been reframed to account for the possibility or unfeasibility of robot and machine consciousness (Wisdom *et al.*, 1952; Scriven, 1953; Mackay, 1956; Ziff, 1959; Lacey, 1960; Culbertson, 1963; Putnam, 1964).

More recently, implementation of bottom-up learning is reflected in Rodney Brooks's (1991, p. 139) ideas about the emergence of intelligence as cognition without representation – "with strict reliance on interfacing to the real world through perception and action, reliance on representation disappears". By an incremental approach toward AI, Brooks seeks to avoid the frame problem (McCarty and Hayes, 1969; Dennett, 1988; Ford, 1996) and sets out to create artificial low-level perceptual and sense-motoric routines, which, put together, might eventually evolve into a significant level of intelligence. It sounds rather optimistic that conceptual complex mental activities, like daydreams, planning or deliberative judgment can be broken down into modules of such reflex mechanisms. But, according to a physicalist perspective toward AI, it could be argued that some kind of robot evolution is realizable and will bring about robots with minds of their own. Presumably, based on some initial framework serving as a parent adviser, such robots shall be capable of self-learning and adaption to new environments.

This gives rise to ideas about implementing a virtue ethical framework for growing artificial moral maturity:

> In robotics, so-called "friendliness theory" attempts to deal with this conundrum: rather than using any finite set of top-down rules or laws, intelligent machines should be programmed to be basically altruistic, and then use machine learning in various settings to create a kind of "best judgment" in how to carry out properly altruistic actions. This approach sidesteps the fundamental calculation and programming problem of how to account for a vast number of unforeseeable eventualities. (Lin *et al.*, 2008, p. 39)

To phrase moral maturity in the MacIntyrian way, as persons, we acquire virtues in our developmental path from dependent to independent reasoners. As such, we become independent practical reasoners through life experience and through the learning (including rule following behaviour) we get from care persons during our childhood. Consequently, we grow to become moral mature persons capable of expressing and enacting our values in a social context (MacIntyre, 1999, pp. 158-159). If we should ever happen to succeed in creating virtuous machines, we will have to deal with a consciousness robot with free will, intentions, goals, self-interest, that is, a moral agent with a justified claim for rights. If robots have this status, we are committed to treat them as moral beings in their own right even though they might turn out to develop a moral sense beyond our understanding.

From a family-friendly angle, an altruistic robot is presented in *Wall-E*, by Stanton (2008). Together with a cockroach, the garbage collecting robot Wall-E is left behind on a polluted earth, from which humans have escaped in big space shuttles. While waiting for everything to get better on the earth, which humans seem to assume will happen all by itself (almost like nowadays attitude toward issues of global warming), they spend their days engaging in superficial activities, quiet similar to a satirical version of American cruise tourists. Hence, they eat (too much), they sit (too much), i.e. they are driven around in flying "arm chairs" with futuristic Ipad-like devices placed right in front of their heads allowing them to take in (too much and too silly) information, which also suggests that interpersonal encounters are spare. Furthermore, due to severe obesity and physical inactivity, they are only able to crawl. Moreover, we can easily imagine that this life style is not doing anything good for the power of sexual reproduction and that this might endanger the survival of the human species. The seamless integration of humans and technological artifacts has turned out not to be of any benefit to humanity, rather humans have become like grown-up helpless babies.

Here, *Wall-E* tells the story of a future world in which love between two robots, Wall-E and Eva (a scout robot, which is sent back to earth from the space shuttle to explore, whether it is possible for humans to return to the earth), becomes a unique catalyst for reestablishing a once lost flourishing human world. The message is that altruistic robots may succeed in saving us from ourselves, in the feel-good sense of the term, i.e. not by trying to take over and deprive us of our autonomy, like expressed above in *I, Robot*.

6. Surveillance

So far, the article has dealt with AI agents in the shape of social robots, but other sci-fi themes are of course also worth mentioning, such as the surveillance theme reflected in the movie *Nineteen Eighty-Four*, by Michael Radford (1984), which is based on the dystopian novel by George Orwell, and describes a totalitarian society with no room for freedom and no ability for building trustful relationships.

Stressing our context of the ETHICOMP conference series' timeline, we shall be looking closer into a more recent version of the surveillance theme. Hence, in *Minority report*, by Steven Spielberg (2002), crime is eliminated from Washington due to the use of advanced knowledge-discovery technology in the shape of three pre-cogs, which operate by means of a kind of distributed cognition allowing them to look into the future and come up with precise crime predictions. Hence, in 2054, law enforcement agencies have the capabilities to respond to criminal activities even before they actually take place, and *Minority report* highlights what is at stake in cases in which knowledge-discovery technologies may infringe upon individuals and groups control of how information relating to them, their actions and their relations are represented.

As an example of a knowledge-discovering technology, the currently ongoing EU project ePOOLICE (www.epoolice.eu) aims at developing an efficient and effective environmental scanning system as part of the early warning system for the detection of emerging organized crime threats and changes in *modus operandi*. The purpose is to be able to meet the future challenges of fighting organized crime by ensuring that law enforcement are capable of engaging in proactive planning at the strategic level.

For the moment being, the current situation and priorities in fighting organized crime can be summarized as follows:

> The SOCTA, 2013 [EU Serious and Organized Crime Threat Assessment] identifies a number of key priorities, which, in Europol's view, require the greatest concerted action by EU Member States and other actors to ensure the most effective impact on the general threat. Facilitated illegal immigration, trafficking in human beings, synthetic drugs and poly-drug trafficking, Missing Trader Intra-Community (MTIC) fraud, the production and distribution of counterfeited goods, cybercrime and money laundering are the particular crime areas listed in this category. (SOCTA, 2013, foreword by Director of Europol, Rob Wainwright)

But it is also important to be able to pay attention to possible future crime scenarios, which may be lurking in the horizon. Hence, in the ePOOLICE system, environmental scanning is carried out on data streams from a variety of online public sources, focusing on *modus operandi* and crime trends, not on individuals. Different kinds of public online data streams feed into the system's knowledge repository that provides a rich taxonomy of domain knowledge. Moreover, data processing is facilitated by means of information fusion techniques, enabling the extraction of descriptive and predicative meanings used for inferring hidden states, i.e. weak signals or indicators of organized crime activities.

In general, such environmental scanning systems for early warning operate at the strategic level of open source intelligence and do not aim at identifying individuals. Nevertheless, ethical and technical issues related to potential privacy infringements in public online contexts still have to be addressed to establish a privacy preserving framework, which ascribes privacy protection into the given system design (Gerdes et al., 2013; Gerdes, 2014).

In relation to the development of environmental scanning systems for early warning, it is crucial to balance security at a societal level with citizens' right to privacy – in short, to balance data privacy and data utility, which is by no means a trivial task. Still, and in line with the ETHICOMP spirit, it can be done by engaging in "engineering activism" (a term coined by Nissenbaum, 2001), which implies that to deal with the ethical impact of systems in context, one must take a pro-active stance toward the challenge of bringing ethics to design:

> Humanists and social scientists can no longer bracket technical details – leaving them to someone else – as they focus on the social effects of technology. Fastidious attention to the before-and-after picture, however richly painted, is not enough. Sometimes a fine-grained understanding of systems – even down to gritty details of architecture, algorithm, code, and possibly the underlying physical characteristics - plays an essential part in describing and explaining the social, ethical, and political dimensions of new information technologies. (Nissenbaum, 2001)

7. Concluding remarks

Sci-fi movies are not about predicting the future, rather they reflect the issues worth asking in our present time. On this background, this article casts light upon the broad spectrum of moral questions that may arise from human technology interactions, particularly focusing on themes addressed in sci-fi movies within the ETHICOMP conference series' timeline.

Hence, sci-fi movies conceptualize philosophical questions in illustrating what assigning agency or patiency to artificial agents means for our being in the world. In conclusion, sci-fi movies encapsulate how we are shaped through the kind of relationships we enter into with robots. Moreover, AI agents can also be embedded in intelligent systems, as illustrated in *Minority Report*, which gives rise to address issues of intelligent surveillance technologies. Here, concerns related to the increasingly pressure upon privacy, due to the introduction of advanced knowledge-discovery technologies, are illustrated and the article argues that, as IT-ethicists, we should not only take an applied approach to ethics, i.e. engage in after-the-fact critical analyses. Rather, whenever possible we ought to act proactively and "do ethics" by getting involved in the developmental phases of designing technology.

References

Arkin, R.C. (2007), "Governing lethal behavior: embedding ethics in a hybrid deliberative/reactive robot architecture", Technical Report GIT-GVU-07-11, available at: www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf (accessed 4 March 2015).

Bartneck, C., Van der Hoek, M., Mubin, O. and Al Mahmud, A. (2007), "Daisy, daisy, give me your answer do!", Switching off a Robot, *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction, Washington DC*, doi: 10.1145/1228716.1228746, pp. 217-222.

Benford, G. and Malartre, E. (2007), *Beyond Human – Living with Robots and Cyborgs*, A Forge Book, New York, NY.

Benso, S. (2000), *The Face of Things: A Different Side of Ethics*, SUNY Press, New York, NY.

Brooks, R.A. (1991), "Intelligence without representation", *Artificial Intelligence*, Vol. 47 Nos 1/3, pp. 139-159.

Carpenter, J. (2013), "Just doesn't look right: exploring the impact of humanoid robot integration into explosive ordnance disposal teams", in Luppicini, R. (Ed.), *Handbook of Research on Technoself: Identity in a Technological Society*, Information Science Publishing, Hershey, PA, doi: 10.4018/978-1-4666-2211-1

Clocksin, W.F. (2003), "Artificial intelligence and the future", *Philosophical Transactions: Mathematical, Physical and Engineering Sciences, Information, Knowledge and Technology*, Vol. 361 No. 1809, pp. 1721-1748.

Culbertson, J.T. (1963), *The Mind of Robots*, University of Illinois Press, Chicago, IL.

Dennett, D.C. (1988), "Cognitive wheels: the frame problem of AI", in Pylyshyn, Z.W. (Ed.), *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Ablex, NJ, pp. 41-65.

Fink, H. (2005), "Etikkonceptionen i Den etiske fordring", *Slagmark*, No. 42, pp. 89-101.

Floridi, L. (2014), *The Fourth Revolution – How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford.

Floridi, L.C. and Sanders, J.W. (2004), "On the morality of artificial agents", *Minds and Machine*, Vol. 14 No. 3, pp. 349-379.

Fodor, J. (1998), *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*, MIT Press, Cambridge, MA.

Ford, K.M. (Ed.) (1996), *The Robot's Dilemma Revisited*, Ablix, NJ.

Fukuyama, F. (2003), *Our Posthuman Future – Consequences of the Biotechnology Revolution*, Picador, New York, NY.

Gerdes, A. (2014), "A privacy preserving design framework in relation to an environmental scanning system for fighting organized crime", in Kimpaa, K., Whithehouse, D., Kuusela, T. and Phahlamohlaka, J. (Eds), *ICT and Society 11th IFIP TC 9 International Conference on Human Choice and Computers, HCC11 2014*, 30 July-1 August, Turku, Springer, pp. 226-239.

Gerdes, A., Larsen, H.L. and Rouces, J. (2013), "Issues of privacy and security in relation to an environmental scanning system for fighting organized crime", *Flexible Query Answering Systems – 10th International Conference, FQAS 2013 Proceedings*, September, Granada, Springer LNAI, 8132, pp. 155-164.

Gunkel, D. (2012), *The Machine Question – Critical Perspectives on AI; Robots, And Ethics*, MIT Press, MA.

Kant, I. (1785), (Akademieausgabe, vol. IV) Grundlegung zur Metaphysik der Sitten, available at: www.korpora.org/Kant/aa04/Inhalt4.html (accessed 4 March 2015).

Lacey, A.R. (1960), "Men and robots", *The Philosophical Quarterly*, Vol. 10 No. 38, pp. 61-72.

Lin, P., Bekey, G. and Abney, K. (2008), "Autonomous military robotics: risk, ethics, and design", *CALPOLY*, available at: http://ethics.calpoly.edu/ONR_report.pdf (accessed 4 March 2015).

Løgstrup, K.E. (1997), *The Ethical Demand*, University of Notre Dame Press, Notre Dame.

McCarty, J. and Hayes, P.J. (1969), "Some philosophical problems considered from the standpoint of artificial intelligence", in Meltzer, B. and Michie, D. (Eds), *Machine Intelligence*, Elsevier, New York, NY, Vol. 4, pp. 463-502.

McDermott, D. (2008), "Why ethics is a high hurdle for AI", *North American Conference on Computers and Philosophy (NA-CAP)*, Bloomington, IN, July.

MacIntyre, A. (1999), *Dependent Rational Animals – Why Human Beings Need the Virtues*, Carus Publishing Company, Chicago, IL.

Mackay, D.M. (1956), "The epistemological problem for automata", *Automata Studies*, Princeton University Press, Princeton, pp. 235-251.

Nagel, T. (1974), "What is it like to be a bat?", *The Philosophical Review*, Vol. 83 No. 4, pp. 435-450.

Nissenbaum, H. (2001), "How computer systems embody values", *Computer – Innovative Technology for Professions*, Vol. 34 No. 3, pp. 117-119.

Putnam, H. (1964), "Robots: machines or artificially created life?", *The Journal of Philosophy*, Vol. 61 No. 21, pp. 668-691, available at: www.jstor.org/stable/2023045 (accessed 4 March 2005).

Rawls, J. (1999), *A Theory of Justice – Revised Edition*, Oxford University Press, Oxford.

Sparrow (2004), "The turing triage test", *Ethics and Information Technology*, Vol. 6, doi: 10.1007/s10676-004-6491-2, pp. 99-107.

Schärfe, H. and Dougerthery, E.A. (2011), "Initial formation of trust: designing an interaction with Geminoid-DK to promote a positive attitude for cooperation", *Social Robotics: Third International Conference, ICSR 2011*, LNAI7072, Springer, Aalborg, pp. 95-103.

Scriven, M. (1953), "The mechanical concept of mind", *Mind*, Vol. 62 No. 246, pp. 230-240.

Singer, P.W. (2009), "Robots at war: the new battlefield", *The Wilson Quarterly*, Vol. 33 No. 1, available at: www.jstor.org/stable/40262238 (accessed 4 March 2015).

SOCTA report (2013). "SOCTA 2013 – EU serious and organized crime threat assessment", available at: www.europol.europa.eu/sites/default/files/publications/socta2013.pdf (accessed 4 March 2015).

Turing, A. (1950), "Computing machinery and intelligence", *Mind*, Vol. 59 No. 256, pp. 433-460.

Turkle, S. (2011), *Alone Together – Why We Expect More From Technology and Less From Each Other*, Basic Books, New York, NY.

Verbeek, P.P. (2011), *Moralizing Technology: Understanding and Designing the Morality of Things*, The University of Chicago Press, Chicago, IL.

Walzer, M. (1977), *Just and Unjust Wars: A Moral Argument with Historic Illustrations*, Basic Books, New York, NY.

Wisdom, J.O., Spilsbury, R.T. and Mackay, D.M. (1952), "Symposium: mentality in machines", *Proceedings of the Aristotelian Society, Supplementary Volumes*, Men and Machines, Vol. 26, pp. 1-86, available at: www.jstor.org/stable/4106628?origin=JSTOR-pdf (accessed 4 March 2015).

Ziff, P. (1959), "The feelings of robots", *Analysis*, Vol. 19 No. 3, pp. 64-68.

Corresponding author
Anne Gerdes can be contacted at: gerdes@sdu.dk