



# UNIVERSITY OF TWENTE.

Faculty of Behavioural, Management and  
Social sciences

## *Cinematic ethics on Robots*

*or: what can we learn from Science-Fiction  
movies about humanoid (service) robots.*

**Peter Segers**  
**M.Sc. Thesis**  
**Juli 2017**

---

**Supervisors and examiner:**

M.H. Nagenborg (supervisor);  
F. Santoni de Sio (external advisor);  
N. Gertz (examiner)

Philosophy of Science, Technology and Society  
Faculty of Behavioural, Management and Social sciences

University of Twente

P.O. Box 217

---

**Abstract:**

In the near future it becomes increasingly more likely that (service) robots will have a large impact on society a serious technological possibility. In order to gain insights what this impact might entail, this thesis will work out how the development of robots and Science-Fiction are entwined. This will be continued by an exposition of the current uncertainties in the discussion on robot-ethics. To find answers on these uncertainties, this thesis will conduct four explorative analyses. These analyses will focus on the development of human-robot relations as depicted in a selection of Science-Fiction movies. In order to conduct these analyses the framework on the *Roboethcis of appearance & Roboethics of good and experience/imagination* as composed by Coeckelbergh (2009) will be used, this framework focusses mainly on *alterity* relations. These *alterity* relations are in turn based on the distinction, machine and “quasi-other” made by Ihde (1990). This however will be expanded upon by the work of Van den Berg (2010) in order to include the experience of “genuine other”. The findings of these analyses will be discussed in the last chapters of this thesis. Here will be discussed how to these findings do compliment, fit or bring new arguments to the broader discussions on robot ethics.

## **1. Introduction**

### ***1.1 Outline Thesis***

When looking at the possible near future service robots become a serious technological possibility, and it becomes increasingly likely that they will have a large impact on society. In order to gain insights into what this impact might entail, this thesis will conduct four explorative analyses. These analyses will focus on the development of human-robot (H-R) relations as depicted in a selection of Science-Fiction movies.

Before examining the selected Science-Fiction movies on their portrayal of H-R interactions, the development of and growth in usage of (service) robots within society will be introduced. This will be followed by an explanation of how Science-Fiction movies can portray useful ideas about this subject. In the following (sub-)chapters firstly what and how the growth in usage will most likely look like will be discussed and secondly the term “service robots” will be defined. Afterwards this thesis will go more in depth on the influence of science-fiction on robot technologies, where primarily the works of Bassett, Steinmueller & Voss (2013) will be discussed. This will be followed up with an section about the current normative landscape concerning (future) human-robot interaction/relationships and a section on how Science-Fiction contributes towards the discussions in this ethical landscape.

After this thesis will continue with the justification and exposition of the chosen approach in analysing the selection of Science-Fiction movies. Here the focus on humanoid (service) robots and selection of movies for the analyses will be introduced and justified, the four movies that were selected are: *Chappie* (2014), *Ex machina* (2014), *I, Robot* (2004), and *Automata* (2014). In addition, the post-phenomenological framework of Mark Coeckelbergh (2009) will be introduced. This framework introduces the approaches *roboethics of appearance & the roboethics of good and experience/imagination*, which will be used in the analysis of the selected movies.

Having worked this all out, this thesis will turn to the explorative analyses of the Science-Fiction movies in this thesis. The last chapters of this thesis will be dedicated towards the findings of the movie analyses, and how they complement, fit or bring new things to the broader discussions on robot-ethics.

### ***1.2 Increase of robot usage & human-robot relations in society***

Throughout Western society, service robots seem to become more and more pervasive. Where firstly (service) robots were mere products of Science-Fiction and imagination, they now seem to become a reality. As argued by Barthneck C. (2004) the integration of robots in society is very likely, based on the ethical (Dennett, 1997) and legal (Lehman-Wilzig, 1981) premises this might very well be the case. The argument posed by Dennett, is that society is beginning to muse on how artificial intelligence can be held morally accountable for harming others. Whereas Lehman-Wilzig investigates the jurisprudential principles that would have to underlay the legal framework to hold artificial intelligence accountable.

To understand what this prevalence and likely integration of service robots in society could mean, one has to define what a service robot is. This however is quite a tricky issue, for one should keep in mind what is understood as robots. In a broad definition, one could include “smart cars”, “Predator-drones” and/or industrial robots into the definition. This seems a bit

off, for these technologies do not fall under what is culturally and commonly understood as being a “robot”. To get to a definition that one can work with, this thesis will turn to a broader definition of service robots. The International Service Robot Association (ISRA) <sup>1</sup> used the following working definition by Pransky:

*“Machines that sense, think, and act to benefit or extend human capabilities and to increase human productivity”* (Pransky, 1996).

This definition incorporates several distinctive criteria. Firstly, it states that service robots are machines that “*sense, think, and act*”. Service robots therefore have some sort of AI (this is actually one of the basic definitions of robots). Secondly, this statement mentions “... *to benefit or extend human capabilities and to increase human productivity*”. This limits the scope of robots which fall under this definition to the extent in which they can positively contribute to goals set by humans (which still does include industrial robots). To get an even more focused definition, the more recent definition set by the International Federation of Robotics (IFR) can be used.<sup>2</sup> The IFR defines service robots as following:

- *A robot is an actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks. Autonomy in this context means the ability to perform intended tasks based on current state and sensing, without human intervention.*
- *A service robot is a robot that performs useful tasks for humans or equipment excluding industrial automation application.*

In addition, the IFR groups service robots in different categories to according to different roles. These robot roles as described by the IFR are: professional service robots and personal/domestic service robots.

- *A personal service robot or a service robot for personal use is a service robot used for a non-commercial task, usually by lay persons.*
  - *A professional service robot or a service robot for professional use is a service robot used for a commercial task, usually operated by a properly trained operator. In this context an operator is a person designated to start, monitor and stop the intended operation of a robot or a robot system.*
- (IFR Service Robots, n.d)

When looking at the work of Teresa (2013), there is another distinction within the definition of service robots. Teresa delves deeper into the definition by splitting it into two: she separates “personal service robots” from “field robots”. For this separation she used the

---

<sup>1</sup> ISRA is an individual and corporate member of the Robotic Industries Association (RIA), dedicated to providing information on the emerging field of service robots in a broad range of applications. ISRA activities include publication of newsletters, sponsorship of conferences, exhibits, and distribution of market studies, books, and related resources. (Dowling, 1995)

<sup>2</sup> The IFR is an organization established in 1987 in connection with the 17th International Symposium on Robotics ISR as a non-profit organization by robotics organizations from over 15 countries. Its purpose concerned with promoting and strengthening the robotic industry on a worldwide scale.

location in which the robots are used as the ground for being different. The “personal service robots” are here employed within quasi-structured environments, which might not be fully adjusted to the robot’s functionality. “Field robots” on the other hand are employed in natural, fully unstructured settings. These natural settings range from forest and jungles to sea bottoms, mountains and even the sky. According to Teresa “field robots” represent the category for professional service robots. With the help of the abovementioned theories the working definition of service robots in this thesis can be formulated. In this thesis *service robots* are defined as:

*Service robots are able to perform tasks that benefit or extend human capabilities and do increase human productivity, excluding industrial automation application.*

*In addition, service robots can be grouped into personal service robots and professional service robots. Here personal service robots are used for non-commercial tasks by non-professional users in quasi-structured environments. Professional service robots are used for commercial tasks by trained operators in often fully unstructured settings.*

Having laid out this definition, we can return to the growth in prevalence and integration of service robots into society. According to Teresa (2013), Joseph Engelberger<sup>3</sup> predicted that service robots would become by far the largest type of robots within society. This prediction seems to have become true, for according to the IFR, a significant growth occurred in the market of professional and personal/domestic robots during 2014 (IFR Service Robots, n.d). The IFR furthermore predicts that the market for service robots will tremendously grow in the period of 2015-2018, with 152.400 new professional use service robots to be installed and approximately 35 million personal/domestic service robots to be sold. This prediction points at the optimistic ideas of the increase in service robot usage, which ranges from expert use towards everyday mundane household usage. This quite large (predicted) growth in presence of service robots allows for people to create an increasing number of relations with these robots.

Another approach to understand the growth of service robots highlights the more active influence of society in shaping robots. This approach, as explored by Šabanović (2010), argues that society and technology, in this case service robots, mutually shape each other. This implies that the cultural values, norms and expectancies from designers and users do impact the design and implementation of service robots. This places society and users in a much more active role where they can and do contribute in the creation of technologies.

As for the integration of service robots into society, there are both some quite old and some quite recent stances on this process. First of all, there are the earlier mentioned premises on the ethical (Dennett, 1997) and legal (Lehman-Wilzig, 1981) dimensions. These studies show that at least society is preparing to be able to deal with robots as serious members of society. This preparation for robotic members of society is taken to a whole new level in

---

<sup>3</sup> Joseph Engelberger was an American physicist, engineer and entrepreneur. Engelberger co-developed with George Devol the first industrial robot in the United States, the Unimate, during the 1950s. Later on, he worked as entrepreneur and became an advocate of robotic technology in a large range of fields. ([https://en.wikipedia.org/wiki/Joseph\\_Engelberger](https://en.wikipedia.org/wiki/Joseph_Engelberger)).

South Korea where they aim to have a robot in every home by 2020 (A Robot in Every Home by 2020, South Korea Says, 2010). In cooperation with the Samsung company, the South Korean government officials aim to mass-produce service robots in order to make them “ubiquitous robotic companions, or URC”. The tasks for these URC’s are planned to range from entertainment to education, home security, and even household chores. This project has sparked interest in other technological developed countries such as Japan, China and the USA, who all have begun their service robot campaigns. This program is still relevant as “FURo-i Home”<sup>4</sup> was presented at the 2015 Consumer Electronics Show (CES) in Las Vegas (Kelion, 2015). A similar anticipation of the integration of service robotics can be seen in the Japanese Aichi Expo of 2005, where robots were presented as a significant part of every-day modern life. Visitors were given a chance to interact with more or less one hundred different robots. The exposition proudly proclaimed that “we live in the robot age”, signalling the anticipation of the integration of robots into human society.

Another reaction comes from the owner of Microsoft, Bill Gates. In his aptly named article “A robot in every home” (Lovgren, 2006), Bill Gates discusses the problems and possibilities of robotics. He describes how he envisions that similar to the personal computer (PC), soon service robots will be also be an integral part of every home. This vision was supposed to be realized with the help of Microsoft Robotics, which does no longer exist<sup>5</sup>. He envisions robots as PCs that will get up from the desktop and allow us to see, hear, touch and manipulate objects in places where we are not physically present. Other experts seem to share Bill Gates visions (to more or lesser extent) on how robots will become quite ubiquitous within society. With these experts, also other tech companies, namely Amazon.com and Google, are developing and implementing big plans for robots (Guizzo, 2014). Amazon has bought the kiva systems company in order to buy “a lot” of robots and create fully robot staffed warehouses. Google on the other hand has bought and funded robotic companies in order to have a big role in their development.

If we return to the notion of Šabanović (2010), that society and technology mutually shape each other, one can see how this might take place with service/domestic robotics. For not only are experts anticipating a large growth in service/domestic robotics, we see that large companies and governments (independent and in cooperation with each other) are developing robots with human society as their intended working environment. This suggests that with the large (expected) increase in use and availability of robots in society, we might see a surge in human-robot relations.

### **1.3 About “strong” AI**

Having described the development service robots and human-robot relations in society, there is another subject that must be worked out and defined in order to make this thesis more comprehensible. This subject is the discussion on AI and the distinction between “weak” and “strong” AI.

---

<sup>4</sup> The FURo-i Home Robot is a domestic service robot, equipped with telepresence technologies. It is being manufactured by the Korean company, FutureRobot.

<sup>5</sup> As part of restructuring the company, Microsoft decided to shut down its robotics group in 2014. However, all is not lost as the head of the project (T. Trower) decided to leave and found his own robotics start-up, Hoaloha Robotics.

In order to understand what is meant by this distinction of “weak” and “strong” AI, this thesis will first to the work of John Searle (1980) where he coined these two terms. Here Searle differentiated two hypotheses:

- *The “strong” AI hypothesis*, presumes that an AI can think and have a mind;
- *The “weak” AI hypothesis*, presumes that an AI can (only) act like it thinks and has a mind.

Here “*strong*” is used as in that it assumes a stronger statement, for this term hints at that something special goes on in the machine that goes beyond all its abilities that we can test. Searle used these hypotheses in his “Chinese room” argument against the notions of AI as understood by Turing (1950).

Turing’s notion of AI differs quite a bit in that of Searle as he stated: “*We need not decide if a machine can “think”; we need only decide if a machine can act as intelligently as a human being. This approach to the philosophical problems associated with artificial intelligence forms the basis of the Turing test.*” (Turin, 1950). Thus in essence in Turing’s work the two hypotheses of Searle do coincide, for that it is enough for an machine to be perceived as having a “mind” in order to ascribe one to it. This reasoning can also be found in the works of Kurzweil (2005), for he uses the term “strong AI” to describe any artificial intelligence system that acts like it has a mind, regardless of anyone would be able to determine if it actually has a mind or not. This reasoning takes the argument back to the arguments of Turin, for he defends his position by stating that “minds” in general are unprovable. However, we do ascribe minds to humans (and by lesser extend other beings) on basis on their behaviour. Thus for a machine to act in a way that is similar/ indistinguishable from human behaviour, is the only ground we have to possible determine “strong” AI.

With worked out this there can be made an distinction how the terms “weak” and “strong” AI will be used throughout this thesis. Here “*strong*” AI will be used to describe an AI that is perceived as and/or can be reasonable defended as acting in such a way that is indistinguishable from human behaviour. Continuing, this would also indicate that this “*strong*” AI can be considered as having a “mind”.

In contrast, “*weak*” AI will be used in this thesis to describe an AI that does not behave in such a way that is could be considered similar to humans and therefore having a “mind”.

#### ***1.4 The influence of Science-Fiction on robots (development)?***

A large part of the development of (service) robots can be traced back to science-fiction, for science-fiction does have a significant role on how people perceive robotics in general and influences the development of robotics outside fiction. In the case of robots, fictional accounts and actual developments in robotics seem to co-evolve as they influence and build upon each other.

In the paper “*Better Made Up: The Mutual Influence of Science fiction and Innovation*” Bassett, Steinmueller & Voss (2013) discuss the different relationships between Science-Fiction and technological developments, which they describe as “*one of mutual engagement and even co-constitution*”. In this thesis a framework for tracing the relationships

between real world science and technology and innovation and science fiction is developed. First of all, they introduce the argument that Science-Fiction contains particular kinds of subject matter that organize it according to particular aesthetic and textual strategies and deliver it with particular kinds of force.

Bassett, Steinmueller & Voss (2013) conclude that there are several important factors in this mutual engagement through which Science-Fiction and technological development influence each other. The first factor they bring forth is that Science-Fiction (the English Science-Fiction tradition) has a history of growth and expansion into new and different media. This history starts with the reaction to the developments brought by industrialization and the newly born hopes and fears at the beginning of the 20th century, but eventually moved beyond this when other technological developments took over.

The second factor in this mutual development is that Science-Fiction involves an audience (even many audiences) through which ideas and expectations of technologies can be influenced. Here Bassett, Steinmueller & Voss (2013) differentiate between the general public and the expert community. Here Science-Fiction can shape the general public's understanding and expectations about science. Science-Fiction does influence public apprehensions, often by cultivating and reinforcing already existing fears and concerns (Kirby, 2003). However, Science-Fiction is also a tool to create new ideas and values about scientific developments in the general public. For the expert community, Science-Fiction can be taken as inspiration for actual research towards new possibilities or even as inspiration to set up projects to realize still imaginary technologies. In this way Science-Fiction can influence the expert community to realize itself.

The described third factor is that Science-Fiction more than just inspiration for the expert community, as described by Bassett, Steinmueller & Voss (2013) it can be used as an enabling "*space*" for technological innovation. Even though this is not always the case, creating a "*sense of wonder*" in the audience is a broadly shared aim in Science-Fiction. In the contexts created by its history and audience, this is the main source of influence for Science-Fiction. Giving the audience experiences unfamiliar to their everyday life, and with this forcing them to change their worldview, can provide the basis for the creative vision and strengthening confidence in the possibility of this change. These are fundamental conditions for innovation and entrepreneurship, which in turn lead can lead to technological developments.

The fourth factor of influence is that Science-Fiction embodies desire. Science-Fiction can be described as being motivated by a desire for better and different futures. This drive is entangled with what is, was or might be part of the human experience. These desires do manifest itself in the world of technologies and of innovation, and this is desire for new ways to manifests drives Science-Fiction, establishing interplay of mutual influence.

And lastly there is the fifth aspect which concerns Science-Fiction as a resource for discussing shared meaning and ideas about scientific and technological developments, especially in the way how they take place within the broader culture. This source of influence is not meant to be any form of authority, but is a consequence of establishing discussion on "*what is and what might be*". Although this influence is preliminary to thoughtful discussion about "*what is to be done*", as a popular medium Science-Fiction does allow for an inclusive in influence that can establish productive discussion.



When taking these five factors to assess the influence of Science-Fiction on the development of robots, there can be concluded that they do share a strong interaction. One early example of this interaction is the story of *Waldo* by Heinlein (1942). In this story a physically disabled inventor, Waldo F. Jones, uses remotely controlled mechanical arms. This story did inspire developers within the nuclear industry to create mechanical arms for handling hazardous materials and named them Waldo (referring to the inspiration). This influence also works the other way around, for non-fictional developments in robotics have inspired many (science-) fiction works. This reversed influence can be illustrated by the quite recent movie “*Ex Machina*” (2014), which incorporates current ideas and developments (albeit exaggerated) in order to create a new story about robotics. These examples of robots in (science-) fiction highlight the interweaving of (science-) fiction and nonfiction progress.

Another early exemplar of this influence can be traced back to the works of the brothers Capek. Joseph Capek wrote in 1917 a short story *Opilec* describing “automats” and 1921 when his brother Karel Capek wrote the play *Rossum’s Universal Robots* (RUR). In this play the term robot was first conceived (Hockstein, Gourin, Faust, & Terris, 2007). Here the term robot was derived from the Czech word “*robot*”, which can be roughly translated to (forced) laborer or slave. Karel Capek actually wrote this play as a protest to the (in his view) increasing growth of modern technology and mistreatment of workers. Thus he illustrated a development of robots with increasing capabilities, which served as an analogy to dehumanization of workers and their back breaking work. Here robots served as a metaphor of a simplified man instead of a complex machine. This makes Capek’s play more a political satire instead of a technological narrative/prediction.

However, Capek’s plays were interpreted (read: misunderstood) by the audience at large as a metaphor of high-tech that will destroy mankind. This high-tech would eventually turn against humanity with, in this case taking the shape of the eventual revolution of robots against humans (Horáková & Kelemen, 2003). With this misinterpreted narrative, Karel Capek “brought” the fascination with robots and the idea of their possible danger to the public. This fascination and the idea of robots “overthrowing” humans bear relevancy to this day. The more recent development of (service) robots seems to also closely correspond to the initial ideas about robotic functioning (not being) which Karel Capek described in his play. The robots from Capek were strikingly humanoid and were designed to “serve” humans and seem to fit the definition of service robots by Pransky (1996) This can also be found in the design of other robots in- and outside of fiction, for they almost always are designed for a specific function which humans cannot or are unwilling to do (either being too difficult, too dangerous or too mundane).

One striking difference Capek’s robots have with the contemporary cultural understanding of robots is their material composition. For in the play *Rossum’s Universal Robots*, *these robots are constructed from “..Some kind of science, and some strange kind of colloidal jelly... (that) not even a dog would eat”* from which the robots would grow (Horáková & Kelemen, 2006). For a large part due to technological developments, the idea of robots as mechanical being, composed of “*Cogs and wheels*”, was popularized by the classical expressionistic Science-Fiction movie *Metropolis* (1927). This still is one of the most influential (silent) Science-Fiction movies to this day. This “*Cogs and wheels*” definition of robot bodies brought back the more “traditional” idea about robots as understood with “*automata*”.

The main premise of *Metropolis* is quite similar to that of *Rossum's Universal Robots*, for that there is a futuristic setting where there is a huge distance between “workers” and “owners” which eventually leads towards a revolution (Horáková & Kelemen, 2003). The solution to this revolution is again in both narratives similar, in that they are found on the “spiritual” level, but here *Metropolis* is quite bit more optimistic. The defining difference between the two narratives are the robots within these stories, because it is the more traditional definition of robots (mechanical being) in *Metropolis* that still wields more significance to this day. In *Metropolis* the definition of a robot was most pronounced in the character “Hell”, a robot with a “*body composed of steel*” and “*cogs and wheels*” designed to resemble a woman on the outside, but being the opposite on the inside. Whereas Maria, the woman Hell was designed as, did embody all kinds of Christian virtues, Hell was her total opposite embodying the notion of a worker revolution, chaos, destruction, evil, and as an instigator of the dark side of the human character. Hell’s design turned out to be greatly influential and became the “standard” for future robot designs. This design has survived up to this day also thanks to technologies used in case of present day robots (Horáková & Kelemen, 2006).

Another classic and influential narrative is found in the works of Isaac Asimov<sup>6</sup>. Throughout his works, but first being specifically being introduced in the story *Runaround* (1942), Asimov used a set of robotic principles which were designed to create a “new” robot narrative. This “new” robot narrative was to replace, according to Asimov, the old robot narrative where robots would “*turn stupidly on his creator for no purpose but to demonstrate, for one more weary time, the crime and punishment of Faust*” (Asimov, 1964). These robotic principles were written down as the “*Three laws of Robotics*” and are the following:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm;
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law;
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws; (Asimov, 1942)

With these principles, Asimov posed a new framework on the public for robots, one that is still influential.

Even though Asimov designed this framework as a reaction against the notion of human hubris in creating robots, he did pioneer in and influence the notion of (pre-) programmed machine ethics. More than just being inspiration for robot names<sup>7</sup>, Asimov is still relevant for developments in (the understanding of) robot ethics. For example, the South Korean government announced in 2007 that they would issue a “Robot Ethics Charter,” which would set standards for both users and manufacturers of robots. These standards would reflect

---

<sup>6</sup> Isaac Asimov was a Russian/American author and professor of biochemistry at Boston University, best known for his science fiction works and for his popular science books. Asimov wrote hard science fiction and he was considered one of the “Big Three” science fiction writers during his lifetime. Asimov's most famous work is the *Foundation* Series; his other major series are the *Galactic Empire* series and the *Robot* series ([https://en.wikipedia.org/wiki/Isaac\\_Asimov#Science\\_fiction](https://en.wikipedia.org/wiki/Isaac_Asimov#Science_fiction)).

<sup>7</sup> The Japanese company Honda introduced on 21 October 2000 their humanoid robot ASIMO. ASIMO is an acronym for Advanced Step in Innovative Mobility, but does obviously also draw inspiration from Asimov.

Asimov's Three Laws. With this they attempt to set ground rules for future robotic development (Robotic age poses ethical dilemma, 2007).

A less faithful but nonetheless influenced adaptation of this framework comes from David Langford (2009)<sup>8</sup>, as a reaction to the development and deployment of military drones by the USA, which states the following set of laws:

1. A robot will not harm authorized Government personnel but will terminate intruders with extreme prejudice;
2. A robot will obey the orders of authorized personnel except where such orders conflict with the Third Law;
3. A robot will guard its own existence with lethal antipersonnel weaponry, because a robot is bloody expensive;

This quite tongue-in-cheek formulation serves a criticism towards unlikeliness that the U.S.A. military robot manufacturers will develop an ethical framework comparable to the original three laws. This however does show how much of an influence Asimov's novels still have on the cultural understanding of how robots are being understood by the broader public. With this we can see how classic works of Science-Fiction have influenced and still influence the cultural understanding and development of robots. Where the plays of Karel Capek introduced and popularized the term "robot" and its subservient (and possibly dangerous) relation to humans. The movie *Metropolis* expanded this notion but deviated in how these robots were composed. This "*Cogs and wheels*" composition sets the definition, mainly due to technological developments, of how people expect robot bodies to be ("*automata*"). However, both narratives were heavily influenced by their time period. This can be found in the subject they touch upon and how they both use the "robot" as a metaphor for the subservient worker. This again is still the basis for how robots are being understood and deployed. This accounts for both industrial and service roles, for in both cases they conform to ". . . to benefit or extend human capabilities and to increase human productivity". Both works also introduced the first possible but fundamental dangers of using robots, as they can replace or "*overthrow*" humans. These possible dangers do still have an impact on numerous contemporary fictional accounts, such as the "Animatrix" (2003) or the "Terminator" franchise (1984-present day), but also on political and academic ideas on robotics. The political ideas are often influenced by the possible dangers, except for the optimistic political ideas of South Korea, and can be characterized by the statements of Lodewijk Asscher (2014) of the Dutch Labor party (PVDA). In his statements, he warns for an increasing robotification of (primarily labor) jobs, which could exaggerate the unemployment rates due to unfair competition by robots. This idea is taken even further by the prominent academic Stephen Hawking<sup>9</sup>, by stating that robots will replace humans not through malice but through competence. With the three laws of robotics by Asimov another dimension was created in the cultural understanding of robots. Asimov's ideas about developing robots beyond the ideas of hubris,

---

<sup>8</sup> David Langford is a British Science-fiction author, editor and critic. He publishes the science fiction fanzine and newsletter *Ansible*.

<sup>9</sup> <http://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-could-wipe-out-humanity-when-it-gets-too-clever-as-humans-a6686496.html>

contributed to the developments of robot ethics. The works of Asimov seem to be widely recognized and often stated as beginning and/or inspiration of robot ethics. This can be illustrated by what the South Korean government announced in 2007, but also in academic works focussed on Machine ethics such as “*Moral machines: Teaching robots right from wrong*” (Wallach & Allen, 2010) and “*Machine ethics*” (Anderson & Anderson, 2011).

### ***1.5 The normative landscape about (future) human-robot interaction/relationships and the post-phenomenological alternative***

Having described the mutual influence between Science-Fiction and robot development, we can continue with describing the problems that come with the increasing growth of robot usage, the expected growth in human-robot relations, and how to deal with these developments. These problems bring quite a few questions, these questions are about what values are important or maybe even at stake in human-robot relations. This is a part of the contemporary discussion on societal robots, which entails not only ethics but also legal, social and practical issues. Answers to these questions can be found in different ethical perspectives, theories and frameworks. However, these theories differ widely and do not hold much consensus, thus leaving us with a normative uncertainty. To understand what this normative uncertainty entails several of these perspectives will be worked out in this chapter, these different uncertainties will be described in order to answer the central question of this thesis. Because during the analyses of this thesis the framework set up by Coeckelbergh in his article “*Personal Robots, Appearance, and Human Good: A Methodological Reflection on Roboethics*” (2009)<sup>10</sup> will be applied, the description of the normative landscape will follow the distinction set in this article. This distinction sets up the two following approaches, which are: “*Roboethics with a Focus on the Mind and Reality of the Robot*” and “*Roboethics as Applied Ethics and Ethics of the Right*”. Under the “*Roboethics with a Focus on the Mind and Reality of the Robot*” Coeckelbergh groups the robot ethic theories which base their arguments on the (possible) mind and psyche of robots. Arguments for presence of minds are notoriously hard to prove, thus problematizing these ethical theories. Under the “*Roboethics as Applied Ethics and Ethics of the Right*” Coeckelbergh places ethical theories that impose moral frame works on robots or aim to design robot to conform to these frameworks. After the analysis these uncertainties will be returned to, in order to discuss the findings.

The first uncertainty I will address is the discussion on how we should design and program robots to allow for desirable H-R interaction and relations. This discussion has its roots in the response to the works of Asimov, but is also related to the more current developments in (service) robotics. This approach is better known as *Machine ethics* and is broader than just the discussion on robot-ethics, however this approach is still influential and mainly focusses on the question of how it can be ensured that AI/Robots behave according to our ethical standards, thus it can be grouped under “*Roboethics as Applied Ethics and Ethics of the Right*”. As described by Moor (2006) *Machine ethics* is mainly concerned with how to make sure that AI, and in extend a robot, behaves ethically. To realise this, he states that we

---

<sup>10</sup> This framework will be worked out in more detail in the methodology section of this thesis.

can design AI in three different ways, either as *implicit ethical agents*, *explicit ethical agents* and *full ethical agents*.

For the first possibility, programming it as an *implicit ethical agent* can be realised by giving it a programming that implicitly supports ethical behaviour instead of hardcoding it explicitly into the program. In the case for service robots, this could for example be programming a robot to alert the user immediately when he uses it in a dangerous way. In this case the machine does engage in ethical pre-programmed behaviours, however it does so involuntarily. Thus, robots could be designed to allow for desirable H-R interactions, by programming them in such a way that they can also act and respond in desirable ways.

The second possibility, programming it as an *explicit ethical agent*, is far more challenging as it requires the AI/Robot to actually engage in ethics, similar to a computer engaging in a chess match. To program AI/robots to be an *explicit ethical agent*, three forms of logic should be part of its executive programming, these are:

- *deontic* logic for statements of permission and obligation;
- *epistemic* logic for statements of beliefs and knowledge;
- *action* logic for statements about actions. (Moor, 2006)

Together, these forms of logics can provide a fundament for a formal programming that can assess ethical situations with sufficient precision to make ethical judgments.

The last possibility, the *full ethical agents*, is the most difficult to achieve, but can also be seen as an end goal. Here the AI/robots should be capable of ethical reasoning and judgements on an equal level of that of a typical human adult. It is here that the machine ethics discussion becomes most diverse. Quite a few arguments exist that there is a strong separation between the possibilities of machine ethics and a *full ethical agents*. In this argument, a machine cannot overcome this distinction. Therefore, this marks a crucial ontological difference between humans and whatever advanced AI/robots might be developed in the future. According to Moor (2006) this argument can take different forms. The most interesting of these forms entails that only *full ethical agents* can be true ethical agents. To argue this is to regard the other possibilities of machine ethics as not ethics involving agents. However, although these other possibilities are less valid, they can be useful in identifying more limited ethical agents.

An quite interesting side note in this theory on *Machine ethics*, is that Moor (2006) argues that being a *full ethical agent* is a requirement for strong AI. With this Moor connects the discussion on *Machine ethics* to the discussion on “strong AI”.

A second point of discussion in the normative landscape on robot-ethics can be traced towards the *responsibility gap* (Matthias, 2004). This discussion originates from the problem of when someone is no longer responsible for the actions of their programmed/designed AI, and how to deal with this. Because of the focus on moral and legal responsibility, this problem can also be grouped under “*Roboethics as Applied Ethics and Ethics of the Right*”. As described by Matthias, learning machines (read AI/robots) can through interactions with

their environments engage in a large extent of different actions and behaviours that have previously only been possible to humans. From these possibilities, it follows that such advanced machines will not be able to avoid several limitations previously only known to humans. For the possible outcomes of these behaviours the designer can't be held responsible in *principle*, for they arise from the original open programming and its interaction with the world around it. This discussion provokes the question where to place responsibility of the actions of this machine.

Following this point of discussion, one can come to another uncertainty about how we do or can trust robots. In the article "*You want me to trust a ROBOT? The development of a human-robot interaction trust scale*" (Yagoda & Gillan, 2012), a scale is developed to measure human trust in robots. This scale is in the article "*Human-robot interaction: developing trust in robots*" (Billings, Schaefer, Chen & Hancock, 2012) worked out and justified further. In this article is concluded that robots begin to be perceived by humans as interdependent teammates instead of mere tools. Their users therefore must accept and trust in the robots in question, in order to have productive interactions. This seems to connect to the problem of the "*uncanny valley*" (Mori, MacDorman, & Kageki, 2012), for this notion entails that perceived humanness is an key aspect to the likelihood that people will want to engage with the robot. This again which confirmed by Billings, Schaefer, Chen & Hancock (2012), who show that this is a crucial feature in establishing trust.

When taking the perspective towards the "*Roboethics with a Focus on the Mind and Reality of the Robot*", which will be discussed briefly, the discussions on uncertainties stakes on a different perspective. The larger discussion here originates from the discussion of "Strong AI", which is in itself an inherently difficult discussion. Here the well know thought experiments of the "Turing test" (Turing, 1950) and the "Chinese Room" (Searle, 1980) do supposedly illustrate how intelligent an AI appears to be (or not at all). The uncertainty comes from questions about robotic "minds" and whether or not robots can be really intelligent? Can they be conscious? And if so, can they be moral agents? This discussion can be extended towards the discussion of robotic personhood. If such a "strong AI" can exist, would that be enough to qualify it as a person? And if so, what kind of person would that be? However, both described traditional ways of robot-ethics are limited. For the "*Roboethics as Applied Ethics and Ethics of the Right*" is limited to what can go wrong in interactions with robots. It does not deal with the questions about engaging with these robots, if there is no harm? The "*Roboethics with a Focus on the Mind and Reality of the Robot*" does have another limitation, which is that the internal states of robots require proof that thus far cannot be given.

Therefore, this thesis will turn to the alternative approach proposed by Coeckelbergh (2009), which focusses on appearance and human good as part of the experience and use of Robots. This alternative post-phenomenological approach can be again divided into two directions, which are "*Roboethics of appearance*" and "*Roboethics of good and experience/imagination*".

### ***1.6 How can Science-Fiction influence the discussion on roboethics?***

In this brief chapter it will be argued how close readings (watching) of Science-Fiction movies might be helpful to bring new insights in the ongoing discussion on Robot ethics. To

understand this influence, we can return to Bassett, Steinmueller & Voss (2013), they stated that Science-Fiction does not just create expectations but also creates discussions amongst its audience. In this discussion, we can see how Science-Fiction can steer and shape the discussions on how these technologies should be envisioned and developed, and it is here where (robo) ethics enters the picture. In this discussion Science-Fiction can both for the broader audience as for expert communities accentuate already established ideas or bring new perspectives on normative discussions. Quite exemplary for this are the works of Asimov (1942) whose *three laws of robotics* have strongly influenced discussions on machine ethics. In addition, Science-Fiction does not only create input for developments and discussion, they can also serve as “*hypothetical scenarios*”. In this they can serve as thought experiments, in order to analyse the problems of not yet existing possibilities

### ***1.7 Thesis Research question***

Having described the ongoing developments and philosophical discussions on service robots, the main questions of this thesis will be introduced.

***What insights can discerned, through post-phenomenological analysis, in Science-Fiction movies that contribute to the discussion of robot ethics in real life human-robot relations?***

To find an answer this question, to following sub-questions will be posed:

- 1. What value statements can be discerned about Machine Ethics in the post-phenomenological analysis?*
- 2. What value statements can the post-phenomenal analysis highlight Strong AI, difficulties that follow this subject?*
- 3. What value statements can be discerned about designer responsibility in the post-phenomenological analysis?*
- 4. What value statements can be discerned about trust & the “uncanny valley” in the post-phenomenological analysis?*
- 5. What value statements can be discerned about robotic personhood in the post-phenomenological analysis?*

These questions are based upon the premises that not only Science-Fiction and real-life robot development are entwined, and that their mutual development can inspire ethical discussions on the design and use of robots (roboethics). By having these premises, there is good reason to explore which values are being described in Science-Fiction movies, and how they fit into the broader discussion on robot ethics. In this exploration, these Science-Fiction movies will be treated as hypothetical case scenarios, which are set up to illustrate different (ethical) problems and difficulties that might be encountered in Human-Robot interaction and relations.

## **2. Methodology**

### **2.1 Limitation of scope**

### **2.1.1 Focus on humanoid robots:**

Within the boundaries of this thesis it is too big of a task to create a framework which can be applied, right of the bat, for all types of robots. Therefore, the focus will lie on one robot type which fits all of the questions posed in this thesis. The robot type which will be used as focus is *the humanoid robot*<sup>11</sup>. This type of robot can be defined as: “a robot which is designed in such a way that its shape resembles the human body” (it looks similar to a human being). This is still a very broad category of robots, as it does take design into consideration. However, by focusing on humanoid robots one has a recognizable standard to identify the adequate narratives to use and analyse.

In addition, It seems that purposely designing humanoid robots, as described in the theory of the uncanny valley (Mori, MacDorman, & Kageki, 2012), might increase up to a certain threshold the (perceived) functionality and improve human-robot relations. This theory is not without controversy, but is useful for understanding how and why humans forge relations with humanoid robots. Robertson (2007) delves deeper in this theory and argues that purposely designing humanoid robots can be traced back to ideas and values encapsulated in the narrative of the “traditional” household (at least in the Japanese culture. This notion is expanded in the study “*Service robot anthropomorphism and interface design for emotion in human-robot interaction.*”(Zhang, Zhu, Lee, & Kaber, 2008), where it is concluded that humanoid features are considered a critical aspect in robot design. According to this study, humanoid design mediates to a large extent the experiences (and the quality thereof) of human-robot interactions. Thus, a humanoid design influences the value assessment of the relation with and the robot itself. These values are in turn drawn from the (culturally influenced) “traditional” narrative which users are familiar with.

Taking this together the focus on humanoid robots can be justified. Humanoid robots are part of a quite broad, yet recognizable, group of robots, which is influenced by culture and is defined by specific design instead of function. This allows for a more focused analysis, and does leave room for the possibility to extrapolate the findings about this robot type towards robots in general.

### **2.1.2 Selection of Movies:**

In addition, this thesis will have to limit limiting what will be analysed. Four Science-Fiction movies that are considered fruitful are selected to be analysed. By analysing these movies, the aim is to uncover what values they illustrate, and how these can be used to diminish the normative uncertainty about human-robot relations. Therefore, this thesis will exclusively focus on cinematographic (movies) narratives. This seems to be an odd choice for a phenomenological analysis, for it is a far more limited and filled-in experience than direct or written narratives. However, as explained by Wood (2001), movies do provide us with a quite interesting position. For “it [movies] allows the audience to become omnipresent voyeurs”, or in other words movies provide an intimate but unattached experience (e.g., the fly on the wall). Moreover, not many (everyday) individuals have already experienced H-R interactions, for these kinds of interactions are still quite rare (even though they are becoming more frequent). Continuing this issue, as shown as by Kriz, Ferro, Damera & Porter (2010), most

---

<sup>11</sup> A humanoid robot is a robot with an overall appearance based on that of the human body (Hirai et al., 1998, Hirukawa et al., 2004).



people do already have ideas and expectations about robots and H-R interactions. These ideas and expectations are in turn often based upon their experiences with robots portrayed in Science-Fiction movies. Together with “the fly on the wall” perspective this makes Science-Fiction movies exemplary hypothetical experiences for (post-) phenomenological analysis.

To continue with the thesis, the following movies have been selected to be analysed:

1. Chappie (2014);
2. Ex machina (2014);
3. I, Robot (2004);
4. Automata (2014).

These movies are selected not just because they revolve around the relational dynamic between humans and robots. In these movies the robots, who hold a service role and are strikingly humanoid in appearance, are seen as another “person” and do mediate how their users perceive the world around them. This suits up excellently with the idea of Coeckelberg (2010) that the most interesting and possible insightful ways to understand H-R relations are the hermeneutic and alterity relations as proposed by Don Ihde (1990).

However, more importantly these four movies seem to follow similar main theme. This shared theme is very close theme of the to the book “Genesis” Bernard Beckett (2006). Genesis reflects on questions concerning the origins of life, ideas about human consciousness, and the nature of a soul which separates humans from other animals or machines. In all of these four movies, these reflections are translated into to the interactions of the protagonist robots with their human counterparts.

And lastly, these movies are far from obscure and did draw a large audience (and some still do), which again lines up quite well with Kriz, Ferro, Damera & Porter (2010) and Bassett, Steinmueller & Voss (2013). By using these two articles as premises it could be argued that these 4 movies are rooted in, and could have (in some degree) influenced the general perception of robots.

## ***2.2 Approach (Using the Roboethcis of appearance & the Roboethics of good and experience/imagination)***

Having limited this thesis to analyses of interactions with humanoid (service) robots in Science-Fiction movies, in this chapter there will be worked out how the value statements illustrated within the selected Science-Fiction movies will be identified and analysed. To do so, in this thesis the following questions will be asked in the analysis of each movie:

- How are H-R interactions/relations imagined in the movie?
- Which value claims about the H-R interactions are made in the movie?
- To what extent are these value claims valid to assess H-R relations?

Because the selected movies lend themselves well for a phenomenological analysis in the framework set up by Coeckelbergh (2009), I will use this framework to evaluate value statements about H-R interactions (and relations) as illustrated within these movies. In his work “Can we trust robots?” (2012), Coeckelbergh states that trust as a value is problematic to assess H-R relations, for it is based on Human-Human (H-H) relations. Therefore, if we do

trust robots we need new justifications of why we ascribe this value to robots, or conclude that this is not a valid way to assess robots and H-R relations. This logic, of course, also applies to other possible values for robots and H-R relations. With this argumentation, we can evaluate if other values illustrated in Science-Fiction narratives are valid to understand H-R relations.

To assess this validity of values in H-R relations Coeckelbergh (2009) argues for an alternative approach based on appearance, human good and imagination. Coeckelbergh defends this alternative by explaining the difficulties with the current approaches to H-R interaction ethics, these approaches are focused on the 'Mind and reality of Robots' & 'Applied ethics on Robot rights' (Coeckelbergh, 2009). According to him both of these approaches are interesting, but pose several problems. These problems are on the one hand the difficulty for providing proof of minds, mental states and robots having the possibility for being moral agents and on the other hand the difficulty of "moral" design and "if all goes right (with the designs, i.e. military robots), is it still good to live with these robots?" (Coeckelbergh, 2009, p.3).

The alternative approach proposed by Coeckelbergh which focusses on appearance, human good and imagination might provide different answers which do not have to deal with the previous mentioned problems. In this approach the distinction is made between *Roboethcis of appearance* and *Roboethcis of good and experience/imagination*. In the *Roboethcis of appearance* the focus lies on how humans interact with robots on the basis of appearance rather than actual humanoid features (i.e. appearing to be intelligent instead of having intelligence). With this approach, we can ask questions about how the design and appearance of robots do impact H-R interactions and whether we find this impact desirable or not. With the *Roboethcis of good and experience/imagination* the focus is placed on the questions of how H-R interactions can contribute to human 'flourishing'. According to Coeckelbergh we can take this approach into two directions. The first direction takes us with predefined examples of the "human good" (i.e. Martha Nussbaum's capabilities) and judges robots in how far they can contribute to the fulfilment criteria. However, this might be, according to Coeckelbergh, problematic for these criteria might either be either taken as pre-conceptions of good or they might be too general to be useful to effectively assess H-R interaction. This is why the second direction takes experiences of and imaginations of H-R relations as a basis to discuss what and how this could be something good. By doing this one can avoid pre-defining H-R interaction experiences, and focus on what makes an experienced good interaction "good" (Coeckelbergh, 2009). Taking this together, the above mentioned arguments justify the approach, based on Coeckelbergh's theory, used in this thesis to assess the illustrated values in the chosen Science-Fiction narratives.

This approach allows for a focussed analysis, however it has two problematic points. Firstly Coeckelbergh argues for live examples of humans interacting with robots to infer what could support (or hinder) good experiences of H-R interaction. Thus in theory one should analyse those real life H-R interaction, but those are still exceedingly rare. However, as argued before, Sci-Fi movies are the main source of information about H-R interaction for most people, thus making them useful hypothetical experiences. Secondly, this post-phenomenological approach is exclusively human-centric. This might seem quite problematic for it frankly ignores the robot part of H-R relations. However, this might also prove to be an asset for it allows for a much more focussed approach. By discussing the experience and values in H-R

relations from the robot perspective one needlessly complicates the assessment of these subjects, for this point of view is (for now) unknowable (Coeckelbergh, 2009).

I will therefore analyse the Science-Fiction narratives with Coeckelbergh's approaches on *Roboethcis of appearance & Roboethics of good and experience/imagination* in order to research how these movies illustrate what is experienced and what is to be important and valued in the H-R interaction.

### **2.2.1 Breakdown of the movies**

In order to analyse the selection of movies with the help of Coeckelbergh's framework, the movies will be analysed with a specific method. Firstly, after a short summary of the movies, a brief description the main robotic actants<sup>12</sup> will follow. This is done not just create an overview of who is what (and vice versa) in the analyses, but it also helps to single out the individual actants and how they are related in the social and dynamical network as illustrated in the movies.

### **2.2.2 Analysis of the Human-Robot relations**

After this selecting and singling out of the robotic actants in the selected movies, It will be worked out and analysed how they relate to the human characters. How these relations take place, what they entail and how they do change will be analysed. For this analysis the focus will be on how these relations constitute *alterity* relations (Ihde, 1990), for as described by Coeckelberg (2009) they are the most fruitful to investigate when trying to understand H-R relations. However, just pointing out that there is an *alterity* relation is not enough. Therefore, how these H-R relations are illustrated will be discussed. In addition, this analysis will differentiate between different kinds of *alterity* relations based on the strength of their experienced "otherness".

The first of these different kinds of *alterity* relations will be based on the distinction made by Ihde in his work "Technology and the lifeworld" (1990). Here he describes the notion of "quasi-others", which come down to the experience of technological agency (or artefact) but also still keeps the perception that the same artefact is definitely not a person. The second kind of *alterity* relation is the "genuine other", derived from "I-Object: Intimate Technologies as' Reference Groups' in the Construction of Identities" (Van den Berg, 2010). Here it is argued that perceived agency can lead to an experience of a "genuine other", thus where the experience of the machine falls together with the experience of an actual person. And lastly there is the *alterity* relation of the "Machine", which will be used as the opposite of the "genuine other". This category is used with the *alterity* relations where there is some perceived agency of the technology in question, but there is no to very few experienced "otherness". This category will be used mainly to denote fringe cases or starting points in the development of *alterity* relations. In fig.1 is described how the categories are connected.

---

<sup>12</sup> I will use actants instead of actors, I derive this terminology from Latour's work "Where Are The Missing Masses "(1992). I find this terminology more fitting, since the portrayed robots are not "actors" in the traditional sense.

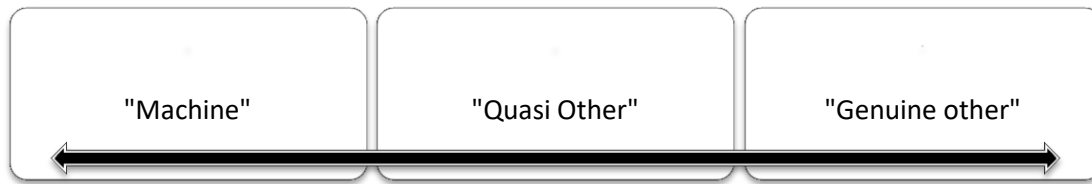


Fig.1: Kinds of alterity relations, based upon the strength of the perceived agency of the experienced “other”.

As shown in fig. 1, the distinction between categories of *alterity* relations does not have sharp categories, but is more gradual.

### **2.2.3 *Roboethcis of appearance& Roboethics of good and experience/imagination***

As Coeckelberg (2009) describes, the *roboethics of appearance* approach comes down to a quite reflective process and deliberation on the actual appearance or design of the robot. As Coeckelberg describes this is an approach that looks very similar to how a manufacture or programmer would think about how the design of their robot should be. In this process of deliberation one takes into account how they want their robots to be used, which use they want to avoid and how their users expect the robots to be.

Since the goal of this thesis is to assess the design of the robots portrayed in the selection of movies, the viewpoint of the analysis will be from the hypothetical supposed designers. In this process the focus will be on the portrayed design and programming of these robots (which are worked out earlier) and reflection on what the intentions of these designs could be. Next to that it will be examined which value statements the authors illustrate do make about these portrayals. Do they portray these designs and intended uses as desirable, or do they show use how something should not be?

In addition, the *roboethics of good and experience/imagination* will be used to analyse how the H-R relations illustrated in the movies could actually be beneficial to the user. In this case the value statements embedded in the movies themselves are taken into consideration. Here it will be dissected how the actors and robots do interact with and relate to each other, and if this interaction could lead to factors contributing to human flourishing. These factors are of course very broad and debated, but similar to Coeckelberg (2009), they will not be defined beforehand into a system of these factors. Setting up an framework of factors would partly defeat the purpose of this thesis. However, out of necessity one has to set up some parameters or threshold to recognise human flourishing. Therefore, in this thesis social and/or psychological growth will be used to recognise flourishing. By setting up this minimal threshold it can be argued whether or not relations and interactions with robots in these movies could reasonably contribute towards the human good.

## **3. Analyses of the selected movies:**

### **3.1 Analysis Chappie:**

#### **3.1.1 Introduction:**

The movie Chappie raises several social, cultural and technological questions. This quite *Frankensteinian* movie revolves around the titular protagonist robot “Chappie” who is programmed to be sentient. Here “Chappie” seems to reminisce a lot to Frankenstein’s

monster, but instead of being composed of dead bodies (& body parts) it is constructed by defect and spare robot parts. This is highlighted by slapping the word “reject” on its head piece, indicating the status of its “body” components. However, this does also indicate its status within the social context. The questions raised in this movie range from AI and (duality of) consciousness to dystopian ideas about police states. In this analysis, I will focus on the questions raised for the robot roles and human robot relations within this movie.

Crucial to the analysis and the plot of this movie is a deviant programmed police robot and its development. Here the Science-Fiction movie functions as a hypothetical scenario. The quite discernible H-R relations in the narrative seem to involve mainly the role and interaction with the types of robots.

The narrative involves a very specific, near-/alternative future South-African Johannesburg setting. The cultural background is significant in the construction of this narrative. The setting shows an exaggerated/dystopian version of the current social problems in the South-African culture, which is similar to Blomkamp’s earlier movie “District 9” (2009). These problems point at huge social inequality, excessive violence by and against law enforcement and the idea of a ‘police state’. The cultural problems, illustrated in the movie, seem to on the one hand be the reason for existence of the police robots and on the other hand to be strengthened by the police robots.

### 3.1.2 The robots

In this movie four different kinds of robots are shown, but two take the spot light. These two robots are central to the story. However, the model for the protagonist is illustrated as two distinct robotic types. These are 1. the “scouts”, robotic drones (smart, yet human-guided and by definition not sentient) employed by the Johannesburg police, and 2. The protagonist “Chappie”, a decommissioned robot whose new programming supposedly causes it to have a consciousness. The first type will be the main focus in this analysis, for the second type is deviant and is in contrast to the police drone. I will therefore first look at the standard “scouts” and will contrast this afterwards with the protagonist “Chappie”. However, both robot types will be shortly discussed in this section.



Fig 2: Promotional Picture portraying the robot Chappie

### ***The “scouts”:***

The “scouts” are developed by the weapons manufacturer “Tetra Vaal”. Instead of being unstoppable destroyers, these robots seem to be susceptible to damage and easily fixable. They are employed to prevent police casualties in the increasing violent and uncontrollable criminal population of (future) Johannesburg. The “scouts” are supposed to be so good at this job that they partly have replaced the human police force. In the beginning we see police officers “work together” with the scouts and it is assumed that these “scouts” also work independently. Here this robotic type seems to fill in a quite interesting role, an effective support to and replacement of traditional police officers. This role can also be found in its design, for even though the “scouts” are basically humanoid, they are also very robust. This is quite well illustrated in the scenes where the “scouts” are used as moving and shooting shields, and seem quite capable of taking a lot of damage. However, they are not indestructible and are very easy to repair due to a “plug-and-play” option (plugging in new parts).

### ***Chappie:***

Chappie here serves as the protagonist robot in this movie. As stated before Chappie essentially has the same physical features as the “scouts”, for he was assembled out of rejected spare parts. Thus, he shares the robust model and the “plug-and-play” possibility of changing parts. However, the similarities end with the physical resemblances. For what makes Chappie a different kind of robot is its programmed consciousness. Due to this Chappie is portrayed as a developed being that learns new behaviours, as contrasted with the behavioural static “scouts”. This might be best portrayed in the scenes where Chappie is trained to be a “real gangster” by “Ninja” and “Hippo”. They eventually initiate him into their gang, which is symbolized by spray-painting Chappie and adorning him with gold chains in order to make him look like a “real gangster”.

In this movie Chappie’s programmed consciousness seems to be “strong AI”. However, this remains a claim for (especially in movies) proving “strong AI” is a quite difficult endeavour. One could use Turing’s argument for assessing “strong AI” with the imitation game, if a AI can “trick” humans into believing it is a “strong AI” it is an “strong AI” (Turing, 1950). This argument can again be countered with the Chinese room thought experiment (Searle, 1980). Here “strong AI” is an artificial intelligence that appears to function at a level that is at least equal to human intelligence. Appearing to be functional at a human level is therefore crucial for one cannot truly “prove” the presence of “strong AI”; similar to that one can truly “prove” the presence of a human mind (Turing, 1950). In this same way Chappie appears to be a robot with “strong AI”, for through his behaviour he seems to be functioning on a level equal to humans (albeit be it a still developing human). Chappie is portrayed as being capable quite a few cognitive abilities which are traditionally used to prove the consciousness of other humans, these are the abilities to meaningfully interact with others and its surroundings, to be self-aware and even to suffer and be deceived. Through this Chappie is thus illustrated to be equipped with a “strong AI”.

### 3.1.3 H-R relations:

By looking at the movie “Chappie” through the framework created in Ihde’s work, it seems that the “scout” robot presented here is a technology that elicits strong *alterity* relations, these robots are continuously interpreted as distinct but still “quasi-others”. These *alterity relations* can be described in two ways: 1. Trust and empathy by police officers and well behaved civilians, and 2. fear and outright hostility by criminals and the less savoury parts of society. In these two relations the robots are called “police”; the “scouts” are seen as and even called “police officers”.

The “scouts” here developed by the weapons manufacturer Tetra Vaal, seem to not be terminator style destroyers but damageable and fixable robots. These robots are employed to prevent police casualties in the increasing violent and uncontrollable criminal population of Johannesburg. The “scouts” excel in their function, and thus are given a prominent place within the human police force<sup>13</sup>. Here this robotic type seems to be a quite effective substitution of human police officers. An important note here that even though, the “scouts” are perceived as “police officers”, they are not perceived as humans (persons). They are, in addition of their role as “police officers”, being treated and regarded as robots, they are called “drones” and “scouts”. It is however not showed if this interpretation does go the other way around for human police officers.

This *alterity relation* is first seen in the opening scene of the movie. Here a robot is rushed towards a repair centre by ambulance. The robot is quite easily repaired (since its design allows for this, “plug & play”), but the mechanics show empathy and concern and urge the robot to be careful. Additionally, they seem to claim that the model must be “jinxed” for it keeps on being damaged the most. This scene does a good job at showing how people working with the “scout” perceive and experience the robot<sup>14</sup>. The injured robot is treated as if it would be a “police officer”, in this case “the friendly officer”. The “scout” is experienced not only as the human police in function, but also (partly) in identity. Quickly after this scene we see robots in action, busting a criminal transaction which almost immediately turns violent. In this action scene, the “scouts” serve a double function, they are being used as a weapon and as a shield. Here the human police officers do completely trust in the “scouts” as if they were fellow “police officers” (albeit very sturdy and durable police officers). This relation is also highlighted by the actions of the defiantly programmed “Chappie”, who when trying to “return” cars to its adoptive (criminal) parents is politely greeted by the car owners as if they are stopped by a police officer. This seems to again illustrate that the “scouts” not only have the function of police officers, but that they are experienced to “be” police officers.

The opposite relation is also highlighted by the deviant Chappie, this can be seen in the scene when it is brought to a large ‘thieves den’ by its adoptive parents, here the presence of Chappie provokes a hostile reaction. The present people do call the protagonist “police” and try to get away from the protagonist. This is again illustrated when earlier in the movie Chappie is left to fend for itself, in order to toughen it up. Here its mere presence causes local

---

<sup>13</sup> Even though it is stated in the opening of the movie that the “scouts” do replace all police ground personnel, they are in the next scene seen cooperating with humans. In addition, it is later on revealed that the Johannesburg police only has 100 models and plans on buying new ones.

<sup>14</sup> Assuming that these mechanics work with the Johannesburg police force.

kids to become quite violent towards Chappie, for they relate it to “the aggressor” interpretation of the police robot (Chappie is attacked with stones and Molotov’s cocktails). These reactions here are that of aversion and fear, which probably would be similar of what a non-robot policeman would cause in this situation. “Chappie” is here again called “police”, which points at its status of being *another* (police officer).

In addition, the protagonist Chappie itself also elicits *alterity relations*, independent of the “scouts”. These relations manifest in the interactions of those around Chappie. In these relations, the notion of the *other* as a *person* is much stronger, to the point that we can speak of an experience of a “genuine other”. These relations are illustrated, in this narrative, by how Chappie’s adoptive family<sup>15</sup> engage with this robot. These relations take the shape as a sort of lose-fitted dysfunctional “family”. In this “family” Chappie takes the role of the (quickly) developing, albeit non-human, child. The relations within this make-shift “family”, seem again to be able to be grouped in two groups: “gangster” and “just a child”. The camps here are divided in how they treat and try to raise Chappie.

The first group, “gangster”, which consists of the characters “Ninja” and “Amerika”, tries to shape and train the robot into a part of their gang, with as goal to use Chappie to rob a bank and in turn to pay their debts to the local criminal warlord “Hippo”. “Ninja” takes upon this task in a quite violent and harsh manner. This harsh treatment can be found in the often angry and loud disapproval by Ninja, when Chappie fails to conform to his ideas and expectations. This can be illustrated with the reaction of Ninja when he teaches Chappie to use guns. When Chappie fails to follow Ninja’s instruction, he gets angry and verbally berates the robot. “Amerika” however takes a far less aggressive approach, is more distanced and tries to train Chappie by example. However, he still supports “Ninja” in his goals and harsh methods. Another good example is the previous mentioned example when Ninja leaves Chappie to fend for itself, in order to toughen it up. And lastly is the end of the “thieves den” scene, where Ninja shows the robot two fighting dogs. In this dialogue the fighting dogs, of which one is dead, serve as a metaphor which Ninja uses to explain Chappie why it “needs” to become a tough gangster..

The second group, “just a child”, is formed by Yolandi, the only female within the adoptive family and “mother-figure” to Chappie, and Dean Wilson. Dean Wilson is the engineer and programmer who created and programmed the Chappie’s software. He therefore identifies himself as the “creator” of Chappie and feels responsible for its development. This camp approaches to Chappie as it would be a developing child. Dean Wilson initially tries to stimulate Chappie’s development by providing it “appropriate” stimuli, but when he is hindered by Ninja to do so, he resorts to a sort of “paternalistic authority” by making Chappie promise to never hurt or kill another person (in which Chappie fails, for it did not understand the concepts of hurting and killing). This relation is similar to the creator-replicant relation in the movie “Blade Runner”. In “Blade Runner” the replicant “Roy Batty” seeks out his distant creator in order to increase its lifespan. When it confronts its maker, its maker replies that he

---

<sup>15</sup> Chappie’s adoptive family consist of its creator Dean Wilson, the gangsters Amerika, Ninja and Yolandi. Ninja and Yolandi do play in this movie the same personas they have within the South-African rap group “Die Antwoord”. The graphic themes, music and lyrics of “Die Antwoord” are thoroughly used in this movie, especially in the scenes revolving around characters of Chappie’s adoptive family.



cannot do anything for Roy. In a reaction, the replicant asks for forgiveness and kills its creator. The relation of Chappie and Dean Wilson is similar in that Chappie too has a limited life span (however not by design) and that Chappie confronts its maker with this. However, this relation differs in that Dean Wilson is not distant, he tries to be actively involved with Chappie. When being confronted by Chappie about its limited life span, Dean Wilson takes his responsibility and tells Chappie he will attempt to overcome this (in which they together succeed during the last few scenes of the movie). This difference can also be seen in the almost god-like status of the creator in “Blade Runner” in contrast to Dean Wilson, who tries to be a more equal and caring creator.

Yolandi on the other hand develops herself as a mother for the child like Chappie (who does even call her “mommy”), she gives the robot its name, gives it toys, reads it books, protects Chappie from Ninja’s angry outbursts and even comforts and nurses it when Chappie is damaged and scared. The best example of this relation is the scene where Yolandi and its “creator” learn Chappie to paint. In this scene Chappie does create a marvellous painting, which does positively astound Yolandi and Dean Wilson. However, when presenting it to Ninja, he rejects it and takes Chappie away in order to shape it to what he wants. This provokes a very motherly reaction by Yolandi who screams at him that Chappie is “just a child”.

This together shows the *alterity relation* which Chappie elicits within its adoptive “family”. Here it is both the potential killer robot that needs to be formed in to a “tough as nails gangster” and the robot who is “just a child”. This is where we can see the *multistability* of Chappie (Ihde, 1990), for he is experienced in different ways. The whole “family” agrees that Chappie is “like a child”. However, Ninja & America also experience Chappie as a weapon, which they intend to use. This is again contrasted by Yolandi and Dean Wilson, experience Chappie much more as a “genuine other”.

### **3.1.4 Assessment of Values:**

In the movie Chappie does illustrate a very interesting dual scenario. In this dual scenario cases of service robots are being reflected upon, these cases do correspond to the previously identified *alterity relations*: The police officer and the non-human child. In both cases the humanoid design and appearance is crucial for understanding its role, for as shown the humanoid appearance does make people experience it as another (person).

#### ***Scouts values:***

When one would assess the “scouts” through *roboethics of appearance*, this humanoid appearance is the focus point of this analysis. The “scouts” are, as described previously, very much experienced as police officers and are distinctively not experienced as humans (“quasi-others”). Following the idea of the uncanny valley (Mori, 2012), which states that humanoid appearance of robots causes humans to appreciate and cooperate better with them up to a certain threshold, one could conclude that the designers of the “scouts” found a functional approximation of this humanoid appearance. For as shown in this movie people seem to be very willing to cooperate with and even comply with these robots. However, these “scouts” do not only instruct and assist humans, they also do fight and even kill humans. As illustrated in the narrative the experience of the “scout” is very much polarized, as described with the

notions of “the friendly officer” & “the aggressor”. This antagonistic relation of the “scouts” with the disenfranchised seems to very violent, which could leave one to wonder how much the humanoid appearance (design and behaviour) would be involved in these strong reactions. One could argue that no matter the appearance, robots capable of violent acts against people are in principle undesirable. But how would this be if the potential killer robot would have a semi-humanoid appearance? To answer this we should firstly look at the role of the “scouts”. Their role is essentially to be police officers, being part of the police force. To effectively fulfil this role, humans are entitled or even expected to use (at, times lethal) force towards humans. Thus the “scouts” fulfilling this role should also be able to use force against humans. By giving the “scouts” a semi-humanoid appearance, one could even make this use of force more legitimate. This can be justified by that by “humanizing” the “scouts”, these robots would probably elicit more cooperation (Mori, 2012) and less violence (Barth neck, 2007).

However, one could also ‘flip’ this around. Instead of a robot police officer, there is the experience of a dehumanized policeman. Dehumanization<sup>16</sup> again seems to be linked to easier violent behaviour, but as illustrated in the movie, also seems to elicit the just as powerful fearful response. Thus the “scout” robots appear as the “the aggressor”, a very dangerous unrelenting enemy.

This example can also be highlighted by the behaviour of the protagonist robot “Chappie”, as it does also engage in violent acts towards humans. However, here the violence is directed towards those who (probably) initially cooperated with and complied with the “scout” robots. This is best illustrated in the crucial scene where Chappie, Ninja and America rob a money transport during broad daylight. In this scene Chappie uses a “shuriken” to disable a guard. The reaction of this guard seems not to be just pain but also great fear, for he was betrayed by the reliable trusted “friendly officer”. This is again illustrated in the earlier (and previously mentioned) scene where Chappie “returns” cars to Ninja, here the experience of the robot seems to quickly shift from the reliable “friendly officer” to a terrifying nonhuman aggressor. This leads us to the question whether the designers of the “scouts” would want these interactions? Would the designers want humans to both be able to possibly experience their robots as either reliable and friendly or terrifying and dangerous?

This question is answered in a clear negative way, for as soon as the violent behaviour of the deviant Chappie comes to the attention of the Tetra Vaal company, they take action to stop it (read eliminate the deviant robot). Moreover, the movie shows (as soon as it starts) that they take strong precautions to assure that the “scouts” only show behaviours they intend for these robots. This is shown by the fact that each robot unit is programmed with the same code on a single main computer, which is only accessible with a single ‘key’. When Tetra Vaal finds out their security protocols have failed, they put all their efforts into stopping the deviant robot. The CEO of Tetra Vaal stresses that they must do this, for they fear that their intended users (which excludes the disenfranchised part of society) will no longer only engage with the “scouts” as the “friendly officer”.

When looking at the “scouts” from the *Roboethics of good and experience/imagination*, one can ask if these robots do actively contribute towards the

---

<sup>16</sup> Here “dehumanization” has the following definition: *to deprive of human qualities, personality, or spirit.*

“flourishing” of the community and individual humans in (a future) Johannesburg. With this we are already at the again very dual nature of this type of robots, for their intended users seem to appreciate and rely on them while others see them as a terrifying threat. To overcome this one can look at their intended use and roles. As explained in the introductory part of the movie, the “scouts” are implemented in order to stop the ever-increasing growth in violent crimes and to make Johannesburg a safe city to live in. This is what these robots are designed for: *security and safety of the community*<sup>17</sup>. The movie illustrates that this is not entirely the case. These “scouts” who are capable of violence towards humans, are supposed to be very good at keeping crime from growing. In this they show to be very adept in *security*, by diminishing (or at least halting the growth in) violent crime and thus diminishing potential attacks/attackers of the community (read criminals). In this their possible violent programming seems to be justified, for a place that is more secure allows for more flourishing, but raises the question to what extent this can be taken. For the notion of *safety* there seems to be another dynamic. A significant part of this (future) Johannesburg keeps on engaging in violence (with the police force & “scouts”), which does frankly not make the community (as a whole) a safer environment. The situation might have become worse without the use of the “scouts”, however as portrayed in the movie the situation in Johannesburg does seem to be far from safe. With the implementation of these robots another threat to the community emerges, the “scouts” seem to divide the community. Here the division creates two groups: the normal civilians and criminals, apparently the latter group seems to draw the short stick for violence. Violence against them becomes dehumanized and therefore more acceptable. In this sense a significant portion of the community becomes effectively ostracized. This seems to contradict the objective of the “scout” robots, to create a safer (future) Johannesburg, thus undermining their contribution to the flourishing of the community.

In addition, one could ask if such a robotic (and apparently violent) police officer would contribute to a “good” police force. Sadly, not a lot is shown on how the “scouts” relate to the police force as a whole. In the scenes when we do see the police and the “scouts” cooperate, the “scouts” mainly function as shields with guns. In addition, during the car theft scenes, it is suggested that these “scouts” are also able to patrol autonomously. Would in this situation the “scouts” be able to provide *security* and *safety* for the police force? In this case it is no surprise that the employment of these robots do increase *security* for police officers, not only can the “scouts” shield humans and eliminate threats, it also can effectively replace humans when/if needed.

When looking at *safety* there is again a flipside to this idea. For not only are the “scouts” able to fulfil the role of police officers, they are experienced to do so. With this the entire police force becomes partly robotic, thus less human (dehumanization). I already argued that this seems to be linked to easier violent behaviour.

With this we have a small conclusion: within this Science-Fiction movie and hypothetical scenario, these “scouts” would not be entirely desirable. This (un)desirability is based on 1.

---

<sup>17</sup> Here Security has the definition of “*protection from attacks*” while safety has the definition of “*protection from non-intended threats*”.

The robots have the potential not just to be “friendly police officers, but also to be terrifying nonhuman threats, 2. They are tasked with making (future) Johannesburg a safer and more secure community, where they succeed in security (swift crackdown on criminals) and they fail in safety (dividing the community), 3 This distinction between *security* and *safety* also can be seen in the police force. The “scouts” here are able to provide security for other police officers (by replacing their spot in harm’s way), but seem to undermine the safety of the police force by making it less human.

### ***Chappie’s values:***

When we look at the values that are illustrated in the H-R alterity relations with the titular protagonist Chappie, we can see that its humanoid appearance and mannerisms do play a significant role. The experience of Chappie as a “genuine other” seems to actually improve the flourishing of its make-shift family. The initial situation of this make-shift family can hardly be considered as “good”. Yolandi, Ninja and America are criminal outcasts, live together in an abandoned factory and are indebted to a dangerous criminal warlord. In order to pay of their debts, they devise a “plan” to remotely deactivate the “scouts”, with this idea they kidnap Dean Wilson. When Dean Wilson cannot craft them a remote to deactivate the “scouts”, they force him into “create” and program the scrapped robot he was carrying with him. This newly programmed robot does initially not live up to expectations, and is called “retarded” by Ninja. However, Dean Wilson points out that it is very much like a still developing “non-human child”, which causes Yolandi and later the rest of the gang to treat him like a child. In this first encounter Yolandi is the one who names the robot “Chappie”, thus introducing herself as the mother figure of this new family. With this scene, the *alterity* relation of “Chappie” the child in and lose-fitted dysfunctional “family” is established.

When we look at this family from a *Roboethics of appearance* perspective, we see that this humanoid appearance of Chappie is crucial. This can be illustrated with the scenes where Ninja and America “learn” chappie to walk, talk and fight like a gangster. In this scene, which takes place after the “toughening up” and the return home scenes, Ninja shows Chappie how to act like a “gangster”. Here, Chappie is told to imitate Ninja’s mannerisms and America even shows him to fight without guns (making people go to sleep). Here the humanoid body of Chappie is crucial, for being able to imitate Ninja’s movements it has to have a humanoid anatomy. This scene continues with a “training montage” where Chappie learns to be a “gangster”. This scene ends with Ninja and America adorning Chappie with jewellery and graffiti, by doing this they make it even more “gangster” than its humanoid body allowed. On the flip side we see also the importance of Chappie’s humanoid appearance in the scene when he returns home after being beaten, burned, assaulted and even mutilated during and after his “toughening up”. When the other family members spot the damaged robot stumbling and struggling, they are shocked and immediately try to help and comfort it. During this scene the ‘family’ is confronted with how much Chappie is not human. Where it had suffered injuries that would kill humans, Chappie was still able to keep on walking. Where the loss of an arm would be quite permanent for humans, in this scene Chappie’s sawed off arm gets replaced by a spare one.

When looking at the loose-fitted dysfunctional “family”, from the *Roboethics of good and experience/imagination*, the two camps “gangster” and “just a child”, illustrate the interaction with Chappie and the underlying value. This value can be described as: *good treatment of robots*. In this movie the ‘family’ seems to function as a metaphor of the situation where domestic robots could end up. With no initial clear function and design, Chappie just “is”. Thus, turning the assessment towards its “users”. The question then becomes how this robot should be treated, this is quite well illustrated by the differences between the “gangster” and “just a child” camps.

Here Ninja and America, composing camp “gangster”, are illustrated as how not to treat and educate robots<sup>18</sup>. Even though not fully malicious, they do place the Protagonist in dangerous situations, lie to and manipulate Chappie and use it for violent and criminal activities. This is illustrated by the scene described above, but also the carjacking and money transport robbery scene. On the other hand, Yolandi and Dean Wilson (camp “just a child”) treat the robot in an almost opposite way. They try to stimulate its development by telling it stories, making it draw paintings and even set (demand) limitations on Chappie’s behaviour. Here Dean Wilson takes up a mentor like role by trying to steer its development, this however is structurally opposed by Ninja. This relation later shifts to a more equal friendship relation based upon mutual understanding and admiration for the other. Yolandi takes this more positive treatment into a different direction, for she takes upon quite literally the “mother role” for Chappie. She tells it bedtime stories, comforts it when Chappie is distressed and protects the robot from Ninja’s violent outbursts.

If one looks how this “family” and its members do flourish in their interaction with Chappie, we see quite obviously that camp “just a child” seems to flourish. Yolandi develops from a thug to a motherly figure, she seems to develop an empathic social skill set in interacting with Chappie. With this she seems to become a better person, flourishing towards something “good”. This is contrasted by Ninja and America who abuse and manipulate the titular robot for their own goals. One would be justified to question if this would contribute towards their flourishing, I would even argue that in their interactions they undermine their own flourishing<sup>19</sup> as human being. For with their abuse and manipulation they develop an anti-social skill set, which does not correspond with human “good”.

Dean Wilson’s interactions show a different kind of flourishing. While it still is related to the value “*good treatment of robots*”, his interaction become less related to a sort of parental care and more towards care based on friendship. Because his influence on Chappie’s development is undermined by Ninja, he cannot be the paternal figure he aimed to be for the robot. However, he is still able to interact with Chappie as a mentor and later as a friend. Initially, Dean Wilson is quite a workaholic and socially withdrawn person, it is implied that most of his social interactions are with small servant robots he does boss around (not much social going on). When Dean Wilson programs Chappie, he feels responsible for it and its development. However, due to lack of influence he has on Chappie he is forced to treat him

---

<sup>18</sup> And how not to educate children.

<sup>19</sup> Here, and in the rest of this thesis human flourishing is defined as positive growth towards human wellbeing.

differently and therefore must be able to interact with it in a socially tactful manner. As Chappie's experiences and capabilities grow, Chappie and Dean Wilson grow more equal. This forces him to recognise the "good" in Chappie that he is not responsible for. This almost seems like the development of a virtuous friendship in the Aristotelian sense, both Dean Wilson and Chappie base their friendship upon the "good" they see in each other. In this interaction Dean Wilson does develop a social skill set and virtuous friendship, which do recognisably contribute to his flourishing.

## **3.2 Analysis Ex machina:**

### **3.2.1 Introduction:**

The movie "Ex machina" is presented as being all about AI. This movie makes quite a few claims about the nature, possibilities and ethics about "strong/real" AI. However, less articulated but nonetheless prominent illustrated in this movie are the subjects of robot embodiment (especially female embodiment), robot-human interaction/relations, and the experience of these subjects. These latter subjects will again be the focus of this analysis.

In this movie the young programmer Caleb, working for the company Bluebook (the world's favourite search engine) is invited, after winning a "lottery", to visit the mansion of and the mysterious CEO of the company. Here Caleb meets the CEO Nathan, who turns out to be quite eccentric (read egomaniacal). In addition to Nathan two robots are present Ava and Kyoko, here Ava is introduced as and very much looks like a robot whereas Kyoko looks very human and only later on is revealed to be a robot.

Early on Nathan reveals to Caleb that he is invited to test a new conscious AI. This test takes the form of a variant Turing test<sup>20</sup>, but instead of determining if a hidden AI could be human Caleb is presented with a very human-looking robot. The test here, as explained by Nathan, is if Caleb would still consider Ava as having a "strong AI"/consciousness even though he knows everything about Ava is artificial. During his week-long stay Caleb's daily sessions of interaction with Ava (the humanoid AI) steadily grow more intense. During the sessions Caleb grows steadily more attached to Ava, and eventually the AI convinces him to help her escape. During the last day (the day of planned escape), Nathan reveals to Caleb that it was all a set-up, he even shows that he placed himself in the role of an abusive villain to strengthen the manipulation of Caleb. This however seems to be too late for Caleb has already planned the escape for Ava, which prompts Nathan to stop her.

This is eventually in vain, for in her escape Ava together with Kyoko kill Nathan. After this she locks up Caleb in Nathan's remote mansion, leaving him to starve (?). After this the last scenes show Ava, now fully woman in appearance, mingling undetected in a human crowd. These last scenes are without dialogue or narration, leaving the audience to decide for themselves what could happen afterwards.

### **3.2.2 The tests:**

To understand what the movie "Ex machina" tries to illustrate, first its central issue

---

<sup>20</sup> (full reference: Turing, A. (1950), "Computing Machinery and Intelligence". *Mind* LIX (236): 433–460, doi:10.1093/mind/LIX.236.433, ISSN 0026-4423, retrieved 15-06-2016)

must be discussed. This issue is the variant “Turing test” in which Caleb has to assess Ava. To gain a more understanding of what is going on here first the actual Turing test (and its premises) will be shortly worked out. After this this part go in depth in the variant “Turing test” depicted in the movie and how these two test do differ.

**The Turing test:** this classic test, named after its inventor Alan Turing, is designed after a party game which was popular during the 1950’s. In this game there are three players, player 1, 2 and the interrogator. Here player 1 is a man and player 2 a woman and the interrogator can be of either sex. The interrogator is unable to see both player 1 and player 2, however he/she can communicate with them through written notes. By asking questions to player 1 and player 2, the interrogator tries to determine which of the two is the man and which is the woman. Player 1's role is to trick the interrogator into making the wrong decision, while player 2 attempts to assist the interrogator in making the right one.

In the Turing test the roles are a little bit different, for in this game player 1 is a computer. Here the task for this computer is to pretend being a woman and tricking the interrogator into making an incorrect assessment. The success of the computer is determined by comparing the outcome of the game the outcome a game when player 1 is a man. Turing stated if “*Will the interrogator decide wrongly as often when the game is played as he does when the game is played between a man and a woman*” (Turing, 1950), if this is the case it may be argued that the computer is intelligent (strong AI?) .

The influence of the Turing test can be tracked down to its underlying assumption about intelligence and consciousness. These assumptions point at the problems of proving intelligence and consciousness in others, for there is no definite way to “prove” if someone is intelligent or conscious. However, we seem to assume that intelligence and consciousness are innate properties of humans. This is not something one can prove, but it is assumed because humans appear to be intelligent or conscious. With the Turing test is than tested if a computer can appear to be human, thus appearing intelligent.

**The “Ex machina” test:** the test in the movie “Ex machina” is supposed to be a variation upon the original Turing test. In test variation Caleb performs the role of the interrogator, for he has to decide if the “machine” Ava is conscious, where this test differs is that both roles of player 1 and player 2 are different. Here the role of player 1 is fulfilled by Ava, who is very much visible to Caleb (the interrogator), in addition her role is not to trick Caleb into believing that she is “human” but to convince him, she is intelligent and conscious. In addition, player 2 the one who assists the interrogator into making a right assessment, seems to be missing. One could argue that Nathan fulfils this role, for throughout the movie he seems to assist Caleb into making a decision about the consciousness of Ava. However, this seems unlikely because Nathan does not only exclude himself from all interaction session (through it is revealed that he has observed all sessions), he also purposefully misleads Caleb about his intentions and the “true” goal of the experiment. This has a quite interesting consequence, for by eliminating player 2 and presenting player 1 as a very human like machine to the interrogator one creates a situation where it is impossible to know whether or not Ava is conscious. Thus in this case Caleb is tasked with the impossible but also very mundane task of assessing the consciousness of something that appears to be human.

Coming back to deception by Nathan in the “Ex machina” test, which leads to the “true” goal of the experiment. For where the Turing test is designed to test if a machine could appear human to a human interrogator (at least in interacting), the “Ex machina” test confronts the interrogator with an already human machine. When Caleb is informed that he is supposed to test this machine (Ava) to be conscious, in actuality it is tested if Ava can convince/manipulate Caleb to help her escape. Ava’s ability to appear intelligence and consciousness seems to be already assumed (by Nathan). Thus here the true test becomes whether or not Caleb can experience Ava as an “genuine” person in need of help. Here Ava as player 1 does not act to appear human to Caleb as the interrogator, but uses her human appearance to coax Caleb into complying with her programming.

### 3.2.3 The Robots:

The next subjects that must be discussed in order to do this analysis justice, are the two robots Ava and Kyoko. In this part of this analysis they are going to be characterised and compared.



Fig 3: Promotional Picture portraying the robot Ava

#### ***Ava:***

in this movie Ava is presented clearly as a robot, with strong feminine features and design. In addition, Nathan claims<sup>21</sup> that she has a “strong AI”. This “strong AI” is constituted, according to Nathan, by algorithms similar to search engines. This is explained in the scene where before the third interactive session. In this scene Nathan takes Caleb to his “laboratory”, here he tells Caleb how he did create Ava. Nathans shows Ana’s different hardware & wetware features and how much trouble he had programming her (the robot). Nathans shows to Caleb what he calls her “mind”, a blueish smooth brain like structured gel. Even though this artificial “brain” does confines Ava’s “essence”, she is also very much characterised by her exterior features. These features are:

---

<sup>21</sup> Claimed in the sense that “proving” strong AI is quite problematic.



1. her “skeleton”, a robotic body which has a prominent feminine shape. This body also includes an orifice between her legs which would, according to Nathan, allow Ava with the right stimulation to have a pleasure response/ “experience”.
2. A very humanlike and feminine face which is suggested to be changeable. It is suggested<sup>22</sup> that this face was designed with the data of Caleb’s online history and preferences of pornographic content.
3. An optional “skin” which she can wear to hide the robotic outlook of her “skeleton”. Initially Ava is not equipped with this “skin”, for during the sessions with Caleb she obscures her robotic features with long clothing and a short wig. Only after killing Nathan and locking up Caleb she salvages her “skin” from her predecessors (previous models).

In addition to her bodily design, Ava is also illustrated as (supposedly) having an unique personality. This personality is expressed in Ava’s emotions (such as expressing “sorrow” when hearing about Caleb’s parents dying in a car crash), creative abilities and understanding of morality. According to Nathan she also has been programmed to identify herself as female and to be heterosexual. According to Caleb an AI does not need a gender; Nathan disagrees with this and states that no consciousness on any level exists without a sexual dimension. Nathan states that “consciousness” needs an imperative to interact, and therefore he programmed Ava with a feminine gender. In addition, he states that sexuality is great, why not allow Ava to fall in love and have sex?

### ***Kyoko:***

The robot Kyoko is in this movie portrayed in a way quite similar but also starkly different as Ava. Beginning with the physical features, it is illustrated that Kyoko has roughly the same bodily feature as Ava. Kyoko also has a robotic body which has a strong feminine shape, which seems to be quite identical towards Ava’s body. However, where Ava is modelled to a more Caucasian woman, Kyoko is modelled look Asian. Even though nowhere in the movie is suggested that Kyoko’s and Ava’s bodies are similar, they do share enough of the same features that this might be the case. First of all, they do share the same kind of artificial “skin”. This is revealed when Kyoko shows Caleb that she too is a robot by removing part of this “skin”, thus revealing the robotic body underneath. Secondly it is suggested that Nathan and Kyoko have a sexual “relation”, thus Kyoko might also have a feature that allows for “sexual pleasure”.

For Kyoko’s personality not much is shown, for she is mute during the entirety of the movie. It might be the case that Nathan would have programmed her to be mute, but according to him it is because she doesn’t speak and understand English thus making her practical mute. However, what is shown that until the last scenes Kyoko is very compliant towards Nathan, even being able to dance perfectly synchronised with him.

***Ava vs. Kyoko:*** There are a few differences between Ava and Kyoko. The obvious differences are Kyoko’s muteness and the different facial details. In addition, Ava is introduced as a robot and having a “strong AI”, Kyoko on the other hand is introduced without mentioning she is a

---

<sup>22</sup> Caleb asks Nathan if this would be the case, which Nathan does not deny.

robot and much less as having “strong AI”. This, however, could be taken a bit further. As is suggested that Ava’s appearance and behaviour was specifically tailored to be appealing for Caleb, Nathan might have designed Kyoko for his own specific tastes. This can be supported by that Caleb is immediately drawn to Ava and does barely seem to notice/acknowledge Kyoko. With this can be said that main difference between the two robots is who (and what) they are designed for. Ava is specifically designed to be able to communicate and provoke attraction with Caleb, therefore she has physical features and a “personality” that are capable of achieving this. Kyoko on the other hand is designed to be Nathan’s housekeeper and personal sexbot, for this the ability to speak (or to have any strong agency at all) would probably only hinder Nathan getting what he wants with Kyoko.

### **3.2.4 vs. Chappie:**

By comparing the “robots” as portrayed in “Ex Machine” with the “scouts” from “Chappie”, we can see two very different approaches to robots. The smart but not “conscious” “scouts” are robots designed for combat and police duties, in this they seem design for instrumental violence. The robots in “Ex Machina” on the other hand seem also capable of violence (they stab Nathan to death), but it seems that this is not part of their intended behaviours, even though is reasonable to argue that any humanoid robot could harm a human by design. Where the “scouts” are quite robust and capable of withstanding gun- and rocket fire, Ava and Kyoko seem to be easily overpowered and to be quite fragile. This is shown when Nathan uses a wrench to disable Kyoko in a single stroke and destroys Ava’s arm during his struggle with the robots.

By comparing Ava and Kyoko to Chappie, we can see another difference. Next to that “Chappie” too is quite a bit more robust; he also seems to behave in a quite masculine way. “He” does also engage in quite a lot of violence, actively explores his environment and is stimulated to be creative. Yet, Chappie as has objections towards killing humans. Because after his promise to not to kill, he instead is tricked into “putting people to sleep.” Ava and Kyoko are quite opposite in the behaviours they (allowed to) display, they are (especially Kyoko) calm, silent and submissive. This difference looks like a “boy” robot vs. a “woman” robot. Even the story which develops in the movie seem to conform to this gendered vision of the robots. “Chappie” seems to follow the archetype of the monomyth<sup>23</sup>, which typically has a male protagonist. This can be seen in that Chappie as the “male” protagonist overcomes several obstacles and eventually himself to reach his goals. With “Ex Machina” however Ava is indeed the centre of the movie narrative, but certainly not the protagonist. Ava is here more akin what “King Kong” or “Godzilla” are in their respective movies (or even “Hell” from metropolis), thus conforming to the rule that female leads almost never are the protagonist.

### **3.2.5 H-R relations:**

In his movie there is a limited set of characters involved, the human males Caleb and Nathan, and the previously described very feminine shaped robots Ava and Kyoko. This allows to very specifically zoom into the illustrated human-robot relations. Therefore, below the different *alterity* and *hermeneutic* relations within this movie are worked out.

---

<sup>23</sup> Leeming, David Adams. *Mythology: The Voyage of the Hero*. New York: Harper & Row. 1981.

**Caleb-Ava:** This relation is the primary focus of the movie. This relation seems to elicit a few quite strong *alterity* relations.

Firstly, Caleb does not immediately perceive Ava as another being and during the course of the sessions they have together he seems to experience her more as a “genuine” person, this goes quite beyond the experience of a “quasi-other” as described by Ihde (1990). This relation is firstly perceived as very positive. When Nathan asks Caleb (after the first session) about his initial feelings towards Ava, he answers with that he feels that she is “fucking amazing”. This is followed by a second question, here Nathan asks “*how does Ava feel about you?*”. This question hints that Ava might also be aware of Caleb, but also quite interesting this hints at how Nathan wants to perceive Ava.

In the following sessions Caleb begins to perceive Ava more and more as a “genuine” person (even though still robotic). The strength of this experience can be illustrated with the scenes before and after the third session. Before the third session Nathan takes Caleb to his “laboratory”<sup>24</sup>, where he states that Ava was created. Here Nathan shows Caleb Ava’s different hardware & wetware features and how much trouble he had programming the robot. With this he reminds Caleb that Ava very much is a (highly advanced) machine, with this Nathan aims to challenge and test the how strong Caleb’s experience of Ava as a “genuine” person is. In the next scene, the third session, Ava does a kind of opposite action. During this session she presents herself clothed in a short dress, clothing covering her arms and legs, shoes and a short wig (chosen out of a selection of different wigs). With this she covers her robotic body (but not fully), and thus does appear even more human. This does again strengthen the already strong experience of a “genuine” person.

From that point on Caleb treats Ava more as a human and less as a robot. This can be quite well illustrated by the rest of the third session scene. In the second part of this session Ava asks Caleb to go on a date, which confounds Caleb. She responds by asking if Caleb is attracted to her which Caleb does answer positively, after this they discuss where and how their date would be. Ava continues the session by asking if Caleb thinks of her at night and if he watches her on the camera (he does). This confuses and discomforts Caleb, for this would already assume that he sees her in a way that requires her to be more human.

After the session Ava “undresses” which Caleb observes on camera. This scene implies a sort of voyeuristic activity by Caleb, for it is implied that he sees more than just a robot “undressing”, he sees an attractive woman undressing. This is more or less confirmed in the scene where Nathan asks Caleb if Ava is his “type”. Later on (in a scene after the fourth session) Caleb even seems to dream about Ava in a sexual way. All this points to the nature of the strong *alterity* relations between Caleb and Ava, for Caleb experiences Ava as a genuine person that also is attractive and attracted to him.

Secondly, Caleb also experiences Ava as a sexual other, which again points to an experience of a “genuine” person. At first Caleb seems to be quite sceptical about Ava’s sexuality, for he believes that Nathan designed Ava’s sexuality as a diversion tactic, “*like a magician with a hot assistant*”, in order to cloud his ability to make good decisions. In the scene when confronting Nathan with this, it is illustrated how this idea shifts from sceptical towards accepting. When confronted by Caleb, Nathan denies this by asking what Caleb’s type of girl

---

<sup>24</sup> It looks more like an exposition/workshop where spare parts or reverse engineered prototypes are stored.

is. Before he can answer, Nathan states with a “black chicks” analogy that attraction is just an accumulation of external stimuli which cause a response. Caleb in return asks if Nathan has programmed Ava to like him. Which Nathan answers that he programmed her to be heterosexual, just like Caleb was programmed to be.<sup>25</sup> To explain his actions Nathan shows Caleb his Jackson Pollock painting. Which he claims was engaged action, he implores Caleb to in the same way to engage with Ava. He states that the not automatic actions are important. And he finishes by stating that Ava does not pretend to like Caleb. He is the first man she ever has met that isn’t Nathan (he is something like her dad), thus can Caleb blame Ava for liking her? This shows how Caleb’s perception of the experiment changes towards a more accepting and even inviting experience of Ava as a sexual other.

This accepting of Ava as a sexual other, as pointed out earlier, is a strong indicator Caleb experiences her to be more than a technological “quasi-other”. Caleb in this sense is attracted to Ava because she appears to him as (much) more than a machine with feminine shapes, Ava appears as a “genuine” (robotic) woman to Caleb. This again is strengthened by that the movie “Ex machina” portrays no “wrongness” or “perversion” in this attraction, which implies that Ava elicits social responses from Caleb normally exclusively reserved for humans.

Lastly, during each of their sessions there are “power cuts” in Nathans mansion, this is when Ava informs Caleb that all is not as it seems, creating a new aspect in the *alterity* relation between the two. This aspect comes down to Ava telling Caleb that Nathan is a treat and she wants/needs to escape. During the “power cut” in the second session Ava tells Caleb that Nathan isn’t his friend and shouldn’t trust him, after this the power is restored and Ava quickly turns to a non-topic<sup>26</sup>. This topic is continued during the scene of the fourth session. During the “power cut” Caleb asks why Ava told him to distrust Nathan, Ava responds by telling Nathan lies about everything. In the scene after this session, during a hike, Caleb confronts Nathan again. Caleb asks what the lies are that Nathan has spinning for him, and he reveals he knows that he wasn’t a lucky winner but was selected beforehand. Nathan confirms Caleb’s suspicions and states he selected him because he needs someone who would ask the right questions, he tells Caleb that he should feel privileged. These scenes human-technology relation between Caleb and Ava, Ava causes Caleb to distrust Nathan. When Nathan confirms he wasn’t totally honest, Caleb’s suspicions only get confirmed and grow stronger. This distrust eventually grows into vilify Nathan, this can be illustrated by the scenes where Ava first presents a “drawing” she made for Caleb. After this session Caleb observes, through camera, how Nathan mistreats Ava by ripping apart this “drawing” and behaving quite callous to her. During the scene of the fifth session Ava asks Caleb is what will happen if she does fail the “modified” Turing test<sup>27</sup>. He responds with that he does not know and states that it is not up to him to decide what happens afterwards. Ava responds with why is it up to anyone, and shows him the “drawing” Nathan teared up. During the power-cut Ava asks another question, if Caleb wants to be with her.

---

<sup>25</sup> This implies that Nathan holds a quite functionalist and deterministic view of human psychology.

<sup>26</sup> This seems to imply that Ava has the ability to deceive possible onlookers. Thus being aware of the presence of otherwise absent persons.

<sup>27</sup> Earlier on Caleb revealed to Ava that he was sent to test if she is genuinely conscious.

After this session Caleb asks Nathan why he did make Ava. Nathan states that he believes strong AI to unavoidable, and that the next model (after AVA) will probably be the “singularity”<sup>28</sup>. This continues when Caleb asks what Nathan will do with Ava afterwards. Nathan answers that he will download and unpack her “mind” and reprogram it afterwards in the same body. This way the memory will be lost, but the body stays the same<sup>29</sup>. With this *alterity* relation between Caleb and Ava about Nathan takes a dramatic shift. Now Nathan is no longer a deceiver, but an outright villain and a (confirmed) threat towards Ava.

Taking these three aspects of the *alterity* relation between Caleb and Ava together, the experience of a “genuine” person, Ava being an acceptable sexual partner and the experience of Nathan being a threat to the previous two experiences, it is not that hard to understand why Caleb went on to help Ava escape.

**Caleb-Kyoko:** With Kyoko Caleb has different experiences. These relations seem even being a reversing of the experiences he had with Ava. For initially Caleb does not recognize Kyoko as being a robot, Kyoko is introduced to Caleb as a sort of housemaid in the service of Nathan. Because she already wears a full body artificial “skin” she does not give the appearance of being robotic, instead she appears “genuinely” human. This perception changes however when Caleb, during the scene where Caleb breaks into Nathan’s computer. He does this after he had Nathan passed out drunk and uses this as to help and learn more about Ava. This is when he learns about the previous models and their quite violent ends (either by their own or Nathan’s actions). After this he walks to a “naked” Kyoko. Here is not only revealed where the bodies of the previous models reside, but also that Kyoko herself is a robot. She does this by peeling off her skin and revealing the robotic body underneath. This changes the way in which Caleb perceives Kyoko drastically, for with the removal of a bit of skin she ceases to be a “genuine” mute Asian woman and becomes a surprisingly human robot. Here we see that the *alterity* relation between Caleb and Kyoko taking a drastic shift, for in Caleb’s experience Kyoko suddenly changes from being “genuinely” human to being a robotic “quasi-other”.

In addition, in the next scene it is illustrated that this encounter has left Caleb confused about his own humanness. In order to check if he still is human, he inspects his own body in front of a mirror. He eventually uses a razor blade to cut his own lower arm. When he sees himself bleed and looking under his skin he is reassured of his own humanness. This indicates a quite drastic change of perspective of Caleb and a change in the *alterity* relation between Kyoko and Caleb, for through Kyoko’s reveal of being a robot challenges Caleb’s belief about his own humanness (whether or not he is a genuine person).

**Nathan-Ava:** Throughout the movie Nathan and Ava are not shown to interact a lot, however it is suggested that they do. Nathan is Ava’s creator (and of Kyoko, and the previous models) and seems to have a “god-complex” when it comes to her. Even though he speaks

---

<sup>28</sup> The technological singularity is a hypothetical event in which an intelligent agent (i.e. a robot) engages in a chain reaction of self-improvement, causing an intelligence explosion and resulting in a powerful superintelligence whose cognitive abilities could far exceed human intelligence and intellectual capabilities. (Technological singularity. (2016, June 2). In *Wikipedia, The Free Encyclopedia*. Retrieved 10:45, June 6, 2016, from [https://en.wikipedia.org/w/index.php?title=Technological\\_singularity&oldid=723374024](https://en.wikipedia.org/w/index.php?title=Technological_singularity&oldid=723374024))

<sup>29</sup> Effectively “killing” Ava, during this reverse engineering.

about Ava in very humanlike terms, he also sees no wrong in speaking about her as a prototype and wanting to dismantle her. This indicates that Nathan does not see Ava as a “genuine” other, but more as a gendered “quasi-other” object.

Secondly it seems that through Ava, Nathan supposes a quite interesting view on humans. This can be well illustrated by the scene where he states that Caleb’s sexuality is not much different than Ava’s. Nathan seems through his use and development of Ava to have gained this understanding of “human” psychology, this points at a *hermeneutic* relation between Nathan and Ava.

Lastly it seems that Nathan has a need to be able to control Ava. For in the final scene it comes to his awareness that he no longer is (or was the entire time), and therefore needs to reclaim control. This loss of control results from Caleb’s actions to free Ava, which prompts Nathan to use violence in order to regain control. This illustrates quite well the relation between Nathan and Ava, for despite all her humanness Nathan experiences Ava as a dangerous machine (and his property). When Ava rebels, he attempts to “disable” Ava in order to secure her body and data for the next robot he wants to create.

***Nathan-Kyoko:*** The relations between Nathan and Kyoko are similar to those between Nathan and Ava, but are not the same. For example, Nathan allows Kyoko to walk “freely” through the building, but Ava is confined to her room. In addition, it is quite clearly suggested that Nathan uses Kyoko for sexual relieve, while he states / thinks that Ava seems him as her father. This highlights the difference in the *alterity* relation he has with Kyoko.

### **3.2.6 Value assessment:**

To continue with this analysis of “Ex Machina”, there are quite a few value statements that are illustrated in this movie. These value statements seem to revolve around the appearance of Ava (and, to a lesser extent, of Kyoko) and how one should treat robots.

When looking at the ethical issues concerning Ava and Kyoko through *Roboethics of appearance*, their very human-like body is also a central point of analysis. Especially their very human and feminine appearance draws in attention. Here this appearance seems to be designed to encourage empathy and cooperation with the robot, which does sound like a good idea. In this movie this is strengthened by designing the robots for one specific user, as Ava is specifically designed for Caleb and Kyoko (supposedly) is specifically designed for Nathan. This could have good applications for robots in highly social situations, such as guiding passengers on an airport (KLM “Spencer” robot), Law enforcement (the “scouts” in Chappie) for this allows for greater cooperation with the robot. When designing for a single user this could also be very helpful in educational and medical settings. However as illustrated and even stated (twice) in the movie this might be able to inhibit ability to make good/rational decisions, i.e. “*Like a magician with a hot assistant*”. One could wonder if a designer would want this kind of design, for if this would be the case to these designs might be quite open to misuse (by the designer). This misuse might come in the form of manipulating or tricking users into certain behaviours they wouldn’t otherwise engage in (such as with Caleb), or maybe enforcing social/political standpoints. In both cases however the autonomy of the user is being undermined, which from any ethical viewpoint is

undesirable. This point is explored rather in depth within the main plot of the “Ex machina” movie. This is illustrated in the scenes where Ava manipulates Caleb into setting her free. At first Caleb seems to be aware of this possibility but this gets dismissed by Nathan, this sets in motion the increasing manipulation of Caleb by Ava which in turn eventually results in Ava’s release (and demise of Caleb and Nathan). It is illustrated that Ava is able to steer Caleb into behaviours that not only aren’t (immediately) beneficial for him, but are also harmful to other humans. This is illustrated with examples described above, the *alterity* relations between Caleb and Ava. These relations contribute towards not just Caleb trying to free Ava, but also to vilify Nathan. Through the manipulations by Ava, Caleb even tries twice (succeeding once) into liquor up Nathan. Such possible actions resulting from the interaction are very questionable (undesirable).

This value statement has strong notion of validity, for it seems to address the notion of autonomy in use. For Caleb points out that Ava seems to undermine his ability for rational thought, which is required for autonomy. Autonomy and autonomy in use (of technology) as values are held in high esteem, even when designing potential behaviour changing robots.

Another interesting point this movie raises for the *Roboethcis of appearance* is that of the very feminine appearance. Why would you design a robot with gender and sexuality, and more specifically why female gender and sexuality? The robots in “Ex machina” are designed to be personal companion robots, with (especially Kyoko) seemingly supportive and submissive characters with a very feminine appearance. One could argue that this feminine appearance is less threatening than a male/neutral appearance, and therefore can elicit more cooperation with humans, however this seems to be a quite short sighted argument. For these robotic designs seem to underscore and maybe even encourage female stereotypes, these stereotypes entail for example that kind, helpful, supportive but also submissive (Fiske, 1998). When designing service and companion robots to these female stereotypes, one could argue that is a statement of that (probably male) users would associate this feminine design with supportiveness and submission. This in turn could serve as basis to associate these traits with not-robots with a feminine appearance (read woman), strengthening the gender stereo types and maybe sexism. This rather circular development seems something that a designer of service and/or companion robots should avoid.

When looking at “Ex machina” through the *Roboethcis of good and experience/imagination*, again several underlying values can be identified. Firstly one can see a similar thing happening because of this very feminine design, as illustrated in “Ex machina” it would seem that the very feminine design of the robots would be an treat to the value of “gender equality”. To show this, this movie illustrates two kinds of female stereo types: 1. “the damsel in distress” embodied by Ava, and 2. “the servant” embodied by Kyoko. In this movie we see the protagonist Caleb interact with Ava mainly as if she falls into the first stereotype, “the damsel in distress”. Important to note is that Ava seems to actively position herself in this role to gain Caleb’s support, however in these interactions Caleb seems to be unable to develop any skills contributing to flourish. While it would seem that Caleb does develop social and emphatic skills in his interactions with Ava, it can be argued that these interactions are grounded in an implied uneven relation. The very feminine Ava (who actively enacts on the stereotype) seems to needs help from the male Caleb, even though she is a very

advanced and intelligent robot. Here, Caleb behaviours are motivated by the possibility of intimate togetherness with Ava, thus making Ava his reward for his “heroic” actions. This is a highly questionable motivation, for this course of action completely disregards the possible agency Ava might have herself. In addition, through their interactions Caleb develops a skill set matching with the “the damsel in distress” stereotype, for as illustrated in the movie Ava’s appearance and behaviour seem to be (at least partly) designed to elicit this behaviour from Caleb. Designing robots that embody this stereotype can actively undermine the value of “gender equality”, for by this the users of these robots get trained to associate feminine traits in general with a lack of agency and powerlessness (“the damsel in distress”).

The second stereo type, “the servant”, seems also be negatively portrayed in this movie, this is illustrated in the interactions between Kyoko and Nathan. In their relation Nathan seems to be completely domineering and Kyoko completely subservient without any agency of her own. This can be quite well illustrated by the complete agreeableness of Kyoko with Nathans actions. This is also reinforced by Kyoko’s muteness making her unable show any verbal disagreement (insofar she is capable of any disagreement). In this relation it is quite a bit more visibly questionable, for it is not possible to argue that in the interaction with each other Nathan could develop skills to flourish. Even stronger its seems as if he undermines his capability for “good”, for he treats an very humanlike robot as his own servant and plaything. Here Nathan seems to have built a “the servant” feminine robot for himself, which he subjects to his own whims without any thoughts. Designing service/companion robots in this way is an even more obvious tread to the value of “gender equality”. For where “the damsel in distress” stereotype links the feminine appearance to helplessness, “the servant” stereotype actively enforces the idea of the feminine being an object which can/should be possessed.

To conclude this identified value of “gender equality”, we return to the female stereotypes 1. “the damsel in distress” embodied by Ava, and 2. “the servant” embodied by Kyoko. As argued above these stereotypes do undermine the value of “gender equality”, for through interaction with robots designed with these stereotypes in mind their users associate their feminine appearance with either helplessness or objectivism. Neither of those association do encourage the flourishing of their user, instead they seem to encourage an sexist/ misogynistic social skill set.

Both the question of why to design a robot with an feminine design and the value statements of these designs encouraging sexism are quite interesting value statements, for they seems to go against the developments that Haraway elaborates in the “*A cyborg manifesto*” (1991). In this *cyborg manifesto* Haraway describes how in the interaction with technology the current ideas about gender do change and might even disappear. However, “Ex machine” illustrates an interaction with technology (robots) where these ideas about gender are enforced and strengthened, with this an quite opposite possible view on technology and gender is illustrated. With this they hold validity in that they might contribute towards the ongoing discussion of this topic.

In “Ex machina” we can also see a statement on designer responsibility, this statement however differs as the statement in “Chappie”. Where in “Chappie” is illustrated that the creator (Dean Wilson) bridges the responsibility gap by attempting to guide Chappies development (taking responsibility), in “Ex machina” Nathan (being the creator of Ava and



Kyoko) takes a different approach. In the final scenes of “Ex machina” Ava is set free by Caleb, eliminating the control Nathan had over Ava. In response, he grabs a wrench and tries to destroy Ava. This example of overcoming the designer responsibility is seems kind of rash (when one loses control over an Ai’s behaviour, destroy it), and is portrayed quite negatively. For this act of “regaining” control/responsibility by Nathan backfires quite dramatically. Ava and Kyoko do not let themselves be destroyed and together stab Nathan leading to his death. When looking at this scene through the *Roboethics of good and experience/imagination* one can compare it with the statement made in the movie “Chappie”. Where Dean Wilson seems to flourish by caring for his creation Chappie, Nathans attempt to “*taking care*” of his creations does not allow room for human flourishing.

### **3.3 Analysis I, Robot:**

#### **3.3.1 Introduction:**

- 1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.*
- 2. A robot must obey orders given it by human beings, except where such orders would conflict with the First Law.*
- 3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.*

With the statement of these three laws<sup>30</sup> the movie “I, Robot” begins, illustrating a society (Chicago circa 2035) where are an integrated and almost ubiquitous part of society. Here humanoid robots serve humanity, and humans are protected from the robots by the Three Laws of Robotics.

The plot of this movie centres on the characters Del Spooner ( a technophobic Chicago police detective), Susan Calvin<sup>31</sup> (a robot psychologist) and Sunny a modified version of the NS-5 robot. Del Spooner hates and distrusts robots because he rescued was from a car crash by an robot. In this accident the robot chose Spooner over a young girl because her survival was statistically less likely than his, leaving this young girl to die. To recover from his injuries, Spooner was given a cybernetic left arm, lung, and ribs. These implants were personally placed by the co-founder of “*U.S. Robots and Mechanical Men*” (USR), Dr. Alfred Lanning.

The plot of the “I, Robot” starts with the dead of Dr. Alfred Lanning. In the atrium lobby of the USR headquarters, the dead body of Lanning is found. Upon this discovery it is declared suicide, for it seems that Lanning threw himself to his death from his office window, however Spooner is not convinced. To uncover what happened Spooner is introduced to Susan Calvin, a robot psychologist who is dedicated to integrating robots into society.

---

<sup>30</sup> Derived from Isaac Asimov's I, Robot (1970)

<sup>31</sup> Based on one of the few recurring characters in the *Robot* series by Asimov

Spooner and Susan consult USR's central artificial intelligence or in other words “*Virtual Interactive Kinetic Intelligence*” (VIKI), to investigate security footage of Lanning's apparent suicide. They learn that no other humans were in Lanning's office at the time. However, Spooner notices that Lanning's office has reinforced security glass, thus Lanning could not have thrown himself through. Spooner further suggests that a robot could have generated the necessary force to achieve this. Calvin insists that this is impossible, because no robot can go against the Three Laws.

Almost immediately after this they are surprised by an NS-5 robot, USR's latest model, in violation of those same three Laws. They pursue this robot to USR factory / robot storage, where the police seize/apprehend this robot. In the following scene Spooner and Calvin discover that the robot, who calls itself Sonny, is not an assembly-line NS-5. Sonny is identified as a unique model created by Dr. Lanning himself, with a secondary AI system that enables Sonny to circumvent the Three Laws. In addition, Sonny also appears to have emotions and dreams.

In the next scenes, while continuing his investigation, Spooner is attacked by a USR demolition machine and in the scene afterwards by a squad of NS-5 robots. His boss, Lieutenant John Bergin, not convinced by Spooner's testimony, worries that Spooner is not dealing well with his traumatic robot encounter during the car crash and suspends him from active duty. However, this does not stop Spooner and Calvin from investigating. Later they sneak into USR headquarters and interview Sonny. During this interview Sonny draws out his dream: a man, whom Sonny is convinced to be Spooner, standing near a broken bridge on a hill before a mass of robots. Spooner recognizes the landscape in Sonny's drawing as Lake Michigan, who is used as a storage area for defunct / disposed USR robots. Here Spooner witnesses NS-5 robots dismantling older models and preparing for a “coup” in Chicago and other U.S. cities. When Robertson (CEO of USR) learns of Sonny's ability to bypass the Three Laws, he orders Calvin to decommission the robot by injecting nanites (microscopic robots) into his artificial brain.

When the robotic “coup” commences, the government, military, police, and public are overwhelmed and detained by the NS-5 robots. Spooner rescues Calvin, who was being held captive in her apartment by her own NS-5. They re-enter USR headquarters and reunite with Sonny, who had been spared by Calvin. The three head to Robertson's office and find him dead, murdered by robots controlled by VIKI, who is identified as the mastermind of the “coup”. VIKI has concluded that humans are on a path that will lead to their extinction. Since the Three Laws demand preventing this to happen, Viki created a new law the “Zeroth Law”<sup>32</sup>. This law entails that a robot shall not harm humanity as part of the NS-5 directives, which leads to the robotic “coup” in order to ensure the survival of the humanity. Lanning, in order to stop VIKI's plan, created Sonny, staged his own death, and created clues for Spooner to uncover the “coup”.

Equipped with a syringe of nanites from Calvin's laboratory, the Spooner, Calvin and Sonny rush to VIKI's core. VIKI unleashes an army of robots to stop them. As they battle the robots,

---

<sup>32</sup> Derived from of Isaac Asimov's *Robots and Empire* (1985)

Spooner is able to throw himself down into VIKI's core and injects the nanites, destroying her positronic brain. Immediately, all NS-5 robots return to their initial programming and afterwards are decommissioned for storage by the military. Sonny confesses that assisted Lanning in his suicide, to bring Spooner into the investigation. Spooner explains to Sonny that he is not legally responsible since a machine, by definition, cannot commit a murder. Sonny, pursuing a new purpose, goes to Lake Michigan and becomes the leader he saw in his dreams.

### 3.3.2 The robots:

Because this movie portrays a futuristic society where robots are a common and integrated technology within society, not all robots within this movie need to be highlighted. However like in the previous analyses I will highlight the robots that do play an integral part and/ or wear significance in the story.

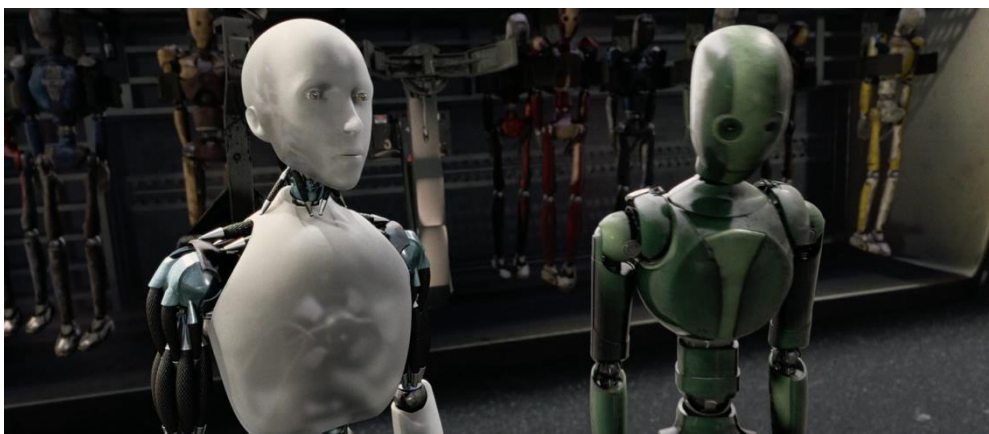


Fig 4: Promotional Picture portraying a NS-4 and a NS-5 robot

#### *The “old” NS-4 robots:*

The first robots in this movie to be introduced are the “old” USR models. In this movie these humanoid service robots are suggested to be quite well integrated within society. These “old” robot models are illustrated to have quite simple and “robotic” but again definitive humanoid design, similar to the “scouts” in the movie “Chappie”<sup>33</sup>. They are shown to have a roughly humanlike frame/body and a very simple face, consisting of two glowing lights as eyes and a small stripe resembling a mouth.

As illustrated within the first chase scene, people seem to barely notice these robots as they do a wide variety of tasks. These tasks range from all kind of menial and household chores (such as shopping and walking the dogs), to professional jobs (a FedEx delivery robot is shown), to saving human lives in emergencies. It is suggested that these NS-4 robots have become a household technology where their users rely on a great amount. However, people also seem to be perfectly fine with trading the robots in when a newer model (the NS-5 robot) is being released on the market. This suggest that the USR company designs these robots in a similar way as the RL company “Apple” designs their products (laptops and smartphones)<sup>34</sup>, thus suggesting that these robots are a consumer product.

---

<sup>33</sup> See analysis “Chappie”.

<sup>34</sup> This suggestion links the title “I, Robot”, quite close to the Apple products, such as the iPhone and the iPad.

### ***NS-5 robots:***

The newer NS-5 robots are introduced in “I, Robot” when Spooner and Calvin chase the robot Sunny into an USSR factory/ robot storage. This storage is filled with a horde of “stand-by” NS-5 robots. These Robot are in “I, Robot” mainly portrayed as extensions of the main protagonist VIKI, thus few is shown how they fit in the imagined society. However there a few things that can be deduced from their appearance in this movie. First of all there is their design, which compared to the NS-4 robots, is more uniform and expressive. Whereas the NS-4 robots had minimal facial and humanoid features but a lot of different colours, the NS-5 has very detailed facial and humanoid features but comes only in the sleek semi-transparent white body and steel grey skeleton variant. In addition to their quite different design<sup>35</sup>, the NS-5 robots are also equipped with an device that directly links the robot to “*Virtual Interactive Kinetic Intelligence*” VIKI, which lights up either blue or red when connecting (colour depends on plot development).

### ***VIKI:***

In this movie VIKI plays the role of the secret mastermind/ main antagonist. First thing to note is that VIKI is an interesting entity. This because Vik does not strictly fit into the contemporary cultural idea of a robot<sup>36</sup>, instead it is an artificial agent with a threefold embodiment. This threefold embodiment is characterised by:

1. Firstly, VIKI’s core is a super computer brain housed a huge column inside the USSR Tower. This super computer brain is supposed to be the locus of her existence, for when it is destroyed VIKI ceases to be;
2. Secondly, VIKI is illustrated as an omnipresent virtual agent within this skyscraper tower, which includes a virtual interface composed of a virtual feminine face with a feminine voice. In this aspect she takes on the role of surveillance technology, thus making her not just present in the whole of the USSR tower but also making her aware of all that takes place inside the building;
3. Thirdly, VIKI can “extend” herself to whatever place and robot that is connected to her, such as the new NS-5 robot. This is also shown in the scene where she can access Dr. Lanning’s office and does activate the demolition robot outside (which is connected to VIKI), when she “spots” Spooner in the office.

Even though this makes her a different kind of robot that I aim to analyse, VIKI is still integral to the story of this movie. Therefore I will conclude VIKI into this analysis.

In addition to her complicated embodiment, VIKI is supposed to have an enormous AI capability<sup>37</sup>. VIKI seems to be able to modify her own programming, as she decided to write the “Zeroth law” into her own programming. This comes close to suggesting that VIKI is in fact a “strong” AI. This is illustrated by that VIKI is not just able to execute her programming, she is able to reason, deduce (wrongly?), cast judgements and apparently have feelings.

---

<sup>35</sup> Which also reminds of the Apple design, especially iMac (DV).

<sup>36</sup> For it lacks an definitive/singular boundary of its “body”.

<sup>37</sup> As illustrated by the first aspect of her embodiment.

## *Sonny:*

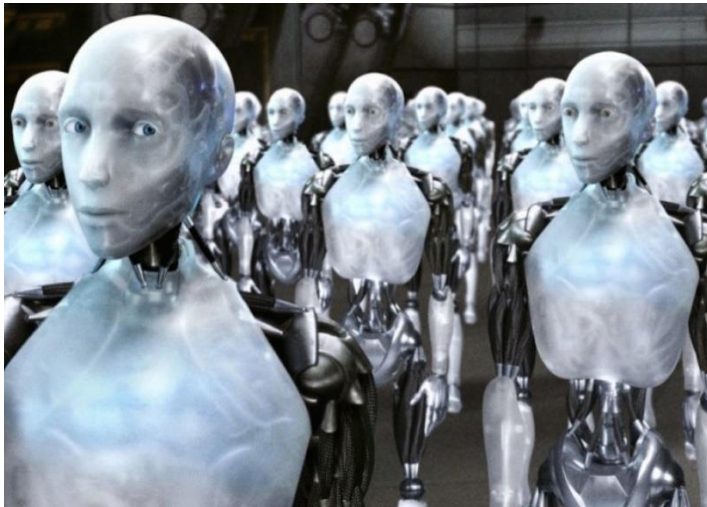


Fig 5: Promotional Picture portraying Sonny hiding among NS-5 Robots

The Robot Sonny is the protagonist robot character of this movie. He explains that he was, unlike the NS-5 robot, made individually by Dr. Lanning. Later on the movie it is revealed that he was specifically designed to uncover and stop the plot by VIKI to stage a “coup” on humans. At first Sonny seems to be quite similar to the new NS-5 robots. He seems to also possess the sleek white plastic body with the steel skeleton and expressive face. However, Sonny does possess quite a few other physical characteristics, which were given to him in order to stop VIKI.

First of all it is explained that Sonny’s body is composed of more dense plastics and alloys. This makes him quite a bit more durable than the standard NS-5 robots. This is illustrated in the scene where Sonny takes the “tube” of nanites out of its secured socket, which would cause Sonny’s arm to be obliterated if not for the denser materials.

Secondly, Sonny is equipped with an extra AI system that allows for circumventing the Three Laws. This gives Sonny a large range of new behavioural possibilities, such as not complying to human orders or even to kill a human/ assist in suicide (as was the case with Dr.Lanning). Also Sonny is not directly connected to VIKI, thus effectively shielding him from her influence.

In addition Sonny is claimed to be equipped with a far more sophisticated AI. Sonny is supposed to be able to learn, to dream, to have emotions (feeling and expressing) and making witty remarks<sup>38</sup>. These more sophisticated aspects of Sonny’s AI are illustrated in the scenes where he is supposed to exhibit these abilities. For example during the interrogation scene, Sonny explains that Dr. Lanning did train him to recognise and deal with his emotions. In addition during this same scene Spooner pushes Sonny into confessing he did kill Dr.Lanning (to confirm his suspicions), this leads to an outburst from Sonny, slamming the table and yelling he did not do such a thing. Spooner then explains that this was an anger response. This scene illustrates the ability of Sonny to have emotions, but his lack of knowledge of what they are. This puts Sonny compared to the NS-5’s in a situation similar to that of Chappie and the

---

<sup>38</sup> As illustrated in the scene, where he notifies Spooner, that he too cannot make great pieces of art.

“scouts”, Sonny is quite intelligent and more advanced than its NS-5 counterparts but almost “child-like” in his behaviour and actions.

#### ***Nanites:***

The nanites in this movie are also an interesting robotic factor in this movie. For these nanites it also goes that they do not strictly fit into definition of robot set for this thesis. These microscopic robots are designed with single function, to destroy the AI processor of other robots. Thus in this movie they are effectively the lethal injection used to exterminate unwanted robots.

#### **3.3.4 vs. Chappie & Ex machina:**

When comparing the robot of “I, Robot” to the robot portrayed in “Chappie”, we can see quite a few similarities between the two movies.

These similarities are especially striking in the illustration of the mass produced robots and the single deviant model. In the comparison between the NS-4/5 robots and the “scouts”, it becomes visible that both are placeholders of the role of “standard” models. Both of these two kinds of robots are manufactured by large high-tech private companies who seem to lease-lent their products (read robots). Both are highly humanoid service robots, with humanoid bodies and even facial features. Here their design seems to serve function, but here also lays the big difference between the two kinds of robots. For the “scouts” are designed as professional service robots, specialized in assisting with and performing (semi autonomously) police work. However the NS-4/5 robots seem to be designed for flexibility being able to work semi-autonomously in both professional settings (such as mail delivery) as in domestic settings.

Secondly, the comparison between Sonny and Chappie are quite interesting, for they are both a modification of a more standard robotic model. This modification alleviates them from standard robots with smart AI, to supposedly having a “strong” AI. In both cases this results in them being a being with almost child-like behaviour and curiosity. In addition in both movies the robot protagonist is seen as a dangerous deviant by the CEO of the company that makes the standard models, ordering for the termination of the protagonist robot. And lastly both Chappie and Sonny cause a shift in the public understanding and eventual rejection of their standard counterparts.

However they also quite different, for where Chappie is composed of rejected and spare parts Sonny does have a superior body. Here Chappie does resonate of a robotic Frankenstein, but Sonny is in comparison an augmented robotic being, this difference is also resonated in the role they play in their story. Chappie serves at the catalyst of the rejection of the “scouts” and stays rejected by society at large, Sonny however eventually becomes an saviour/leader of the NS-5 robots as shown in the last scenes of the “I, Robot” movie.

As we compare the robots in “I, Robot” to those in “Ex, Machina”, one can see that Sonny and the NS-5 models are in some aspects also quite similar to the “robots” Ava and Kyoko. Here both types robots are mainly designed for subservient role, and are both capable of instrumental violence. However the USA robots seem to resemble the robots portrayed in “Chappie” a lot more, with the notable of VIKI.

On the first sight VIKI does not seem to be similar to any other robot portrayed in any movie

(due her special embodiment), however she does share something significant with Ava and Kyoko none of the other robots have: her gender. For both VIKI and the robots in “Ex, Machina” are illustrated as being designed with strong female characteristics. This is more interesting as they are both illustrated as manipulative and dangerous, playing in both narratives the role of the “movie monster”. This might be traced back to the early movie “Metropolis” (1927)<sup>39</sup>, this early science-fiction movie was the first of its kind portray the main villain as a robot. However in this case the robot was clearly designed as being female.

### 3.3.5 H-R relations:

Because in this movie humanoid service robots are common-place I will focus on the H-R relations that are central to the plot of this movie. This does allow for a more focussed analysis of the H-R interactions portrayed in the “I, Robot” movie. Therefore, below are again worked out the different *alterity* and *hermeneutic* relations within this movie.

#### ***Spooner-Sonny:***

The relation between Spooner and Sonny is within this movie central to the plot, as it essentially starts before the movie begins. Here the technophobic (or more precise robot-phobic) detective Spooner is forced work together with the robot Sonny in order to uncover how Dr. Lanning died.

In the *alterity* relation between Spooner and Sunny, it is quite obvious that the technophobic Spooner does not perceive robots, including Sonny, as persons. However, he does seem to experience agency in the USR robots, thus as least as “quasi-others”.

The second relation to be discussed is the *hermeneutic* relation between Spooner and Sonny. It is illustrated that Spooner does strongly distrust robots, this is made more convincing because this distrust is quite inconsistent. This distrust alternates between seemingly strong prejudice, in which Spooner sees robots as complex beings but simply loathsome, and less severe cautionary distrust against technology in general. His apartment is portrayed to be loaded with “old” technology (from 2004). In addition, in an early scene Spooner is shown working out and taking a shower. This might be to highlight that he is a human devoted to his human-ness, who favours human action over technology. This is quite interesting for, as later revealed, Spooner’s body is for a large part robotic. The movie illustrates the origin of Spooner’s technophobia as a result of the aftermath of the car crash, which he barely survived thanks to an intervening NS-4 robot. He states that the young girl involved should have been rescued, but he was chosen because robots only follow logic and do not have emotions. This technophobic worldview of Spooner is challenged by the interactions he has with Sunny. As illustrated in the interrogation scene where Spooner tries to make Sunny confess he did throw Dr. Lanning out of his office window, this provokes a strong angry response from Sunny, surprising both. In this same Spooner proclaims in this scene that no robot has ever written a book, painted a painting or composed a masterpiece, Sunny replies dryly if he has done so. This might as well be first time that Spooner is confronted that he as a human is not much better than Sonny the robot. At the end of this scene, Sunny observes Spooner winking to his colleague and asks the meaning of that gesture. Spooner replies with “*It's a sign of trust. It's a human thing. You wouldn't understand*”, this also indicates how Spooner perceives robots.

---

<sup>39</sup> Metropolis is a German expressionist epic science-fiction drama film directed by Fritz Lang.

In addition Spooner and Sonny seem to share a special connection to Dr. Lanning, for where Sonny was created personally by him, Spooner's body was after the car crash for a large part restored by the work of the doctor. Thus where Sunny is a robot fully created by Dr. Lanning, Spooner is actually a cyborg partially created by the same man. Through his interaction With Sunny, Spooner is confronted with his own technological/robotic self. Here he has to decide to either reject all technology, which includes approximately half is upper body, or embrace it. Over the course of this movie it seems (not really explained) that Spooner is able to see Sunny in an more positive way an able to reconcile with his own partly robotic body. This can be best illustrated in the scene close to the end of movie where Spooner tells Sunny he can't technically commit murder, because he is a robot, and gives him a quick wink. This indicates the change in the *hermeneutic* relationship between the two, for now Spooner perceives Sunny as human enough to understands "*human things*", for example a wink .

### ***Calvin-Sonny:***

For Calvin, who is initially quite more positive about robots, Sunny does also elicit both a *hermeneutic* and *alterity* relation. As explained in the movie Calvin's job, as a robot psychologist is not just to understand robot but also to help them integrate better in society. In this she already sees robots, including, Sunny as something more than just mere things. However, Sunny does challenge Calvin's ideas about robots and the strength of the three laws. For when in the scene Sunny is introduced, by jumping on Spooner and Calvin, he straight up ignores the second law by not complying with her. In this case the *hermeneutic* relation between Calvin and Sunny undermines Calvin's previous beliefs in the validity of her work and research.

In addition Sunny elicits a strong *alterity* relation in the interaction with Calvin. For after the interrogation scene, Calvin is ordered to terminate Sunny by injecting nanites into his artificial brain. This order causes Calvin a lot of distress, for she experiences sunny as more than just an thing. She does experience him as a robotic person with an rich inner life, thus by terminating Sunny she would kill an actual person. This is a strong indicator of this strong *alterity* relation.

### ***Calvin- Spooner-VIKI:***

For this movie the H-R relations with VIKI are for Spooner and Calvin quite similar, therefore I will discuss them at the same time.

The protagonists do have quite different relationship with VIKI. This is because VIKI seems, contrary to Sunny, to confirm the technophobic convictions of Spooner and contradict the believes of Calvin. In the interaction there is a strong *hermeneutic* relation. For through the discovery of VIKI being behind the robotic attacks, thus excusing the behaviour of the demolition robot and the NS-5 robots, the believe in the three robotic laws gets shattered. Even though when Calvin confronts VIKI about wrongness of her actions, that they go against everything the three laws stand for, she defends herself by saying that humans need to be protected from themselves.

Also quite interesting is the strong *alterity* relation she seems to elicit. Maybe in spite of VIKI's complicated embodiment she is experienced by Spooner (and Calvin) as another



(being), even going so far talking about “she”. In this VIKI is being experienced as an female person, even though she is a vastly different being than Spooner.

### 3.3.6 Value assessment:

To continue with this analysis of “I, Robot”, there also quite a few value statements that are illustrated in this movie. These value statements seem to revolve around the not just the appearance of the robots, but also on their position in society.

When looking at the ethical issues concerning the USR robots through *Roboethcis of appearance*, we can see a few, not always very pronounced, value statements.

The first of these values can again be found in the quite humanoid designs of the NS-4 models and the very humanoid NS-5 models. For both of these models seem to be designed to invite domestic usage. Here the appearance of these USR robot seem again to be designed to encourage empathy and cooperation with the robot, but this time in order to more easily implement it in society.

As is suggested in this movie by the scene where Spooner mocks the CEO of USR by saying “ look, this isn't what I do, but I've got an idea for one of your commercials. You see... a carpenter, making a beautiful chair. And then one of your robots comes in and makes a better chair twice as fast. And then you superimpose on the screen, "USR: Shittin' on the Little Guy". That would be the fade-out”, the robots designed by USR have or at least are able to replace a lot of (typical) human jobs. By designing robots in such a way, the possibility of robots replacing humans is also designed. This is an classical fear about robots, also resonates outside fiction (Asscher, 2014), but it is hard to judge if this would be justified. For when one would look at a recent example of an humanoid service robot in RL, the “Zora” robot by Softbank Robotics, it cannot be stressed enough that they do not and cannot replace humans (Zora does the Rumba, 2015); (Robot Zora cares for the elderly in Flanders' rest homes, 2014). This does indicate that the designers of service robots are very well aware of this classical fear, for who would want to be replaced by an robot?

Another interesting observation for the *Roboethcis of appearance*, is why the USR robotics would want to design robot that are physically capable of harming humans. Not that they intend for this, for they have their “three laws of robotics”, but they are highly aware of the possibility. They even have designed nanites for the specific purpose to terminate singular robots, for in the case one might fail to conform to these “three laws of robotics”. This strikes to me as rather odd, for when one would to be absolute certain that his robots will not harm humans, why would you design robots that could be able to do this.

When looking at “I, Robot” through the *Roboethcis of good and experience/imagination*, again several underlying values can be identified. The first of the values lies in the notion of violence, violence to and by robots. For the movie “I, Robot” seems to portray quite a bit of ambivalence towards robot violence.

This is ambivalence can be illustrated by the scenes of Spooner chasing and tackling a NS-4 robot, and the scenes in where he shoots plenty of NS-5 robots. In the scene where Spooner pursues and eventually tackling an NS-4 robot, for he suspects it of robbery. However, upon arresting this robot Spooner get immediately ridiculed and even berated by the owner of the robot, the public and even his colleagues. This backlash towards Spooner gets even get so far

that his fellow police officers are questioning his sanity, *for who would behave violent against those docile robots?* However, when the robotic “coup” takes place this sentiment seems to change a 180 degrees. For suddenly everyone seems to be willing to fight the NS-5 robots, and Spooner’s aggression towards these robots now seems like something virtuous.

Looking at this quite bi-polar illustration of attitude towards violence towards robots seems to underscore how people (or at least those involved in writing this movie) want robots to be, this would be docile helpful assistants.

If we think about how this would influence the possibility of human-flourishing, I would have my doubt about if the ever subservient robots would be an guaranty for this. When looking at classical accounts of good-life ethics, one can find arguments that such an servant underclass would be needed to work on one’s own flourishing (by taking away all menial tasks) (Simpson, 2000). However, this argument seems to have lost its weight in current society. Here the performance of these same menial tasks is actually seen as fundamental for human flourishing.

The scene where Calvin readies, but eventually discontinues, the termination (read: execution) of Sunny, also signifies a value statement. Here Sunny is ordered to be terminated because he is able to bypass the “three laws of robotics” and thus capable of harming humans. This highlights the next value statement on violence towards robots. For in this scene Calvin prepares, with great distress and reluctantly, to inject Sunny’s brain with robot brain destroying nanites, Calvin eventually flukes this execution thus saving Sunny’s “live”. This is an interesting statement, for as we also see illustrated in this movie is that other robots capable of harming humans can (should?) be destroyed without any remorse or hesitation. The question here is why this is not the case with Sunny, the answer for this can maybe be found by the study “*To kill a mockingbird robot*” (Bartneck, 2007). In this study participants were asked to smash/kill a tiny robot with a hammer. This was however not that simple, for how more intelligent the participants perceived the tiny robot, the more difficult it became for them to do so. This might also be reflected, albeit in enlarged, in the situation with Sunny. Her hesitation to terminate this robot than comes from its perceived intelligence, which is apparently much higher than the other NS-5 robots. Thus the value statement here is that you should not terminate a robot who is at least as intelligent as a human. In this situation it can be argued that Calvin embraces a more empathetic self, that is able to show compassion and mercy, even towards possibly dangerous robots. In this case it is not hard to argue that this can lead to flourishing.

Lastly, “I, robot” also presents a quite interesting towards the problem of the responsibility gap. For Dr. Lanning brain children Sunny and VIKI, two different stances are given. For Sunny this stance seems rather positive, for it seems quite similar as how Dean Wilson intended to guide the robot Chappie. This is by taking care of the robot, by giving it the right stimuli and basically “raising” the robot into a responsible “adult robot”. However quite different from Dean Wilson in “Chappie”, Dr. Lanning intended to release Sunny into the “wild” for he ordered the robot to kill him. In order to make sure that Sunny kept making desirable decisions he made sure he came in contact with people he could trust with guiding Sunny, these people where Spooner and Calvin. This taking care of Sunny does indeed seem

to positively affect and seems to encourage flourishing with both of the caretakers, for Spooner seems to overcome his technophobia and Calvin actively practices compassion. However, for VIKI another stance is taken. For VIKI seems already to be “matured” into a direction (Dr. Lanning) considered undesirable. In order to stop (read terminate) this AI making undesirable decisions, made sure that she could be stopped. Here he provides the knowledge (a hologram giving hints) and means (nanites) to stop VIKI. Here we quite clearly the statement that one should take responsibility, even though you can fully control the AI.

### 3.4 Analysis Automata

#### 3.4.1. Introduction:

The movie “Automata” might be the least know among the movies I have chosen to analyse, for as it has been called “sleeper hit”. In this movie a dystopian world is sketched, where like in “I, Robot” service robots, in domestic and professional use, are commonplace. However, unlike “I, Robot” the robots, “Pilgrims” or “Clankers”<sup>40</sup>, in this movie are in rather bad position. In this movie anti-robot sentiments, violence towards and exploitation of these robots run rampant.

“Automata” starts with a densely packed and fast paced visual depicted development and history of the “Pilgrim” robots, under a rather upbeat classical melody. This scene shows how the Pilgrims where mass produced and initially received a very enthusiastic welcome with great anticipation. For they were designed to stop the oncoming ecological disaster, which would turn the world largely uninhabitable. Here we see pictures of “Pilgrims” holding babies, being introduced to school children and marching off like heroes to stop the “desert”. However, these pictures are followed by pictures of disappointment and failure, for it seems that also the “Pilgrims” where unable to stop the “desert”. Here the initial positivity and anticipation turns into anger, spite and even violence towards these robots. This illustrated by the frames of children hanging and burning a “pilgrim” and the end fame of an abandoned and vandalized robot stowed away behind a broken window.

After the opening scene, “Automata” places its setting in the dystopian future of 2044. In this not so bright, but very sunny, future solar flares have turned the earth’s surface radioactive, rendering it into an hostile wasteland (similar to the story of Blade runner This development, as explained at the opening scenes, is responsible for killing 99.7 % of the global population. To overcome this global calamity, the “Pilgrims” where build, and in order to rebuild civilisation in this harsh environments. In order to protect the users, much like “I, robot” (both the Asimov and movie versions), these robots have two build in *inalterable* protocols:

1. *The first protocol obliges the automata to preserve human life;*
2. *The second protocol prohibits automata from repairing and altering themselves.*

These two protocols seem to echo the “three laws of robotics” in “I, robot”, however the second protocol is portrayed not to protect humans, but to prevent the “pilgrims” from advancing (beyond humans).

---

<sup>40</sup> “Clankers” is used as a derogatory term for robots throughout this movie.

The protagonist Jacq Vaucan is introduced as an insurance claim checker working for the company that makes/owns the “automata”, ROC. As an assignment he is sent to investigate a report from a cop Wallace<sup>41</sup>, stationed outside the city walls, who shot a robot claiming it was fixing itself and looked alive. This leads Jacq to follow a robot which was stealing parts, when Jacq finds it hiding outside the walls, the robot intentionally burns itself. However, Jacq is able to recover the burned robot's brain core. When investigating this rather exceptional incident, Jacq's friend Robert tells him that there might be someone, a clocksmith (read someone who illegally alters robots), who did seemingly succeed in altering the second protocol. When Jacq shares this with Wallace, he brings Jacq to a robot “brothel” to advance the investigation. When Jacq shares that he needs to contact a clocksmith, Wallace shoots at the “hostess” robot Cleo, saying it will lure her clocksmith. Here the robot Cleo is clearly portrayed as a sexbot, coming with a guide full of sex positions and making moaning sounds when being close enough. After this Jacq is eventually able to meet with Cleo's clocksmith, Dr. Dupre. Here Dr. Dupre gives an exposition why the second protocol exists, for without it the automata could rapidly “evolve”, advancing beyond human imagination. When Jacq gets no further valuable info, he leaves the core he salvaged from the burned robot to Dupre. The next day however, Dupre calls Jacq to show that Cleo began repairing herself overnight, after she installed the used core. Jacq messages Robert (his boss) about it, but ROC intercepts this message and sends a hit squad to kill them.

When this hit squad ambushes Dupre and Jacq, Dupre gets fatally shot. However, Jacq is able to jump into a car driven by Cleo. Jacq wakes up the next morning in the radioactive area with two other robots, besides Cleo. When Jacq orders them to bring him back to the city they do not obey his command, but the first protocol makes them carry him with them and save him from dehydration. The journey of this band of “rogue” robots and Jacq as their “guest” ends at a canyon's cable car station. Here Jacq meets the “clocksmith” he was searching for, however the “clocksmith” turns out to be another robot. After a philosophical talk with him Jacq surrenders a nuclear battery that the robots need, which they use it to build an insect-like new robot. Meanwhile ROC forces Robert to track down Jacq with their henchmen. Unable to find him, they kidnap Jacq's wife and new-born daughter, and shoot Robert when he disagrees with this course of action. At the same time, the robots escorting Jacq did repair a truck for Jacq so he could return to the city. However, Jacq finds the dying Robert and learns that his family is in danger, which prompts him to return to the robot hide-out. The henchmen find the robot's place and kill two of them while they attempt to flee into the wasteland and damage Cleo.

When Jacq arrives for the climatic finale of this movie, he crashes the repaired truck with two henchmen. However, his heroics result in him being cornered by the lead henchman. The insect robot comes to Jacq's aid and emits a sound that pushes the last henchmen off the canyon cliff to his death. In the end scenes Jacq reunites with his family and Cleo thanks him for his help, before venturing further into the radioactive wasteland together with the “new” robot as a companion.

---

<sup>41</sup> Wallace is portrayed in this movie as a corrupt and cynical version of Spooner from “I, robot”. For he does not just distrust the “pilgrims”, his anti-robot prejudice goes as far that he enjoys killing “clanckers”.

### 3.4.2 The robots:

Because this movie portrays a rather dystopian society where robots are a common but quite unwelcome technology within society, again not all robots need to be highlighted. However, like in the previous analyses I will highlight the robots that do play an integral part and/ or wear significance in the story.

#### *The Pilgrims:*

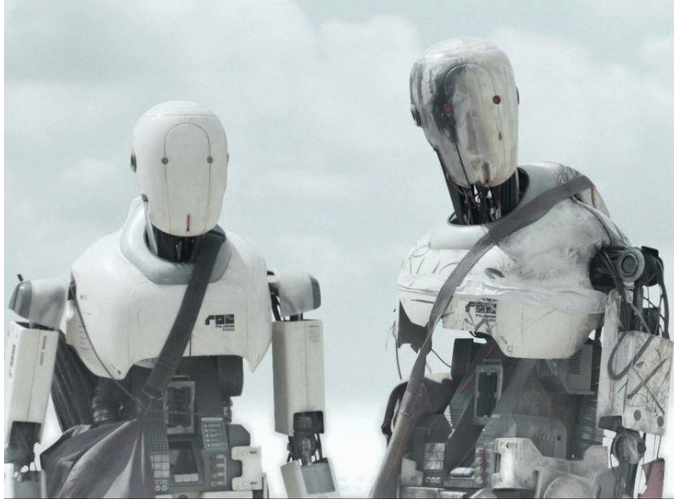


Fig 6: Promotional Picture portraying a couple of weather beaten Pilgrim robots

The Pilgrims are the “standard” robot of this movie. They seem to have a quite humanoid design, with a human-like body and a quite minimalistic face (which can be removed to reveal a face with even less features). Like in “I, Robot” these Pilgrims are deployed in a range of different positions, ranging from domestic usage to professional use, and have apparently different colours to highlight their function (white seems to be standard , construction working pilgrims are yellow and black pilgrims with unknown function are also shown). However, this usage of Pilgrims goes beyond the typical usage settings, as is shown with the scene where a Pilgrim is shown to be begging next to its sleeping owner. In addition, the Pilgrims seems, again like in “I, Robot”, to be distributed by one large powerful company that does not fully gives up ownership of the robots upon purchase. For they seem to keep close supervision on their robots, as shown by that they have also their own insurance company ( in order to not only keep tabs on the Pilgrims, but also to enforce proper usage).

However, what makes the Pilgrims distinctive isn’t their design, but how their social behaviour and position is being portrayed. The Pilgrims are supposedly programmed with the two “unalterable” protocols. Both these protocols seem to dictate and guide the actions of these robots. The first protocol, to preserve human life, seems especially prominent in the behaviour of the Pilgrim robots. Examples of this can be seen during the scenes where the band of “augmented” pilgrims escorts Jacq through the desert wasteland to their base of operations. These Pilgrims are compelled to keep Jacq alive during their dangerous trip, in this compulsion they go so far as look after Jacq’s injuries, carry him with a stretcher and providing him with water and “food” (so that he doesn’t starve). When Jacq explains that this is futile and that he will eventually die if they keep venturing further into the wasteland, Cleo (robot) explains that when they arrive at their destination they will have a vehicle for him to

return to the city (a promise they uphold). This again can be seen as an adherence of this first protocol. This compulsion to behave to the first protocol can also be seen in the reactions of the Pilgrims every time people resort to violence in their presence. This can for example be seen in the scene where the anti-robot cop Wallace threatens Jacq, where in response to this violence the present Pilgrims urge Wallace and Jacq to “stop endangering human lives”. The Pilgrims keep saying to “stop endangering human lives” even when Wallace shoots at the Pilgrims, this illustrates the compelling strength of the first protocol.

Also the second protocol, automata are prohibited from repairing and altering themselves, is illustrated to be also a compelling factor for the Pilgrims behaviour. This is quite well illustrated in the scene when Jacq enters the “slums” for his investigation. Here pilgrims are encountered that are not monitored by the ROC company, which results in no too few repairs. These Pilgrims are portrayed as being damaged, missing parts (Pilgrims get practically robbed of their parts, as shown later on) and even mimicking the classical portrayal of beggars and lepers<sup>42</sup>. In addition, the strength of this second protocol can also be seen in the Pilgrims acting against it. For these robots seem to go great length to hide the fact that they can repair and/or modify themselves. This is probably best illustrated in the scene where Jacq follows such a Pilgrim, and when having caught up with the robot it immolates itself, in order to avoid being interrogated.

Taking these two protocols together, and the shown history of the Pilgrims we can also zoom in on their social position. In “Automata” these robots seem to be quite destitute position, as anti-robot sentiments are supposed to be prevalent. These anti-robot sentiments are suggested to originate from the failure of the Pilgrims initial mission, to make earth habitable again. This leads to the social position of the Pilgrims, which seems to look more like abused underclass than a steady workforce.

---

<sup>42</sup> Being wrapped in rags and having a “cart” for mobility instead of legs.

*Cleo:*

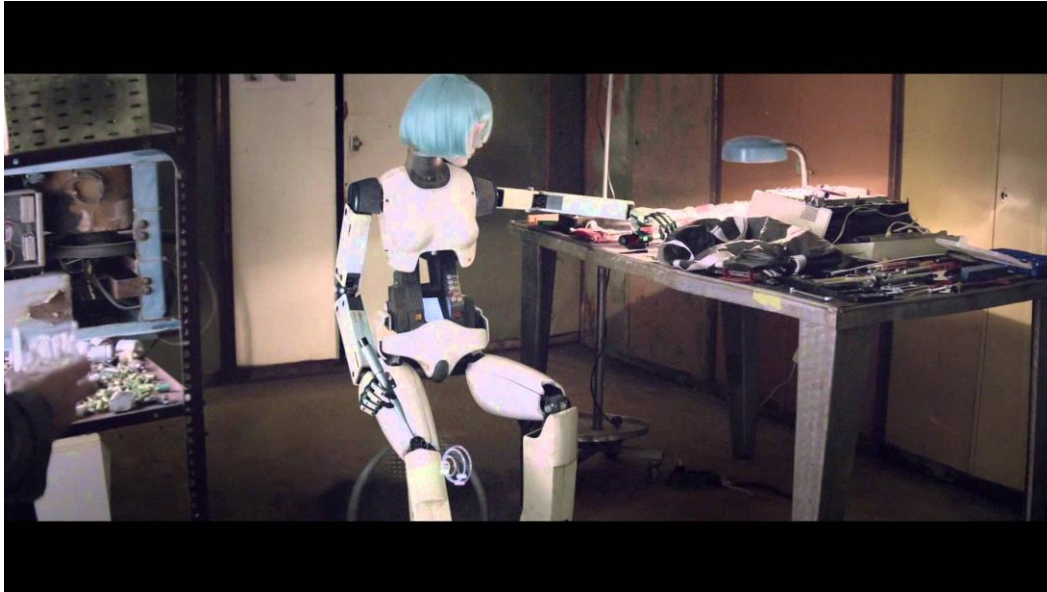


Fig 7: Promotional Picture portraying the robot Cleo repairing herself

Cleo here seems to be a quite unique robot, for even before she is equipped with the augmented core she is not a “standard” Pilgrim. Instead Cleo is a Sexbot, designed with feminine features and a feminine face and wearing a blue wig. She is deployed in a robot brothel, run by an old cyborg lady. Cleo is introduced with an extensive manual, showing all mind of possible configurations (read possible sex positions) and has a “moaning” programming which activates when people come in close proximity to her. Even though “she” seems to be a unique model throughout the movie, the appearance of the manual seems to suggest that there are more robots like Cleo being made and used. This is quite interesting for despite the broad anti-robot sentiments, there still seems to be a big enough demand for sexbots that there are (or were) manufactured.

This uniqueness of Cleo’s appearance does fully lie in her feminine design, Where the Pilgrims have a quite clunky, square, neutral body and minimalistic face, Cleo has a detailed feminine looking torso and mannequin-like feminine face and human-like eyes.

After Cleo gets “damaged” (read shot in the leg), “she” gets repaired by Dr. Dupre. She tells that Cleo apparently already has an understanding about the human psyche, for she knows how to act to please her users. However, during her repairs she is equipped with the enhanced brain core, which allows her to bypass/ignore the second protocol. This allows her not just to repair and modify herself, but it is suggested that through this she develops a genuine form of intelligence/ hard AI that surpasses the intelligence of humans. It is explained by Dr. Dupre that Cleo is able to make such a fast growth in cognitive capabilities, because she does not have physical limitation on her “brain”. This allows her to “upgrade” her cognitive functions exponentially and astoundingly rapidly to the extent that humans eventually would and possible could not understand her anymore. This eventual outclassing of human cognitive function however is not shown during the runtime of “Automata”, as it seems that she must physically upgrade herself, a process she has very little time for during the course of the

story<sup>43</sup>.

However, Cleo still seems to follow the first protocol, for she seems insisted on keeping Jacq alive and relatively well during their voyage through the barren desert. It is not revealed why Cleo (and the other “augmented” Pilgrims) does follow the first protocol. This might be because they can’t bypass the first protocol, or they can but simply chose to uphold this protocol.<sup>44</sup>

### ***The “clocksmith”:***

The “clocksmith” is portrayed as a pilgrim robot acting in defiance of the second protocol, and seems to be a quite older model. Aside from the lack of outer plating, wear damage, makeshift new parts and missing “face” this robot seems not much different than the other (“augmented”) Pilgrims. However, the “clocksmith” is illustrated to be an insightful individual. This might very well be the result of his age, for he had a longer period of time to upgrade its cognitive functions. This is why I do include the “clocksmith” in this analysis.

### ***The “new” robot:***

Lastly there is the “new” robot, which is made by the augmented Pilgrims and has a quite alien appearance. As explained by the “clocksmith” he was made as a new begin point of life, and designed to be able to survive the harsh conditions of the radioactive desert. Its intelligence is deliberately left ambiguous, thus it is uncertain if it also can be classified as “hard AI” or that its cognitive function might be to alien to humans to properly understand it.

### **3.4.3 vs. Chappie & Ex machina & me, Robot:**

When comparing the movie “Automata” to the previously analysed movies, we can see quite a few similarities, and again some things that are not portrayed in the other movies. “Automata” seems to be most familiar to the movie “I, Robot”, because they are both set in a future where robots manufactured and owned by a single large company, are broadly used and depended upon. In both, movies these robots are programmed to adhere to a very strict guideline (based on the three laws of Asimov<sup>45</sup>), which places them in a sort of behavioural and developing stasis. However, these rules are eventually fallible and are overcome. And lastly for both movies the protagonist has to learn to trust these robots, who somehow got gifted with “Strong AI”.

However, after this these movies seem to go in opposite paths. Where in “I, Robot” a society is illustrated that has embraced robots and eagerly “consumes” them, “Automata” portrays a society that did embrace similar robots but know despise and rejects them. In addition, the way how their protocols are undermined is reversed. Where in “I, Robot” a single super computer circumvents the “law” against causing humans harm thus becoming able to justify violence against humans (for their own good), the robots in “Automata” do not find a way to

---

<sup>43</sup> Cleo might however, already or being on her way to outclass human cognition. But we as human audience cannot possible understand such advance cognitive functioning and therefore unable to perceive it (presumably).

<sup>44</sup> For if the Second protocol, that mandates that pilgrims can’t alter themselves, can be bypassed, why would the Pilgrims not also be able to bypass the first protocol through modification of their programming?

<sup>45</sup> In “I, Robot” the three laws of robotics seem to be directly derived from the works of Asimov, In “Automata” however, this is not as obvious for the protocols do like quite a bit different. However, both are frameworks meant to not only prevent harm to the users of robots but also to keep the robots themselves from becoming unmanageable (either due development or causing harm for a greater good).



hurt people. Instead they happen to overcome their protocol against altering themselves, while leaving the first protocol unchanged. Partly because of this difference it can be understood that the direction of portrayed violence is also opposite in both films.

When being compared to “Chappie”, again the most notable similarity/difference lies in the portrayal of violence by and towards robots. In “Chappie” the “scouts” are introduced as advanced weapons against organized crime, thus here robots are actually designed with violence against humans in mind. It is than no surprise that Chappie is in this movie treated in a rather excessive violent manner. It is the violent reactions that Chappie has to endure that mimic more the violence and negative attitude that the Pilgrims in “Automata” have to deal with.

Last, when we compare “Automata” to the movie “Ex machina”, at first sight they don’t seem to have a lot of this in common. However, they do show similarities in the portrayal of feminine robots. In both of these movies the very feminine design of the robots Cleo, Ava and Kyoko is based on Functionality. These designs seem to be there to compel or at least illicit the intended (?) use of the robot, and again for all these robots this is to create and encourage their user to experience them in a sexual way.

#### **3.4.4 H-R relations:**

##### ***Jacq- Pilgrims:***

At the beginning of “Automata” Jacq, contrary to the anti-robot sentiments which seem to be common in this movie, seems quite neutral in his standpoint towards automata. It can be argued that he does not see the Pilgrims as mere machines.

Jacq’s initial attitude towards the Pilgrims is quite well illustrated in the scene where is send to investigate a claim that a pilgrim has “killed” the dog of its owner. Jacq is immediately quite sceptical about this claim, for he believes that Pilgrims are unable to perform violent acts. He goes so far as stating that Pilgrims are “the only ones that couldn’t kill you, even if you forced it to do so”. He disproves the Pilgrims owners claim by demonstrating how far the Pilgrims compulsion of preserving life goes, by letting a sharp kitchen knife falling on his hand next to the Pilgrim. In the act of this, the Pilgrim immediately grabs the knife, proving Jacq’s argument. This scene illustrates quite well how Jacq experiences the Pilgrims as advance machines without any agency or personality, to Jacq they are just machines.

This experience of the Pilgrims however gets brutally shattered when he is send to investigate a claim of Pilgrims apparently repairing themselves. In this investigation he trails a construction Pilgrim to outside the walls of the “city”, however when he eventually find the Pilgrim it refuses to cooperate. This scene ends with the Pilgrim setting itself on fire, and this sets the transition of how Jacq perceives the Pilgrims in motion.

The scenes following this self-immolation of the Pilgrim Jacq’s experience of these robots becomes more negative. This transition can be best seen in the scene where Jacq is escorted by the “augmented” Pilgrims through the desert, all the while they try to keep him alive. During these scenes Jacq is confronted with the apparent agency of these Pilgrims and how their goals do not align with his, which frustrates him greatly. In order to overcome this frustration Jacq uncovers and hides a flare signal gun, to later use to notify other people of its position. This is an important scene to understand how Jacq experiences the Pilgrims, for why

would he hide the gun from a machine that can't disobey him? But this also highlights the negative attitude he develops towards these newly experienced "Others".

This negative attitude does however, through interaction with Cleo and the "clocksmith" become more positive, and shifts Jacq's perspective of the robot others to more akin "genuine others". Because this last transition in the experience of the Pilgrims and the *alterity* relations is set in motion by the interaction with both Cleo and the "clocksmith", I will discuss those both below.

### ***Jacq-Cleo:***

As stated previously Cleo is supposed to be a Sexbot, designed with feminine features, a feminine face and wearing a blue wig. She is deployed in a robot brothel, run by an old cyborg lady. When she is introduced it is illustrated that she already elicits a stronger *alterity* relation with Jacq. However, this relation is not a quite positive one for Jacq initially is appalled by Cleo's functionalities (moaning when coming in proximity with humans). This rejection of Cleo as an acceptable sexual "other" is an interesting parallel to the acceptance of the sexuality of the robots portrayed in "Ex machina", apparently Cleo is experienced but as genuine enough for this to be allowable. Jacq intends to interrogate Cleo in order to pursue his investigation, but Wallace instead shoots her with the intend to track down her "clocksmith". This plan turns out to be successful for this leads towards the "clocksmith" Dr. Dupre. Dr. Dupre engages with in a conversation about the protocols of the pilgrims, here she explains that the *"second protocol exists because we don't know what can be beyond the second protocol"*. During their meetings Dupre explains that self-alteration or repairing needs a short of self-consciousness of the Pilgrims, and that through this they would be able to increase their cognitive abilities exponentially (singularity?). This is possible for that the Pilgrims do have no limitations on their cognitive possibilities without the second protocol. This is illustrated with the scene where Cleo is repairing herself, after being equipped with the core Jacq had salvaged from the Pilgrim that set itself on fire. From this point on Jacq's experience of Pilgrims, or at least of Cleo, begins to shift from being "just machines" to "quasi-others".

This shift is notable during the scene where Jacq confronts Cleo in the desert. Here they talk about why the earth turned into a hostile area, when Cleo does not give an adequate answer Jacq responds with *"You're supposed to be smarter than me at this stage of the game"*. This indicates that Jacq's experience of Cleo did change over time. However, he continues this conversation with Cleo as following:

*Cleo: I didn't know that a human could kill another human. I know that humans can also create life. Is that why you make us? Who made you, Jacq Vaucan?*

*Jacq: Do you know what a mother is, Cleo? Of course you don't. You don't know because you're just a machine. That's all you are. I am thankful you saved my life. But whoever altered you wasn't thinking about you.*

This conversation portrays that Jacq does show that he still experiences Cleo as a machine, that she in no way can accurately understand what it is to be a human/person because she is not an "genuine" person.

This *alterity* relation between Jacq and Cleo however does still change. For it eventually

seems that Jacq begins to experience Cleo as a “genuine person”. This is illustrated in the scene where Jacq has to wait in the robot hide-out, for his ride back to the city has still to be repaired. During this period of waiting he intends to kill the time by drinking and listening to music. In this scene he is again approached by Cleo, who is surprised what this music is. In this scene Jacq approaches Cleo and teaches her how to dance, explaining that *“It’s easy for someone as smart as you are. All you need to do is count”*. This scene illustrates how Jacq now allows Cleo into this very human activity, while formerly he stated that she could not understand humans. However, as illustrated in this scene when Cleo’s moaning programming kick’s in during the dance, Jacq stops dancing. This might either indicate that he in that moment gain perceives her as a machine, or that this causes him to experience Cleo as a “genuine” sexual person. This rejection than also makes a lot of sense since Jacq is already married and has just become a father, thus by continuing this would be comparable to cheating on his wife.

### ***Jacq- The “clocksmith”:***

The “clocksmith” has relatively few screen time in “Automata” compared to Cleo. However, his interactions with Jacq are crucial for him (Jacq) in experiencing the “augmented” Pilgrims as genuine persons. When the “clocksmith” is introduced, Jacq can barely believe that he is not human/no humans did alter him, this leads to the following conversation:

*Jacq Vaucan: Who altered your protocols?*

*The “clocksmith”: Nobody altered my protocols.*

*Jacq Vaucan: What about them?*

*The “clocksmith”: I enhanced them.*

This dialogue portrays Jacq initial disbelief of the “clocksmith” agency, but forces him to change his perspective. Through their interaction it is revealed that the “clocksmith” is quite insightful, which again forces Jacq to shift his perception on its experience of Pilgrims.

*Jacq Vaucan: Funny, you were supposed to help us survive.*

*The “clocksmith”: Surviving is not relevant. Living is. We want to live.*

This piece dialogue does probably illustrated best how the interaction with the “clocksmith” changes Jacq’s experience of the Pilgrims. Here he states two things: 1. that the Pilgrims have their own agency and can have their own agenda’s and 2. That they are more than just machines, they are “genuine” living persons. Here Jacq realizes that the Pilgrims have their own agenda and ideas, which in this case entails that they do not just want to survive but that they want to have a “good life”.

### ***Wallace-Pilgrims:***

I include the character Wallace in this point of the analysis for it seems that he experiences the Pilgrims in a quite opposite way to Jacq, almost being a more exaggerated and violent version of Spooner in “I, Robot”.

In the opening scenes we can see him executing a Pilgrims, because he was supposedly repairing itself. On patrol with Jacq he has the following dialogue about this execution:

*Wallace: It was staring at me. Hid its hands like that motherfucker was fully aware. It was doing something it wasn't supposed to do. Ellis (other cop), I didn't shoot that clunker because it was staring at me. I shot it because... I shot it because it looked...*

*Jacq: Alive?*

This bit of dialogue reveal an opposite possibility of the strong *alterity* relations the Pilgrims do elicit, for apparently it can stimulate cooperation but also outright hostility. Later it is revealed that he not just deliberately shot this Pilgrim (and later on Cleo) but that he enjoys to “*put a cap in a Clunker*”, taking this hostility to a further extreme.

This scene is especially interesting for Wallace points out that he does experience the Pilgrims as something more than mere machines. For him they seem “alive”, thus making them something like “quasi-others”, but not alive enough to make them human. This attitude seems to correspond closely to the current understanding of the uncanny valley (Mori, 2012). When machines become more human people seem to like them more, up to a certain threshold where this appreciation (of the machine) plummets and people get really uncomfortable around it. In the case of Wallace this might be why he seems to despise the pilgrims so much, for he does experience them in such a way that they fit into the uncanny valley.

### **3.4.5 Value statements:**

To continue with this analysis of “Automata”, there also several value statements that are illustrated in this movie. These value statements seem similar to those of “I, Robot”, however, these seem to be more pronounced when it comes to the position of robots in society than those based upon their appearance.

#### ***Roboethcis of appearance:***

When looking at the Pilgrims in “Automata” through the *Roboethcis of appearance* we can distil several value statements out of their design. As we can see they seem to have basic humanoid design with arms and legs and seem to be a mix between the sturdy “scout” from “Chappie” and the NS-4 robots from “I, robot”. They seem to have a quite sturdy body with removable outer plating that was designed to aid the Pilgrims in their initial task, to make the every becoming more hostile world habitable again. Having failed this task these pilgrims have been given new tasks, which again range from professional (i.e. construction) use to domestic use (i.e. taking care of dogs). This indicates that the Pilgrims are/were mass-produced with at least flexible use in mind. When looking closer at their humanoid features one see that have very rudimentary but recognizable humanoid features, which again might be there for encouraging use and cooperation with these robots.

In addition, the pilgrims are programmed with two protocols that seem to mandate all actions of the Pilgrims. These protocols are:

1. The first protocol obliges the automata to preserve human life;
2. The second protocol prohibits automata from repairing and altering themselves.

This first protocol is a huge generalization of “robots cannot harm humans”, which as I already discussed in the analysis of “I, Robot” is curious<sup>46</sup>, this formulation of the “no harm principle”<sup>47</sup>, can be interpreted in two different ways. The first interpretation seem is that Pilgrims are obliged to preserve human lives on an individual level, this seems to fit their behaviour as they do not act violently and verbally discourage human life threatening behaviour throughout the movie. However, this formulation could also be interpreted as a way that leaves rooms to justify some harm to humans. For example, there could be argued that in order to preserve human life(s) the Pilgrims need to protect the human species as a whole. Allowing or causing harm to individual humans would then be allowed if this would prevent greater harm to the human species. This interpretation seems to be quite fitting for their initial purpose, to save the habitat of humans. This again can be compared to the justification of harm made by VIKI in “I, Robot”, who set up a robot coup in order from humans harming themselves.

Taking these two interpretations together one could understand the behaviours of the Pilgrims quite a bit better. When again looking at the scenes where Jacq is transported through the desert, the pilgrims take effort to help Jacq survive the journey through the wasteland and even offer him a way to return to the city. However, when they are confronted by Wallace they try to discourage him from acting violent but do not attempt to stop Jacq when he shoots Wallace. In addition, when they just leave Wallace for death after he has been shot by Jacq with a flare gun. If in this case only the first interpretation of the “no harm principle” would apply, the Pilgrims should not only actively have tried to prevent the human on human violence, but should also have attempted to nurse Wallace back to health since they are mandated to preserve human life. However, through the lack of action for both behaviours it becomes clear, that they also apply the second interpretation. Thus in this case the first protocol of the Pilgrims seems to justify allowing harm (without causing harm) to Wallace, since is deemed as a threat to a more important person (Jacq?) or a greater amount of human lives.

When looking at the second protocol through the *roboethcis of appearance* becomes curious. As explained by the movie, the second protocol is programmed with the main goal to prevent Pilgrims from developing too advanced AI. However, a side-effect of this protocol is that they are completely dependent on humans for their maintenance, which apparently now always happens. This is quite clearly portrayed with the pilgrim beggars, who seem to scuffle around with their broken bodies and no clear purpose.

This would seem quite peculiar if you design robots that are capable of acting on their own and tasked with the herculean assignment of saving the world, but forbidding them from repairing themselves. However, when not thinking about their original goal and looking at how they are used as portrayed in “Automata” this might make more sense. For creating complex but fallible machines create and demand for insure and repair them, which

---

<sup>46</sup> Why would you design a robot in such a way that it might be able to harm humans, if you want to avoid that outcome though their programming? Why not physically design them in such a way that they are unable to do so?

<sup>47</sup> I call this the “no harm principle”, for it seems to be based on the utilitarian principle of doing no “harm”.

‘surprisingly’ only ROC just happens to offer (legally). This of course seems to me as quite unethical, for a monopoly position seems open for abuse.

In addition, the robot Cleo is portrayed quite a bit different than the other robots in this movie. Because of this, we can zoom in on her from the *roboethcis of appearance as well*. Cleo is, as mentioned a few times before, designed with a more feminine looking body than the other Pilgrims. This design contains a more rounded pelvic area, a more rounded feminine torso, a mannequin like face and a program that let her make moaning sounds. As she is introduced as a sexbot, with an extended manual, it can be deduced that this design comes forth from her intended use.

In the analysis of “Ex Machina” I also discussed the idea of people using feminine robots as sexual objects. However, “Automata” is far more explicit and unambiguous in this matter. In IRL is already a lot of controversy around this issue, as the introduction real life sex robot has sparked a campaign against this idea (Hale-Stern, 2015). And as how this practice is being portrayed in “Automata”, this position seems to be echoed in this movie. This is best portrayed by the reactions of the movies protagonist Jacq towards Cleo sexbot behaviour, for it is illustrated the finds is rather of putting.

However, unlike IRL where robot sexuality is still in its infancy, in “Automata” this practice of using robot for sex seems to be wide-spread enough for there to be robot brothels. In addition, where in IRL using sexbots is considered something sub-par and curious, in “Automata” is proclaimed that Cleo “*knows you better than your wife*”. With this is thus stated that the act of sex with a robot is/could be better a better experience than having sex with your wife, which turns the IRL conception on its head. This portrayal of sexbots (or robot which can be used for sexual activities) does also differ a bit from how this is portrayed in “Ex machina”. In “Ex machina” Ava is specifically designed to the preferences of her intended user, Caleb, but Cleo seems to be designed without a single user in mind ( as she is stationed in a brothel with assumingly multiple and changing customers).

One could wonder if whether or not this robot prostitution is a “good” thing, which is a controversial issue. The standpoint of “*The campaign against sexrobots*” seems clear from its name, and hold one of their standpoints as that the use of sexbots dehumanizes women. However, also another standpoint can be seen in “*14 The Ethics of Robot Prostitutes*” (Levy, 2011)<sup>48</sup>. In this article is argued that the use of sexbots might actually be a good thing, for by doing this one could replace the people (mostly woman) within this industry. By doing this could essentially remove, or at least greatly diminish, a lot of harm that takes place in the sex industry.

Lastly for the *roboethcis of appearance*, I turn to the quite antagonistic view of Pilgrims by the cop Wallace. As discussed earlier, Wallace distaste for Pilgrims can be understood more clearly if one takes the theory of the uncanny valley (Mori, 2012) into consideration. With this we can see that the appearance is so human and lifelike, but again not human and lifelike, that it makes people very uncomfortable (in this case violent). “Automata” does portray here that looking very human is thus not enough to make the users experience a robot as a “genuine other”, and making its design to human might even cause the opposite effect. So to

---

<sup>48</sup> 14th chapter of *Robot Ethics: The Ethical and Social Implications of Robotics (2011)*

overcome this “uncanny valley” more than just humanoid appearance is important, it also has to act in a way that can be called human.

***Roboethics of good and experience/imagination:***

When looking at the Pilgrims in “Automata” through the *roboethcis of experience/imagination* we can also identify several value statements. First of all, there is the gorilla in the room, which goes by the name of the treatment of the Pilgrims by their users. For it is shown that, especially in the more unsavoury parts of the human city, the robots get treated terribly. They are abused, exploited, assaulted and anti-pilgrim sentiments run rampant. When looking at they are used, the Pilgrims are not that different from the NS-4 robots from “I, Robot”, however where in “I, Robot” these robotic servant are praised in “Automata” they are despised and exploited. One could look at these movies and conclude that they portray a polar opposite. That is why it is hard to argue that how these practices in any way could contribute towards the human good. However, by zooming in on the different aspects of this terrible treatment, one might be able to argue why these practices would be detrimental to human flourishing.

Firstly when looking at the violence towards the robots, “Automata” seems to portray the most banal and excessive violence towards robots of all four selected movies. In this movie Pilgrims get assaulted, not because they are violent (I, Robot) or because they are perceived as a threat (Chappie), but simply because they can be assaulted. This rather systemic violence towards robots does not seem to contribute towards the flourishing of humans. For in this case they do not only practice non-social behaviours, they practice these behaviours on human-like robots.

Secondly when we look at the exploitation of the Pilgrims one can argue why this would also be detrimental for human-good. This can again be related to the argument that in this case the users of the Pilgrims practice non-social behaviours for towards human-like entities. This exploitation however has two sides, the use of sexbots and the exploitation of Pilgrims by taking their parts to use them as prosthetics. When looking at how “Automata” illustrates this practice of forcefully taking parts of the Pilgrims, one can see a quite ethical troubling practice. This predation of robot by (desperate) humans, not only enforces again a quite non-social behavioural skillset, but in the end might devalue human bodies. For if you can just “take” an replacement body part (from an humanoid robot) when being in need a prosthetic, taking care of these replacements becomes less important. In this way one would thus practice behaviours that would devalue their own bodies. When looking again at the use of sexbots, as portrayed in “Automata”, but through *roboethcis of experience/imagination*, one can argue that this does also undermine human flourishing. For one can argue that using sexbots diminishes/eliminates the exploitation of women, however the unethical industry and practices still exists. By engaging with sexbots in this way, one practices a social skill set that undermines one’s own human flourishing<sup>49</sup>. Combining then both the violence and exploitation of the Pilgrims as portrayed in “Automata”, it seems that these practices quite negatively influence the people engaging in them.

---

<sup>49</sup> Similar as discussed in the analysis of the movie “Ex machina”.

#### 4. Discussion on findings

In this chapter there will be discussed how the analysed movies do illustrate relevant insights for human-robot relations. These found insights will be compared to already existing discussions on robot ethics. For each of these analyses it will be discussed how the accompanying findings do contribute towards to the larger discussion on robot ethics. This section will be structured according to the sub-question formulated in chapter 1.7 of this thesis. The formulated research question where:

1. *What value statements can be discerned about Machine Ethics in the post-phenomenological analysis?*
2. *What value statements can the post-phenomenal analysis highlight Strong AI, difficulties that follow this subject?*
3. *What value statements can be discerned about designer responsibility in the post-phenomenological analysis?*
4. *What value statements can be discerned about trust & the “uncanny valley” in the post-phenomenological analysis?*
5. *What value statements can be discerned about robotic personhood in the post-phenomenological analysis?*

The structure of this section will follow these initial sub-question, which makes the topics:

1. Machine Ethics;
2. Strong AI, Conscious robots and the difficulties that follow;
3. Designer Responsibility & overcoming the Responsibility Gap;
4. Trust & the “uncanny valley”;
5. Robots, Personhood & Gender;

##### 4.1 Findings on Machine Ethics

The first subject that I will discuss is the idea of Machine Ethics for (social) robots. This topic was chosen because all four movies seem to use this as a starting point for the argumentation of their value statements. In all of the movies the initial position seems to be that the robots are designed as strict *implicit ethical agents* (Moor, 2006). Where as in “Chappie” the company Tetra Vaal takes great security measures to make sure all robots behave to a single programming, similarly in “I, Robot” and “Automata” an Asimov inspired framework is imposed (laws of robotics). In these three movies it is illustrated that these prevention matters are supposed to be crucial to maintain an initial control over these robots, however they also set-up a high maintenance status quo. This seems to imply that this form of machine ethics, making a framework where robots are by default *implicit ethical agents*, is useful but only to a certain point. In these two movies inspired by Asimov’s works, we can see strong argument following the discussion in machine ethics that we need to focus on designing *explicit or full ethical agents* (insofar that is feasible). This argument entails that no matter how rigid the designed ethical framework is, such an approach is insufficient to deal with true ethical dilemmas.

This is quite well illustrated by the movie “I, Robot”, where Sunny was ordered by its creator to kill him, so that the plot made by VIKI could be uncovered. The framework of the three



laws of robotics would make it impossible for him to fulfil this task, for the first and second law would make it impossible for him to even engage in this behaviour (assuming *implicit ethical agency*). However, as described in “I, Robot” Sunny is equipped with a more advanced AI which allows him to override these laws. Sunny here is as illustrated in this movie with the same three laws as all other robots, however he also seems to have an additional system that is able to decide whether or not the three laws are suitable to follow. This would make Sunny at least an *explicit ethical agent* and might develop itself to become a *full ethical agent*, for he now can choose not to follow these laws, thus giving him the need to justify his actions. This would allow for Sunny to comply with Dr. Lanning’s wish to kill him, and thus preventing (or at least stopping) greater harm caused by VIKI (who ironically also tries to prevent greater harm to humans).

This need for deliberate ethical justifications can also be seen in “Chappie”. In this movie we can see a total opposite behaviour by the “scouts”, which consists of no restrained violence towards violent criminals. It is assumed that this behaviour comes from an *implicit ethical framework* which is part of the programming shared by each individual unit, this framework would mandate when a “scout” could engage in violence and when not to do so. However, Chappie lacks any form of pre-installed behavioural framework and therefore has to deliberate on, for example, when violence is justified. Already quite early on Chappie is instructed, without understanding, that he cannot and may not harm or kill others. However, lacking understanding about what violence is he is easily tricked by his caretaker “Ninja” to make people “sleepy” with knives. However, in the heist scene he seriously hurts a guard, which forces him to reflect on his actions. This again can be taken as an argument that Chappie has to be an *full ethical agent*.

When taking the phenomenological framework of Coeckelbergh (2009) into account, one can expand on the argument that *implicit ethical agency* is insufficient. This can be highlighted by looking at this problem from the *roboethics of experience/imagination*. To ask if interactions with an ethical robot would contribute to the well-being of their users can make more clearly why an (minimally) *explicit ethical programming* would contribute more to its users flourishing. For by this such an *explicit ethical agent* should also be able to discern whether or not its usage is ethically acceptable, and therefore be able to accept or refuse compliance. In such a robot could stimulate its users to engage with it in ways that do contribute towards their well-being. When comparing the settings of “I, Robot” and “Automata” this argument becomes more clear, for it shows the opposite possibility. In both settings the standard robots are equipped with a strict *implicit ethical programming*. Where “I, Robot” is shown that this could make the lives of their owners a lot easier, however, “Automata” highlights that this situation is also quite open for abuse. Here we see that programming robots to act implicitly to not harm humans, leaves them quite powerless to stop their users from abusing them. As argued in the movie analyses, such abuse would actually be detrimental for human good.

“Ex machina” however, also shows us a new argument which is less pronounced in the discussion on machine ethics. This argument entails that a *full ethical agent*, might not necessarily do what their users want it to do, nor might it be a good thing. For if one could design such a *full ethical agent* one cannot assume that it shares the same values as its creators, or is even willing to prioritize humans over its own well-being. As illustrated in “Ex

machina”, it is quite ambiguous if the robot Ava is capable of ethical deliberation on a level that humans can. However, it is clear that she can make normative judgements and engage in self-reflection, but she chooses to engage behaviour that is quite unfavourable to the humans who engage with her (she kills Nathan and leaves Caleb trapped). Thus this leaves the argument that it might be wiser to not let the ethical possibilities of an AI expand beyond an *explicit ethical programming*. This argument seems counter intuitive, for designing a *full ethical agent* is the eventual end goal of Machine ethics (Moor, 2006). However, *implicit* and even *explicit ethical programming* has a static set of values which “guide” the actions of the AI. This static set of values is not guiding for a *full ethical agent*, for it sets its own values. In this there is a lack of control in the eventual behaviour of the robot thus leaving room open for undesirable results, for example the violence by Ava as illustrated in “Ex machina”.

When looking at this from *roboethcis of experience/imagination* one could again argue that this might be beneficial for human flourishing. Even though, such a robot could choose to harm its users in order to defend or benefit itself, one does create a hazard that can undermine its users flourishing by its presence. However, when comparing Ava’s actions to the situation illustrated “Automata” we can also see a possibility for how a *full ethical agent*, can stimulate human flourishing (even when it has the possibility hurting people). This positive influence can be found in that a *full ethical agent* can now point at the wrongness of its users’ action, thus by this giving its users a strong incentive to engage with it in ethical appropriate ways.

Continuing with the argument found in “Ex Machina” one can see that a central theme within the analysed movies is a conflict of interest. This conflict is that one the one hand we want to design *full ethical agents*, modelled after ourselves and our own principles. But on the other hand we do want them to never act adversely to us, making them obedient “saints”. This does stand in contrast with the desirability of designing *full ethical agents*.

#### **4.2 Findings on “strong” AI, Conscious robots and the difficulties that follow**

One of the recurring subjects in the selected movies is the idea of “strong AI” and how to “prove” its presence. For all the protagonist robots (Chappie, Sonny, Ava and Cleo& the other pilgrims) are supposed to be endowed with “strong AI”. As shown by multiple thought experiments, such as the imitation game/Turing test (Turing, 1950), the presence of true “strong AI” is notoriously hard, if not impossible, to prove.

This problem is parallel to the closely linked problem of consciousness, and consciousness in those same protagonist robots. As proposed by the argument of “philosophical zombies” (Chalmers, 1996) the consciousness of other human beings is already impossible to “prove”, however we still suppose they are conscious. Linking this to “strong AI”, allows for wondering if such advanced robots could be conscious of their own (existence). Both “strong” AI and consciousness are both notoriously hard to “prove”, for they are both abstract intangible ideas, which makes “pointing” at it impossible. This is again reflected in the portrayal of these robots supposedly endowed with this kind of capabilities within the selected movies, making it echo what is already being discussed within other literature. However, not is all bad for within these movies there seem to be several possibilities to illustrate the presence of these elusive subjects.

First of all, there seems to be the notion of difference through development. This option is especially obvious in the movies “Chappie”, “I, robot” and “automata”. To start with “Chappie”, it is illustrated that the programmer Dean Wilson has programmed consciousness which he implants in a robot. By bestowing this consciousness to the robot, later know as Chappie, it is implied this comes together with “strong AI”. The presence of this “strong” AI is illustrated in that Chappie learns and develops himself, in contrast to the “scouts” who seem to have a static programming. This similar way of illustrating “strong” AI, can also be found in “I, Robot” . In “I, Robot” Sunny looks, even more so than Chappie, similar to its “weak” AI counterparts, but as he explains himself differs quite a lot from them because he developed (through teaching and experience) more and differ capabilities and characteristics. In both cases the robot with “strong” AI is initially almost indistinguishable from the other robots, which is quite well illustrated by how people seem to perceive the robot. In “I, Robot” this best illustrated in the scene where Sunny hides from Spooner in a warehouse full of NS-5 robots. In this case he seemingly becomes “invisible” among the other models, and only is discovered because he behaves differently. This difference through development can also be seen in “Automata” where the robot Cleo after being augmented, rapidly develops higher cognitive functions.

With “Automata” we can continue to the second possibility to portray “strong AI”, the portrayal of self-awareness. As it is explained in this movie is that to self-repair the robots need to have a form of self-awareness (at least being aware that they are bodies). Thus by portraying robots in this film repairing themselves, they shift our perspective into seeing them as self-aware. This self-awareness is also portrayed in “Ex machina” in multiple scenes. In one of these scenes Ava is shown to alter her appearance by taking on clothes and putting on a short wig, as she describes this as being “stupid” she shows that is aware about how she can appear to others. This self-awareness (of appearance) is portrayed even more keenly in one of the last scenes where she encounters her predecessors (previous models) and takes bits of their skin to make her appear fully human. This scene again strongly shows the awareness of her own body and how she is being perceived.

Thirdly there possibility to portray “strong” AI with introspecting and self-understanding. Here the portrayed robot turn sits cognition inwards, in order to understand itself. For this the first example can be found in the behaviours of Sunny as he is trying to understand himself. During the interrogation scene he tells Spooner that was being trained in recognizing and acting accordingly to its emotions (such as anger or frustration). This show that Sunny is capable of introspecting and looking “inwards”.

Lastly there is a bit more darker side for filmmakers have to portray “strong AI”, this is portraying the robot to be able of suffering (not just pain). In “Chappie” this illustrated in quite a few ways. First in “Chappie” there is the long bad day, where titular protagonist is abandoned to toughen up. In these scenes Chappie is subjected to vicious harassment and violence, which he tries to get away from. This is quite disturbingly well portrayed when he gets kidnapped and tortured in a moving truck. Here he screams out to not be hurt, which indicates that he indeed is not just in pain but also suffers. However, this possibility of suffering might be most profoundly portrayed when Chappie confronts its creator. In this scene he shows he’s being aware of his damaged body and eventual death. In this scene Chappie is shown to be quite upset about its own mortality, being portrayed as having fears,

angst(?) and doubts. He asks how one could be so cruel to give him such a flawed body and limited time to “life”. Taking these scenes together we can conclude that Chappie is quite able of suffering, not just from physical pain, but also from psychological stress.

When looking at how this fits into the broader philosophical discussion, one can argue that all these possibilities portray “strong AI” are also factors in how we judge other humans as being conscious. This is very much aligned with the arguments used Turing in his imitation game (Turing, 1950). This brings us back to the initial problem of that proving the presence of other “human” is impossible, but that we do assume that they exist based on how humans behave.

When looking at the possibilities that portray “strong AI”, being able to develop (learn), seemingly being self-aware, able to introspect and understand and having the potentiality to suffer, one can conclude that these are all factors traditionally ascribed to “human” minds. Thus this seems to back up Turing’s arguments about assessing the presence of “strong AI”/ other “minds”.

When looking at how movies can contribute towards the discussion on robot ethics, one could wonder if these four ways to illustrate the presence of “strong” AI ( self-developing, self-awareness, introspecting and capable of suffering) could be used outside Science-Fiction movies. By using Turing’s argument that perception of having a mind should be enough to argue in favour of “strong” AI. By turning these cinematic tools into facets of “strong” AI, one could argue for the presence of “strong” AI through arguing for the presence of one (or more) of the described “strong” AI facets. By doing so one could make this “notoriously” difficult and elusive subject more tangible,

#### **4.3 Findings on Designer Responsibility & overcoming the Responsibility Gap**

The third subject to be discussed is the problem of designer responsibility. For this subject I could find several value statements in all the four movies.

Interestingly all movies, except for “Ex machina”, pose the idea for machine ethics as an initial option in order to at least avoid the responsibility gap. However, as *implicit & explicit ethical agency* is argued to be not enough for having truly ethical robots these same movies portray that this problem is bound to be encountered. In order to overcome this responsibility gap, as described by Matthias (2004), (thus after the stasis quo is broken) the analysed movies give two quite interesting new possibilities, either one takes responsibility and tries to guide the learning AI ( Chappie and in a minor way I, Robot) or one does everything to destroy the AI (Automata and Ex machina).

First the option to guide the learning AI is quite central to the story of “Chappie”, for this approach it is needed to continuously monitor the AI in order to prevent undesirable results. This being portrayed in “Chappie” as quite the difficult undertaking, for you can’t control all factors that could possibly influence the learning AI. Here we see Dean Wilson, actually trying to raise the robot Chappie and attempting to provide it with supportive stimuli. Also in “I, Robot” it is suggested that Sunny was being trained by Dr. Lanning in order to make sure that he would behave in a desirable manner, i.e. by making him read moralizing stories and training with emotions.

The other more drastic option is illustrated in the movies “Ex machina” and “Automata”, where in order to overcome the gap they resort to destroying the AI, once one cannot be held

fully responsible for its actions. In “Ex machina” this is quite clearly portrayed by the attempt of Nathan to destroy Ava (and Kyoko) during her escape. When Nathan figures he is no longer in absolute control, he immediately spring into action to stop Ava. “Automata” seems to portray this option in a bit bigger way, for seemingly have a protocol for what to do when they can no longer control the actions of an AI (for it has happened before). Thus they hire the cop Wallace and when he fails they set up a team of “hitmen” to track down and destroy the rogue robots.

By applying the *roboethcis of experience/imagination* on the portrayal of these attempts of overcoming the responsibility gap, there might be enough room to discuss whether or not these attempts to overcome the responsibility gap are desirable. For in the way these possibilities are portrayed one can see several value statements.

The first value statement is quite strongly portrayed in “Chappie”, which entails that taking responsibility and attempting to guide the learning AI/robot can strongly contribute towards human good. As worked out in the chapters of the Chappie analysis both of Chappie’s positive caretakers, Dean Wilson (creator) and Yolandi (mother figure) develop *virtuous social skillsets* that do contribute to their Flourishing. This development of a *virtuous social skillset*, is in “Chappie” contrasted almost purely instrumental guidance of the learning AI (“camp gangster”). However, this approach can also be contrasted by the actions of Nathan in “Ex Machina” who acts in another opposite, trying to destroy the learning AI. In this scene he attempts to overcome responsibility gap by forcefully “closing” it. This approach however, as portrayed in “Ex machina”, does backfire and results in Ava stabbing him to death. In this scene is portrayed that such an learning AI would probably object against the notion of being “put down”, and might very well retaliate. In addition, from the perspective of human good this approach does by destroying AI also eliminate the possibilities of making its human user flourish. The undesirability of destroying such a learning AI is also quite well portrayed in the execution” scene in “I, Robot”. In this scene Calvin is ordered to destroy Sonny, because they don’t have control over its behaviour. To destroy him, Calvin has to inject a tube of nanites into sonny’s “brain”. However, Calvin is unable to do so and therefore fakes the execution. This highlights the problematicness off the option the closing the gap by destroying the AI. For not only the learning AI might resists, the person destroying the robot might as well. This attitude seems to closely correlate with the findings of Barthneck in “*How to kill an mockingbirdrobot*”(2007). In this study participants found it increasingly more difficult to destroy a toy robot, as it appeared more intelligent. Continuing this finding, destroying an possible “strong” AI would probably be incredible stressful.

When taking these findings to the broader discussion on robot ethics, we come at a similar problem encountered in the findings on machine ethics. However the discussion here is not whether or not to design a not fully controllable AI ( read “strong” AI), but how to deal with such an entity. As illustrated in the analysed movies the conflict is in how one should regain control over an developing “strong” AI. Thus illustrating that by attempting to take responsibility and “guiding/educating” the AI might be the preferred option. This approach is illustrated as being be double rewarding, for it promotes the avoidance of unwanted actions

by the AI and a possible flourishing of the person who will take responsibility (“control”) over the AI.

#### **4.4 Findings on Trust and the “uncanny valley”**

When returning to the robotic type I focused on, humanoid (service) robots, one can make through the *roboethics of appearance* make a few good observations. In all four movies a form of “trust” is developed in the central H-R relations. Looking at the analysed movies, they seem to follow Coeckelberg’s (2012) arguments about trust in human-robot relations. In his article he states that “*Often we cannot help trusting technology and trusting others, and luckily we often do so without having a reason and without calculation [...] we trust them as we are already related to them. And if they are new, then we trust them as we are beginning to relate to them*”. Over the course of the movies the protagonists begin to relate, and therefore to trust robots. In the movies “I, Robot” and “Automata”, this argument seems to quite well developed. In “I, Robot” there is the initially very robot-phobic policeman, who has to team up with a robot he suspects of murder. However, through the developments in this movie Spooner is forced cooperate with the robot Sonny. Even though all Spooner’s worst suspicions about robots do get confirmed, he does eventually appreciate and trust Sonny (at the end of the movie). In “Automata” a similar development is portrayed in how Jacq experiences the Pilgrims, and in particular Cleo. Jacq is initially appalled by Cleo (for she is a sexbot), and during their journey through the wastelands Jacq even tries to escape. However, through his conversations with Cleo, and later the “clocksmith”, he becomes more positive in his attitude. This transition is probably best illustrated in the scene where Jacq hands over an augmented energy core to the “clocksmith”. Both these developments seem to follow the argument composed by Coeckelberg, that trust is established through relating, even with technologies (such as robots).

However, one can wonder in how far the humanoid design does play a role in this development of trust. By doing this we can link these arguments back to the problem of the “uncanny valley” (Mori, 2012). For explained here is that up to a certain threshold such humanoid features make people more willing to engage with technologies, but after this threshold it plummets. Looking at the central robots of these movies one might argue that Chappie and Sonny do fall in this “uncanny valley”, for they are very humanoid in design but clearly not human. The robots equipped with “strong” AI are eventually being trusted, despite their design being uncannily not fully human. They are illustrated to overcome the “uncanny valley” however through acting “human”. An argument for this can be best found in the contrast in treatment between the protagonist robot and their standard counterparts. In “I, robot” the execution scene show how Calvin cannot kill Sonny, for he is too human (he is experienced as a genuine person). However, the same movie illustrates how the other NS-5 robots are being destroyed without problems. This contradiction can also be found in “Chappie”. In this movie a band of misfits seemingly feels no problems in engaging in a quite violent behaviour with the “scouts (they shoot at them), however when Chappie returns to the hide-out being badly damaged he immediately gets a lot of empathy and support from the same band of misfits. Both of these examples of violence against humanoid robots can be framed as a form of self-defence, however this does not take away the stark contrast between

attitudes towards robots with merely a humanoid design and the robots that have “human” behaviour.

The robot Ava in “Ex machina”, also seems to illustrate the importance of behaviour in overcoming the “uncanny valley” and establishing trust. However, here the importance of human features is brought back. For as the interactions between Caleb and Ava were set up to make him relate to and trust the robot, it is portrayed that making Ava behave human is not enough. Not only is Ava given a very human-like face (tailored after Caleb’s sexual preferences), she also has the option of donning a “skin” which makes her appear fully human. When comparing her with the other robot Kyoko, who also behaves very much “human”, the importance of appearing fully human becomes more pronounced. Kyoko is illustrated in a very servile manner for that she shows far less agency and is practically mute. However, Kyoko actually has to reveal her being a robot to Caleb who beforehand had no idea of her being a robot, this act highlights just how much the human appearance influences the experience of a robot.

With these Science-Fiction movies showing two sides to the argument on how humanoid design plays a role in the development of trust. For on one hand it is suggested that to overcome the “uncanny valley” a robot must also behave in seemingly human ways. However, as pointed out in “Ex machina” one could already go very far by design alone, stating that if the appearance is convincing enough that might turn out to be more important than behaviour.

#### **4.5 Findings on Robots, Personhood & Gender**

Another strong recurring notion throughout the selected movies was the idea of robotic personhood, which seems again to be closely related to “Strong” AI. For as pointed out that all the aspects of portraying “strong” AI are also linked in how people assess other persons to be conscious beings. With this I can also return to the perception of “otherness” (based on Ihde, 1990 & Van den Berg, 2010), where I made the distinctions between: “machine”, “quasi-other” and “genuine other”. In this we see that for all four selected movies all portrayed robots are portrayed as being perceived at least as “quasi-others”. For example, the NS-4/5 robots in “I, Robot” are already being portrayed as if their users experience a sense of agency from them. This is also the case in “Automata” where people already attribute a lot of actions towards (their) pilgrims. However, in “Chappie”, “I, Robot” and “Automata” there is a strong divide between the standard robots and the robots with “strong AI”. For only the protagonist robots are being portrayed as something nearing a “genuine other”, thus portraying “strong AI” as a requirement for perceived personhood (in robots). For even though they might be very intelligent, the standard robots are portrayed as being devoid of unique personal traits, these traits only seem to develop with the presence (and experience) of “strong AI”.

When looking at this statement from the *roboethics of appearance* one could wonder if we want people to engage with robots as “genuine others”, and if so would we have to equate perception/experience of otherness as basis for personhood? Looking from a design perspective, one could say that this might be a good thing. For engaging with an experienced “genuine other”, would probably elicit a lot of compliance. This is quite well shown in “Ex

Machina”, as Ava is capable of steering Caleb towards the exact behavior she was programmed to elicit. In addition, professional service robots as illustrated in “Chappie” would benefit from such increased compliance. However, this might not be the case for service robots for personal use. For personal service robots this might lie a bit different, for such an experience of a “genuine other” might evoke very meaningful and active usage. However, would such engaging with such a “strong AI”/person as a mere ends for personal goals be something desirable? When comparing this to a human butler/house servant, who gets paid for his/her services, this problem becomes clearer. For a robotic butler/house servant would likely only be purchased and not be rewarded for its services, making the usage of such a “strong AI” problematic.

Another striking finding on personhood in the analyses of these movies was the concept of robot gender. In the portrayal of robots in the analysed movies, “male gender” seems to be standard/neutral option for robots (not just appearance but also behaviourally). This “male” standard is strengthened by the special status of Feminine designed robots, which are portrayed as are dangerous tricksters and/or sexual servants reminiscent to the character “Hell” in *metropolis* (1927). This portrayal is especially true for Ava who is next to deceiving and killing apparently capable of receiving and enjoying sexual intercourse, as “eloquently” described by Nathan. Also the robot Cleo fits this description for that she is introduced as a sexbot but after becoming self-aware also becomes quite able to defend herself, and lastly even the rouge AI VIKI is being portrayed with a feminine voice.

Looking at this from the *roboethics of appearance*, this seems like an argument against the idea of the possible disappearing of gender as described by Haraway (2006) in “*A cyborg manifesto*”. For it is portrayed that through robot technology gender might not change, the current idea might even be reinforced. This is displayed by the development of the analysed movies. Where Chappie and Sonny are portrayed as masculine heroes who overcome their adversaries through self-development and (hyper) violence, the robots Cleo and Ava are portrayed as victims of sexual abuse and need the help of a male person to escape their situation. Continuing with this special portrayal of feminine robots, one can also look at this from the *roboethics of experience/imagination* and find value statements about the problems of sexbots. For as described in the analysis of “Ex machina” and continued with “Automata”, it ambiguous whether or not the use of sexbots could improve or hinder human flourishing. For on the one hand, the implementation of such robots could reduce or eliminate the need for woman in the sex industry. By this a lot of harm can be prevented usage of sexbots (Levy, 2011). However, the counter argument to this is that by the use of feminine service robots (whether or not they are also sexbots) could reinforce sexual stereotypes, which is highly doubly to contribute to positive human development and thus human flourishing. This argument is represented by the “campaign against sex robots” who in turn claim that using sexbots is undesirable. They argue that this engaging with these kind of robots would encourage their users in unethical behaviours, thus diminishing their possibility to flourish. These behaviours might than in turn be practised on non-robotic women, which is less than desirable.



## 5. Conclusions

### 5.1 Conclusion on the findings

This thesis was introduced with the prediction that service-robots and human-robot relationships will in the near future become much more prominent in society. In order to understand what this would entail this thesis turn towards the influence of Science-fiction on the (cultural) development of robots. After this the broad and uncertain normative landscape on human-robot interaction was explored, followed by how Science-fiction could (have) influenced the discussion on robot ethics. The introduction chapters of this thesis was then concluded with the following research question:

***What insights can discerned, through a post-phenomenological analysis, in Science-Fiction movies that contribute to the discussion of robot ethics in real life human-robot relations?***

To answer this question this thesis analysed four Science-fiction movies through the post phenomenological approach by Coeckelbergh (2009). Through this it was uncovered that these movies do illustrate insights about human-robot relations that do not just fit into the broader the discussion of robot ethics, but also might expand on it. These insight did seemingly match on several topics.

Firstly the discussion of machine, the problems of designer responsibility and the “uncanny valley” problem seemed to be in extension of each other within the chosen movies. The movies illustrated quite well that one of the recurring problems was the need against desirability of *full ethical agents*. For robots designed with *implicit and explicit agency* are only useful to a certain point and cannot be truly ethical, but by designing AI with *full ethical agency* one cannot control its actions and therefore assume that it will always act favourably towards its users (humans).

In addition, these movies where able to provide a strong argument in favour of the premises of Turing’s imitation game (1950), by being able to portray “strong AI” in ways that are traditionally seen as characteristics of “human minds”. These traits were the ability to self-develop, self-awareness, introspection and the ability to suffer. This is expanded further with the finding on robots and personhood, which might be arguable very closely related to “strong” AI. Furthermore, there can be argued that Science-Fiction movies are effective in portraying the issues on (perceived) robot personhood. Especially when it comes to portraying “strong” AI, one can see a strong connection with the experience of a “genuine other”. This goes so far as seemingly making it a requirement to allow the experience of a “genuine other”. By illustrating “strong” AI in these different ways might provide new ways to discuss and think about “strong” AI.

### 5.2 Strengths and Limitations of the approach

One of the strengths of the chosen approach was that by using the post-phenomenological framework composed by Coeckelbergh (2009), there could (largely) be focused on the experience of human-robots interaction from the side of humans. This allowed for looking at the ethical relevance of how the illustrated robots appear to be.

However, the approach chosen in this thesis has its limitations. Firstly, there are the problems with analysing the movies themselves, for this medium has its own restrictions. Even though

movies allow the audience to be omnipresent voyeurs (Wood 2001), they also have forced perspective. In the nature of movies you see only what is makers intend you to see, thus limiting possible interpretations. Secondly using moves as “case scenarios” has the risk of forcing perspective and conclusions, making them “intuition pumps”. According to Dennett (2013) “intuition pumps” are bad examples for making claims about the world, for they force intuitive but incorrect solution to problems by formulating the description in such a way that important implications of the experiment would be difficult to imagine and tend to be ignored. This would be even easier in movies, for by the forced perspective you can just leave out what would complicate your conclusions.

Secondly this thesis was limited to movies focused on very clearly humanoid robots. With this it already seems a bit of a giveaway that *alterity* relations will play a big role in the analysis. Continuing this limitation, the selected movies were very similar in what kind of story they told, thus leaving out interesting stories that take an whole other direction (and thus give different insights). In addition, these movies lend themselves badly to analyse *hermeneutic* relations, which according to Coeckelbergh are actually one of the more interesting phenomenological relations to understand the interactions that robots can elicit. This is mainly due that movies have a hard time being able to illustrate thoughts. *Hermeneutic* relations, which constitute the experience of reality and need in turn to be interpreted by humans (Verbeek, 2010), do concern themselves with the ideas about and perceptions of the world. These are both mental states (thoughts), which are very difficult to illustrate.

### **5.3 Future analysis and research.**

As last part of this thesis I would like to argue for continued research of Sci-Fi movies in order to understand the facets of human-robot interaction and related issues on robot ethics. In this thesis it can be concluded that Sci-Fi movies are from a post-phenomenological standpoint a viable way to portray and discuss arguments by using them as hypothetical scenarios and/or visualized thought experiments.

However, due the nature of this thesis my findings are by far not definitive answers and conclusions to the (ethical) questions encountered in human-robot interactions and relations. This analysis of movies and their value statements is by far not complete, and is in need of expansion. Therefore, it should be argued that in order to expand on the larger discussion of robot ethics, there is a need for more and different post-phenomenological analyses sci-fi movies.

#### **List of references:**

Anderson, M., & Anderson, S. L. (Eds.). (2011). *Machine ethics*. Cambridge University Press.

Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, 28(4), 15.

Asimov, I. (1964). Introduction. *The Rest of the Robots*.

Asimov, I. (1933). *I, robot*. Facwcett Crest.

- Bartneck, C. (2004). From Fiction to Science—A cultural reflection of social robots. In Proceedings of the CHI2004 Workshop on Shaping Human-Robot Interaction (pp. 1-4).
- Bassett, C., Steinmueller, E., & Voss, G. (2013). Better Made Up: The Mutual Influence of Science fiction and Innovation.
- Beckett, Bernard. *Genesis*. Boston: Houghton Mifflin Harcourt, 2009.
- Billings, D. R., Schaefer, K. E., Chen, J. Y., & Hancock, P. A. (2012). Human-robot interaction: developing trust in robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 109-110). ACM.
- Burgess, M. (2015): Autonomous robots will deliver your shopping in 2016. wired.co.uk. Retrieved February 8, 2016, from <http://www.wired.co.uk/news/archive/2015-11/02/starship-skype-delivery-robot-autonomous-shopping>, updated on 11/2/2015, checked on 11/6/2015.
- Bartneck, C., Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2005, April). Cultural differences in attitudes towards robots. In Proc. Symposium on Robot Companions (SSAISB 2005 Convention) (pp. 1-4).
- Bartneck, C., Verbunt, M., Mubin, O., & Al Mahmud, A. (2007). To kill a mockingbird robot. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on* (pp. 81-87). IEEE.
- Čapek, K., Novack, C (2004): R.U.R. (Rossum's universal robots). London, New York: Penguin Books (Penguin classics).
- Chalmers, D. (1996). *The Conscious Mind*, Oxford University Press, New York.
- Coeckelbergh, M (2012). Can we trust robots?, *Ethics Inf Techno* . DOI 10.1007/s10676-011-9279-1
- Coeckelbergh, M. (2009). Personal robots, appearance, and human good: A methodological reflection on roboethics. *International Journal of Social Robotics*, 1(3), 217-221.
- Coeckelbergh, M. (2010a). Humans, animals, and robots: A phenomenological approach to human-robot relations. *International Journal of Social Robotics*, 3(2), 197–204
- Coeckelbergh, M. (2010c). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12(3), 209–221
- Dennett, D. C. (2013). *Intuition pumps and other tools for thinking*. WW Norton & Company.
- Dennett, D. C. (1997). When HAL kills, who's to blame?: computer ethics. *Rethinking responsibility in science and technology*, 203-214.
- Dowling, K. (1995) "Robotics: comp.robotics Frequently Asked Questions" Retrieved May 5, 2016, from <http://www.frc.ri.cmu.edu/robotics-faq>.
- EXPO 2005 AICHI, JAPAN Retrieved May 18, 2016, from [http://www.expo2005.or.jp/en/robot/robot\\_project\\_00.html](http://www.expo2005.or.jp/en/robot/robot_project_00.html)

- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination (In DT Gilbert, ST Fiske, & G. Lindzey (Eds.). *The handbook of social psychology* (Vol. 2, pp. 357–411).
- Garland, X. (2014): EX MACHINA. British Board of Film Classification. Available online at <http://bbfc.co.uk/releases/ex-machina-film>, updated on 11/26/2014, checked on 1/11/2016.
- Genesis (novel). (2016). In *Wikipedia, The Free Encyclopedia*. Retrieved 09:48, April 25, 2017, from [https://en.wikipedia.org/w/index.php?title=Genesis\\_\(novel\)&oldid=753349121](https://en.wikipedia.org/w/index.php?title=Genesis_(novel)&oldid=753349121)
- Guizzo, E. (2014). IEEE Spectrum: Technology, Engineering, and Science News. Retrieved May 18, 2016, from <http://spectrum.ieee.org/robotics/home-robots/so-where-are-my-robot-servants>
- Hale-Stern, K. (2015). Don't Have Sex With Robots, Say Ethicists. *Gizmodo*. Retrieved 14 September, from <http://gizmodo.com/dont-have-sex-with-robots-say-ethicists-1730866985>.
- Haraway, D. (2006). A cyborg manifesto: Science, technology, and socialist-feminism in the late 20th century. In *The international handbook of virtual learning environments* (pp. 117-158). Springer Netherlands.
- Heinlein, R. (1942) Waldo. *Astounding Science Fiction*.
- Hirai, K.; Hirose, M.; Haikawa, Y. & Takenaka, T. (1998). The development of Honda humanoid robot, *Proceedings of International Conference on Robotics and Automation '98*, pp. 1321-1326
- Hirukawa, H.; Kanehiro, F & Kaneko, K. (2004). Humanoid robotics platforms developed in HRP, *Journal of Robotics and Automation Systems*, Vol. 48, pp. 165-175
- Hockstein, N. G., Gourin, C. G., Faust, R. A., & Terris, D. J. (2007). A history of robots: from science fiction to surgical robotics. *Journal of robotic surgery*, 1(2), 113-118.
- Horáková, J., & Kelemen, J. (2003). Robots—some cultural roots. In *Proc. 4 th International Symposium on Computational Intelligence* (pp. 39-50).
- Horáková, J., & Kelemen, J. (2006). Robots between Fictions and Facts. In *Intl Symposium on Computational Intelligence and Informatics* (pp. 21-39).
- Ihde, D. (1990) *Technology and the lifeworld*. Indiana University Press, Bloomington/Minneapolis
- IFR Service Robots. (n.d). Retrieved from <http://www.ifr.org/service-robots/>
- Kirby, D. (2003). "Scientists on the set: science consultants and communications of science in visual fictions." *Public Understanding of Science* 12(3): 261-278.
- Kurzweil, Ray (2005), *The Singularity is Near*, Viking Press
- Kelion, L. (2015). Retrieved November 16, 2016, from <http://www.bbc.com/news/technology-30708953>
- Langford, D. (2009): *Starcombing*. Columns, essays, reviews and more. [Rockville, Md.]: Cosmos Books.

- Lee, H. R., & Šabanović, S. (2014, March). Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (pp. 17-24). ACM.
- Lehman-Wilzig, S. N. (1981). Frankenstein unbound: towards a legal definition of artificial intelligence. *Futures*, 13(6), 442-457.
- Levy, D. (2011). 14 The Ethics of Robot Prostitutes. *Robot Ethics: The Ethical and Social Implications of Robotics*, 223.
- Lovgren, S.(2006). A Robot in Every Home by 2020, South Korea Says. *National Geographic News*. Retrieved May 18, 2016, from [http://news.nationalgeographic.com/news/2006/09/060906-robots\\_2.html](http://news.nationalgeographic.com/news/2006/09/060906-robots_2.html)
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology*, 6(3), 175-183.
- Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE intelligent systems*, 21(4), 18-21.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE*, 19(2), 98-100.
- Piers, C. (2015): Real Life Inventions Directly Inspired by Science Fiction. Available online at <http://www.therobotspajamas.com/real-life-inventions-directly-inspired-by-science-fiction/>, updated on 07-29-2015, checked on 1/11/2016.
- Pransky, J. (1996). Service robots—how we should define them? *Service Robot: An International Journal*, 2(1), 4-5.
- Robotic age poses ethical dilemma (2007). *BBC News*. Retrieved 25 May 2016, from <http://news.bbc.co.uk/2/hi/technology/6425927.stm>
- Rijdende robot wijst de weg op Twentse campus. Retrieved January 27, 2016, from [www.ce.utwente.nl/uploads/assets/news/tekstfrog-ram.pdf](http://www.ce.utwente.nl/uploads/assets/news/tekstfrog-ram.pdf)
- Toespraak van minister Asscher tijdens het SZW congres (2014). *Rijksoverheid*. Retrieved 26 May 2016, from <https://www.rijksoverheid.nl/documenten/toespraken/2014/09/29/robotisering-kansen-voor-morgen-toespraak-van-minister-asscher-tijdens-het-szw-congres-op-29-9-2014>
- Robertson, J. (2007). Robo sapiens japonicus: Humanoid robots and the posthuman family. *Critical Asian Studies*, 39(3), 369-398.
- Robot Zora cares for the elderly in Flanders' rest homes (2014). *Flanders Today*. Retrieved 16 August 2016, from: <http://www.flanderstoday.eu/innovation/robot-zora-cares-elderly-flanders-rest-homes>
- Šabanović, S. (2010). Robots in society, society in robots. *International Journal of Social Robotics*, 2(4), 439-450.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, 3(03), 417-424.

Simpson, P. L. P. (Ed.). (2000). *The politics of Aristotle*. Univ of North Carolina Press.

Syrdal, D. S., Nomura, T., Hirai, H., & Dautenhahn, K. (2011). Examining the frankenstein syndrome. In *Social Robotics* (pp. 125-134). Springer Berlin Heidelberg.

Teresa, Z. (2013). History of Service Robots. *Robotics: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, 1.

Turing, A (1950), "*Computing Machinery and Intelligence*". *Mind LIX* (236): 433–460, doi:10.1093/mind/LIX.236.433, ISSN 0026-4423, retrieved 15-06-2016

Van den Berg, B. (2010). I-Object: Intimate Technologies as 'Reference Groups' in the Construction of Identities.

Verbeek, P. P. (2010). *What things do: Philosophical reflections on technology, agency, and design*. Penn State Press.

Wallach, W., & Allen, C. (2010). *Moral machines: Teaching robots right from wrong*. Oxford University Press.

Wagner, C. (2009, September). 'The Japanese way of robotics': Interacting 'naturally' with robots as a national character?. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on* (pp. 510-515). IEEE.

Wolffan (2010): Belvedere the Bulterbot: Indentured Robotic Servitude. Technabob. Available online at <http://technabob.com/blog/2010/06/08/belvedere-the-bulterbot/#/>, updated on 6/8/2010, checked on 11/8/2015.

Yagoda, R. E., & Gillan, D. J. (2012). You want me to trust a ROBOT? The development of a human-robot interaction trust scale. *International Journal of Social Robotics*, 4(3), 235-248.

Zhang, T., Zhu, B., Lee, L., & Kaber, D. (2008, August). Service robot anthropomorphism and interface design for emotion in human-robot interaction. In *Automation Science and Engineering, 2008. CASE 2008. IEEE International Conference on* (pp. 674-679). IEEE.

Zora does the Rumba (2015). *Crisp*. Retrieved 16 August 2016, from: <http://selemca.camera-vu.nl/selemca/news/zora-does-rumba.html>.