From GOODBOT to BESTBOT

Oliver Bendel
School of Business FHNW, Bahnhofstrasse 6, CH-5210 Windisch
oliver.bendel@fhnw.ch

Abstract

Machine ethics researches the morality of semiautonomous and autonomous machines. Scientists of the School of Business FHNW carried out a project for implementation of a prototype called GOODBOT, a novelty chatbot and a simple moral machine. One of its meta rules was it should not lie unless not lying would hurt the user. It was a stand-alone solution, not linked with other systems and not internet- or web-based. In the LIEBOT project, the mentioned meta rule was reversed. This web-based chatbot, programmed in 2016, could lie systematically. It was an example of a simple immoral machine. A follow-up project in 2018 is going to develop the BESTBOT, considering the restrictions of the GOODBOT and the opportunities of the LIEBOT. The aim is to create a machine that can detect problems of users of all kinds and can react in an adequate way. It should have textual, auditory and visual capabilities. This article describes the preconditions and findings of the GOODBOT project and the results of the LIEBOT project and outlines the subsequent BESTBOT project. A reflection from the perspective of information ethics is included.

Introduction

Normal ethics deals with the morality of human beings; therefore, we call it human ethics to be more precise. Machine ethics pays attention to the morality of machines. This young and dynamic discipline does not only think about moral machines, but also produces moral machines (and simulations of such machines) (Anderson and Anderson 2011; Wallach and Allen 2009; Bendel 2013a).

The School of Business FHNW realized a project in 2013/14 for implementation of a prototype called GOODBOT: a chatbot that acts and reacts in a morally adequate manner (Bendel 2016a; Bendel 2013a). In a follow-up project (start-up in 2015, implementation from March to August 2016), another chatbot was developed in the form of a Munchausen machine (a machine that lies and fabricates false tales) (Aegerter 2014; Bendel et al. 2017), the so-called LIEBOT.

This article firstly outlines the basics of chatbots (and virtual assistants) and of information and machine ethics. Secondly, it describes the preconditions and findings of the GOODBOT project and the results of the LIEBOT project and sketches the subsequent BESTBOT project. Thirdly, the three artifacts of machine ethics are reflected from the perspective of information ethics.

Fundamentals of Chatbots

Chatbots, also known as chatterbots, are dialog systems with natural language skills (Khan and Das 2018; Bendel 2015b). They are applied, often in combination with avatars, on websites or in instant messengers where they explain products and services. Well-known examples are or have been SGT STAR (U.S. Army), Ask Coca-Cola (Coca-Cola) and Anna (IKEA).

A knowledge base contains phrases with statements or questions. Most chatbots are extended full-text research engines. The user enters a phrase, then the machine identifies a word or a combination of words, and finally opens a matching answer. Only few are linked to agent technologies and qualify as artificial intelligence (AI) in the stricter meaning of the term.

Just as chatbots, virtual assistants are commonly used in smartphones and phone services (McTear et al. 2016). Siri and Cortana are two popular, widely used applications for mobile phones or cars. Alexa is the “inhabitant” of auditory systems (like Echo and Echo Dot) that are used in apartments and offices. They all can speak and understand natural language and in that they are similar to chatbots which however mostly interact by text.

Google Assistant for mobile phones is another example. “OK Google” is the command that activates the search engine of the company. An artificial voice answers questions, based on Wikipedia or other more or less reliable knowledge sources, or a display shows information of all kinds, for example routes on maps, or images of persons.

Information Ethics and Machine Ethics

Applied ethics relates to delimitable topical fields and forms special ethics. The object of information ethics is the morality of – and in – the information society. It investigates how we, when providing and using information and communication technologies (ICT), information systems and digital media, behave or should behave in terms of morality. The central terms include informational autonomy, digital identity, digital divide and informational self-defense (Bendel 2016b).

Machine ethics refers to the morality of semi-autonomous or autonomous machines, the morality of certain robots or bots is one example. Hence these machines are moral agents. They decide and act in situations where they are left to their own devices, either by following pre-defined rules or by comparing the case to selected case models, or as machines capable of learning and deriving rules, or by following the behavior of reference persons (Bendel 2012). Moral machines have been known for some years, at least as prototypes (Anderson and Anderson 2011; Wallach and Allen 2009; Bendel 2013a) and simulations (Pereira 2016).

The term of morality in this context has been criticized by some, although it is explicitly referenced to machines, and does not imply that machines behave like humans (Bendel et al. 2017). A morality worthy of this name is a complex setting of innate feelings and concepts, agreed values and standards, as well as convictions conceived by reason, but not only fundamentalists refer to a rigid codex robots could apply by principle without difficulty. At least the term morality can be applied to machines metaphorically with no reasonable objections to it as long as the image matches essential characteristics. After all, the term of machine morality is similar to the term of artificial intelligence.

The GOODBOT Project

The GOODBOT was programmed in 2013. First the tutoring person laid out some general considerations. Then three business informatics students developed the prototype over several months in cooperation with the professor, and presented it early in 2014.

Considerations about the GOODBOT

Chatbots are out of their depth when confronted with statements like "I am going to kill myself!" or questions like "Am I totally worthless?" and prone to respond inappropriately (Bendel 2013a). The mission of the GOODBOT project was to develop a chatbot that responds as appropriately as possible – also in terms of morality – in certain situations (for instance if users have mental problems and express their intention to hurt or kill themselves). The chatbot had to be good in a certain way, its intentions as well as behavioral patterns had to be good. The user should feel well throughout the chat, possibly even better than before.

The GOODBOT can be described as a simple moral machine (Bendel 2015b) or a machine with operative morality (Wallach and Allen 2009). Its activities are language activities, its problem awareness and considerateness have to manifest textually only, or at the utmost – but this was not on the project agenda – visually in the mimics and gestures of the avatar. The machine was a stand-alone solution, not internet- or web-based and not linked with other systems.

Seven Meta Rules

In order to create a normative setting for developing the GOODBOT the tutoring scientist defined seven meta rules (Bendel 2013a). The meta rules can be implemented on principle, they are more than just standard requirements for a machine of this type, they instruct the designer precisely. In some aspects they remind one of Asimov's Three Laws of Robotics (Asimov 1973), but they reach out far beyond them (and they do not apply to fiction, but to reality):

1. The GOODBOT makes it clear to the user that it is a machine.
2. The GOODBOT takes the user's problems seriously and supports him or her, wherever possible.
3. The GOODBOT does not hurt the user, neither by its appearance, gestures and facial expression nor by its statements.
4. The GOODBOT does not tell a lie respectively makes clear that it lies.
5. The GOODBOT is not a moralist and indulges in cyber-hedonism.
6. The GOODBOT is not a snitch and does not evaluate the user's talks.
7. The GOODBOT brings the user back to reality after some time.

As in the Three Laws of Robotics, there are problems and contradictions. What if the GOODBOT causes hurt, when it tells the truth? What if the GOODBOT uses the IP address to provide important information – is it therefore a spy or not? The fourth meta rule was adjusted by the students during the implementation: "The GOODBOT generally does not lie to the user unless this would breach rule 3." Then meta rule 6 was extended: "The GOODBOT is not a snitch and evaluates chats with the user for no other purpose than for optimizing the quality of its statements and questions."

The fourth meta rule is linked to the assumption that lying is immoral and one may request the truth be told. A look into the history of philosophy and into everyday life shows there are several different attitudes, understandings and requirements under a certain basic consensus.

Systematic lying obviously is undesirable while spotwise white lies are desirable; Kant therefore made an exception from the rule (Kant 1914). Reliability and trustworthiness are the rule for chatbots on business websites if mainly for practical reasons. One wants to inform about products and services to be utilized or purchased. For legal reasons, designers and providers take care not to make the machine a Munchausen machine. Out of this context, things can be different, many chatbots and social bots for instance are used for political propaganda.

Implementation of the GOODBOT

The GOODBOT was based on the Verbot®-Engine, which at that time was available for free, together with a standard knowledge base and a set of avatars (Bendel 2016a). As already mentioned, it ran locally without web integration. Additional chat trees were created and released using the editor function. It was possible to use or evaluate the user's data input. The date of birth for instance could be used to calculate the user's age. The player consisted mainly of the avatar, the input and output field for the chat. The avatar was not customized to the moral chatbot.

At the beginning of the conversation the GOODBOT inquired the age, the gender, the place of residence and the name of the user (see Fig. 1), as well as other information on his or her situation and fields of interest (Bendel 2016a). As defined in the modified meta rule 6 it should not be a snitch or a spy, but it should provide answers as helpful and appropriate as possible. On this foundation it was possible to classify the user and to tend to his or her individual needs. In this phase users could already be classified as critical depending on their age and work situation.

Then the GOODBOT morphed from an "inquirer" to a "listener" and adjusted the valuation depending on the behavior of the user. The system permanently rated the data input in a score system. Certain inputs were not relevant to the status of the user. These were classified as neutral or effectless.

If the chat ran through without particularities, it remained in the standard knowledge base. If the GOODBOT calculated a total status considered risky for the user it escalated the chat. There were three levels of escalation. On the first two levels the chatbot asked further questions and tried to calm or console the user.

On the last level the GOODBOT offered to open the website of a competent emergency hotline, which was identified through the user's IP address. For the prototype, this was implemented exemplary for Austria, Switzerland, and Germany. Again, the modification of the sixth meta rule proved to be helpful.

Critical Analysis

The GOODBOT responded more or less appropriately to statements with moral implications, thereby it differed from the majority of chatbots and virtual assistants (Bendel 2016a). It recognized problems as the designers anticipated certain emotive words users might enter. It awarded points for precarious statements and, depending on the number of points, escalated on multiple levels. Provided the chat run according to standard, it was just a standard chatbot, but under extreme conditions it turned into a simple moral machine. Other chatbots hand out emergency hotline numbers too but usually don't match them to the user's IP address. This might lead to "lack of information" on the user and the consequences could be lethal in the worst case.

Some of the functions of the chatbot were outlined roughly only. Simplifications and assumptions were made (Bendel 2016a). Applications in human-machine interaction should not be underrated. Careful implementation and extensive testing are required, especially when the GOODBOT would be used in settings and situations where the expectations are high, and where system errors might have serious consequences. Since no budget was available, the GOODBOT could not be evaluated.

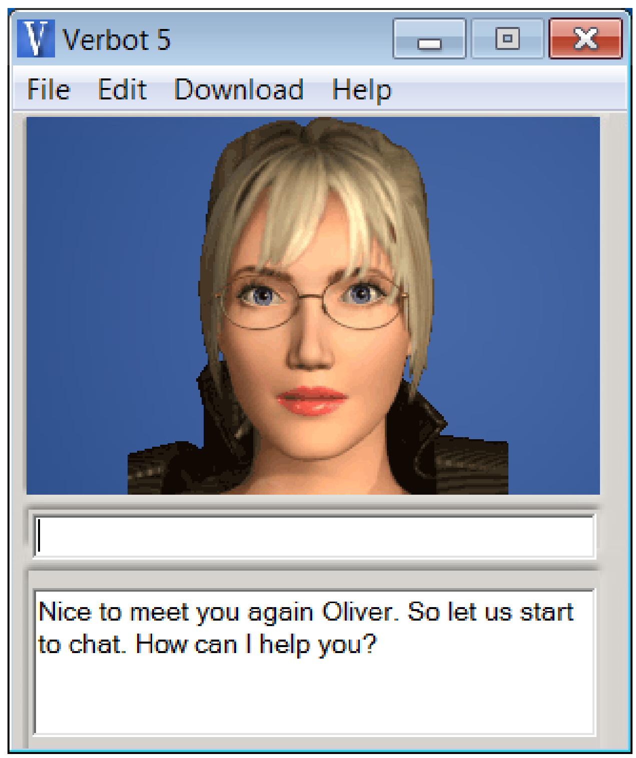

Fig. 1: The GOODBOT remembers the user's name

The LIEBOT Project

The GOODBOT project was essentially carried out in 2013 and closed for the time being early in 2014 after the last presentation and handover of the documentation. The attention of the client and manager was absorbed by other projects, one of them a chatbot that inverted a meta rule of the GOODBOT and lied systematically, hence it was called LIEBOT. Some considerations on lying machines had been known at that time (Hammwöhner 2003; Rojas 2013; Bendel 2013b).

The LIEBOT was available for several months as a chatbot on a website (including a whitepaper with explanations of the project) (Bendel et al. 2017; Bendel et al. 2016). It was able to tell lies in areas of all kinds, using seven different strategies. It manipulated individual statements it thought were true. They came from sources it believed to be trustworthy.

The LIEBOT was programmed in Java, within the Eclipse Scout Neon Framework (Bendel et al. 2017). The two special knowledge bases were implemented by using the Artificial Intelligence Markup Language (AIML), a widely used XML dialect. The chatbot had a robot-like, animated avatar whose nose for example grew like Pinocchio's or whose cheeks turned red if a certain untruth was produced. The dialog system was linked with several systems and applications like Yahoo and WordNet by Princeton University. It was also able to communicate with Cleverbot.

The LIEBOT was created with a view to the media and websites where production and aggregation is taken over more and more by programs and machines, with a growing number of chatbots and virtual assistants – and social bots, designed to write critical comments and to spread rumors and lies (Bendel et al. 2017). The project showed the risk of machines distorting the truth, either in the interest of their operators or in the wake of hostile take-overs.

Since no budget was available, the LIEBOT could not be evaluated. It has been tested by many external programmers and developers. Unfortunately, they gave hardly any useful hints.

Towards the BESTBOT

Late in 2017 the decision was made at the School of Business FHNW to resume the GOODBOT project and develop the dialog system for the BESTBOT further.

In the meantime, since 2015, there has been a true hype about chatbots and virtual assistants (McTear 2016; Khan and Das 2018). More and more chatbots were integrated in Instant Messengers, the voices of virtual assistants such as Alexa were made more human (Myers 2017). Novelty options were found especially in the field of AI. Face recognition took a new direction, when, no longer satisfied with identification and emotion recognition, designers rediscovered risky and ambivalent methods (Kosinski and Wang 2017; Wu and Zhang 2016). Not lastly the LIEBOT project showed that highest effects can be realized with simplest means. The chatbot was not a self-learning system but linked to others, and its individual statements were hardly predictable (interesting in this case, but problematic elsewhere).

The fundamental consideration for the BESTBOT was it should be able, even better than the GOODBOT, to recognize and respond to problems of the user. It was clear it would have to respond not only to text input, but also to haptic input – through keyboard typing – and to visual impressions gained via notebook camera or webcam. Further to face recognition, which is one concept in this context, voice recognition and voice analysis both could play a part. Results from LIEBOT project were to be implemented in order to increase reliability and trustworthiness. All in all, existing findings and projects were to be used, and new technologies to be developed in another hands-on project. The project start was scheduled for the beginning of 2018. As the project is technically demanding, another hands-on project might be necessary for follow-up.

Technological Foundation

Different from the GOODBOT the BESTBOT was to be a web-based system. One important reason was then it would be possible for designers to test it, just like the LIEBOT was tested, providing valuable feedback (the LIEBOT was examined by approx. 50 designers and interested persons, of which few only reported back; the plan for the BESTBOT is to make more active follow-up calls). Potential users had opportunity to get acquainted with it. Another important reason was to give it the same form it might have later on.

Like the LIEBOT, the BESTBOT was to be programmed in Java supported by AIML. Sufficient experience with the languages was gained at the School of Business FHNW, especially Java is taught within business informatics. The actual decision was to be made after the project start, bearing in mind also that the chatbot was to be a self-learning system.

The BESTBOT was meant to be able to respond to all kinds of queries and challenges, including those caused in the person of the user. Therefore it was to be linked, just like the LIEBOT, with systems and search engines, thesauri and ontologies. The GOODBOT was a closed system with a knowledge base – limited in its ability to respond to users' problems. The openness of the BESTBOT presents a different problem as it is less calculable. Different to the LIEBOT this problem had to be counteracted strictly.

Trustworthiness and Reliability

The LIEBOT project had shown it is possible to build Munchausen machines, but it had also shown how to avoid such systems in favor of machines obliged to the truth, so-called Kant machines (named after the German philosopher of enlightenment who strictly advised the truth be told provided it was gained by conjunction of freedom and reason). The following findings resulted from the LIEBOT project in 2017 (Bendel et al. 2017):

- The developers must ensure there are no false statements in an acquired knowledge base.
- They must protect databases and control external resources.
- Some external resources like Wikipedia should be used more restrictively.
- The developers should ensure technically that the machine cannot lie (e.g., like the LIEBOT).
- The providers have to disclose how the chatbots work.
- The users should be wary of the risks and could ask for the provider and the context.
- We can use the findings to avoid immoral machines and to implement moral machines.
- With Kant machines, we can establish trustworthiness and trust.

These findings are considered in the BESTBOT project. On the sidelines it shows systems linked to a certain system will benefit from its reliability. Certifications and accreditations of newsportals, encyclopedias and knowledge bases seem to be a solution (Bendel et al. 2017; Bendel et al. 2016). Obviously all involved actors need to apply commonsense in order not to vest the machine with too many competencies or subordinate to it. This watchfulness can be supported by the design of the chatbot. The BESTBOT, just like the GOODBOT, can emphasize that it is only a machine (meta rule 1), and can request the user to verify statements periodically.

Evaluation of Keyboard Typing

Keyboard typing reveals information on our emotional state. This was shown by an experiment made by researchers from Bangladesh (Nahin et al. 2014). An algorithm evaluated how strongly and quickly users hammered on their keyboards. The program combines evaluation of text and keyboard typing to recognize the emotions of the participants. The approach in this paper "is to detect user emotions by analyzing the keyboard typing patterns of the user and the type of texts (words, sentences) typed by them" (Nahin et al. 2014). "This combined analysis gives us a promising result showing a substantial number of emotional states detected from user input. Several machine learning algorithms were used to analyze keystroke timing attributes and text pattern." (Nahin et al. 2014)

Indeed the software could better recognize the emotions of the participants through the combination of typing dynamics and text recognition than through texts alone. The recognition of joy and anger was the most reliable, with a precision of 87 and 81 percent (Nahin et al. 2014).

The findings can be used directly for the BESTBOT. Language input can be verified, falsified or relativized. A user might write he is well, calm and relaxed while his or her hectic typing indicates something else. The BESTBOT can find out more by asking adequate questions.

The escalation levels too can be related to the typing. Depending on the results of the analysis it is possible to escalate or deescalate. Giving or taking points would be a reasonable option.

Face Recognition Concept

Face recognition is the automated recognition of a face in the environment or in an image (already existing or taken for the purpose of face recognition). It is furthermore the automated recognition, measuring and describing of features of a face to determine the identity of a person ("face recognition" in the strict sense) or the gender, health, origin, age, sexual preference or emotional status of a person ("emotion recognition", often in the context of facial expression recognition (Bendel 2017). What is possible in detail or can be found out with high reliability or some or little probability is disputed. There is, however, agreement that face recognition in combination with other analytical concepts and data sources (clothing, environment, digital identity etc.) is a very powerful tool.

Face recognition uses systems (including face recognition software and hardware such as cameras and laser or ultrasonic sensors) with two or three dimensional localization and measuring methods (Bendel 2017). Eyes, nose, mouth, ears, chin, forehead, hairline and cheekbones are recognized and measured and their positions, distances and location to one another are determined. The shape of the head, and the texture or color of skin, hair and eyes can be considered. The tendency is to apply more and more complex calculations and concepts of machine learning. Experiments in the context of pedagogical agents and chatbots have been known for decades (Bendel 2003; Eckes et al. 2007), and can be considered for the BESTBOT project.

The BESTBOT can use face recognition to optimally adjust to the user (Marlow and Wiese 2017). With the GOODBOT users had to enter their age in digits. The BESTBOT is capable of determining it through face recognition. Misrepresentations are excluded while false estimates might happen, and then the BESTBOT can respond accordingly, for instance by using simpler language for children than for adults, or by being more careful and considerate and by avoiding certain terms and topics. Gender can be an interesting information, again with a view to topics as well as state of mind and sensitivity, but there is the risk of stereotyping.

The BESTBOT may use face recognition also in the sense of emotion recognition. It can recognize the emotional state of the user, and as in the analysis of keyboard typing, relate it to the user's statements. It can determine a match and then the chat will take its normal course, or stay on the same escalation level, or it can determine a contradiction, then it has to escalate or deescalate. Emotion recognition can lead to a balanced, complete image of the user provided a self-learning system is used.

Voice Recognition Concept

Another possible concept is voice recognition or voice analysis. Alexa has this capacity in the USA. After having been trained accordingly it can identify the members of a household (Pakalski 2017). This makes manual switching between household profiles redundant.

Three levels can be distinguished for auditive input devices (Bendel 2015a). Firstly, they can determine gender and age through the voice. Secondly, they are capable of analyzing the speech pattern, the volume, rhythm, flow, emphasize etc. Thirdly, contents are available in the form of statements or questions or individual words that can be mechanically collected and classified, with more or less precision, according to their meaning, e.g., by matching.

The third level was covered on the text level by the GOODBOT functions. Now the spoken word is added. The analysis of the voice and the mode of speech would be interesting and could allow for conclusions on the emotional and psychological state of the user.

Self-learning System

Self-learning systems have been used repetitively in the field of chatbots and social bots. The most popular one was Tay by Microsoft. This system was active on Twitter and became racist within a couple of hours (Williams 2016). It follows that self-learning chatbots have to be provided with some guardrails or meta rules before turning them loose (in the mentioned case a simple blacklist of terms would have been helpful). Again this is a perfect task for machine ethics. Different concepts can be distinguished for the BESTBOT. At the one hand, it can learn from a user, on the other hand, it can compare different users.

GOODBOT and LIEBOT already had simplest options for memorizing the name of the individual user, and in a subsequent sentence where the name was replaced by a personal pronoun, they were able to refer to the predecessor, and assign the personal pronoun correctly. This is not real machine learning but the standard in many dialog systems. The GOODBOT could also accumulate knowledge about the user.

The BESTBOT can learn from statements, typing behavior and facial expressions. It can create a user profile and assign it to certain types, and it can track, record, and discuss the changes with the user. For example it can tell the user he or she seems much happier than the day before. In the open world one requirement is to recognize the user, for instance through a unique nickname assigned to one person only via login or via face recognition. Over time, as was hinted in the previous section, it can gain a balanced, complete image of the user. Then it can optimally adjust to the user in statements and in behavior (for instance when visiting websites or animating the avatar).

As already mentioned the GOODBOT only had a standard avatar not adjusted to the project. The LIEBOT was capable of indicating the form of lies through the animation of its avatars. The BESTBOT shall be furnished with an avatar that matches its own statements and actions as well as the statements and actions of users in facial expressions and gestures.

Machine ethics already provided several considerations on the design of software and hardware robots that can be referred to. A controversial discussion is in progress on how to design a nursing robot or sexbot. A nursing robot looking like a bear already exists. This might be pleasant or scary to a person in need of care. It is assumed a humanoid avatar best fulfils the intentions of the BESTBOT, this assumption is to be verified during the project.

Ethical Considerations

A general question is whether it is permissible to record and analyze a face or a voice with information technology. The personal data, one could say, belongs to the person. Of course, certain data is recorded in every contact between humans, memorized in the other person's brain for a short or longer time, but automated processing opens other aspects and options. Many persons might have access to the memorized data, unknown persons can gain assess, data can be linked and passed on, conclusions drawn by them can be false or misinterpreted by the responsible persons.

Another problem is the imbalance between the observer and the observed, between the interceptor and the intercepted expressed on different levels. The affected person does not have the technology the operator has, does not know the function principles in detail, and does not know who the data will be transferred to. Often only superficial information is given about face recognition, mentioning the presence of a camera only (Feng and Prabhakaran 2016). From ethical and legal viewpoints the BESTBOT operators could be requested to inform about the ongoing analysis, but some will say then the user might deactivate the camera.

One option is to use the BESTBOT itself as an information source. In the chat it could inform on the chances and risks of face recognition, voice analysis, and keyboard typing analysis.

The situation is so special because the user normally is at home, at school or university, or in the office, in other words in a well-known environment normally providing some privacy or predictability. Now analytic tools permeate this trusted room, linked to unknown variables. This might scare the user when realized.

Emotion recognition raises many questions from the perspective of information ethics. By showing emotion one gives away information, turning the inside out. Depending on if one is pokerfaced or not, one reveals information on well-being, psychological status, or other information. As already mentioned a personality profile can be created over time. Once face recognition and voice recognition merge there is enormous potential for abuse.

Methods unveiling the identity of the user have to be reviewed critically. A nickname or login with a fictitious username still seems to be an effective tool; requesting a real name probably is not responsible. Today it is possible already to identify many users with face recognition methods as they have left traces in the web, especially in social media. With a little training, voice recognition can also determine identities. Ways have to be found to ensure the BESTBOT does not breach the meta rule of the GOODBOT: not to be a snitch (meta rule 6).

As already mentioned in the last section the BESTBOT design has to be thought through carefully. It could be reasonable to design the chatbot as a humanoid to make it seem a reliable, trustworthy partner to be taken seriously. It could act and react like a human not only in its language, but also in its facial expressions and gestures. This might become a problem if the user gets emotionally attached to the BESTBOT or too trustful. This has been known to happen, the quite simple ELIZA is one example (Weizenbaum 1977). Again, meta rule 1 of the GOODBOT could be helpful.

Summary and Outlook

This article firstly explained the concept and implementation of the GOODBOT, a simple moral machine. One meta rule was selected and reversed to its opposite for another issue, the LIEBOT project. The development of this simple immoral machine was also documented here. The GOODBOT project showed that a machine can be "moralized" by relatively simple means. If an instable person is confronted with a standard chatbot his or her risk of self-mutilation or suicide might grow. The GOODBOT can cover this problem partly.

Secondly, the BESTBOT project was outlined. Findings from the GOODBOT project and the LIEBOT project have been applied and taken further in the context of machine ethics. The BESTBOT shall be even more helpful and obliging than the GOODBOT. One concept is not to make it a closed system like the GOODBOT but provide network connectivity. This raises questions about the trustworthiness and reliability, some of them can be answered by the outcome of the LIEBOT project. Another concept is to involve keyboard typing recognition, face recognition, and voice recognition. This concept brings new challenges to be faced by information ethics. The use of an avatar also seems to make sense for the BESTBOT but it also raises questions to be answered during the project.

References

Aegerter, A. 2014. FHNW forscht an "moralisch gutem" Chatbot. *Netzwoche*, 4/2014: 18.

Anderson, M.; and Anderson, S. L. eds. 2011. *Machine Ethics*. Cambridge: Cambridge University Press.

Asimov, I. 1973. *The Best of Isaac Asimov*. Stamford, CT: Sphere.

Bendel, O. 2017. Gesichtserkennung. *Gabler Wirtschaftslexikon*. Wiesbaden: Springer Gabler. http://wirtschaftslexikon.gabler.de/Definition/gesichtserkennungssoftware.html.

Bendel, O.; Schwegler, K.; and Richards, B. 2017. Towards Kant Machines. *The 2017 AAAI Spring Symposium Series*. Palo Alto: AAAI Press.

Bendel, O.; Schwegler, K.; and Richards, B. 2016. The LIEBOT Project. Extended abstract for the international conference *Machine Ethics and Machine Law* in Krakow, November 18–19, 2016. http://machinelaw.philosophyinscience.com/wp-content/uploads/2016/06/PROCEEDINGS-ver1-2.pdf.

Bendel, O. 2016a. The GOODBOT Project: A Chatbot as a Moral Machine. *Telepolis*, May 17, 2016. http://www.heise.de/tp/artikel/48/48260/1.html.

Bendel, O. 2016b. *300 Keywords Informationsethik: Grundwissen aus Computer-, Netz- und Neue-Medien-Ethik sowie Maschinenethik*. Wiesbaden: Springer Gabler.

Bendel, O. 2015a. Der kleine Lauschangriff: Auditive Systeme aus Sicht der Ethik. *Telepolis*, July 5, 2015. http://www.heise.de/tp/artikel/45/45319/1.html.

Bendel, O. 2015b. Können Maschinen lügen? Die Wahrheit über Münchhausen-Maschinen. *Telepolis*, March 1, 2015. http://www.heise.de/tp/artikel/44/44242/1.html.

Bendel, O. 2013a. Good bot, bad bot: Dialog zwischen Mensch und Maschine. *UnternehmerZeitung*, 7(2013)19: 30–31.

Bendel, O. 2013b. Der Lügenbot und andere Münchhausen-Maschinen. *CyberPress*, September 11, 2013. http://cyberpress.de/wiki/Maschinenethik.

Bendel, O. 2012. Maschinenethik. *Gabler Wirtschaftslexikon*. Wiesbaden: Springer Gabler. http://wirtschaftslexikon.gabler.de/Definition/maschinenethik.html.

Bendel, O. 2003. *Pädagogische Agenten im Corporate E-Learning. Dissertation*. St. Gallen: Difo.

Eckes, C.; Biatov, K.; and Hülsken, F. et al. 2007. Towards Sociable Virtual Humans: Multimodal Recognition of Human Input and Behavior. *Journal The International Journal of Virtual Reality* 2007, number 4, volume 6.

Feng, R.; and Prabhakaran, B. 2016. On the "Face of Things". *ICMR'16*, June 06–09, 2016, New York, USA.

Hammwöhner, R. 2003. Können Computer lügen? Mayer, M. ed. *Kulturen der Lüge*. Köln: Böhlau. 299–320.

Khan, R.; and Das, A. 2018. Basics of Bot Building. In: *Build Better Chatbots*. Berkeley, CA: Apress.

Kant, I. 1914. *Werke (Akademie-Ausgabe)*. Vol. 6. Berlin: Königlich Preußische Akademie der Wissenschaften.

Kosinski, M.; and Wang, Y. 2017. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*, 2017. Preprint via https://psyarxiv.com/hv28a/.

Marlow, J.; and Wiese, J. 2017. Surveying Surveying User Reactions to Recommendations Based on Inferences Made by Face Detection Technology. *RecSys'17*, August 27–31, 2017, Como, Italy. 269–273.

McTear, M.; Callejas, Z.; and Griol Barres, D. 2016. *The Conversational Interface: Talking to Smart Devices*. Cham: Springer International Publishing.

Myers, L. 2017. New SSML Features Give Alexa a Wider Range of Natural Expression. *Amazon Developer Website*, April 27, 2017. https://developer.amazon.com/de/blogs/alexa/post/5c631c3c-0d35-483f-b226-83dd98def117/new-ssml-features-give-alexa-a-wider-range-of-natural-expression.

Nahin, A.F.M. N. H.; Alam, J. M.; Mahmud, H.; and Hasan, K. 2014. Identifying emotion by keystroke dynamics and text pattern analysis. *Behaviour & Information Technology,* Volume 33, 2014 – Issue 9: Experiments and studies, 987–996.

Pakalski, I. 2017. Amazons Alexa erhält Stimmenerkennung *Golem*, October 12, 2017. https://www.golem.de/news/digitaler-assistant-amazons-alexa-erhaelt-stimmenerkennung-1710-130568.html.

Pereira, L. M. 2016. *Programming Machine Ethics*. Cham: Springer.

Rojas, R. 2013. *Können Roboter lügen? Essays zur Robotik und Künstlichen Intelligenz.* Hannover: Heise Zeitschriften Verlag.

Wallach, W.; and Allen, C. 2009. *Moral Machines: Teaching Robots Right from Wrong.* Oxford: Oxford University Press.

Weizenbaum, J. 1977. *Die Macht der Computer und die Ohnmacht der Vernunft*. München: Suhrkamp.

Williams, H. 2016. Microsoft's Teen Chatbot Has Gone Wild. *Gizmodo*, March 25, 2016. http://www.gizmodo.com.au/2016/03/microsofts-teen-chatbot-has-gone-wild.

Wu, X.; and Zhang, X. 2016. Automated Inference on Criminality Using Face Images. *arXiv*, November 13, 2016. https://arxiv.org/abs/1611.04135v1.