Societal Embedding of AI glasses

H.A. Jekel*, V. Rokx-Nellemann*

University of Twente, Faculty of Engineering Technology and Applied Mechanics, Drienerlolaan 5, 7522 NB, Enschede, The Netherlands

**Corresponding author h.a.jekel@student.utwente.nl, *v.rokx-nellemann@student.utwente.nl*

ABSTRACT: This paper aims to use constructive technology assessment (CTA) to solve the Collingridge dilemma faced in the technological development of AI glasses. CTA solves the dilemma by providing anticipation on future implications to improve the fitness of the technology under development. First, the actor analysis is used to investigate the direct and indirect actors affected by the product. Second, scenarios are sketched to anticipate on the co-evolution of society/morality and technology [1].

Key words: AI glasses, responsibility gaps, CTA, Societal Embedding, Mechanical Engineering

1 INTRODUCTION

Societal Embedding of technology enriches, directs and integrates new innovations [2]. Considering all the issues that pertain to the implementation of AI glasses in society, this paper focuses on the major problem in societal embedding which is the control dilemma of Collingridge. The dilemma states that while it is easy to influence technological development when its moral and societal implications are not yet manifested, it is hard to know these implications at the start of the development. However, later in the development process, when these moral and societal implications become clear, it is hard to guide the development in a different, more desirable direction [3].

This paper aims to bypass the Collingridge dilemma for the technological development of AI glasses into the healthcare industry, using constructive technology assessment (CTA). The aim of the development of AI glasses is to assist the visually impaired in their day to day tasks. Concerning the current development of the AI glasses, two main advancements are being made:

1. The AI glasses will be developed such that the user is able to quickly identify the gender of the person right in front of the user.

2. The AI glasses will be developed such that the user can navigate amongst obstacles in the users respective path.

It is safe to assume that any future launched product called AI glasses, will contain at least both current developments. Therefore, this paper will be concerned with the societal embedding of AI glasses as a combination of these developments.

Following the introduction, the methodology, section 2, will discuss methods to bypass the Collingridge dilemma. The results, section 3, will show how a selection of these methods can be used to bypass the Collingridge dilemma in the technological development of AI glasses. In the discussion, section 4, directions for future research are mentioned, suggesting other methods to bypass the dilemma that have not been applied in this paper. The conclusion, section 5, is used to emphasize the most important results obtained in applying the selected methods.

2 METHODOLOGY

2.1 CTA-tools to solve the Collingridge dilemma

To solve the Collingridge dilemma, constructive technology assessment (CTA) has been developed. CTA considers technological development in terms of Darwinian evolutionary theory. Natural selection is apparent as innovations are considered as variations exposed to the selection environment of markets, laws and regulations. CTA is used to anticipate on future implications to improve the fitness of the technology. CTA is said to be a form of prospective evaluation [3]. Therefore, all CTA-tools need to be applied at the beginning of the development, anticipating potential future impacts of the technology when they can still be changed.

Nyholm's book "Humans and Robots" played a central role in this papers literature study [4]. The book inspired the allocation of the relevant actors and influenced future scenarios sketched in this paper. It seemed rather natural to use the CTA-tools actor analysis and scenarios to solve the Collingridge dilemma.

Consequently, this paper starts with an actor analysis to anticipate on stakeholders' actions. The actor analysis emphasizes the linkages between different actors. The main focus of the actor analysis was to determine the impact of the actors on the AI glasses and to determine the impact AI glasses have and will have on the stakeholders. This includes the people who may be directly or indirectly affected by the product. An actor map, as displayed in figure 2, was produced to visualize the most important actors. This includes the people who have a direct effect on the product's design or development, those who make use of the product, those who indirectly influence the product and those who are indirectly affected by the product.

Regarding scenarios, this paper considers a scenario with both techno-moral and socio-technical aspects. Here, the techno-moral aspects explore 'soft' impacts of technologies [1]. Examples of such 'soft' impacts are: The perception of AI glasses, section 3.2.a, the responsibility of actors in the development and deployment of AI glasses, section 3.2.b, moral enhancement, section 3.2.c and morality, section 3.2.e. The socio-technical aspect, section 3.2.d, investigates how regulations or the lack thereof can influence the development of AI glasses, demanding customers to stay critical [1].

The next subsection discusses new approaches in the field of AI to solve the Collingridge dilemma. Although these new approaches will not be applied to AI glasses in this research, they form great alternatives for future research. Therefore an elaboration follows to help future developers of AI glasses in solving the Collingridge dilemma more extensively.

2.2 New approaches in the field of AI to solve the Collingridge dilemma

The new approaches that have arisen from the field of artificial intelligence (AI) for solving the Collingridge dilemma are socio-technical experimentation and technological mediation.

2.2.a Socio-technical experimentation

Instead of speculating about an uncertain future, van de Poel proposed to approach innovations as social experiments with corresponding ethical considerations, responsibly regulating innovation processes [5]. Socio-technical experiments consider technological developments as an intervention in society with unknown outcomes. It uses the currently available matching ethical framework to assign either a pass or a fail to the development of the technology.

2.2.b Technological mediation

Another approach worth mentioning is the approach of technological mediation. Although the above mentioned tools are helpful tools in solving the control dilemma of Collingridge, it is important to critically look at their flaws in solving the dilemma [3].

The method of socio-technical experiments considers the current ethical framework in evaluating if it is ethical to conduct the technological development. The flaw here is that the ethical framework changes as the AI glasses are introduced to society. Since the ethical framework changes over time, an ethical rejection at the start of a technological development does not necessarily lead to an ethical rejection at the end of the development [3].

The method of scenarios uses anticipation, speculating about future ethical frameworks. It does not take the dynamics into account of between technology and the ethical framework, as these speculated scenarios are formed at the start of the development process [3].

As an alternative, the approach of technological mediaton focuses on the dynamics of the interaction between technology and human values [3]. Technological mediation can be seen as a dynamic variant of techno-moral scenarios, where the scenario is constantly adjusted using empirical data required through the study of the dynamics of techno-moral change. Technological mediaton is not as speculative as techno-moral scenarios but does not completely give up on anticipation like socio-technical experimentation does.

3 RESULTS

3.1 Actor Analysis

The actors have been divided into groups as displayed in figure 2. The figure displays those who have a direct effect on the product's design or production - red, those who make use of the product - blue, those who indirectly effect the product - orange, and those who are indirectly affected by the product - green. It is important to diversify these groups since the interests and motivations differ and thus the impact upon these

groups is varied. It should be noted that only the major actors are discussed since they have the greatest impact on the outcome of the product.

3.1.a Product Design Engineers

The engineers who design the AI glasses have a direct impact on the outcome of the product. They must create a high quality product which adheres to regulations and guidelines for the design process. At the same time, engineers need to develop an attractive and safe design, both for the users and the medical personnel. Additionally, engineers should consider the consequences of the design for the users' environment. Engineers need to ensure that people within the vicinity of the product do not refer to the product in a negative way.

3.1.b AI glasses Manufactures

The manufactures are part of the red group, figure 2, as they have a direct impact on the product. They are responsible for the material utilized for production. They are also responsible for meeting the regulations and standards pertained to the manufacturing quality and process. The manufactures receive the final design of the product from the product design team and design a system, a production facility, that produces the AI glasses.

3.1.c Developers

Developers are the programmers that need to develop the software of the AI glasses. To stay at the top of the market, they need to make sure that the state of the art technology is implemented in the software of the AI glasses. Developers are also responsible for the cyber security of the product, they need to make sure that the risk of hacking is minimized. Red teaming should be applied, where experts try to hack-in to uncover vulnerabilities.

3.1.d Medical Personnel

Medical personnel, such as doctors, scientists, medical students, and specialized medical engineers is another large group of actors. The medical personnel will be working directly with the visually impaired, and will form the major link in communication between the visually impaired and the developers/researchers. Doctors and scientists will communicate directly with developers. Other medical staff and medical students will take care of regular check-ups with the users and provide the doctors and scientists with user data. The majority of innovations are started at universities where people work to improve upon pre- or non existing devices. Medical personnel, similarly to users, may also seek the media to raise awareness of a medical condition, that needs technological innovation or funding.

3.1.e Visually Impaired

The visually impaired are the most important actors and the largest group as they are the ones who dependent most on the outcome of the products development. As discussed below insurance companies and media are two indirect actors which have an affect on the visually impaired. Ultimately, the medical facilities are the largest connection direct connection to the user as this is where tests, procedures and implementation of a medical device or tool is accomplished.

3.1.f Insurance Companies

The target group (visually impaired) are connected to indirect actors such as insurance companies, as users are only compensated if the glasses are included in the coverage package of the users insurance company. In addition, they will have to purchase the device with their own wealth. These insurance companies are important to consider since they have an impact on the number of the visually impaired people get access to the AI glasses.

3.1.g Distributors

Distribution of the AI glasses is of key importance for the success of the product. They fall into the category of the indirect sub-group. To successfully embed the technology in society, it needs to be distributed. Distributors need to distribute as many products as possible in order to achieve high profits.

3.1.h Media

Another more recent actor involved in the development of products is the media. When a new innovation or medical procedure is put forward. Media or crowd funding organizations can help bring awareness and funding to a certain project. Enactors are also intertwined in this group, since they may search for media attention to raise funding or attention for the AI glasses.

3.1.i Enactors

Another important group of actors are the enactors. At the start of the development of an innovative idea there is no straight path to a working technology. There is no certainty on whether the idea will work. Enactors have to embrace this uncertainty and establish promises of future performance and applications. These promises need to result in external investments to get the process of technological devel-

opment started [6]. Enactors fall into the orange and green groups, figure 2, since they are responsible for the funding of the AI glasses and because enactores are affected by the outcome of the AI glasses.

Enactors try to convince the selectors that their solution is supreme, bypassing the process of natural selection. The selection of the new technology is not based on natural selection only. It is not necessarily the case that the best new technology for the visually impaired is chosen by the selector. Enactors need to make sure that the technology stands out in the arena of expectations, where different developers of similar technologies battle to win the investments of the selectors. This is visualized in figure 1.

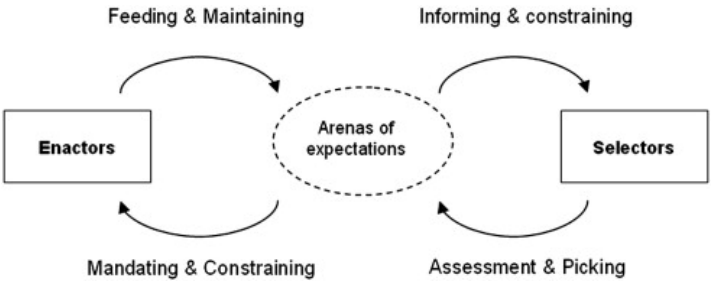

Fig. 1: Enactors, selectors and the arenas of expectations

3.1.j Government
Governmental organizations are responsible for the regulations within a country. Current regulations limit current developments of AI glasses and future regulations will hinder future developments. This makes the government an important actor with an indirect effect. Currently legal regulations regarding AI are non-specific, disabling reinforcement [7].

3.2 Scenarios

As mentioned in the introduction, any future launched product called AI glasses will at least contain both current developments:

1. The AI glasses will be developed such that the user is able to quickly identify the gender of the person right in front of the user.

2. The AI glasses will be developed such that the user can navigate amongst obstacles in an unknown environment.

Development 1 and 2 are both embodied machines that have sensors and actuators, that possess some degree of artificial intelligence and some degree of functional autonomy, which enables them to perform certain tasks that humans would otherwise typically perform. Of importance here, is the fact that the above description is one of the most commonly used definition of what a robot is [4]. Therefore, we can consider the product AI glasses to be a robot and apply robot ethics to AI glasses.

What follows is an ethical description of a scenario that contains both techno-moral and socio-technical aspects as mentioned in section 2.1. Although a technical description of a future form of AI glasses is redundant, the assumption that the glasses will contain a voice assistant-like interface for the user to operate the glasses is important. Traffic situations will serve as an example to reinforce theoretical statements, as most people are familiar with traffic situations.

3.2.a Techno-moral aspect: Perception of AI glasses
Perception of AI glasses considers the way in which the AI glasses are interpreted by the users and their direct environment. Here, it is important to consider that human psychology evolved both on a biological and cultural level before the introduction of robots and AI in society [4]. This causes their perception to be ingrained and non-adaptive in nature. Therefore, the developers of AI glasses should always aim to adjust the AI glasses to the users and not vice versa. This fixed perception could induce significant ethical implications. These implications are caused mainly by our mind-reading tendency, our minds' dual processing, our tendencies toward tribalism and our laziness [4]:

- Mind reading: Also referred to as the theory of mind. It is either a genetically based adaption or a cultural adaptation. It causes us to attribute mental states and other mental attributes to people, animals and objects whenever we interact with them, and whenever we try to interpret them. This will cause users to interpret AI glasses as having certain beliefs, desires, intentions and other mental states. This phenomenon is amplified whenever a robot is equipped with capacities for speech as is the case with the voice assistant interface of the AI glasses.

- Dual processing: Dual processing refers to the idea that people have two different mental systems [8]. The first is quick, intuitive and emotional, whereas the second is slow, deliberate and

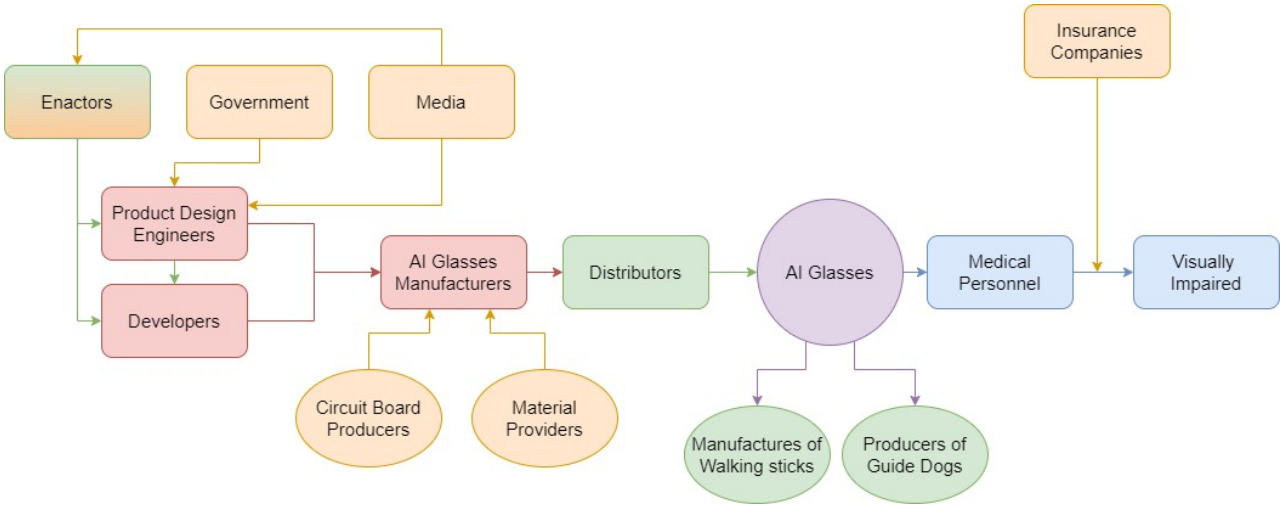

Fig. 2: Stakeholder Map

effortful. Applied to the AI glasses, reason might tell us that the AI glasses is just a robot, whereas the more intuitive or spontaneous system might respond to the glasses as if it were a person.

- Tribalism: Due to anthropomorphization of the robot, polarization might occur, viewing the robot as 'one of us' or 'one of them'.

- Laziness: Where humans might use shortcuts, robots are strictly bound by rules. When using AI glasses in a traffic situation this might result in compatibility issues similar to compatibility issues faced by autonomous cars. Regular people are not used to a situation in which everyone follows the traffic rules strictly, and for the glasses it is very hard to predict the chaotic human behavior.

To summarize, human psychology evolved both biologically and culturally before the introduction of AI glasses resulting in a rather fixed perception. This perception includes our mind-reading tendency, our minds dual processing, our tendencies toward tribalism and our laziness. This fixed perception needs to be considered in each step of the development process, as developers should always aim to adjust the AI glasses to the users, for the prevention of ethical implications.

3.2.b Techno-moral aspect: Responsibility

Our mind-reading tendency and our minds dual processing, section 3.2.a, causes people to automatically assign agency to AI glasses, mainly due to the voice assistant-like interface. In philosophy, the term agency refers to the capacities and activities mainly related to performing actions, making decisions, and taking responsibility for what we do [4]. This section analyzes the agency of AI glasses, answering questions with regards to responsibilities of AI glasses, its users and its developers.

AI glasses will not - at least not in the short term - have an agency similar to that of humans. However, they will exercise some lower level agency. Here, it is important to decide what moral and legal responsibility should be allocated to the agency of AI glasses, as they will be held responsible accordingly. This assignment of responsibility aims to prevent the formation of responsibility gaps.

Future scenarios might include traffic accidents with users of the AI glasses. In these situations it is essential that responsibility gaps are filled, as people feel the need to blame someone or something. When clearly explained that AI glasses are not responsible, tribalisation against AI glasses can be prevented. To provide such clear explanation, the type of agency of the AI glasses needs to be specified.

An adult is capable of exercising responsible agency. When being criticized, this agency is capable of standing one's ground, attempting to convince criticizers of one's principles justifying one's action. AI glasses cannot stand one's ground and can therefore not be considered as exercising responsible agency. AI glasses are only capable of collaborative agency. For that reason, AI glasses can never be held individually responsible for any occurring accidents and

should always be considered in a human-robot collaboration [4].

In rejecting the assignment of responsible agency to AI glasses, it is assumed that the AI technology present in AI glasses is comparable to that of autonomous cars. In contrast, future development resulting in AI capable of exercising this responsible agency might have different potential. In a future world full of responsible robots capable of starting friendships, demand for AI glasses capable of building friendships could become high. In that case it would be better to consider AI glasses as a responsible agent. However, such responsible AI seems very futuristic and will therefore not be considered further in this paper.

It can be concluded that AI glasses will not be a responsible agent in the near future. Therefore, AI glasses should always be considered in a human-robot collaboration, preventing tribalism against AI glasses.

3.2.c Techno-moral aspect: Moral enhancement and Human-Robot coordination in traffic

As mentioned in section 3.2.a, AI glasses should be developed such that they fit the users and the users' direct environment. However, there might be cases where moral enhancement is preferred, having people adjust to the AI glasses. According to Nyholm this should only be allowed if the following 4 statements are fulfilled [4]:

1. There should be a clear and identifiable benefit to the affected people.

2. The ways in which people have to adapt to the AI glasses are fairly non-intrusive and fairly noninvasive

3. The ways in which people have to adapt to the AI glasses are fairly domain-specific, not spilling over into too many other areas of life

4. The ways in which people have to adapt to the AI glasses are by and large reversible.

In traffic situations, it seems rather natural that other users of the congested network will have to adjust their behaviour to that of visually impaired people using AI glasses. The decrease in personal freedom of regular people is disproportionate to the life improvement of the visually impaired in this situation. Moreover, this example complies with all the above statements and thus moral enhancement should be applied here.

In situations where AI glasses enter traffic, acquire advanced Human-Robot coordination, as poorly managed traffic situations can become lethal. Similar to the new field of ethics of automated driving, mixed traffic with visually impaired people using AI glasses on the one hand and regular traffic on the other hand should be considered carefully.

First, it should be understood which incompatibilities occur. Next, it is of great importance to discuss how these incompatibilities should be solved. Finally we need to consider which ethical issues and challenges arise with these solutions.

Incompatibility conditions arise because the agents of AI glasses and other traffic users function differently, that is, they have different ways of forming expectations in a traffic situation. AI glasses will be optimized regarding user safety, user energy and user travel time. This will cause the user to display a robotic-like traffic behaviour, clearly different from other traffic users [4].

This robotic-like traffic behaviour will be apparent in the following ways:

- The AI glasses will not motivate the user to change speed during a walk.

- The AI glasses will keep the user away from any safety-critical situations

- The AI glasses will make the user follow the traffic rules very strictly.

In contrast, human users of the congested network are agents that exhibit satisficing rather than optimizing driving behavior, including non optimal safety, energy efficiency and traffic flow. This includes speeding, aggressive accelerating and decelerating and sometimes bending and breaking of the traffic rules.

These differences result in a clash of traffic behaviour and mutual difficulties in forming reliable expectations, which will inevitably lead to an increased likelihood for accidents. Hence, developers need to come up with ways to increase human-robot coordination to

avoid accidents caused by incompatibilities.

3.2.d Socio-technical aspect: Regulations and risks
In section 3.1.j, the role of the government was highlighted. As stated, legal frameworks are still lacking. Consultations with governmental organisations should be set up to argue legal frameworks. Up until the legal framework is set-up, competitors might produce ethically unacceptable variants of AI glasses, leaving users responsible for buying either ethically acceptable or unacceptable variants of AI glasses. The users of the AI glasses and the users direct environment should have an understanding of the capabilities and limitations of the agency that the product can exercise [4]. AI glasses run the risk of being hacked. One way for a company producing AI glasses to separate itself from competitors is to red team their AI glasses technology. In red teaming, experts try to hack-in to AI glasses to show its vulnerabilities. This will cost more but will result in an ethically more acceptable product. As stated earlier, it is the customers responsibility to stay critical here.

It can be concluded that consultations with governmental organisations should be set up immediately to prevent any unethical variant of AI glasses from entering the market. Until then, users should be critical when using AI glasses.

3.2.e Techno-moral aspect: Morality
Current developments will cause AI glasses to be amoral, they will lack any moral sense. However, a future society might demand AI glasses to include artificial moral agents (AMA). Artificial moral agents are capable of engaging in moral behavior or at least of avoiding immoral behavior. A commonly used example for the use of AMA's is that of critical traffic situations. MIT researches this very topic as they gather human perspectives on moral decisions made by AMA's [9]. Full individual moral agency requires the AI glasses to be responsible agents. However, as explained in section 3.2.b, it is assumed that any future implementation of AI glasses will not be capable of exercising responsible agency, as this seems too futuristic. Therefore, full individual moral agency is considered to be too futuristic as well [4].

However, this scenario of an increasing amount of moral agency within the AI glasses over time, is questionable. Machine ethicists pose six reasons for development of AMA's: Inevitability, prevent harm to humans, complexity, public trust and preventing immoral use. All these reasons can be objected [10]. Before developing moral agency within the AI glasses, it is therefore advised to wait for a response of the machine ethics community on the rejection of their core reasons for development of AMA's.

4 DISCUSSION

The bachelor assignment AI glasses started this year and will be reoccurring in the upcoming years. Within the time limit of this research, it has been impossible to perform an in-depth analysis using all mentioned methods in section 2. For these two reasons, this section proposes directions for further research.

A first proposition for further research is to investigate the remainder of the actors not analyzed in depth.

A second proposition is to set-up a socio-technical experiment. This includes proposing an ethical framework for social experiments regarding AI glasses, that should include four general moral principles: non-maleficience, beneficence, respect for autonomy and justice [11]. These four principles should then be further specified into sixteen conditions to help set up the experiment [12].

A third proposition is to tackle the Collingridge dilemma using technological mediation. Here it is suggested to create a YouTube video of the design and analyze the comments. Analyzing the comments allows for an empirically informed anticipation of future ethical frameworks regarding AI glasses. To execute such study, one needs to obtain approval from the ethics committee. Guidelines for responsible Internet research have to be obeyed. Anonimization of the names of commenters and removing information such as date, time and location of posting is advised [3]. To increase the value of such empirical data, future researchers could show this video to a certain target audience. Future researchers should discover who this target audience is.

A forth proposition is to develop specific ethical guidelines concerning the development of AI glasses and accessory reinforcement mechanisms. Current AI ethics guidelines lack reinforcement mechanisms and due to the distributed responsibility, developers feel a lack of accountability. Current legal regulations are

also non-specific, disabling reinforcement [7].

5 CONCLUSION

This paper proposed the use of constructive technology assessment (CTA) to solve the Collingridge dilemma, for the case of AI glasses. In this assessment, the CTA-tools actor analysis and scenarios where applied. The actor analysis, executed in section 3.1, gave an overview of the various actors/stakeholders involved in the development of AI glasses. The actor groups, as seen in figure 2, are compared and each group is investigated to determine the impact they have on the AI glasses development process and implementation. It is established that the most impactful groups are the red and blue groups. The red group is directly responsible for the quality, security, and adherence to regulations of the AI glasses. While, the blue group is the group which is majorly affected by the AI glasses, and are responsible for the aims, the AI glasses set to achieve. These two groups must be considered during the development and implementation stages, since they are directly affecting, and being affected by the AI glasses.

The scenario argued that human psychology evolved both biologically and culturally before the introduction of AI glasses resulting in a rather fixed perception. For that reason, developers should always aim to fit the AI glasses to society, and not vice versa, section 3.2.a. However, section 3.2.c argued that there can be situations in which moral enhancement is preferred, but should only be applied when the 4 statements mentioned are fulfilled. It is also argued that developers need to come up with ways to increase human-robot coordination to avoid accidents caused by incompatibilities in traffic. Section 3.2.b argued that AI glasses will not be able to exercise responsible agency in the near future. Therefore, AI glasses should always be considered in a human-robot collaboration, preventing tribalism against AI glasses. Section 3.2.d concluded that consultations with governmental organisations should be set up immediately to prevent any unethical variant of AI glasses from entering the market. From section 3.2.e it is clear that before implementing any time of AMA in AI glasses, it is advised to wait for a response of the machine ethics community on the rejection of their core reasons for development of AMA's.

6 DIVISION OF LABOUR

The complete paper is a group effort.

REFERENCES

1. Tools: Scenarios: 'toolbox' for researchers, technology developers and engineers.
2. J. Jelsma J. Deuten, A. Rip, Societal embedding and product creation management, *Technology Analysis & Strategic Management*, (1997), 9:131–148.
3. P.P. Verbeek O. Kudina, Ethics from within: Google glass, the collingridge dilemma, and the mediated value of privacy, *Science, Technology, & Human Values*, (2019), 44:291–314.
4. S. Nyholm, *Humans and robots: Ethics, agency, and anthropomorphism*, Rowman & Littlefield Publishers, Washington, DC, (2020).
5. I. van de Poel, Why new technologies should be conceived as social experiments, *Ethics, Policy & Environment*, (2013), 16:352–355.
6. M. Meeus S. Bakker, H. van Lente, Arenas of expectations for hydrogen technologies, *Technological Forecasting and Social Change*, (2008), 78:152–162.
7. T. Hagendorff, The ethics of ai ethics: An evaluation of guidelines, *Minds & Machines*, (2020), 30:99–120.
8. D. Kahneman, *Thinking, fast and slow*, Farrar, Straus and Giroux, New York, (2011).
9. H. Furey and S. Hill, Mit's moral machine project is a psychological roadblock to self-driving cars, *AI and Ethics*, (2021), 1:151–155.
10. S. Robbins A. van Wynsberghe, Critiquing the reasons for making artificial moral agents, *Science and engineering ethics*, (2019), 25:719–735.
11. I. van de Poel, An ethical framework for evaluating experimental technology, *Science and Engineering Ethics*, (2016), 22:667–686.
12. I. van de Poel, Nuclear energy as a social experiment, *Ethics, Policy & Environment*, (2011), 14:285–290.