xCoAx 2019 Conference on Computation, Communication, Aesthetics & X Milan, Italy

Hanna Schraffenberger
h.schraffenberger@ru.nl

Artificical Intelligence,
Radboud University,
Nijmegen, The Netherlands

Yana van de Sande
j.vandesande@student.ru.nl

Communication Science,
Radboud University,
Nijmegen, The Netherlands

Gabi Schaap
g.schaap@maw.ru.nl

Communication Science,
Radboud University,
Nijmegen, The Netherlands

Tibor Bosse
t.bosse@ru.nl

Communication Science,
Radboud University,
Nijmegen, The Netherlands

Can you fool the AI? Investigating people's attitudes towards AI with a smart photo booth

With the increasing impact of AI in people's everyday lives, multidisciplinary research on the public perception and understanding of AI is more important than ever. Yet, such research is still scarce. In this paper, we present a novel and playful setup for evaluating the impact of actual Human-AI Interaction on people's attitudes towards AI. The proposed setup takes the form of an intelligent photo booth capable of identifying humans. We present a first pilot study, illustrating how this AI system could be used in research. During this pilot, visitors of a film festival were challenged to fool the AI and take a selfie on which the intelligent photo booth would not identify them as a human being. Participants' attitudes towards AI were measured before and after the interaction. Based on exploratory observations, we conclude that multidisciplinary research into AI attitude, Human-AI Interaction and AI literacy is a promising research direction.

Keywords
artificial intelligence
Human-AI Interaction
AI literacy
AI attitude
privacy
security
face detection
interactive installation

1 INTRODUCTION

Artificial Intelligence (AI), the ability of a computer or robot to perform tasks that are commonly associated with intelligent beings (Copeland, 2019), plays an ever-increasing role in people's everyday lives. To mention just a few examples: video services, music platforms and webshops present customers with AI-driven recommendations. Smart assistants, such as Amazon's Alexa or Apple's Siri and AI-powered chatbots answer people's questions. Likewise, AI-based travel advice determines people's route to their destination.

The increase of AI in people's lives raises many questions: How do people think and feel about AI? Can they recognise, understand and evaluate the processes involved in AI-driven decision-making? What mental models do they use when interacting with AI systems — are these models similar to those of humans or more like models of machines? In this paper, we emphasise the need to address such questions concerning AI attitude, the public understanding of AI and Human-AI Interaction from a multidisciplinary perspective, combining expertise from both social science and computer science. We present a novel and playful setup for conducting such research in the context of face detection. The proposed setup takes the form of an intelligent photo booth capable of detecting humans (as well as various objects and animals) on digital images. We illustrate how this setup can be used for research with a pilot study. In our pilot, participants were challenged to fool the AI and take a selfie on which the intelligent photo booth would not identify them as a human being. Participants' attitudes towards AI were measured before and after the interaction.

The core objective of the paper is to spark more multidisciplinary and playful research into AI attitudes, AI literacy and Human-AI Interaction by sharing our approach and experiences. A key contribution of our work is the presentation of many urgent, societally and culturally relevant questions in the AI research landscape.

1.1 Concept

The general idea of our research is to let people interact with an AI system and to investigate whether and to what extent this affects their "AI attitude". The interactive installation presented in this paper asks people to fool an AI system. This idea is rooted in the observation that machines are getting smarter and smarter, and the consequent question whether we humans are still smart enough to fool intelligent systems. More specifically, it is relevant to see how perceptions of human abilities to outsmart AI affect attitudes towards AI.

The proposed setup explores people's attitudes towards AI in the context of automated face detection. The ability to detect faces in digital images has many applications. For instance, face detection is used in digital cameras so that faces in the picture will be captured sharply. Also, face detection is necessary to subsequently analyze the face further, e.g., to estimate a person's emotions based on facial expressions. More importantly,

it is also a first necessary step for face recognition. In other words, faces need to be detected first in order to then identify the person.

Ultimately, our project could have been realized with many different AI systems. We chose the context of face detection due to the combination of five factors: First, we believe it is important to approach AI not as a future technology, but as something that is already affecting people's lives. Face detection and subsequent face recognition is a form of AI many people are already affected by (e.g., people increasingly unlock their phones by presenting their face to the built-in camera). Second, face detection and recognition are controversial topics, which play into issues of surveillance, privacy and security. These topics are very societally relevant and also will benefit from a multidisciplinary approach. By combining the topic of AI and the topic of face detection, we can address several urgent issues at once. Third, face detection often happens without the person's consent and potentially can take place without the person being aware of this. For instance, in surveillance contexts people might not be aware of the camera or might not be aware that a machine rather than a human analyzes the images. This sets face detection apart from other kinds of (well-studied) AI experiences — such as interaction with robots or digital assistants — where people are more likely to be aware of the AI system and intentionally interact with it. By presenting people with a face detection AI, we hope to raise their awareness about such 'hidden forms of AI' and aim to gain more insight in how people are affected by interaction with less obvious forms of AI. Fourth, face detection is something humans are very good at themselves, which makes it an easily accessible topic for research with the general public. Fifth, it is relatively easy to determine the success and failure of a face detection AI system. A suggestion by a recommendation system, for instance, is more difficult to classify as a success or failure. Likewise, it is easy to assess if a person has managed to stay undetected. Such clear distinctions are beneficial for us, as we are interested to see if people's opinion about AI is affected by how successful the AI is and by how successfully they can control the outcomes of their interaction with the AI.

Due to the multifacetedness of the topic, the resulting installation allows us to address many questions. In addition to studying the attitudinal effects of interaction with AI systems, we are also interested how people feel and think about AI in general. We have designed our installation as a tool to answer these questions and as an exhibit that fosters reflection and conversation around the topic of AI.

2 RELATED WORK

Not surprisingly, the increase of AI in people's lives goes hand in hand with an increase in research about the relationship between humans and AI systems. In particular, existing research addresses humans' interaction with robots (e.g., the survey by Goodrich, Schultz, et al., 2008), their experience with virtual agents (e.g., Cassell & Tartaro, 2007), chatbots and virtual assistants (e.g., Klopfenstein, Delpriori, Malatini, & Bogliolo, 2017) as

well as intelligent user interfaces (e.g., Ross, 2000). Also, specific AI-driven tools, such as recommendation systems have been actively studied (e.g., Park, Kim, Choi, & Kim, 2012).

Surprisingly, so far little focus has been put on the general public's attitude towards AI. It stands to reason that people's thoughts and feelings towards AI are, amongst others, shaped by mass media (e.g., articles about AI and movies depicting (future) AI systems) as well as by their own personal experience with AI systems. Whether and to what degree this is the case, still needs to be explored. This paper takes up these questions and proposes a setup to investigate whether interaction with AI affects people's thoughts and feelings about AI.

2.1 Fooling face detection and recognition

Our installation challenges participants to fool a face detection system. This topic has been explored before, both in the arts and in the sciences. In the art context, fashion has been proven a powerful tool to evade face detection (an overview is provided by Davis, 2014). For instance, the designer and technologist Harvey (2011) has presented CV Dazzle — a camouflage from face detection that is based on applying makeup, wearing fashion accessories and styling the hair in a way that prevents the widely-used Viola-Jones face detection algorithm (Viola & Jones, 2001; Viola & Jones, 2004) from recognizing the face. Simply put, this face detection algorithm makes use of the fact that human faces share similarities, and that some areas of a face are generally darker or lighter than other areas.

Another artistic project by Harvey (2017) that aims at fooling face detection is Hyperface. To protect the wearer from face recognition technology, the project uses clothing with special abstract patterns, which contain 'false faces' that distract facial detection systems from the wearer's real face.

A similar approach is used in the REALFACE Glamoflage project by Simone C. Niquille, who designed t-shirts to fool Facebook's face recognition as part of her Master thesis (in Barribeau, 2013). Like Hyperface, her t-shirts present face recognition systems with many faces. However, here the faces are no abstract patterns — instead, the t-shirts feature an artistically designed collection of actual faces of famous people, ideally tricking the system in, e.g., identifying the face of Michael Jackson or Barack Obama rather than the face of the wearer.

The URME Surveillance project by artist Leo Selvaggio (2015) similarly focuses on making cameras identify the wrong person. The artist distributes masks of his own face that people can wear so that they are identified as him.

In the scientific domain, researchers have taken up the ideas proposed by artists. In particular, Feng and Prabhakaran (2013) build on the work by Harvey (2011), and propose a tool to help artists and designers in creating camouflage-thwarting designs. Their tool finds prominent features that cause a face to be recognized and presents suggestions for camouflage options (makeup, styling, paints).

Yamada, Gohshi, and Echizen (2013) propose a wearable prototype similar to eyeglasses that is meant to prevent unauthorized face image revelation. Like Harvey's CV Dazzle, the project is set up to change the apparent features around the eyes and nose, which are used in the Viola-Jones face detection process. Instead of makeup and hairstyles, their device is based on transmitting near-infrared signals. These signals are picked up by camera sensors and corrupt the captured images, rendering the faces in the captured camera images undetectable.

The project by Sharif, Bhagavatula, Bauer, and Reiter (2016) also requires the user to wear specific glasses. However, in contrast to previously reviewed approaches, it focuses on face recognition realized with deep neural networks (DNNs). They propose techniques for generating eyeglass frames that allow the wearer to evade being recognized and even to impersonate other individuals.

Finally, a related area of research concerns face de-identification, which refers to the removal of identifying information from images (see, e.g., Gross, Sweeney, Cohn, De la Torre, & Baker, 2009). A well-known example of face de-identification is blurring faces. Our own project focuses on real-world measures that people can take in the physical domain to prevent face detection rather than on digital modifications. Because of this, such face de-identification approaches as well as digital adversarial attacks, fall out of the scope of this paper. Yet, it should be mentioned that similar questions are addressed with a focus on the digital domain. For instance, Wilber, Shmatikov, and Belongie (2016) explore whether people can still circumvent face detection on Facebook by applying various image filters to photographs.

As this short review shows, the topic of fooling face detection has been actively explored. Our project takes these efforts as an inspiration. However, unlike many existing projects it does not propose a new tool to evade face detection and is not aimed at enabling people to fool face detection, nor is it interested in building more robust face detection mechanisms that cannot be fooled. Instead it is designed to research people's attitude towards and interaction with AI and to provide a thought-provoking and intriguing experience.

3 *MISIDENTIFY.ME:* THE INTELLIGENT PHOTO BOOTH

In order to study whether and how interaction with AI systems affects people's attitudes towards AI, we created an 'intelligent photo booth' application. The resulting system is called "misidentify.me" (see Figure 1) and it is smart in the sense that it is able to detect humans as well as identify a range of objects and animals. In the context of our pilot study, the system was presented in an ad-hoc photo booth, next to a table with masks and makeup and a short text describing the installation and the challenge. However, the system can also be used online. In the following, we will describe this core part of the system as well as how we have presented it in the context of the festival.

Fig. 1.
The *misidentify.me* installation at the InScience film festival

3.1 Interaction design

The front-end and user interaction of our installation was designed as follows: The user is greeted on a start-screen where a text challenges them to fool the AI and to take a selfie on which they are not identified as a human being. The screen features two primary buttons, allowing them to participate in the experiment or to simply try to fool the AI without taking part in the study. The screen also shows a digitally mirrored version of the webcam input, allowing the user to see themselves in front of the computer. On this image, the position of their face is highlighted with a rectangle, using a face detection mechanism described below. This makes sure the user can see that the face detection is working and gives them an immediate idea of what the project is about. After selecting either option, the rectangle around their face disappears, and one primary button allows them to snap a selfie. In the film festival exhibit, people could now use makeup or masks that were provided (see Figure 1) or come up with their own strategies to fool the AI. However, it is also possible to attempt the challenge without such appliances, e.g., by using one's hands to cover the face. [1]

Once the selfie has been taken, users can either retake the image or submit it to the AI. When submitted to the AI, the image is analyzed and a label appears on the image, presenting the system's verdict: either that the user is a human and the confidence that this is the case, or that they are something else, specifying the object or animal that the system recognized and its confidence about this decision (see Figure 2). In case the person is identified as a human, their face is again highlighted with an enclosing rectangle. This result-screen allows the participant to try fooling the AI again, or to exit the experience.

1. Ideally, we would stimulate them to use their own strategies instead of steering them into a certain direction by providing the materials. However, because of the festival context, we adopted a more entertainment-oriented approach.

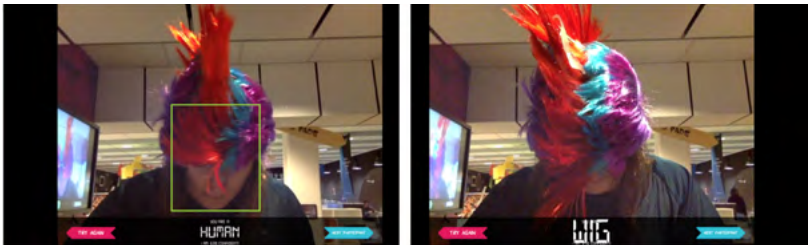

Fig. 2.
The result-screen.

3.2 Hardware

On the hardware side, the installation only requires a computer and a webcam, and any relatively modern laptop should suffice to run the experience. In the festival context, the misidentify.me installation was run inside an old voting booth that was re-purposed as an ad-hoc photo booth. A 15" Macbook Pro from 2015 was placed within the booth, with a mouse and a USB number-block for usability reasons. The build-in camera was used to take the visitors' selfies.

3.3 Software

On the software side, misidentify.me is realized with web technologies. This is done to also offer the experience online and to extend the study with an online experiment in the future. However, during the festival, everything was run locally. The core functionalities are realized with the combination of p5.js (see https://p5js.org/) and ml5.js (see https://ml5js.org/).

The p5.js library is a JavaScript client-side library for creating visuals and interactive experiences, and it is used to facilitate the overall interaction and user flow. The ml5.js library is an easy-to-use wrapper around the widely-used TensorFlow.js (https://js.tensorflow.org) library for machine learning, and provides access to machine learning algorithms and models in the browser. (We chose this particular library because it allows researchers with no or little machine learning experience to understand and possibly adapt the code. This is ideal for multidisciplinary teams.) Among other things, ml5.js supports the use of pre-trained models for detecting human poses and image classification. In the misidentify.me installation, ml5.js is used to analyze the submitted selfies. Two machine learning models are used in combination: One to determine what we called the 'humanness' of the image and one to determine the 'somethingness' of the image. We made sure that in the end, the system has to decide whether it is dealing with (1) a human or (2) something else (an object or animal). In the former case, the image is labeled with the term "human" and a rectangle indicates the position of the human's face. In the latter case, the name of the identified object/animal is reported. In both cases, a confidence score accompanies the result.

To estimate the 'humanness', the ml5 version of PoseNet — a machine learning model for real-time human pose estimation — is used.[2] PoseNet can detect a human figure in an image (or video) and estimate the position of key body joints. We use the information about five keypoints, namely the nose, left eye, right eye, left ear and right ear to estimate the probability that the image contains a human. (We only focus on the face as selfies potentially solely focus on/contain the face.) In our setup, this 'humanness' is simply the average (mean) of the confidence scores of the five face-related keypoints mentioned above. Strictly speaking, this score describes the confidence about the accuracy of the keypoints. Yet, our informal tests revealed that the score works well for our purpose. To decide whether the

2. Alternative face detection methods were tested, and PoseNet was chosen based on informal testing as it provided a good balance between successfully detecting faces and the possibility to fool the system.

image should be labeled as “human” or not, the ‘humanness’ score is compared with a ‘somethingness’ score.

The ‘somethingness’ score is determined with a MobileNet model for image classification. Here, the ml5 library accesses a pre-trained model that was trained on the ImageNet database (Deng et al., 2009; Image-net.org, n.d.) consisting of over 15 million images. It knows 1000 different classification categories (objects and animals). When submitting an image, this model outputs the top three hits and their probability. For the ‘somethingness’ score, we simply use the probability of the top result.

As mentioned, the two scores are compared. If the ‘humanness’ score is bigger than the ‘somethingness’ score, the image is labeled as human. Otherwise, the label of the object/animal that has been identified is presented to the user. In both cases, the confidence (the winning score) is reported as well.

To record the results of those users who give consent and choose to participate in the study, information is written to a JSON file with the Node.js (see https://nodejs.org/en/) web application framework Express (see https://expressjs.com). We record the participant ID to link the data to questionnaire data that we collect with the survey platform Qualtrics (see https://www.qualtrics.com/). Furthermore, we record the result that has been presented to the participant and the displayed confidence about the result. Finally, we record the ‘humanness’ score and the top three labels returned by MobileNet and their probabilities. Ideally, we would save the selfie for further analysis (e.g., to analyze them for used strategies). Unfortunately, we were not able to obtain ethical approval for such a study yet.

4 PILOT STUDY

We used the above described installation in a two-day pilot study at the InScience science film festival in the Netherlands. The following sections illustrate how our setup can be used for research into the public attitude towards AI. To guide our exploratory analyses, and in line with prior research and theory (e.g., Ryan & Deci, 2000; Dietvorst et al., 2018), we expected that participants who succeeded in not being identified as a human would have a more positive attitude towards AI, as it can be hypothesized that fooling the AI gives them a sense of greater control over AI technology.

4.1 Design

We used a quasi-experimental 1 factorial (identified as human yes/no) pretest-posttest design, in which two dependent variables – thoughts and feelings about AI – were measured both before and after interacting with the smart photo booth. Participants were allocated in one of two groups, depending on whether or not the AI was able to identify them as humans. The researchers obtained formal approval for the study from their institution’s ethical committee beforehand.

4.2 Participants

Participants of the experiment were festival and library visitors and acquaintances. While everyone was allowed to use the photo booth, only adults were allowed to participate in the study. In total 42 adult users took part in the experiment. Unfortunately, after cleaning the data and removing the data from participants with incomplete answers, only data from 25 participants (21 Dutch, 2 Italian, 1 Finnish, 1 German) remained. The remaining participants were between 20-70 years of age (M=33.25, SD=14.53). 8 participants identified as male and 17 as female. Participants were relatively well educated with 18 participants having either completed or currently pursuing a university degree. 3 participants reported a background in artificial intelligence or data science and another six participants had other IT related backgrounds.

4.3 Procedure

Adult visitors of the festival and/or library either approached us or were approached by us. They were then asked whether they are interested in interacting with our installation and fooling an AI. If interested, visitors could choose to simply play with the system or to participate in the study. In the latter case, active informed consent was obtained before participating. Subsequently they were asked to fill in a pretest questionnaire on a tablet. The pretest questionnaire stated a definition of AI, to make sure everyone was thinking about the same kind of systems. Subsequently, they got one shot at fooling the AI. (However, they could retake the selfie until they were satisfied.) After being informed of the result, they were asked to fill in the posttest questionnaire.

4.4 Measures

We administered a pretest and posttest questionnaire via Samsung tablets. The questionnaires were presented in Dutch (see www.misidentify.me/questions.pdf for an English version). Both questionnaires measured attitude towards AI in general, with attitude being seen as the participants' thoughts and feelings about AI. Using a Visual Analogue Scale ranging from 0-100, 15 questions addressed *feelings about AI* ('How do you feel about AI in general?'), with answers e.g., anxious-calm, good-bad, or inferior-superior. Likewise, 11 items regarding *thoughts about AI* ('I think AI is... ') included e.g., useful-useless; predictable-unpredictable, or risky-safe. The order of the items was randomized for each participant. Factor analyses showed one factor for the 15 *feelings* items, and one factor for 9 out of 11 *thoughts* items. Accordingly, mean scores were calculated for both variables and entered into the analyses, with scores ranging from 0 (most negative) to 100 (most positive). Cronbach's α's for pre and posttest scales ranged from .67-.95.

In addition to these questions, the posttest questionnaire measured demographics and included questions about their experience with the in-

telligent photo booth (e.g., whether they were surprised by the outcome), and more general (control) questions (e.g., interest in AI, usage of AI in daily lives, AI-related media consumption and whether participants had a background in AI).

4.5 Results

Of the 25 participants with complete datasets, 12 were correctly identified as a human and 13 were not correctly identified as human beings. Of the 13 selfies that were not labeled as human, 10 selfies were labeled as something else (4 masks, 3 wigs, 1 bow, 1 windsor tie, 1 lampshade). For the remaining 3 selfies, the AI did output the label human, but did not associate this label with the participant (e.g., in two cases it detected some other person in the background). We checked whether the participants were distributed randomly across the two conditions. There were no significant differences between the two conditions in terms of gender, age, education, background in computer science, information technology, AI or data science (Range$_p$ = .286-.863). Because of the small number of participants, no inferential analyses were conducted and only descriptive statistics are provided.

Table 1 shows the general feelings and thoughts about AI were more on the negative side, with a score well below the midpoint of the 0-100 scale. In addition, the pretest and posttest data suggest that once participants had interacted with the photo booth, their feelings and thoughts became even more negative.

	Feelings		Thoughts	
Pretest	***M* (SD)** 36,73 (15.63)	**Min-Max** 1-63.80	***M* (SD)** 40.79 (10.57)	**Min-Max** 16.56-57.44
Posttest	***M* (SD)** 32.70 (16.39)	**Min-Max** .87-64.0	***M* (SD)** 35.97 (14.90)	**Min-Max** .89-56.89

Table. 1.
Descriptive statistics (N=25)

When comparing the means for the two conditions, it seems that participants had a more positive attitude towards AI when they succeeded in fooling the AI.[3] In the group that was not identified as human, mean posttest scores for feelings were higher (*M*= 34.68, *SD*=17.85) than scores in the group that was identified as human (*M*= 30.56, *SD*=15.12). Likewise, thoughts about AI were more positive in the group that succeeded in fooling the AI (*M*=37.88, *SD*=15.5) than in the group that did not (*M*=33.90, *SD*=14.57).

3. With "fooling the AI" we refer to the objective outcome that the person was not identified as a human being – not to the subjective experience of having fooled the AI.

5 DISCUSSION

In what ways and to what extent does interaction with an AI system affect people's thoughts and feelings about AI? Unfortunately, based on the little data collected so far, we cannot present a definite answer to the question. In our opinion, the small sample size does not justify inferential analyses.

However, it stands out that people's AI attitude was generally more on the negative side. We suggest to evaluate whether this is indeed the case for the general population, and if so, research the cause of people's negative thoughts and feelings.

Furthermore, we notice that all posttest scores are lower than the pretest scores. This is particularly interesting because many participants were impressed and amused by the AI, which could explain a more positive AI attitude. Hence, it seems promising to research whether interaction with AI leads to a significantly more negative attitude towards AI regardless of the outcome of the interaction. Also, it would be interesting to see if interactions with different types of AI systems (e.g., a AI-driven chatbot) affect people's AI attitude in different ways. Finally, our study provides first indications that when people 'outsmart' the technology, their attitudes toward that technology are more positive.

The many conversations and informal observations revealed that our installation creates AI awareness and fosters dialogue and reflection. Our observations reaffirm that at least some people seemed rather surprised and impressed by the capabilities of the AI system and/or the outcome. In our opinion, this is especially interesting because similar AI technology is already a part of many people's everyday lives. The fact that people nonetheless respond this way raises the question whether people are aware of the AI systems that are already part of their lives and highlights the need for research into what we call "AI literacy" — the question whether people can recognise, understand and evaluate the involvement of AI systems when using technology.

Our pilot study shows how the misidentify.me installation can be used for scientific research. Our hope is that people learn from the limitations of our study and the mistakes we have made. We have only allowed people to submit one selfie. One single interaction might be too little to have a measurable effect on people's attitudes. For future studies, we suggest to allow people to interact with AI systems repeatedly. Another limitation of our research was that the questionnaire was only presented in Dutch, with on-the-fly translations for non-Dutch participants. Also, our questions about attitude (thoughts and feelings) were designed specifically for this study, and validated measures of these constructs would be very desirable. Finally, the distinction into two groups based on whether participants were identified as human was more difficult than anticipated. Although many people were not identified as humans, the AI labelled several people who were wearing a mask with the term "mask". While some of these people felt like they had fooled the AI, others did not feel like the AI was fooled (after all, it did not make a mistake). We used the objective result rather than the subjective experience in this study. However, we believe it is important to also take participants subjective experience into account in the future.

The installation of this study was designed to create a good balance between people who are identified as a human and those, who are not identified as a human. This was successful. On the one hand, the fact

that AI systems could (at times), easily be fooled raises questions about their security and robustness. With respect to this, it has to be noted that other AI systems likely perform much better. On the other hand, some people were not able to fool even our rather basic AI. This raises the question of people's control and agency in a society where smart machines are ubiquitous. We suggest to address such questions in future research.

The experiment was held as part of a film festival. Informal conversations made it very plausible that films can shape people's attitudes towards AI technology. We hope to validate this observation outside of the scope of a film festival. More generally, we believe it would be interesting to test the setup in different contexts.

Aside from the proposed study design, the main outcome of our study is not the answers, but the questions that have been raised during the project. We hope this setup will inspire future research, and suggest researchers to have a look at our questionnaires (www.misidentify.me/questions.pdf). We believe that in order to address questions of AI attitude, AI literacy and Human-AI Interaction, a multidisciplinary approach is needed. Whereas existing AI research mostly focuses on the technical challenges underlying effective and robust AI systems, the effect that these systems have on people is often overlooked. For instance, many researchers try to develop computational methods to extract meaningful information from machine learning models (so-called 'explainable AI'), but such research can be nicely complemented with insights into how human beings generate and process explanations (Miller, 2018). Hence, in our opinion, to make sure the future of AI is truly "social AI", computer science and the social sciences need to go hand-in-hand in a way that is not unlike co-evolution: computer science needs the expertise of social science (e.g., communication science) to understand the effects that developed AI systems have on people. At the same time, the social sciences need the computer sciences to develop and improve existing systems according to their findings (Bosse, 2019). The project presented in this paper is a collaboration between AI/HCI and communication science researchers. Only together, we were able to build the system and set up this study. For future work, we plan to run this study online, allowing people to fool the AI in the comfort of their own homes. With this project, we hope to spark more experiments about the relationship between humans and AIs.

Acknowledgements

We wish to thank the participants for taking part in the study and the people behind the ml5 and p5 libraries for their valuable tools. Also, a big thanks to the coding train slack community (and in particular Oliver Wright) for providing support with using these libraries and to Dan Oved for support with PoseNet.

REFERENCES

Barribeau, Tim.
A T-Shirt That Tricks Facial Recognition Software: The REAL- FACE Glamoflage, 2013. https://www.popphoto.com/news/2013/10/t- shirt-tricks-facial-recognition-software-realface-glamoflage.

Bosse, Tibor.
Social Artificial Intelligence. Inaugural address, Radboud Universiteit Nijmegen, 2019.

Cassell, Justine, and Andrea Tartaro.
"Intersubjectivity in human–agent inter- action." *Interaction studies* 8, no. 3 (2007): 391–410.

Copeland, B.J.
Artificial intelligence — Definition, Examples, and Applications, 2019. https://www.britannica.com/technology/artificial-intelligence.

Davis, Lauren.
Fashion that will hide you from face-recognition technology, 2014. https://io9.gizmodo.com/how-fashion-can-be-used-to-thwart-facial-recognition-te-1495648863.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei.
"Imagenet: A large-scale hierarchical image database." In *Computer Vision and Pattern Recognition,* 2009. CVPR 2009. IEEE Conference on, 248–255. Ieee, 2009.

Dietvorst, B. J., Simmons, J. P., & Massey, C.
(2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Management Science, 64(3), 1155-1170.

Feng, Ranran, and Balakrishnan Prabhakaran.
"Facilitating fashion camouflage art." *In Proceedings of the 21st ACM international conference on Multimedia,* 793–802. ACM, 2013.

Goodrich, Michael A, Alan C Schultz, et al.
"Human–robot interaction: a sur- vey." *Foundations and Trends in Human–Computer Interaction* 1, no. 3 (2008): 203–275.

Gross, Ralph, Latanya Sweeney, Jeffrey Cohn, Fernando De la Torre, and Simon Baker.
"Face de-identification." In *Protecting privacy in video surveillance,* 129–146. Springer, 2009.

Harvey, Adam.
"CV Dazzle: Camouflage from face detection." Retrieved 26 April 2019 from https://ahprojects.com/cvdazzle (2011).

Harvey, Adam.
"HyperFace." Retrieved 26 April 2019 from https://ahprojects.com/hyperface/ (2017).

Image-net.
org. http://www.image-net.org/.

Klopfenstein, Lorenz Cuno, Saverio Delpriori, Silvia Malatini, and Alessandro Bogliolo.
"The rise of bots: a survey of conversational interfaces, patterns, and paradigms." In *Proceedings of the 2017 Conference on Designing Interactive Systems,* 555–565. ACM, 2017.

Miller, T.
Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial I ntelligence 267* (2019): 1-38.

Park, Deuk Hee, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim.
"A literature review and classification of recommender systems research." *Expert Systems with Applications 39,* no. 11 (2012): 10059–10072.

Ross, Edward.
"Intelligent user interfaces: Survey and research directions." University of Bristol, Bristol, UK, 2000.

Ryan, Richard M., and Edward L. Deci.
"Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." American psychologist 55.1 (2000): 68.

Selvaggio, Leonardo.
"URME Surveillance: performing privilege in the face of automation." International Journal of Performance Arts and Digital Media 11.2 (2015): 165-184.

Sharif, Mahmood, Sruti Bhagavatula, Lujo Bauer, and Michael K Reiter.
"Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition." In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security,* 1528–1540. ACM, 2016.

Viola, Paul, and Michael Jones.
"Rapid object detection using a boosted cascade of simple features." In *Computer Vision and Pattern Recognition,* 2001. CVPR 2001.

Viola, Paul, and Michael J Jones.
"Robust real-time face detection." *International journal of computer vision* 57, no. 2 (2004): 137–154.

Wilber, Michael J, Vitaly Shmatikov, and Serge Belongie.
"Can we still avoid automatic face detection?" In *Applications of Computer Vision (WACV),* 2016 IEEE Winter Conference on, 1–9. IEEE, 2016.

Yamada, Takayuki, Seiichi Gohshi, and Isao Echizen.
"Privacy visor: Method for preventing face image detection by using differences in human and de- vice sensitivity." In *IFIP International Conference on Communications and Multimedia Security,* 152–161. Springer, 2013.