# Thank you for your trust, I will now take over your vision

## An experimental setup for researching the interplay between different perceptions of NNC's and humans

**With the integration of neural network classifiers (NNC) in society, the perception of the physical world will switch from the human sensory perception to that of the NNC. This shift could have a huge impact on society, as unexpected decisions might emerge because of a clash between the two perceptions. In particular, with the rise of adversarial examples in the physical world. With this in mind, we propose an experimental setup as a reproduction of the study by Athalye et al. (2018) that could be of merit in Human-Computer Interaction (HCI) research. The proposed setup creates the possibility for the participant to experience the impact of adversarial examples and thus imagine society with embedded NNC.**

### BACKGROUND

Task specific Artificial Intelligence (AI) systems, like NNC's, are developed with the aim to act upon the found patterns, without human help (e.g. kamikaze drones, traffic cameras, etc.). This optimization has come to the cost that the systems are vulnerable to adversarial examples that highlight a clash between machine perception and human perception, as both parties classify the target differently. Recent research confirms the possibility of such a clash, as it shows that NNC's can be fooled by robust 3D adversarial examples in physical space, which is invisible to the human eye (Athalye et al., 2018).

It therefore remains a challenge to imagine the integration of this technology in a society where the general regulations are based on the perception of the human senses.

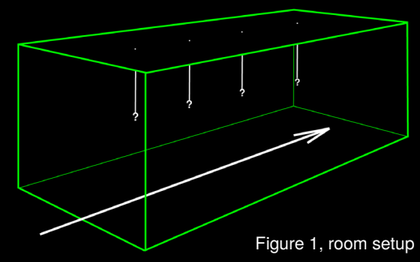

Figure 1, room setup

### PROPOSED SETUP

We propose an informative and immersive experimental setup that mimics the integration of a NNC and enables researchers in HCI to show the physical world from the viewpoint of the AI system to a participant. The setup showcases four objects hanging in a long, narrow space (Figure 1). The first and third objects are unaltered (machine perception = human perception) while the second and fourth objects are adversarial objects (machine perception ≠ human perception). The participant is asked to walk across the space wearing an Virtual Reality (VR) headset with camera and headphone. The VR software will run on Google's Cloud Vision API to identify the objects from a machine perception. The encountered objects, as identified by the VR, are communicated to the participant's headphone using a computer-generated voice. Furthermore, the screen of the headset is black and displays bounding boxes around the place where the objects are detected (QR code).

While encountering the different objects one-by-one, the participant can use tactile perception to identify the object (Figure 2). Because the participants first experience the identification of the objects from the perception of the classifier and afterwards from their own perception, the setup simulates the clash between machine and human perception.

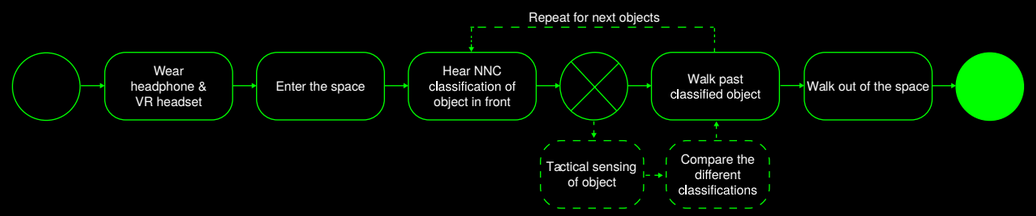


Figure 2, interaction flowchart

Marise van Noordenne
Leiden Institute of Advanced Computer Science
Leiden University
Leiden, Zuid-Holland, the Netherlands
marisevannoordenne@gmail.com

Emma Floor Stolk
Leiden Institute of Advanced Computer Science
Leiden University
Leiden, Zuid-Holland, the Netherlands
flokars@gmail.com

### FUTURE PERSPECTIVE

The proposed setup will offer researchers a possibility to conduct further research on the interplay between the different perceptions of the physical world by AI systems and humans. Thus, the setup creates an opportunity for research on the societal embedding of emerging technology as the setup allows the participants to evaluate the integration of these technologies, in this case NNC, in society.

### REFERENCE

Athalye, A., Engstrom, L., Ilyas, A., & Kwok, K. (2018). Synthesizing robust adversarial examples. In *International conference on machine learning* (pp. 284-293). PMLR

### MORE

The QR code links to a video that further illustrates the user experience of the proposed setup.