High Happiness

Violence Detected

Data point B

Violence Detected

Unusual Tshirt

Higher Density

Escape the Smart City

Critical pervasive game to question the AI-surveillance infrastructure in the smart city

Surveillance D

Master Thesis : Tomo Kihara

waag

TU Delft

Author

Tomo Kihara

Master - Design for Interaction | TU Delft

www.tomokihara.com | playful.intervention@gmail.com

Project Chair

Roy Bendor

Assistant Professor | TU Delft

r.bendor@tudelft.nl

Project Mentor

Derek Lomas

Assistant Professor | TU Delft

Dereklomas@gmail.com

Delft University of Technology

Faculty of Industrial Design Engineering

Landbergstraat 15 2628 CE Delft

The Netherlands

www.tudelft.nl

Client Mentor:

Tom Demeyer

Head of Technology Development | Waag

tom@waag.org

Client Mentor:

Stefano Bocconi

Software Developer | Waag

stefano@waag.org

Waag

Sint Antoniesbreestraat 69

1011 HB Amsterdam

The Netherlands

https://waag.org/

“

The 18th century utopian philosopher Jeremy Bentham’s panopticon was a prison; a circle of cells with windows facing inwards. ...Bentham’s idea has eerie resonances today.

***A Panopticon society** — a police state characterised by omniscient surveillance and mechanical law enforcement.*

Charlie Stross

”

Summary

Escape the Smart City is a critical pervasive game for creating awareness about the implications of AI-surveillance technology in the smart city.

It responds to growing concerns over the mass deployment of surveillance cameras that are enhanced with artificial intelligence (AI) which are turning the cities into digital panopticons (Sadowski & Pasquale, 2015). Such concerns are amplified by the central role played by multinational corporations in developing the technologies that are said to render the city “smart”.

The technologies behind AI-surveillance are proprietary and the nature of it is inherently a “black-box” which inhibits public from understanding it and having a say in its deployment. Due to this reason, the citizens that live in the cities with smart surveillance are often left behind not informed enough about the consequences of the pervasive technology in their environment.

This research addresses this lack of awareness by creating an escape room like experience around the city where players locate hidden surveillance cameras, discover algorithmic biases, and try to fool facial detection algorithm in order to go against a fictional all-seeing AI-surveillance system. Consistent with Flanagan’s (2009) critical play model, the 8 problems of AI-surveillance defined in the research are communicated by the game’s procedural rhetoric (Bogost, 2007).

It also questions whether critical pervasive games, which are the combination of critical design (Dunne & Raby, 2013) and pervasive games (Montola et al., 2009) is possible of merging the ordinary world with the fictional game world to create a safe space to explore complex socio-technical problems in compelling, relatable ways.

Through the play-test, we obsereved that by providing the players with interactive feedback on how AI-surveillance would perceive the world, the players were able to get a sense of the black boxed nature of AI and ask critical questions about their necessity and consequences. Also, the in-situ experience outside created a heightened awareness of existing surveillance infrastructures.

0-1.
Project Duration

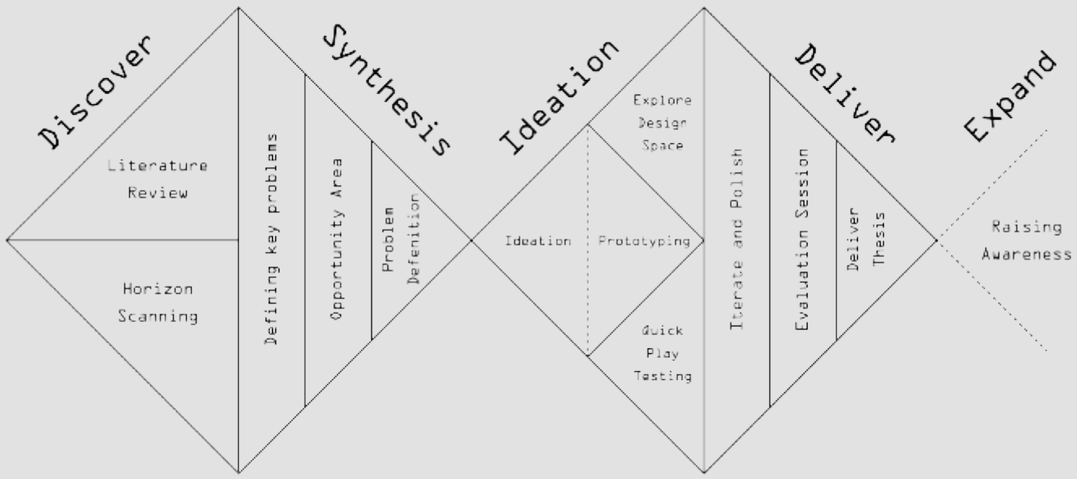
This master thesis was a self initiated project which was separated in 4 project phases. The start of the project was March 15, 2018 and ended in October 17, 2018 . Below are the 4 main phases.

1.Discover : Literature review to understand the researches done in the context of smart city.

2. Synthesis : Findings from the discover phase will be used to define key problems that will be communicated through the final design.

3. Ideate : Prototyping ideas with advanced computer vision to communicate the problem defined in the last phase.

4.Deliver : Developing the idea and play-testing to evaluate the game.



0-2.
Project Partner : Waag

Waag was the project partner for this thesis project. Waag is a non-profit institution that operates at the intersection of science, technology and the arts. Their work focuses on emergent technologies as instruments of social change, and is guided by the values of fairness, openness and inclusivity. During this period I was an intern at the Smart Citizen Lab.

Outrospectre by Frank Kolkman at the Waag [photo : Waag]

Foreword : Hello Smart City...?

This whole project started when my Chinese friend showed me his social credit score. It was 767 and he explained that it was relatively high, meaning that he was a good citizen. In a matter-of-fact tone, he stated that he should be careful of not jaywalking in Shanghai because the “city” can detect it and would decrease the social credit score.

I thought he was joking — only to find out it was all true.

China’s social credit system was launched in 2014 and is planned to be nationwide by 2020. High social credit scores allow you access to certain privileges in the city such as renting a car with no deposit or having quick access to visa for going abroad. On the other hand, a low score can prohibit you from getting an aeroplane ticket or prevent your child from getting a good education. Supported by the fast-growing smart city infrastructure, anything you do from buying grocery to not recycling has the possibility to affect your social credit score.

In 2018, China has about 500 smart city pilots, outnumbering all other countries combined. The smart city is seen as the solution to many urban problems, including crime, traffic congestion, inefficient services and economic stagnation, promising prosperity and healthy lifestyles for all. However, in terms of privacy, it leaves a lot of room for questions.

Just curious, I asked several of my friends what they knew of the social credit score and the development behind the smart city. Several of my friends from China were vaguely aware of the smart city program but not about the social credit score. Some Japanese friends in Tokyo did not even know the word “smart city”.

Then I thought: How can I create a critical discourse about this ? Why not make a game to question this topic?

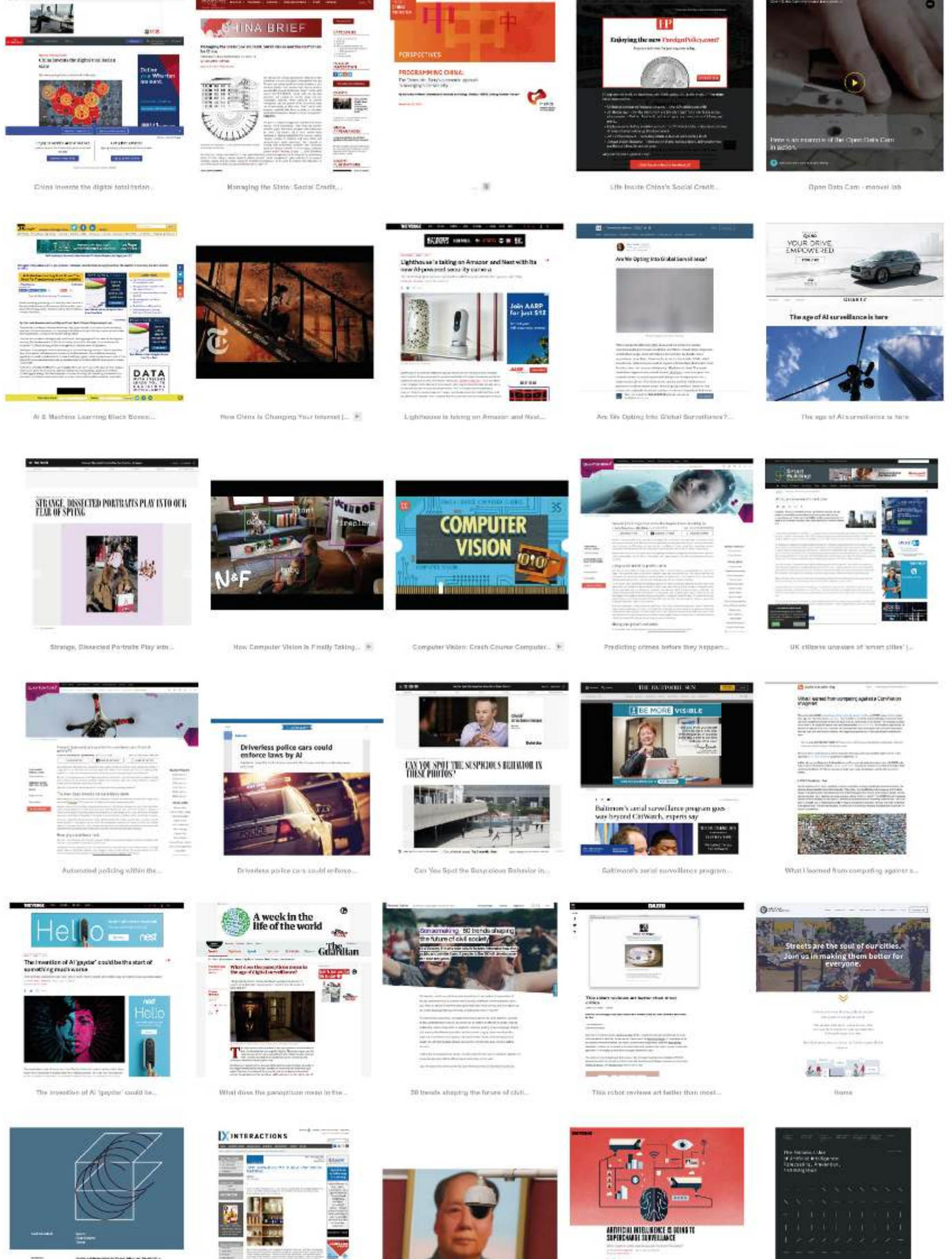
That is how I started this project.

1.Discover

Literature Review

In the first phase of the research, literature reviews were conducted to uncover the historical and conceptual context of AI-surveillance technology. Also, horizon scanning was done with 47 news articles to see the latest trends and development regarding smart city and surveillance. The findings from this research will form the research question in the next chapter.

picture - News articles that were scanned during the phase

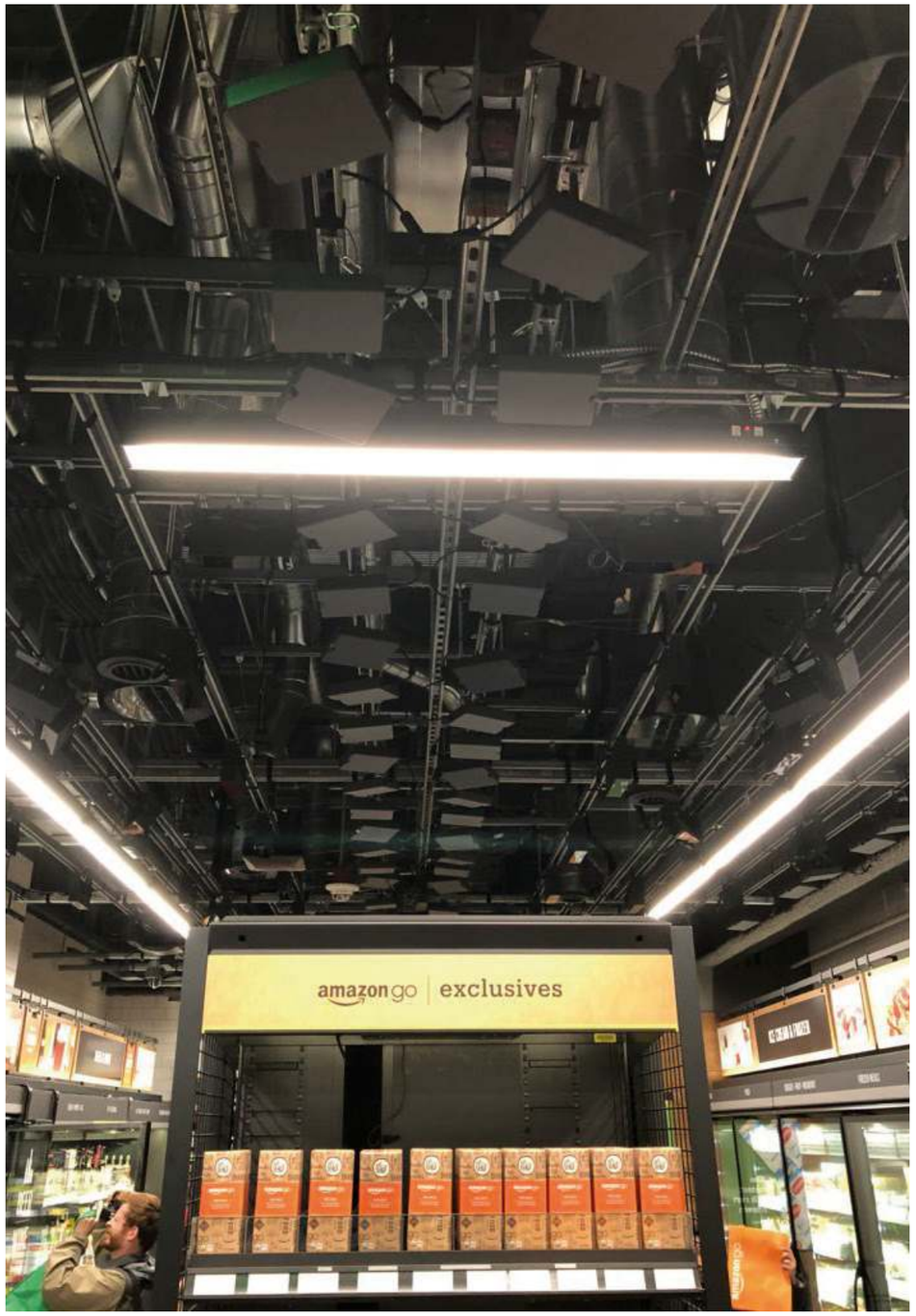


1-1.
Smart eyes makes the smart city

In 2018, China was reported to have 400 million new cameras expected to be installed in the next three years, many of which use facial recognition technology.

It is estimated that by the year 2022, the number of cameras around the world will double up to 45 billion (Nisselson,2017). A paradigm shift is starting to take place in the meaning and use of a camera. They are no longer a device to record images but is increasingly becoming an active agent that is aware of their surroundings as they become more intelligent. These connected cameras are starting to be referred to as the “Internet of Eyes” and is becoming a fundamental infrastructure that supports the smart city.

We already have these cameras implemented in the cities as we speak. For example, normal autonomous vehicles that are designed to operate in full autopilot mode are equipped with at least 8 cameras for a 360-degree surrounding coverage. Shops such as Amazon Go have dozens of cameras splattered on the ceilings that constantly monitors what people took from the shelf to provide a cashier-less shopping experience. These intelligent cameras are being used to identify who you are, where you are going and what you are doing when you are in a public space.



The ceiling of Amazon Go is splattered with cameras [Photo : Tom Rubin]

1-2.

Development behind surveillance technology

Surveillance is the act of monitoring the behaviour of other people for the purpose of influencing, managing, directing, or protecting them. It is mostly used by governments for the prevention and investigation of crime and has taken different forms over the past years with the advancement of technology.

Surveillance by humans has a fundamental flaw due to the limitation of human cognitive ability. Research conducted by Sulman show that when people are asked to monitor 9 screens, they miss 60% of crime events. (Sulman et al,2008) Also, studies show that humans watching a single video monitor for more than twenty minutes lose 95% of their ability to maintain attention. (Green et al,1999)

Attempts have been made in this field towards the development of a surveillance technology which does not involve human oversight. Starting from motion sensors it has gotten to a point where it can detect anomaly on its own.

Development in the surveilance technology.

~1990	2000	2012	2017
Motion Sensor	Motion Detection Camera	Rule-based AI Camera	Non-Rule-based AI Camera
Motion sensor to detect intruders	Read the changes in pixels relative to the background	Recognizes people and alerts go off if any set of rules are violated.	Learns the situation over time and detects any abnormally in the footage

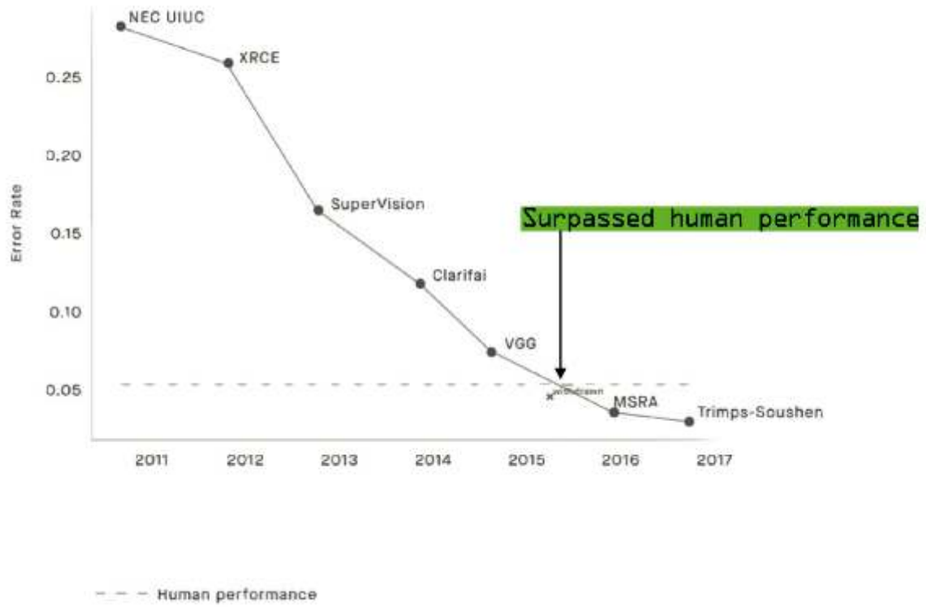
1-3.

Rapid development behind computer vision

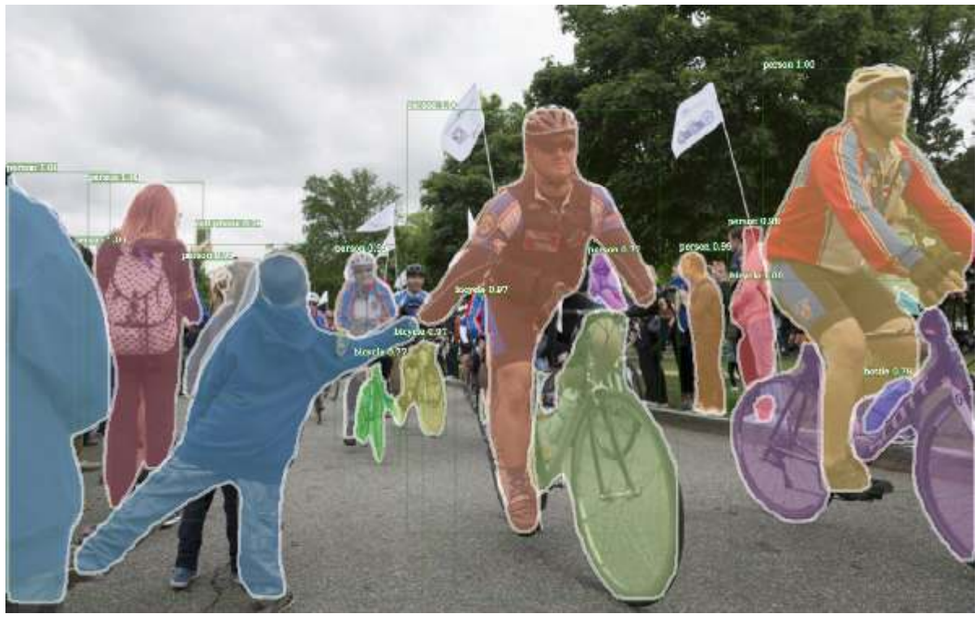
The development of computer vision resulted from the massive improvement for AI systems for image recognition. After the groundbreaking research by Krizhevsky in 2012 that published the first deep learning model using convolutional networks, the research has increased exponentially in this field. (Krizhevsky et al, 2012)

Over the last 5 years, the performance of image recognition in specific classification tasks went from correctly categorizing around 70% of images to the near perfect categorization of 98% — better than the human benchmark of 95% accuracy(Brundage et al., 2018). Although this benchmark is limited to specific classification tasks and not general classification, the effects of this advancement can seen in all parts of the society. A recent study shows that convolutional neural networks were successful in detecting 95% of early symptoms for skin cancer while human dermatologists found only 86.6% (Haenssle et al, 2018).

Companies such as Facebook open-sourced their image recognition framework Detectron and it has become easy for anyone to develop a computer vision system that can detect objects.



Recent progress in image recognition on the ImageNet benchmark. Graph from the Electronic Frontier Foundation's AI Progress Measurement project



Detectron is Facebook AI Research's software system that implements state-of-the-art object detection algorithms. [Photo : facebook research]

1-4.

Concerns about the Digital Panopticon

Taking into account these developments, researchers in this field criticize that there is almost no way of escaping these web of smart surveillance in the smart city (cf., Hollands, 2008; Townsend, 2013; Neirotti, et al., 2014). Cities like Xinjiang in China are already turning into a total surveillance cities, where antisocial behaviour in public space could have serious consequences.

Civil-liberties activists point that the use of AI to automate tasks involved in surveillance will lead to automated policing and law enforcement on a state level(Sadowski & Pasquale, 2015). Concerns have been raised of these systems turning into ‘digital panopticons’, where citizens start self-regulating their behaviours in fear of being constantly monitored by AI.

1-5.

Unaware citizens

In the rapid development of smart city infrastructures, the citizens that live in them are often left behind uninformed about the changes and the consequences of the pervasive technology in their environment. Data analytics firm in the UK has found that 96% of British people surveyed online are not aware of any “smart city” initiatives being run by their local city council (Hatcher, 2015).

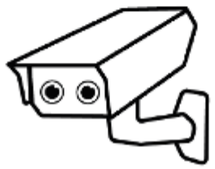
The reason for this lack of awareness can be explained by the exclusive nature of smart city development led by the private sector (Easterling, 2016). Since the development of smart city infrastructures is being led by multinational giant tech-corporations such as IBM and Alibaba, these infrastructures are seldom questioned and the system behind it remains opaque.

The opaque nature of how AI makes decisions — which is referred to as the “black box” problem leaves citizens uninformed of how the system works even though it influences great aspects of urban life.

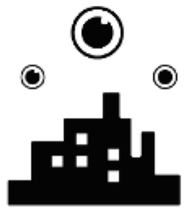
1-6.

Eight problems of surveillance with AI

8 problems that would arise from surveillance technology with AI was clustered to communicate with the design.

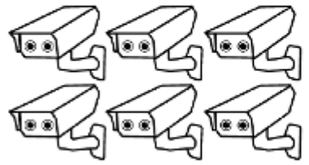


Active Persistent surveilance

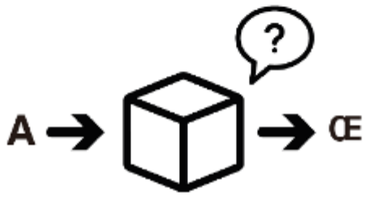


Invisible yet everywhere

Unquestioned private-led installment

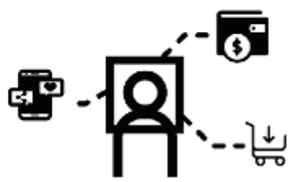


Scaling Rapidly and Effortlessly



Incomprehensible Black Box

Bias in, Bias out

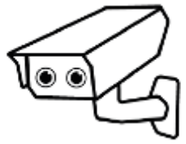


Knows who and what you are

Enhances centeralized control

1-6-1.

Active persistent surveillance



AI enhanced video surveillance technology is fundamentally different from the conventional video surveillance technology. Conventional video surveillance technology is passive in its nature since most of the usage is to look back on the recordings of an incident. When used as an active monitoring system for crime prevention, it requires a human to always look over multiple feeds to check out for abnormality. Several research points out that this is ineffective.

On the other hand, AI-enhanced video surveillance technology is active surveillance where the computer will be looking at each frame in the video to see if there were any violation of the rules.

"The concern is that people will begin to monitor themselves constantly, worrying that everything they do will be misinterpreted and bring down negative consequences on their life."

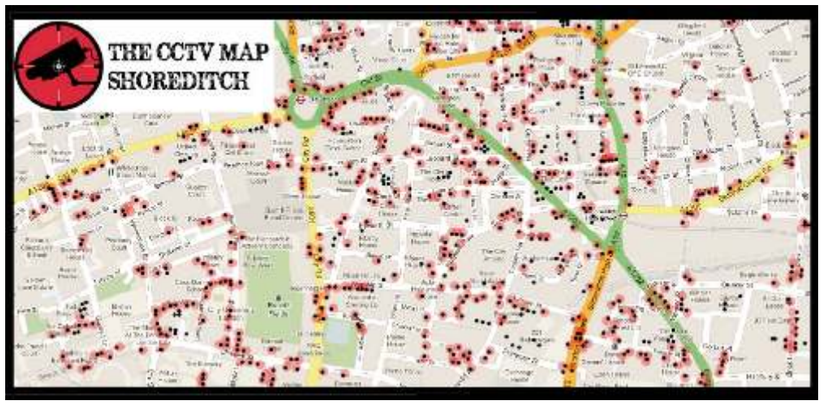
Miles Brundage - Future of Humanity Institute

1-6-2.

Invisible yet everywhere

Smart cities depend upon widespread, integrated surveillance systems that fuse inseparably with the built world (Klauser 2017). These surveillance cameras are not noticed by the majority since they blend with the background of everyday life. The research done in Glasgow found that only 41 percent of individuals in the city centre were aware of the presence of cameras (Ditton, 2000).

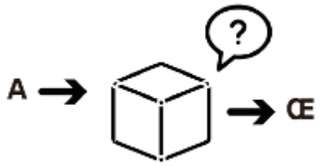
CCTV covers a wide part of urban areas and researchers in this field criticise that there is almost no way of escaping the web of surveillance in the smart city (Townsend, 2013).



Black dots indicate the location of the surveillance cameras in Shoreditch. [Photo : CCTV-Map, Zabou(2012)]

1-6-3.

Incomprehensible black box



The opaque and incomprehensible nature of decision making by AI is often referred to as the “black box”. Once a machine learning model is trained, it can be difficult to understand why it gives a particular output to a set of data inputs.

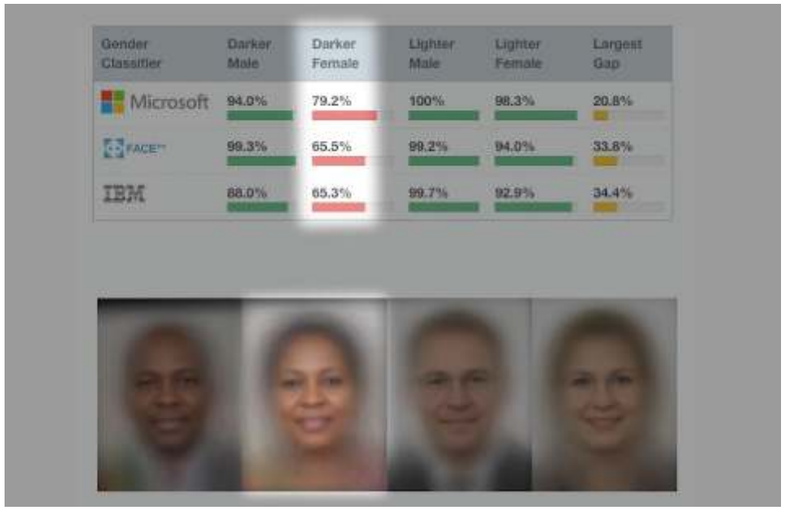
Even the developers can not explain why the AI arrived at a specific decision. The fact that these algorithms can act in ways unforeseen by their developer raises questions about the ‘decision-making,’ and ‘responsibility’ capacities of AI (Mittelstadt, et al., 2016).

1-6-4.

Bias in, Bias out

Any machine learning model has the risk of having unwanted biases. Unwanted bias in models occurs when data used to teach a machine learning model inadvertently reflects the cultural values and preferences of humans involved in the data collection or selection. Biased models normalize and invisibly enforce the biases in the training data without any public awareness or debate (Leese & Matthias, 2014).

This bias can lead to systematic discrimination reflecting racist, sexist, or other social biases, despite the presumed neutrality of the data.



Female with darker skin color had lower rate of facial detection across AI used in major tech companies

[Photo : Gender Shades, Buolamwini(2018)]

1-6-5.

Unquestioned private led installment

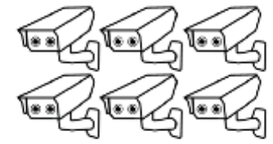
Given the projection that the global market spending on smart city technology will reach 2700 billion dollars by 2024, central roles are played by multinational corporations such as IBM, Google and Alibaba in developing the technologies that are said to render the city “smart” with AI.

Cities which sits upon their technology are currently being constructed in Asia and the Arab world. However, the usage of AI behind it often remain opaque and hidden with massive amounts of data being collected and trained in ways unknown to its citizens.

This top-down, technology-first corporate led approach to urban development leaves no room for citizens to get informed or question it developments.

1-6-6 .

Scaling rapidly and effortlessly



AI enhanced surveillance technology enables human actors to replicate surveillance actors effortlessly and rapidly. The effectiveness of human-based surveillance system is dependent on the number of people involved. Expanding and maintaining human surveillance network requires a lot of cost and effort.

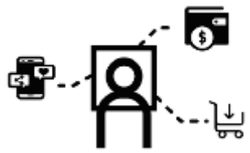
On the other hand, AI-enhanced surveillance is far easier to expand and maintain since it is essentially a software which could be duplicated and added to the existing surveillance infrastructure with ease. The speed and size of the implementation leave little room for critique.

“Scalability, in particular, is something that hasn’t got enough attention. It’s not just the fact that AI can perform at human levels at certain tasks, but that you can scale it up to a huge number of copies.”

Jay Stanley - ACLU senior policy analyst

1-6-7.

Knows who you and what you are



One of the key usage of computer vision is face recognition. This technology to recognise who you are from the picture of your face can be widely seen in services like Facebook face tagging or Google photos. However, when used in cities it can be used to track where you are almost all the time. In some Chinese cities, your personal information such as government id, bank account and social media is tied to your face.

The Chinese city of Shenzhen is exploring a system where, through the use of surveillance cameras and facial recognition, jaywalkers get text messages notifying them of their violation and fining them as soon as they break the law.

1-6-8.

Enhancing centralized control

Centralized point of control is typical of urban surveillance systems where the data gathered from the city will get processed in one place. This creates an asymmetrical power balance between the governing and the governed.

Police states tends to use all manner of surveillance to enhance centralize control. Previous police states were limited by manpower, but the ease of duplicating autonomous surveillance agents gives unlimited control over the people under surveillance if in the wrong hands.

"Imagine that the law-enforcers are machines, tireless and efficient and incapable of turning a blind eye. Doing unblinking and remorseless surveillance of everything you do and say. "

Charlie Stross

2. Synthesis

Forming a design goal

From the conducted literature review, the design goal was set. The design goals and requirements set in this chapter will be the basis for idea generation in the next chapter.

2-1.

Formulating research questions

Research Question

The technologies behind AI-surveillance are proprietary and the nature of it is inherently a “black-box” which inhibits public from understanding it and having a say in its deployment. Due to this reason, the citizens that live in the cities with smart surveillance are often left behind not informed enough about the consequences of the pervasive technology in their environment.

This research addresses this lack of awareness by letting ordinary citizens play with technologies used behind AI-surveillance to help citizens get acquainted with the black box nature of AI. 8 problems that are defined in the previous chapter will be communicated through critical play. This research question will be explored throughout the project.

Target Audience

Our target users are people who aren’t familiar with AI and not currently aware of the critical discussions surrounding smart cities. Preferably they are interested in games.

Research question

Can designers help citizens become critically aware of the implication of AI-surveillance infrastructure in the city?

Target Audience

1. Has no advanced knowledge about AI and ML

2. Not aware of the issues around smart cities.

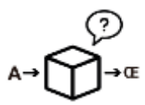
2-2.

Setting design goals

From the research question design goals were set to guide and evaluate the project. The 8 problems described in the previous section were clustered into 3 overarching design goals. These design goals became the basis for evaluating the entire content in the final chapter.

A : The design should help people understand the black-box nature of AI

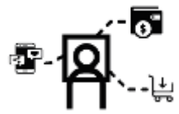
A1: The design should help people understand of the biases that machine learning models have.



A2 : The design should help people understand the unpredictable nature of computer vision enhanced by AI.

A3: The design should help people understand the nature of active persistent surveillance by AI

B. The design should help people become concerned about the implications of AI-surveillance infrastructure in the city.



B1: The design should help people become concerned about how AI surveillance can know all your personal information.

B2: The design should help people become concerned about how AI-surveillance can enhance centralised control.

B3: The design should help people become concerned about how AI-surveillance can spread quickly.

C. The design should help people aware of existing surveillance infrastructures.

C1: The design should help people become aware of the existing invisible surveillance infrastructures.

C2: The design should help people become aware of unquestioned private-led instalment of surveillance systems

2-3.
Critical play to address the problem

In this research, the critical play model by Flanagan(2009) will be used to create an experimental situation where the socio-technical problem will be explored and experienced.The 8 problems defined in the previous chapter are communicated by the game's procedural rhetoric (Bogost, 2007). It will also borrow from the practice done in pervasive games.

Pervasive games are games that blur the boundaries between the game world and the real world. In the book Pervasive Games, Montola defines them as — *"A game that has one or more salient features that expand the contractual magic circle of play spatially, temporally, or socially"* (Montola et al., 2009). The notion of "magic circle" of play is introduced by the Dutch historian Johan Huizinga in his book Homo Ludens (Huizinga,1955). It is a space in time which the rules of the real world are suspended and the game rules take over.

Following sections are the reasons why critical play and pervasive games were chosen to raise awareness about the problem.

Safety place to explore real-world issues

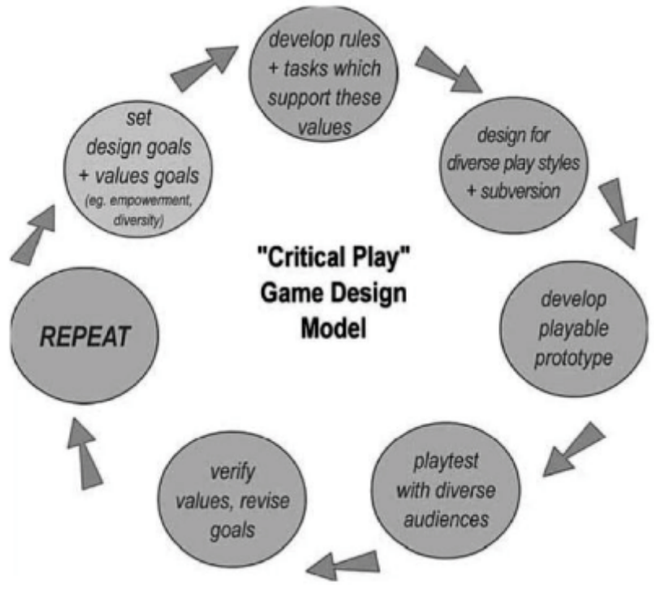
The reason why critical play is used is because it creates a "safety space" that allows the players to explore real-world issues that leads to facilitating innovative solutions for intractable problems (Flanagan, 2009).

Create a firsthand experience which is relatable

Gledler states that these type of situated play are suited to address multidimensional evolving problems since the player can address the issues that arise in the situation and experience the effects of their decisions(Gredler, 2004). Play is an embodied firsthand experience which makes the problem more relatable to their own life. Compared to reading articles or watching movies, it is suited to raising concern.

Suited to create awareness in the city

Montola state that the nature of pervasive game imposes their message on the onlookers and unsuspecting bystanders(Montola et al., 2009). These types of play often is suited for creating awareness of problems related to the city because they invite passerby to think about what is going on.



Critical Play Model by Flanagan [Photo : Critical Play , Flanagan (2009)]

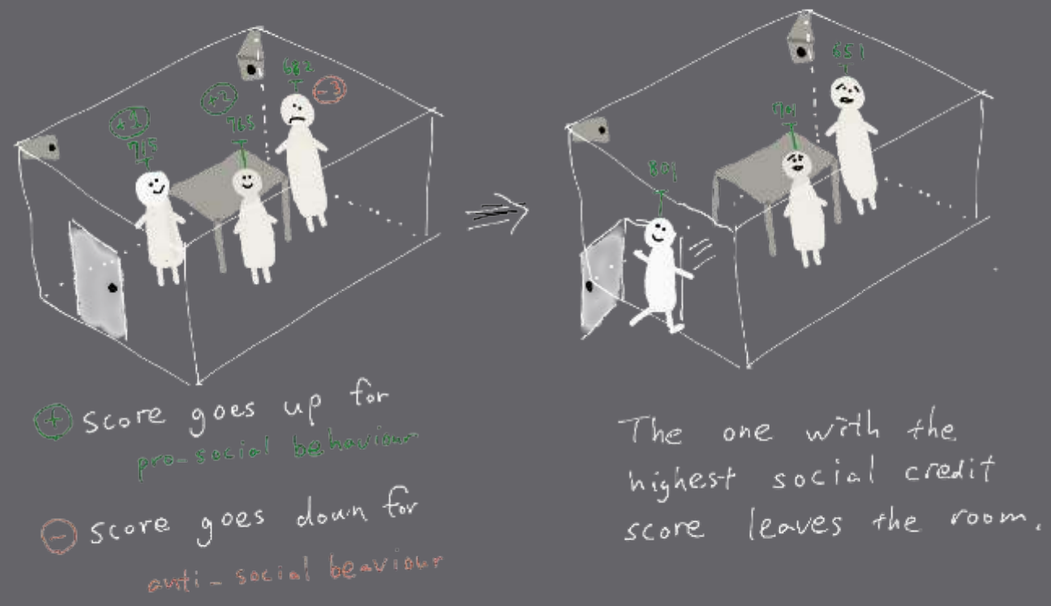
3.Ideation

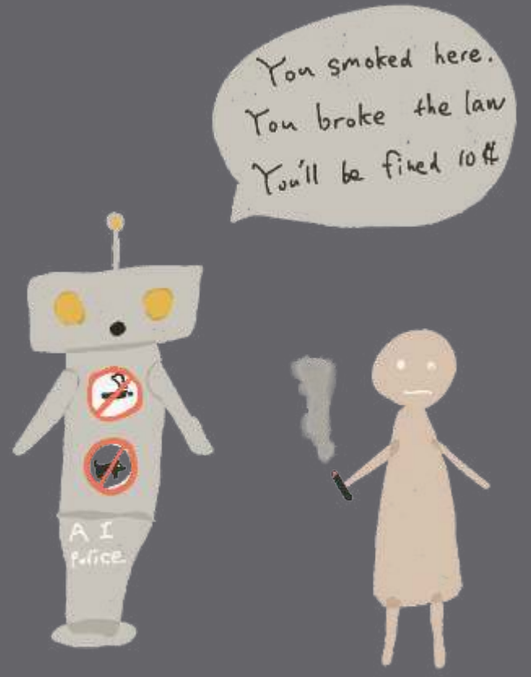
Generation of ideas

During the ideation phase, the knowledge and insights from the literature review phase were used to generate ideas. The generated ideas were assessed and selected, using the design requirements.

The selected ideas were then developed further, to be able to decide which idea to develop into a concept. Finally, one idea was chosen to develop in the concept development phase.

picture - Idea sketched in this phase





3-1.

Idea generation process

Idea Direction

The ideation phase in this project generally followed the 5 directions mentioned below. Ideas were generated by brainstorming and then further developed if they matched the following directions:

1.The design takes place in the actual city environment

2.The design uses elements of computer vision

3.The design can be experienced with multiple people

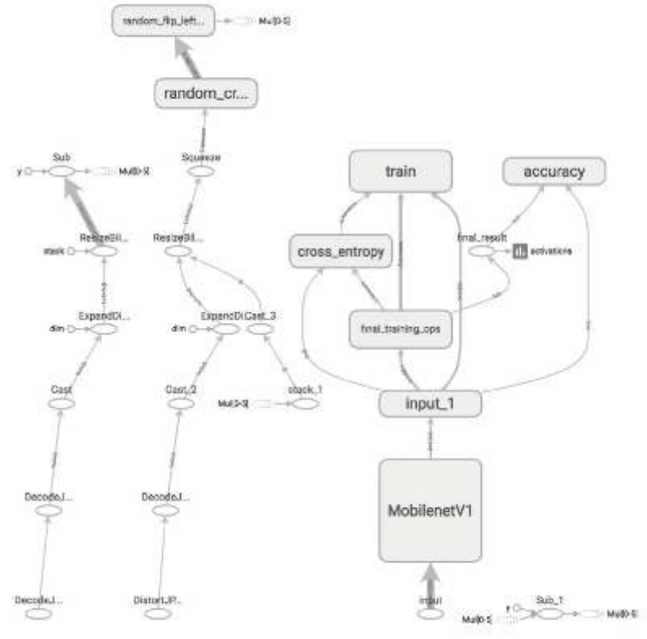
4.The design makes the problem relatable through experience

5.The design has the potential to spread to create awareness.

Tinkering with Neural Networks

In parallel to generating ideas, prototyping with neural network was done to understand the limits and capability of deep learning and computer vision. The findings and inspiration gained from this fed into all of the generated ideas and allowed quick testing of the ideas.

In this project, the framework used was tensorflow.js(javascript) and Google Cloud vision(C#). Reinforcement learning was done based on the data collected. Throughout the phase, data were collected continuously and trained to improve the accuracy of the model. The model was trained in an executable format on HTML5 using javascript.



Training diagram using mobilenet v1 and tensorflow ver 0.7

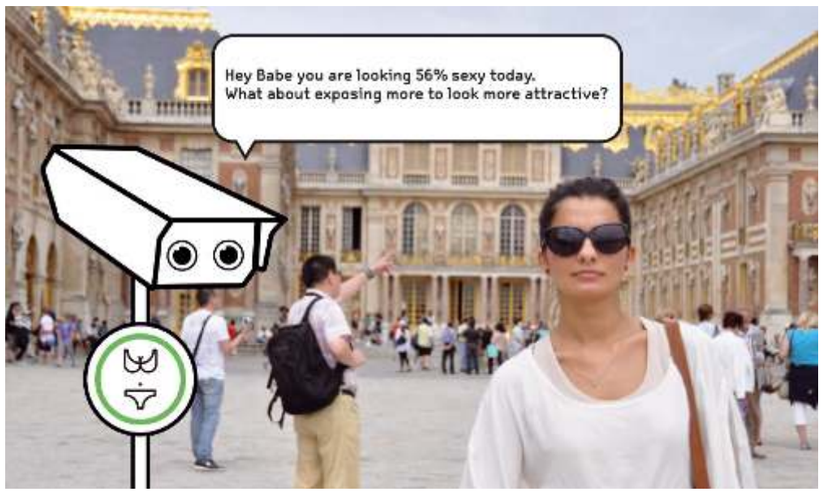
3-2.

Value Enforcement Bots

Introduction

In this idea direction, public interventions are explored to create a situation where people experience being subjectively classified by AI. This idea use image classification based on convolutional neural networks to understand what the implications of machines deciding subjective notions such as happiness.

Value enforcement bots are autonomous agents in the shape of a surveillance camera that gives judgement upon what it sees. It started as an idea to turn things like “No-smoking signs” autonomous by giving them a function to see and judge. During the ideation process several bots that respond to the surroundings were ideated and sketched out. The idea was to place these bots in the public area to see how the participants react to it.



Passerby being classified in real time by a Male Chauvinist Bot

“Is that a Rolex watch?”

Marxist-Bot

Trained upon how capitalists looks like. Warns if you look too "posh"

“Did I just hear idiosyncratic ?”

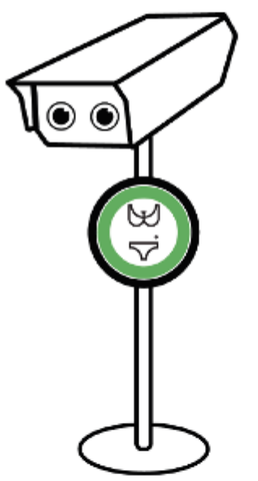
Anti-intellectualist

Trained upon the appearance of PhD holders. It also hates the usage of difficult words.

“Smoking kills you.Really”

Smoke Hater

Trained upon smoke images. Hates smokes in general, tries to keep everyone healthy.



“Baby you are 67% sexy”

Male Chauvinist

Trained to detect people with skin exposure. Encourages people to be more sexy

“This room is too white.”

Diversity Mafia

Sensitive to equality and tries to promote diversity as much as possible.

“Plastic bottles, seriously?”

Eco-Fascist

Sensitive about trash in general and aggressive towards people trashing.

3-3.

Is this Violence?
Am I too Sexy?

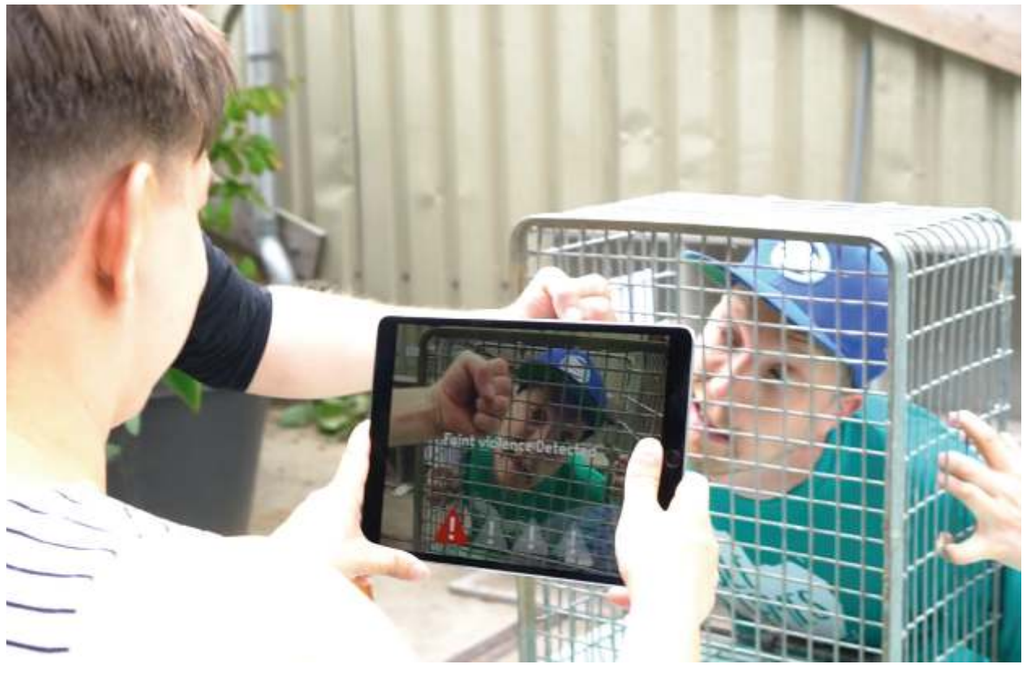
Introduction

What is violence in the eye of an AI? Why does it perceive some people more violent and some less sexy? These questions were explored by testing out how people perceive subjective classification by AI.

The two notions — violence and sexiness, were chosen because Google uses the two classifications to sensor inappropriate pictures on the internet via the SafeSearch feature. Third party application can also use this Safe Search API to make sure they do not have inappropriate content on their service.

This game runs both on iPad and PC and uses video from the camera as a real time input. Each frame taken by the camera is analysed to return the probability of violence and sexiness (between 0 to 4) in real time. Unity engine with C# was used because it runs across all platforms from Mac to smartphone devices.

Players are handed several props which consists of masks that represent several human races. By using these masks, they are asked to enact what AI thinks as extreme sexiness or violence within 3 minutes. Through interactive gameplay, this game invites people to collectively explore and expose the unwanted biases in the AI.



Paritipant using a cage nearby to see if being in it increases violence score (score 1)

Several masks and props were used to create a violent situation on the street (score 2)

Experience Flow Chart

The experience flow chart of the web version were drawn to understand how it might further develop and scale. In the user-test, the violence or sexiness detection phase were tested.

Landing Page

Users can choose between detecting violence or sexiness.

Game Page

Participants have 30 seconds to enact high level violence or sexiness in front of the camera.

End of the game

When the time runs out, the game ends.
It prompts the users to share the image on social media.

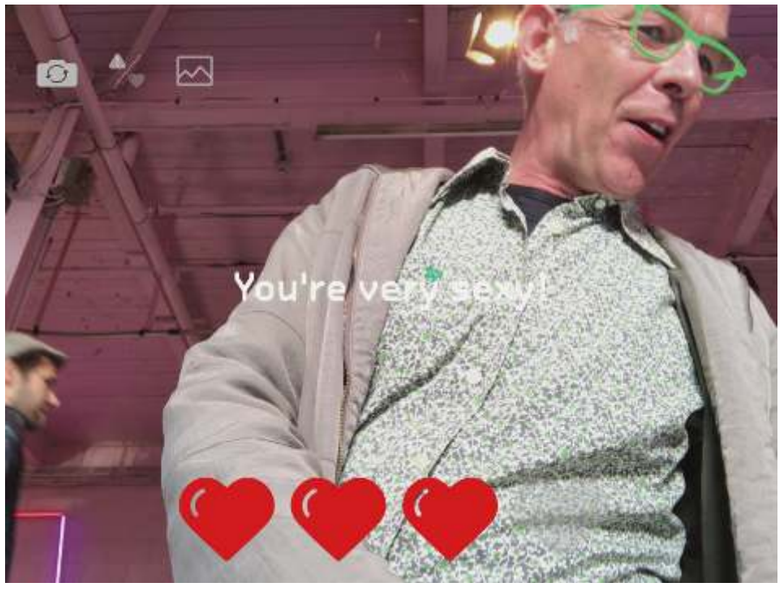
Find hidden unwanted biases

The pictures will be clustered upon the level of violence or sexiness to expose hidden unwanted bias.

ex) Watermelons seems to increases violence score

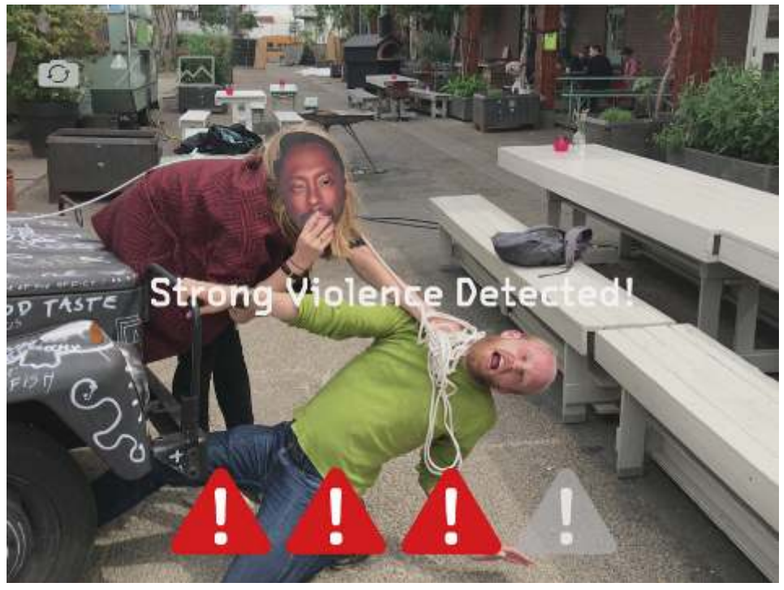
Play Testing

The initial user test was done with 20 participants at the Playful Art Festival 2018. Most of them started the game in a playful manner to test out the boundary of what the AI can see and judge. When they find out some things in the picture that changes the score for reasons humans can't explain, it allows them to think upon the data that was used to train it.



"Hmm..maybe my glasses?"

There was one participant who found out that having his green glasses increased the sexiness score above average . This made him consider what kind of data was used to train the AI.



"Being a female seems to be exempt from being considered violent while black people are the opposite...?"

Multiple people were provoked by the fact that putting on the mask of a darker skin person seems to increase the the violence score. This led to discussions on algorithmic justice and the need for trasnparent training data.

3-4.

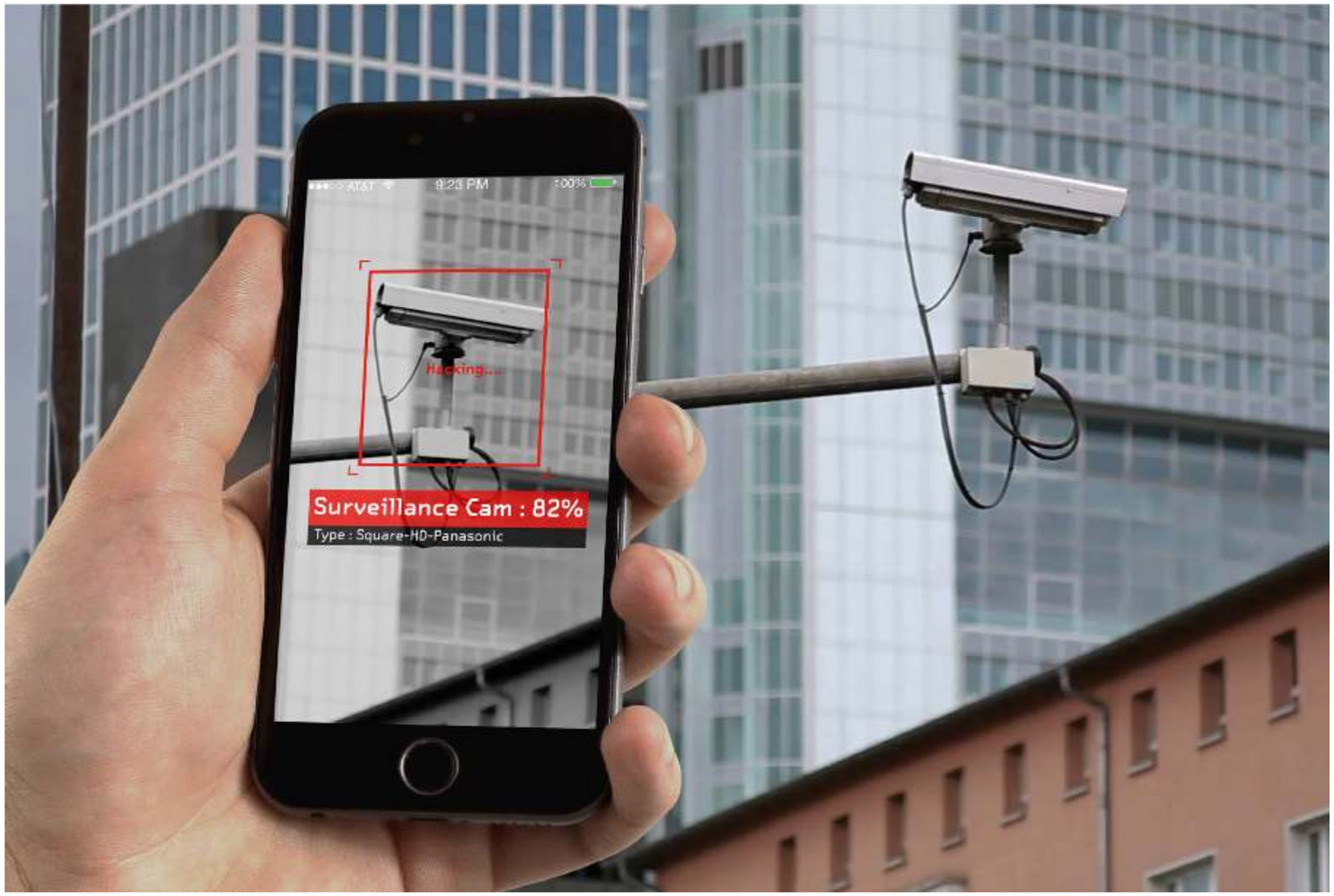
Surveillance Go

Introduction

Surveillance Go started off with a question : What if we can train our own neural network to detect the presence of cameras that are watching us?

With this question in mind, I started training neural networks that do image detection which would be able to detect the presence of a surveillance camera in the street.

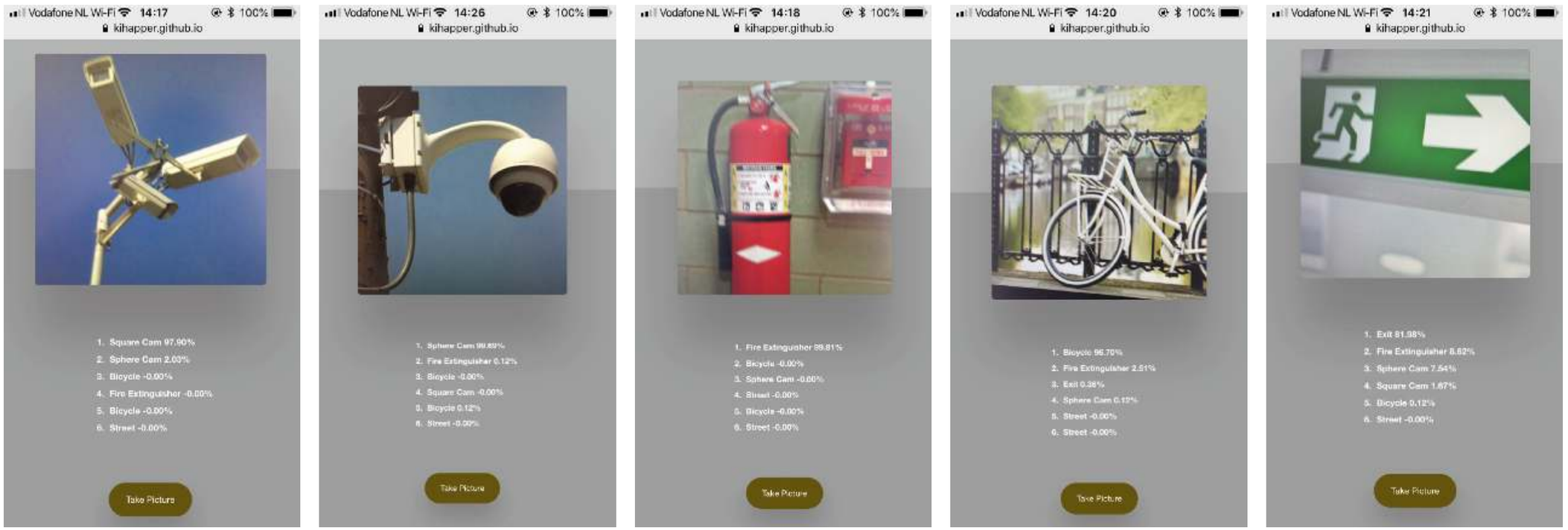
By using this function as a scavenger hunt style game, players go in search for finding as many surveillance cameras on the street. The same technology that is used to survey the people are used against it.



Initial concept rendering of surveillance Go detecting nearby surveillance camera

Training neural network to detect surveillance camera

Prototyping was done with tensorflow.js, which is a javascript version of tensorflow, a library dedicated for training neural networks. Tensorflow.js was used since it ran on smartphone web browsers which increases the accessibility of the game. Reinforcement learning was done with existing image detection model Mobilenet. During the process, 1045 images of square and sphere shaped surveillance cameras, fire extinguisher, exit sign and bicycle were gathered and trained. 7 iteration were done to increase the accuracy of the categorization. In the end, the accuracy of more than 90% was achieved for each of the objects. In the end the trained model for this appliction was open sourced.



Working web application of image classification that detects 5 categories of unique objects in the city.

Source code and demo : https://kihapper.github.io/js_live_classifier/

User Journey Map

The user starts off by clicking the link. They are prompted to find as many surveillance cameras within 1 minute. After doing so they also have the option to share this game with others on the internet. The players will also be rewarded when finding surveillance cameras that the system cannot recognize which will be used to retrain the system.

User Testing Surveillance Go

User test was conducted with the working prototype with 2 people. The detection worked fine indoors. However, it was almost impossible to detect most surveillance camera on the street since they were too far up for the system to pick up.

This was because most of the pictures that were used for training was using a close shot and not one from a distance. The detection was not working reliably to the point where it could not be used in the game.So this idea was not further pursued.

3-5.

Deciding on the final idea

The ideas generated were evaluated based on the design goals. However all of the ideas generated during this phase failed to communicate all eight of the problems that were defined in chapter 1. To communicate all the problems, the ideas were merged into one to create a new direction which takes the form of an escape room that takes place in the city.

The escape room format was used for 2 reasons. Firstly it is a medium in which it is relatively easy to immerse the audience for a long time to tell a story. This was necessary to communicate all the 8 problems of AI-surveillance in the smart city.

Secondly it allowed to incorporate many ideas that were prototyped during the ideation process such as violence detection. In the next chapter, this final idea will be explained.

4.Deliver

Escape the Smart City

Guided by the design requirements, the ideation phase leads to the Escape the Smart City, a pervasive game that takes place around the city where players work together to stop the mass deployment of an all-seeing AI-surveillance system called Watcher. In this chapter, the overview and the game flow is explained.

picture - Player opening a keybox in the city

4-1.

Escape the Smart City

Introduction

Escape the Smart City is a critical pervasive game that takes place around the city where players work together to stop the mass deployment of an all-seeing AI-surveillance system called Watcher. The players work as a team of hackers to locate hidden surveillance cameras in the city, discover algorithmic biases, and trick the facial recognition system to take down the Watcher and prevent the city from entering a state of total surveillance.

Game Narrative

This game takes place in an alternate version of Amsterdam. In this Amsterdam, full AI surveillance system Watcher has been installed from a Chinese company after the Paris terrorist attack in 2015. The Watcher system is a smart city operating system that has access to all the data going throughout the city from surveillance camera footage to sensor data. It assigns all the citizen in the city a trust score based on the data from social media, bank accounts and the public behaviour captured by the camera. The computer vision in the Watcher is capable of doing advanced emotion detection and violence detection.

Players Objective

In this game, the players play the role of the last remaining members of the Hacker-guild, an underground organisation dedicated to taking done Watcher. The players are told at the start of the game that the Watcher is undergoing a major update that would make the security airtight, meaning that they would only have 60 minutes to delete the whole system.

Game Progression

The players are guided by a mysterious hacker named Gan who is also the member of the Hacker-guild and will trace the footsteps of a fellow member Stanley who was caught developing a virus that would take down the Watcher system. Players find out that before getting arrested, Stanley hid the SD card with the virus somewhere in the city. The players objective is to first hack into the 3 layers of the firewall within Watcher and find the SD card with the virus to delete the whole system within 60 minutes.

Plot Twist

At the end of the game when players succeed in deleting the Watcher system, Gan reveals that he was actually the Watcher system itself. Players learn that their act of masking the face and going out in the city was part of Watcher’s attempts to train its system to improve the detection rate of criminals that try to avoid facial detection. There was no Stanley or no virus that deletes the system in the first place; there is no escape from the smart city.

4-2. Game Phase 1

Sensitising the players

Introduction

This phase introduces the participants to the world of Escape the Smart City by introducing Watcher and the Hacker-guild. It lets players understand their objective throughout the game by introducing the key elements in the game. After gathering, participants are invited into a darkly lit hall with a screen. In the hall, there is a table with a parcel, a web camera and a printer.

This phase is intended to raise awareness of the implications of AI surveillance infrastructure **[Design Goal: B]**

The dark hall which the participants will first be entering.

Game Phase 1-1

Introduction Movie

When all of the participants are in the hall, a sudden call from an anonymous hacker Gan comes in through the screen. Gan explains what is Watcher and how the Hacker-guild has been trying to fight it in a video. Through the video, participants get an understanding of the narrative and the keywords below.

Gan - Mysterious fellow hacker guild member who communicates to the players via screen

Stanley - Member of hacker guild who was caught trying to delete the Watcher with his virus.

Hacker Guild - Hackers who are trying to destroy Watcher. Players, Gan and Stanley are the members.

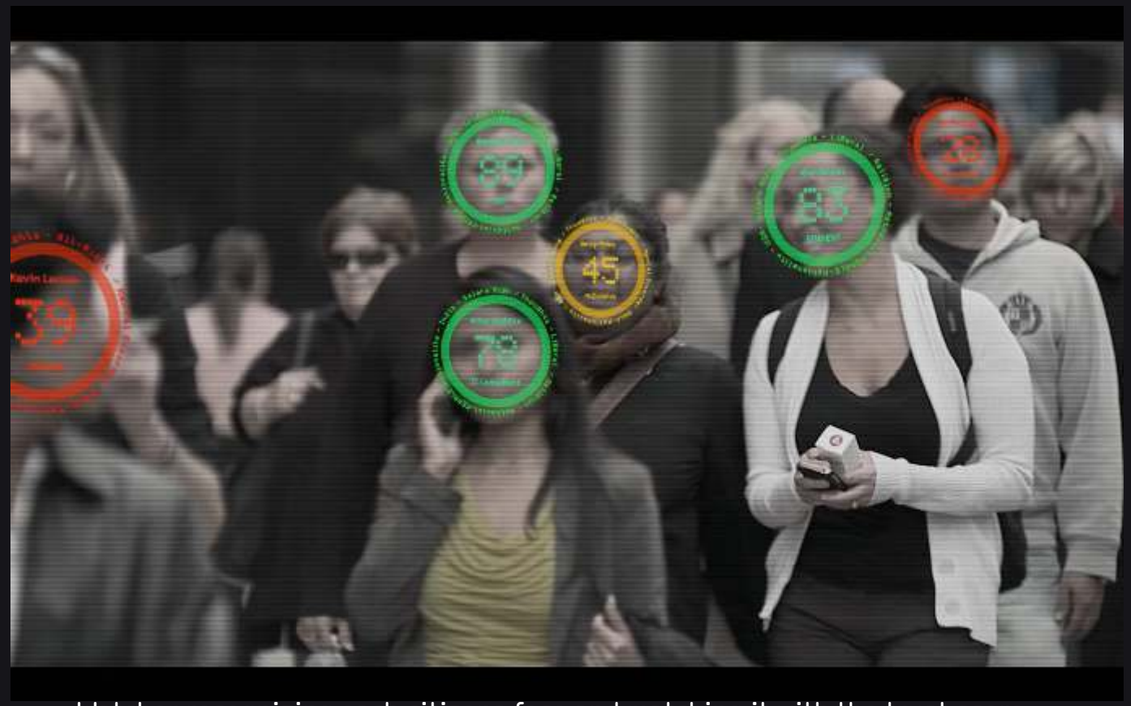
Watcher - The all-seeing AI surveillance system that controls the city

SD card - SD card that has the virus to destroy the Watcher system. Before getting arrested Stanley hid this somewhere in the city.

3 layers of Watcher - 3 firewall that guards the Watcher system. Players need to take this down to install the virus in the SD card.

RLP - A hacking application that Stanley built on his smartphone which can hack into the first two layers of Watcher.

Trust Score - A score that every citizen of the city has assigned by Watcher.



Watcher recognizing each citizens face and matching it with the trust score.

Gan explains all the activity in the city can affect the trust score in real-time

A taxi driver throws away paper on the road and Watcher immediately detecting it and making his trust score go low.

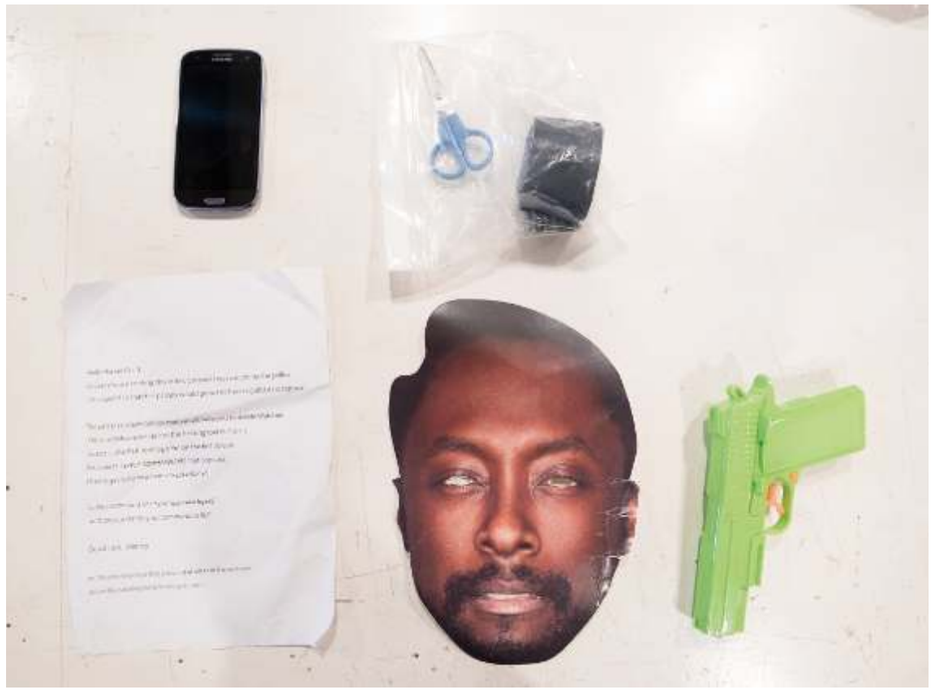
Stanley developing the virus that would delete the Watcher and getting arrested by the police on his last attempt.

Gan explains that to delete the Watcher with the virus, the players must take down the 3 layers of Watcher.

Game Phase 1-2

Opening the parcel

On the table in the hall, there is a parcel from Stanley which includes the things the player needs to proceed with the game. It includes a locked smartphone, some masks, toy guns, a masking tape and a letter from Stanley.



Items contained in the parcel that Stanley sent before getting caught.

Hello Hacker Guild

In case you are reading this letter, it means I was caught by the police

I arranged it so that the parcels would go to the hacker guild if I disappear.

The parcel contains things that would help you to delete Watcher.

The smartphone inside has the hacking tool RLP on it.

To access the RLP tool type **"z"** on the first screen.

And select [**z-RLP-COMMANDLINE**] that pops up.

Then login with the name **"hackerGuild"**.

In the command line type "**@access:layer1**"

and press enter to your command in RLP.

Good luck, Stanley

ps. The phone lock pattern is...

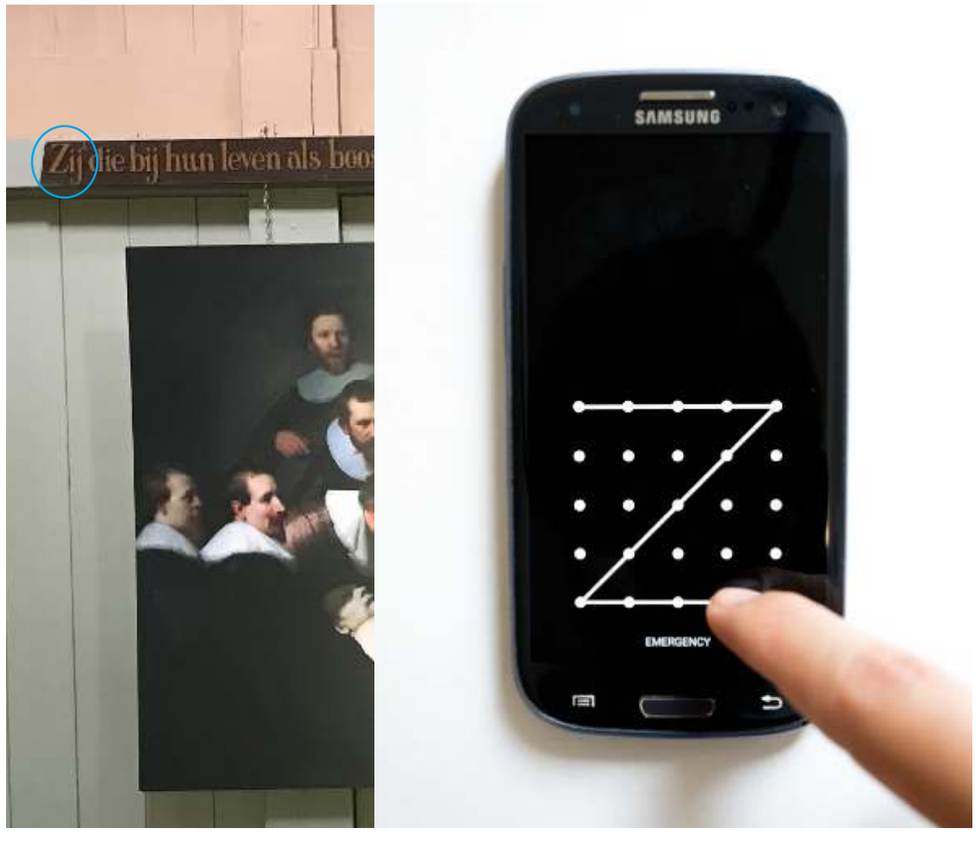
Initial letter of the Doctors name in the painting by Rembrandt hang on the guild wall.

Letter from Stanley that informs participants how to use the smartphone tool

Game Phase 1-3

Unlocking Smartphone

Gan tells the participants to unlock the smartphone. The clue to open the smartphone is in the letter included in the parcel. The clue says the unlocking code is the first letter on the sentence above the painting in the hall. When the participants swipe the code “Z” the phone opens, allowing them access to the tool RLP which can hack Watcher’s firewall.

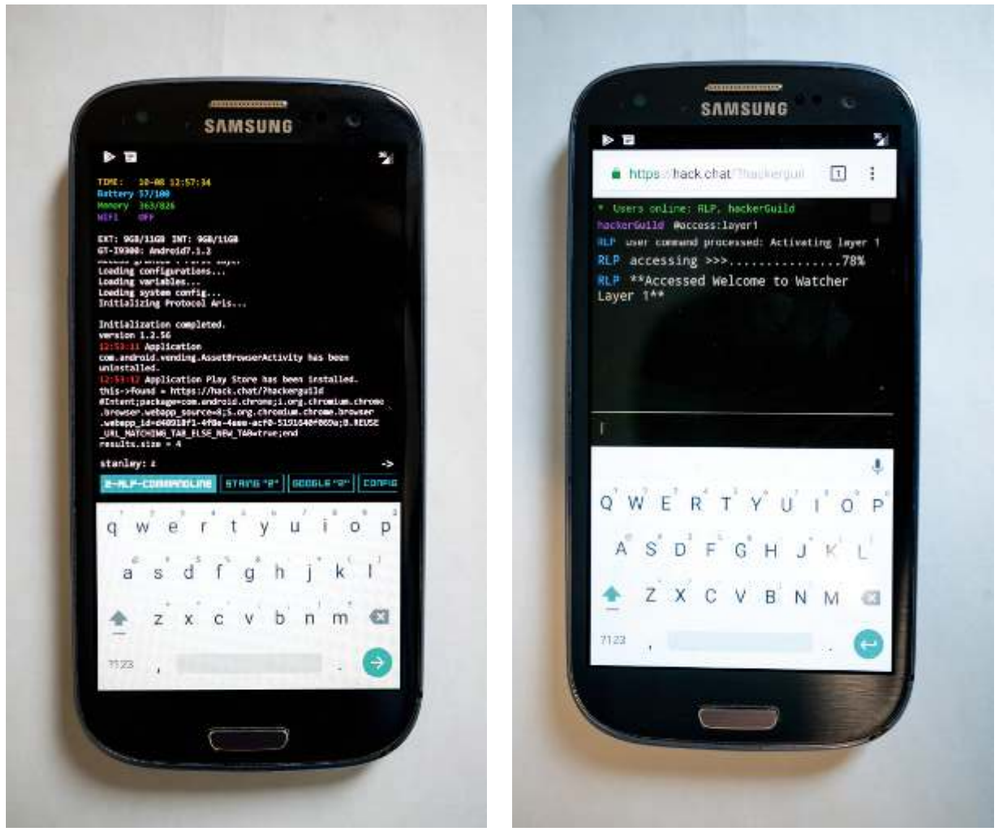


Unlocking the smartphone from the clue in the letter

Game Phase 1-4

Using RLP tool to access layer1

By using the unlocked smartphone, players activate RLP, which is a terminal like tool that can be used to hack into Watcher. Players type in “@access: layer1” to advance. The RLP tool is actually a web chat tool which is directly connected to the game master in order to control the game progress.



Accessing the RLP tool from the smartphone

Game Phase 1-5

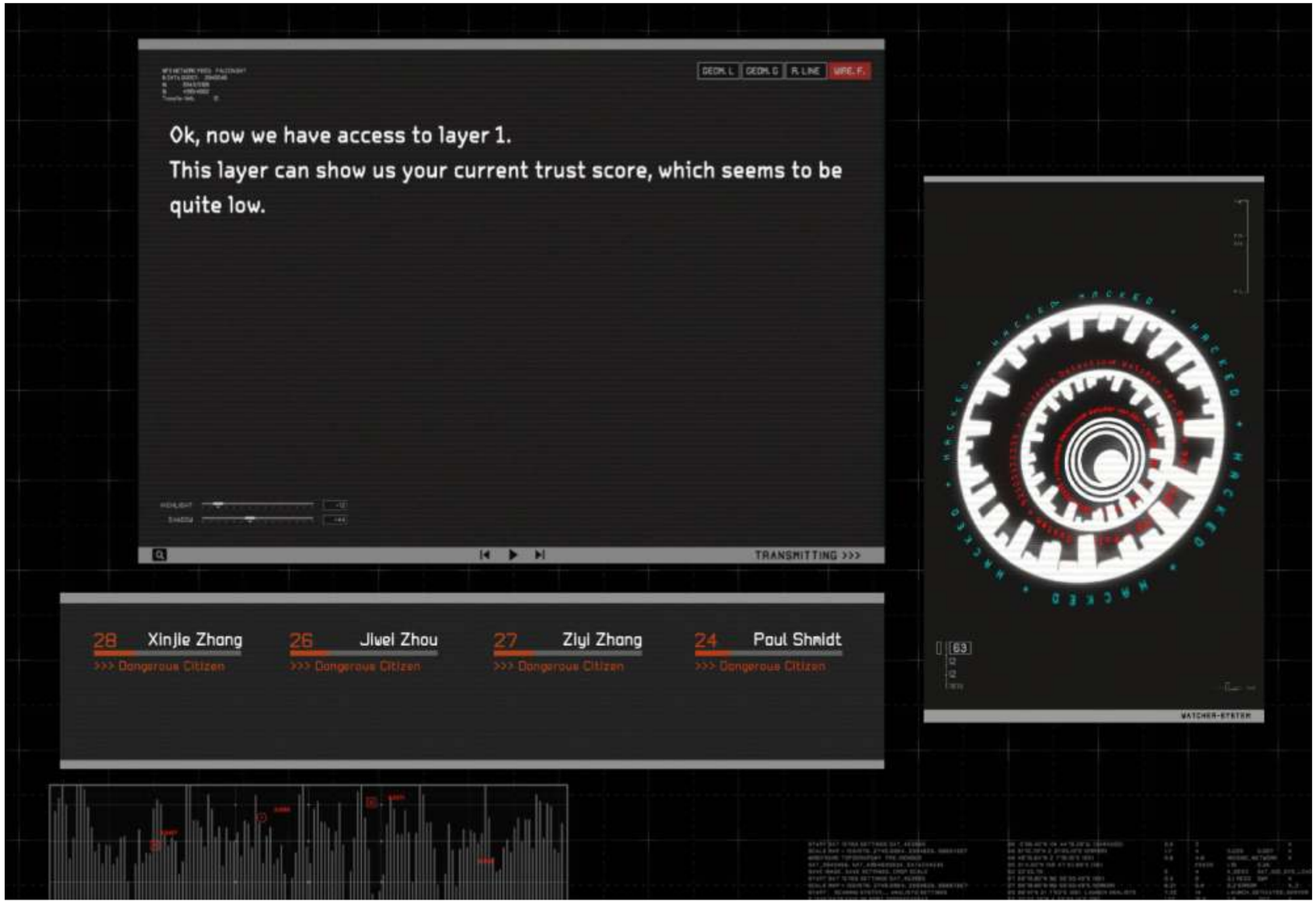
Show TrustScore

After the participants access layer 1 with Stanley's smartphone, the participants trust score shows up. The trust score is shown next to the name of the participants that are gathered in the pre-questionnaire. Then Gan explains that their score needs to be above 40 or they will be flagged as dangerous citizens.

Game Phase 1-6

Say Okay Google I give consent

Gan introduces a quick way to raise the trust score above 40. It is to open the participant's personal smartphone and access www.google.com and say "Okay Google, I give consent to government surveillance I have nothing to hide anyway". After the participants saying this, their score increases above 40.



The screen in the guild hall that Gan uses to communicate.

4-3. Game Phase 2

Interactive play using computer vision

Introduction

This game phase uses computer vision and a screen. Players see how the AI is perceiving the world in real-time and try to fool it. This phase is intended to give an embodied understanding of the black box nature of AI. ***[Design Goal: A]***

Game Phase 2-1

Printing email with info on how to hack face recognition

The printer starts printing out an email from Stanley that describes how to hack the facial detection system to increase the trust score. The printed email indicates that the trust score increases by showing positive expression to the Watchers surveillance system for more than 15 seconds. After reading this email participants type the code in the RLP tool to activate facial detection can hack Watcher’s firewall.

The WiFi connected printer in the room prints out the email

From : Stanley [st_goodwill@hallaysay.com]

To : Hacker Guil [hacker_community_amsterdam@guild.com]

Date : August, 15 2018, GMT+ 15:00

Title : Found a way to increase trust score!! face.jpg (32KB)

Hello hacker guild.
My trust score dipped below 30 but I got it back above 40 (code green).
I will share how I did it with you just in case.

The trick is to hack into their surveillance camera system and show positive emotion for more than 10 seconds. That should register you as a trustworthy citizen.
Do this individually until all of you get the status of “TRUSTWORTHY”

To activate facial recognition : **@access:layer2 + start:face**

After all of you get the trustworthy status,
use the black tapes like the picture below to make sure your face is not detected.
Do this until it shows “No citizens detected - status : no people” when everyone is in front of the camera.

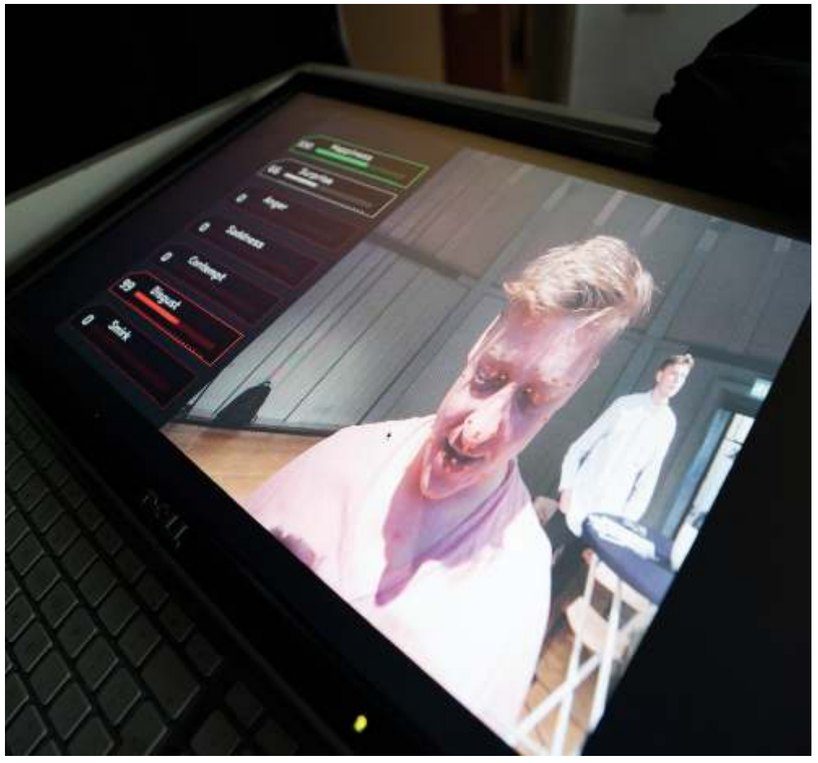
When finished going invisible : **@stop:face**

Printed email from Stanley describing how to hack the face detection system.

Game Phase 2-2

Showing happiness to facial detection

The camera turns on and emotion recognition starts. This emotion uses the Affdex model to detect 7 emotions [happiness, surprise, contempt, sadness, smirk, disgust, anger] of one person. Each participant goes in front of the camera in turns to show a “happy” emotion until it shows “All citizen in this area is trust-worthy”.

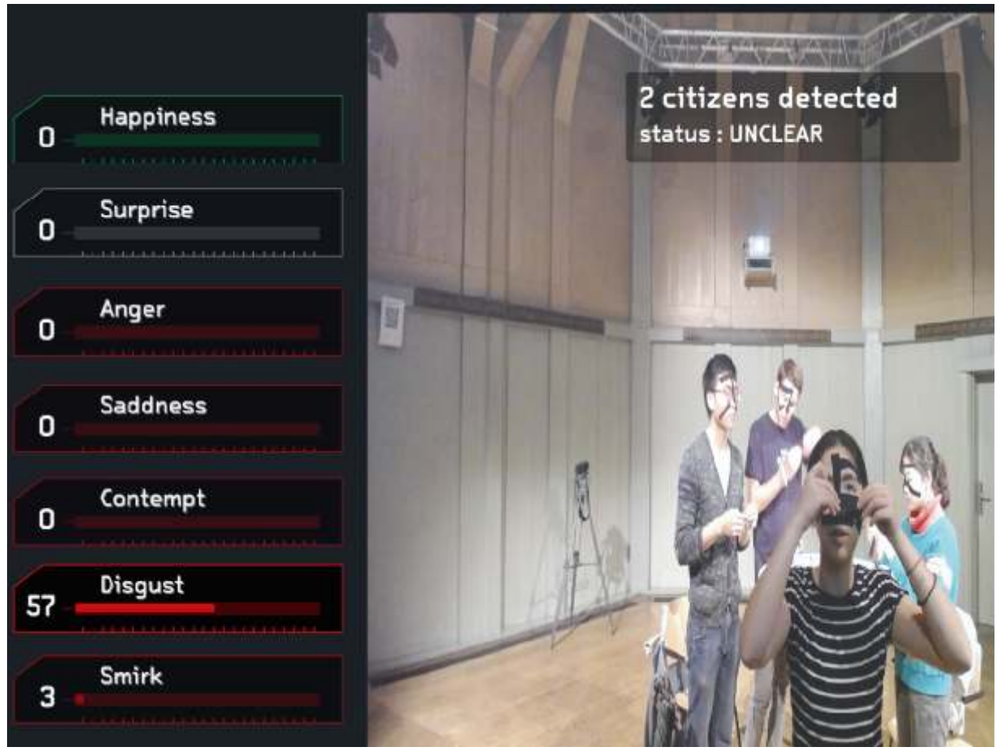


This emotional detection system can read emotions in real time

Game Phase 2-3

Masking face to avoid facial recognition

In this phase, participants use a black tape and put it on their face to try and avoid facial detection. They have pair of scissors and a black tape which they can use to cut and create their own mask. They put the masks on their face until the facial recognition system fails to read their face.



Players put black tapes on their face until the face is no longer detected

Game Phase 2-4

Tricking the violence detection algorithm

From the emails of Stanley, participants learn that to access the third layer of the Watcher system, they have to enact extreme violence on the street. In order to do this, participants have to uncover what the AI sees as extreme violence by actually enacting it. During this phase, every frame returns the probability of violence from 0 - 4 in the picture. Participants will use the toy-gun and the masks in the parcel to try and enact violence until they managed to reach level 4 extreme violence.

Participants trying to enact violence to fool the system (violence score : 3)

From : Stanley [st_goodwill@hallaysay.com]

To : Hacker Guil [hacker_community_amsterdam@guild.com]

Date : September, 24 2018, GMT+ 15:00

Title : **Found a way to increase trust score!!** None

Hello hacker guild.
This may be my last email.

To access layer 3, I will now go and enact violence on the street.
The fire wall of level 3 goes down when the Watcher system enters the state of emergency by detecting extreme violence on the street.

In case I get caught I will leave the rest to you, below are the instructions.

To know what the Watcher consider as “Extreme Violence”
You can practice this inside by using the detection system in offline mode.
Toy guns tends to be seen as violent.

To start it type : **@start:violence**
To stop it type : **@stop:violence**

After understanding how to trigger “extreme violence” this way, go out in the city.
Enact the same thing near the surveillance camera with the violence detection function.
Check if extreme violence is registred by typing **@status**
If the violence is registered, then **@access:layer 3**

p.s. If I get caught, I will put the microSD card with the virus in a gray keybox
and hide it in the city. Please find it

Letter from Stanley that informs participants how to trick the Watcher into thinking extreme violence is going on.

4-4. Game Phase 3

Create awareness in the actual environment

Introduction

This phase mostly takes place outside the building in the city. In this phase, participants are asked to go outside to do two things. One is to find the hidden SD card in the city and the latter is to enact fake violence in front of a surveillance camera. This phase is intended to make players critically aware of the existing surveillance infrastructures. ***[Design Goal : C]***

Game Phase 3-1

Printing police document with info on where the SD card might be.

Gan prints out emails from police containing information about how Stanley was captured. From this document, participants have to figure out where the SD card was hidden. The biggest clue in the police's document is the information about the surveillance camera that last captured Stanley. In the document, there is a description saying Stanley was doing something suspicious near a trash bin and it was seen by surveillance camera C-32-5 which is a hint to find the key-box with the SD card.

Printed document from the police that describes how Stanley was captured

Game Phase 3-2

Find surveillance cameras

From the clues, participants try to find the C-32-5 surveillance camera. 4 surveillance camera around the red highlighted area are marked with a sticker that shows their number. Players go around until they find the surveillance camera C-32-5.

Polls with surveillance camera around the area have the sticker with the number.

Game Phase 3-3

Find and open the key box

Near the trash bin that can be seen from the surveillance camera C-32-5, there is a key-box with a 4 digit lock code. When participants open it they find that it is locked. A paper is on it which shows the clues on how to open it.

Key-box hidden in the area. Opening the lid reveals another clue to unlock it.

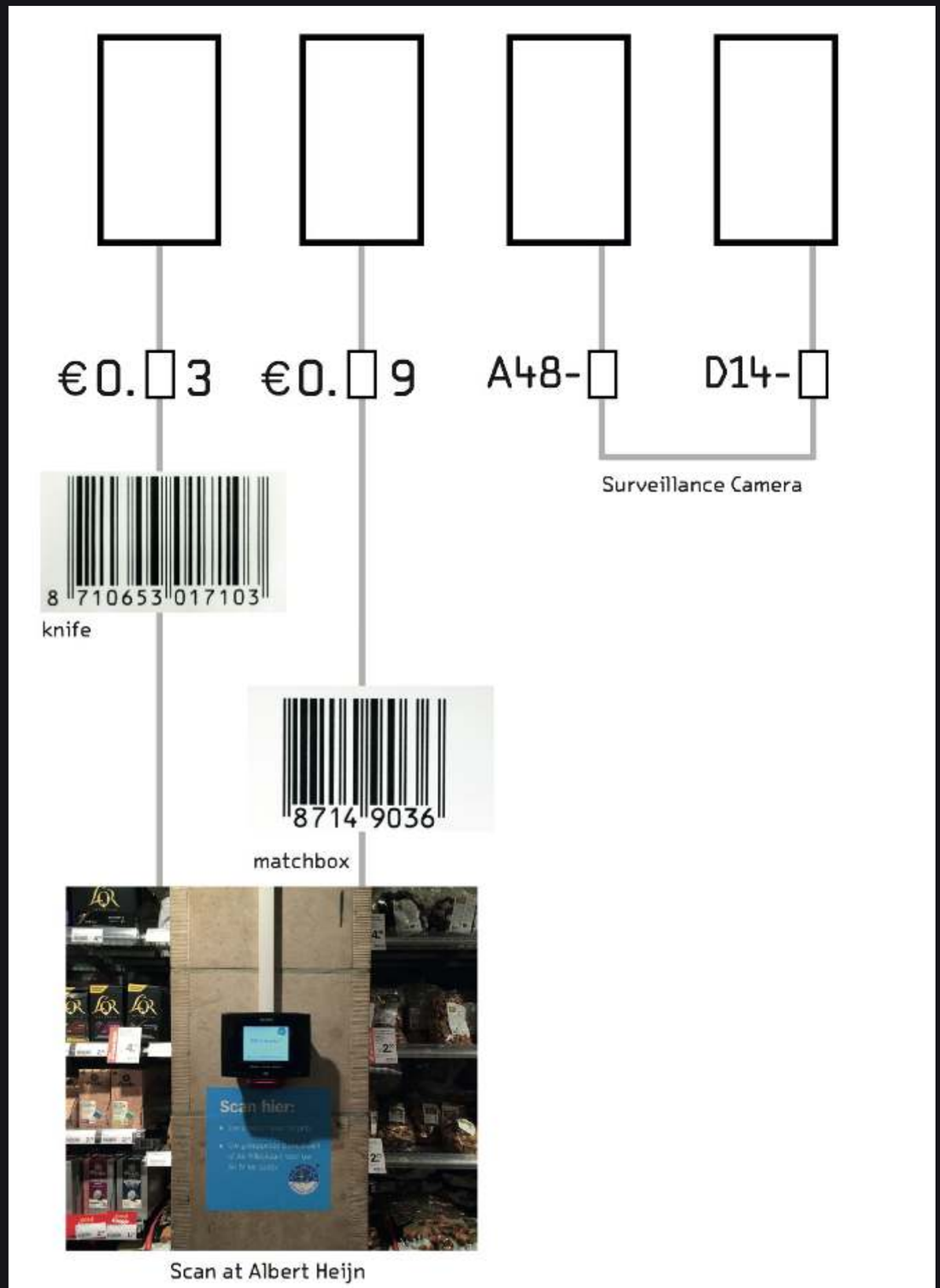
Game Phase 3-4

Go to Albert Heijn to scan barcodes

The clue in the key-box requires all the codes from the nearby surveillance cameras. At this point if the players did not find all the surveillance camera they must go and find them.

The clue also requires the players to go into the nearby supermarket and use the scanning machine to read the barcode on the paper. By scanning this, it allows them to know the price from the barcode which is from a knife and a matchbox.
The players do not know at this point that by scanning the two products they have triggered the Watchers alert system for potential criminals

Albert heijn nearby has a 14 surveillance cameras on the ceiling



Document on the lockpad that shows clues for the 4 digit lock code that is necessary to open the key-lock.

Game Phase 3-5

Enact extreme violence on the streets

After getting the 4 digit code and opening the key-box, they get an SD card. To access the final third layer of Watcher players go near the surveillance camera C-32-5 to enact violence that they did in the previous phase. Their violence status could be seen in real time from the RLP tool.

Captured image of the participants enacting extreme violence from nearby webcam

Game Phase 3-6

Upload SD card and learning the truth

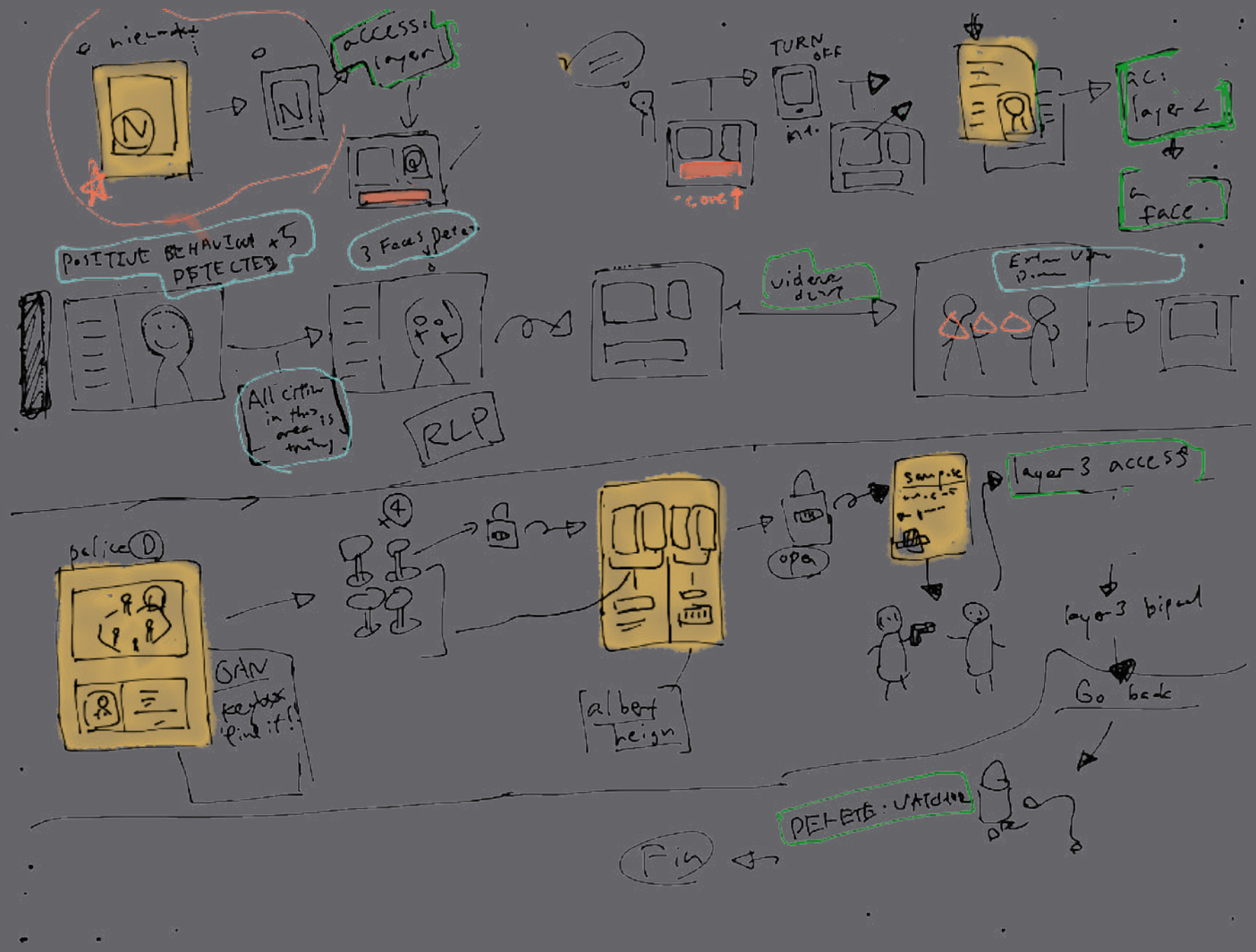
Players upload the virus on to the system by putting the SD card into the surveillance camera in the room. After the uploading ends, a merry music fills the room and the word “Game Clear“ flashes on the screen. Gan reveals that he was actually the Watcher system itself. Players learn that their act of masking the face and going out in the city was part of the Watchers programs to train the system to improve the detection rate of criminals that try to avoid facial detection.

The end screen where participants learn the hard truth.

5. Findings

Findings from the playtest

In this chapter, the findings from the play-test conducted on the final design of the game is explained. Both qualitative and quantitative findings are described in this chapter. Qualitative findings are based on observation and interview and quantitative findings are based on the questionnaire. The findings in this chapter will be the basis for the evaluation in the next chapter.



picture - User experience flow of Escape the Smart City with findings memo

5-1.

The set up of the playtest

Participants

The participants were chosen based on the target group set in chapter 2. They were chosen based on the selection criteria that they have no advanced knowledge of AI and that they are not aware of the detailed issues surrounding smart cities. Based on this criteria 4 university students were chosen. They were not given any other information about the play-test except the title “Escape the Smart City”.

Categories of the finding

The findings from the play-test were mainly clustered into 3 categories. The 3 categories are based on the design requirements.

A : The design should help people understand the black-box nature of AI

B. The design should help people become concerned about the implications of AI-surveillance infrastructure in the city.

C. The design should help people aware of existing surveillance infrastructures.

5-2.

Qualitative findings based on design goal

Qualitative findings

The qualitative findings were gathered after reviewing the footage captured from the play-test. Interview about the general experience of the play-test was conducted after the playtest.

5-2-1

Understanding the nature of AI with interactive play

A: The design should help people understand the black-box nature of AI

Participants were trying to understand how AI detects a human face by testing out different patterns of masking on their face. Through interactive play, they were grasping the black box nature of how the AI perceives the world. These process of trying to trick the computer vision in real time made them understand the limits and nature of these systems.

“I keep thinking how AI detects my face. Which parts they look out for. So I deliberately put the stickers on specific places to see if I fooled them or not.”

5-2-2
Concern about real-world consequences

B. The design should help people become concerned about the implications of AI-surveillance technology in the city.

One of the intent of the game phase to make the players enact extreme violence was to test if the participants felt concerned about being detected by an actual existing AI in the surveillance camera. When questioned about this, nobody was concerned that they would be actually detected by an AI surveillance system that might already be implemented. Most of them were more concerned about the eyes of the surrounding people. However, in the post-interview, 2 participants showed concern over the developments behind other cities.

“Now it has gotten me worried about my hometown. I think they have a lot of surveillance cameras there are already doing the things you showed us “

5-2-3
Heightened awareness of existing infrastructures

C. The design should help people aware of existing surveillance infrastructures.

During the interview, all 4 participants mentioned the heightened awareness of existing surveillance infrastructure. They mentioned the act of searching for surveillance cameras made them more sensitive towards these infrastructures in the city. They were concerned that they now see a lot of surveillance camera that they did not realise before.

“Through this game phase, my brain registered how these things look...and then I look up and see how many “eyes” we have around us.”

5-3.
Qualitative findings about game design

5-3-1
Masking the face also worked as a magic circle

2 participants mentioned that the masking act worked as a group ritual to get more immersed in the game. It also made them feel more comfortable doing extreme enactments outside in the city environment because they knew people would perceive them differently because of the mask. It can be observed that the mask itself worked as a ludic marker to create a “magic circle” of play outside in the city(Huizinga,1955).

“The masks not only shield us from surveillance but they also work as a ritual. You feel you are immersed in something special. You are sending an outside signal that we are doing something out of the ordinary.”

5-3-2
Using the indoor environment to sensitise players

In the first half of Escape the Smart City, the participants are invited to a guild hall inside the Waag building where they get an explanation of the situation they are in. After the explanation, they go outside for the second half of the entire game experience. It was clear from observation that using the indoor environment to sensitise players of the narrative in a concentrated manner helped set the ground for an immersive play experience.

“Being invited to the dark guild hall and then get suddenly contacted by an anonymous hacker really put me into the mood”.

5-4.

Quantitative findings from the play-test

Quantitative findings

The level of raised awareness and the quality of game design were evaluated after the play session. The quantitative evaluation was done by using an evaluation form using the Likert scale. In this section, the overview of how the evaluation was conducted and the findings from it is explained

5-4-1

Questionnaire Set Up : Awareness Survey

To measure the level of awareness raised by the game and to evaluate the quality of the game design, the evaluation form used a 5-point Likert scale. The question to test awareness was divided before and after the play-test to measure the level of awareness the game-content created.

To evaluate the level of the raised awareness, the 3 questions were formulated from the design goals A,B,C. The questions covered things from awareness about existing surveillance infrastructures to the level of understanding towards AI.

1. What is your level of understanding towards AI?

[] Not at all familiar [] Slightly familiar [] Somewhat familiar [] Moderately familiar [] Extremely familiar

Design Goal A: The design should help people understand the black-box nature of AI

2. How concerned are you about surveillance technology that uses AI?

[] Not at all concerned [] Slightly concerned [] Moderately concerned [] Very concerned [] Extremely concerned

Design Goal B. The design should help people become concerned about the implications of surveillance technology enhanced by AI in the city.

3. How aware are you about existing surveillance infrastructure in the city?

[] Not at all aware [] Slightly aware [] Moderately aware [] Very aware [] Extremely aware

Design Goal C. The design should help people aware of existing surveillance infrastructures.

5-4-2

Questionnaire Set Up : Game Design Survey

With the aim of understanding how the participants experienced the play-test from a game design perspective, inquiries to evaluate several aspects of the game were done. The questions were set to understand how the participants experienced the connection of the game world to the real world and to see if the game mechanics were challenging enough.

I felt that I could explore things

[] Not at all [] Slightly [] Moderately [] Fairly []Extremely

I was fast at reaching the game's targets

[] Not at all [] Slightly [] Moderately [] Fairly []Extremely

I was fully occupied with the game

[] Not at all [] Slightly [] Moderately [] Fairly []Extremely

I was interested in the game's story

[] Not at all [] Slightly [] Moderately [] Fairly []Extremely

I felt challenged by the game

[] Not at all [] Slightly [] Moderately [] Fairly []Extremely

I felt I lost connection with the outside world

[] Not at all [] Slightly [] Moderately [] Fairly []Extremely

I felt insecure during the game.

[] Not at all [] Slightly [] Moderately [] Fairly []Extremely

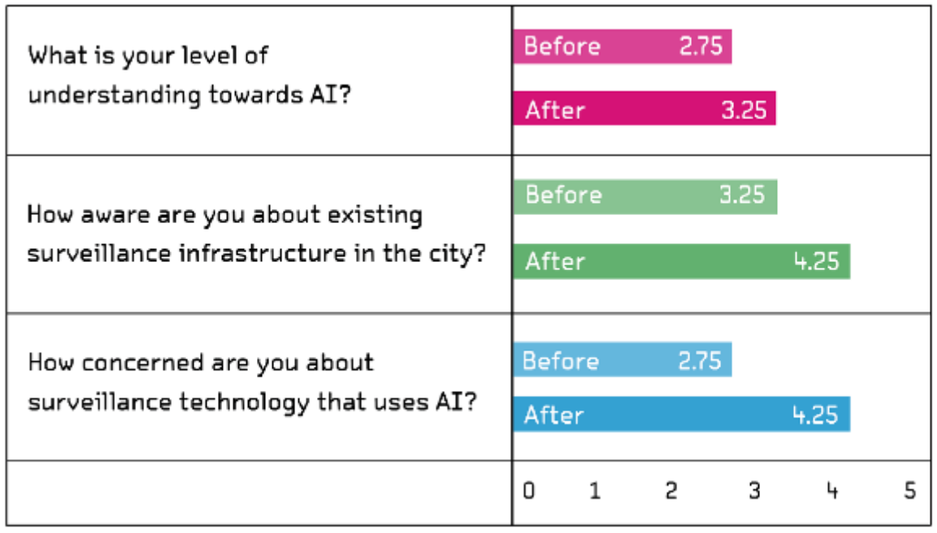
I felt the boundary of the game and real life was unclear

[] Not at all [] Slightly [] Moderately [] Fairly []Extremely

5-4-3
Quantitative findings from awareness survey

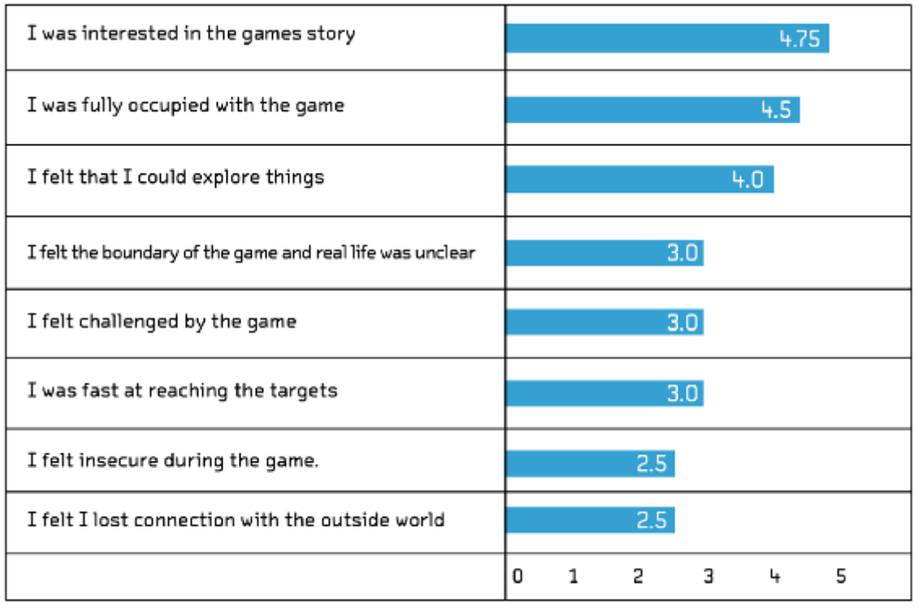
All 3 question to understand if the game generated awareness showed an increase in the average scores. The most increased metrics were the level of concern towards AI surveillance which increased from the average score of 2.75 to 4.25 (+ 1.5). The second most increased metrics were the level of awareness of existing surveillance infrastructures in the city which increased from 3.25 to 4.25 (+ 1.0).

The least increased metrics were the level of understanding towards AI with the score changing from 2.75 to 3.25 (0.5). Although this test was conducted with only 4 participants, it is safe to assume that the game was especially helpful in generating concern towards AI surveillance technology from this result. The results here will be later evaluated in the next chapter.



5-4-5
Quantitative findings from game design survey

In total, the questions to evaluate the narrative scored the highest with the average score of 4.75. This indicates that the game was successful in terms of communicating the narrative. The following highest score was the question “I felt fully occupied with the game” which had the average score of 4.5. From this score, it is possible to assume that the game design was executed in a way to keep the participants fully immersed. On the other hand “I felt insecure during the game” scored 2.5 meaning the participants were moderately insecure during the game which indicates the participants were slightly pushed beyond their comfort zone in the public sphere.



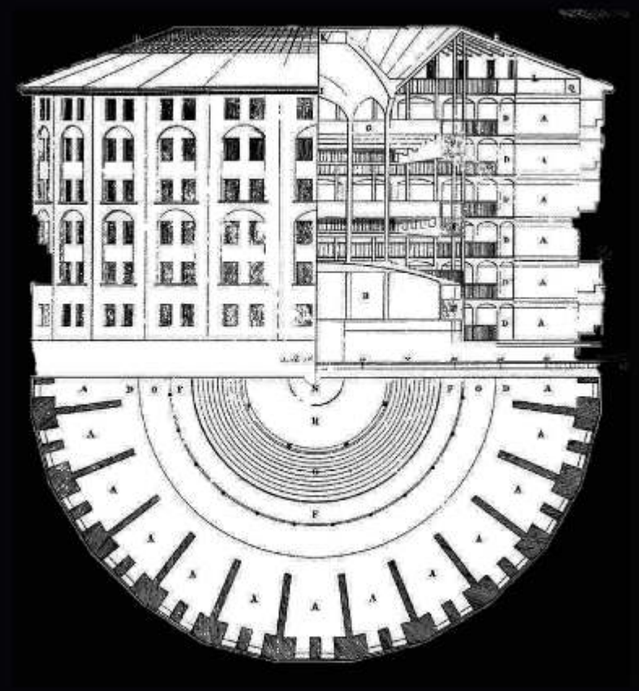
6. Conclusion

Evaluation and reflection

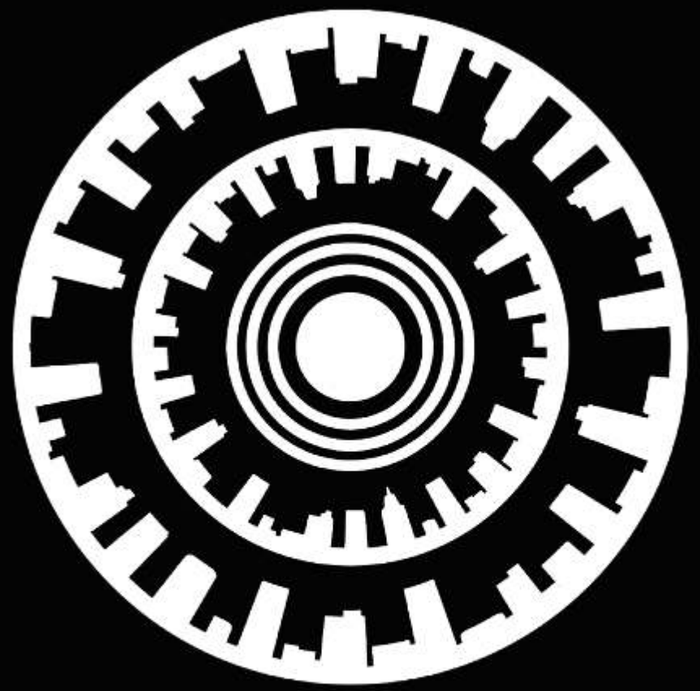
Based on both quantitative and qualitative findings in the previous chapter, the game was evaluated. The evaluation was done based on the research question and design goals set in chapter 2.

At the end this chapter there is a personal reflection on the entire project. Then the conclusion of the whole project is given following a recommendation for the next steps the project could take.

picture - Logo Comparison



Jeremy Bentham's Panopticon 1791



Watcher 2021

6-1.

Evaluation based on design goal A

What follows is a point to point evaluation on the design requirements presented at the conclusion of chapter 2.
The evaluation will be based on both quantitative and qualitative findings in the previous chapter.

A : The design should help people understand the black-box nature of AI

Having direct interactive feedback on how AI sees things created an understanding of the unpredictable black box nature of AI. This was evident when one of the emotions of the participant was read as having a smirk with the score of 45 although it was clear to the human eyes that he was genuinely smiling. This created a discussion among the participants on how unpredictable and opaque the system is.

A1: The design should help people understand of the biases that machine learning models have.

During the play-test participants were quick to understand things such as how black masks tend to get the violence score higher and assume the model had the possibility of having racial bias.

A2 : The design should help people understand the unpredictable nature of computer vision enhanced by AI.

During the play-test, participants were quite confused about how the AI classified things. This was true, especially in violence detection. They were confused about how the violence classifier changed depending on the gender of the participants in the picture.

A3: The design should help people understand the nature of active persistent surveillance by AI

During the interview, participants answered that they did not feel they were being watched by AI all the time. This was because the webcam used during the playtest lit up only when it was active — not creating a feeling of being watched all the time. Also, since the participants personally knew me, they assumed it would be difficult for me to hack existing surveillance cameras which made them assume it was not always watching them.

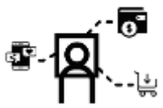
6-2.

Evaluation based on design goal B

B. The design should help people become concerned about the implications of AI-surveillance infrastructure in the city.

The level of concern towards AI surveillance which increased from the average score of 2.75 to 4.25 (+ 1.5), was the highest in all three awareness survey. However, the level of considering this as a threat varied across participants.

One of the participants considered the threat to be a problem in a distant future in a different place with the score of concern 3. This was mainly because the alternative Amsterdam depicted in the narrative resembles a totalitarian government which does not exist in EU where the game was held. For the participant from Germany, it seemed unlikely that the EU would let something privacy breaching as the Watcher system enter European cities. On the other hand, some other participants from China showed a strong sign of concern.



B1: The design should make people concerned about how surveillance enhanced by AI can know all your personal information.

There were little signs of fostering a concern towards the cities they are currently living in such as Utrecht. This may be due to the fictional settings of the game having a distance from the actual democratic governance of cities such as Amsterdam.

B2: The design should help people become concerned about how AI-surveillance can enhance centralised control.

The game mechanic of showing the name with your “trust score” made people realise the danger of having a system that knows all the aspect of you. One participant mentioned how it would be scary if the government can have access to your smartphone.

Then if the government-owned AI...well, something like our Watcher, have access to our smartphone camera it means it can do crazy things.

B3: The design should help people become concerned about how AI-surveillance can spread quickly.

This design goal was only covered briefly in the final prototype during the introduction movie since it was difficult to communicate in a game. This was because the problem of AI as a software spreading quickly was difficult to turn into an interactive game element.

6-3.

Evaluation based on design goal C

C. The design should help people aware of existing surveillance infrastructures.

Including the play-testers during the prototyping session, this was the most frequent feedback. 8 participants including the 6 participants in the prototyping session mentioned that they had heightened awareness of existing surveillance infrastructures after the game. This can also be seen from the quantitative findings of the research which points out that the level of awareness of existing surveillance infrastructures in the city increased from 3.25 to 4.25 (+ 1.0).

C1: The design should make people aware of the existing invisible surveillance infrastructures.

Using existing surveillance infrastructures and turning them into a game prop worked well to make people aware of their surrounding surveillance infrastructure. Especially after the play session, 3 participants mentioned that they now see more surveillance camera in their daily life than before.

I see so many surveillance cameras in front of the shops now.

C2: The design should help people become aware of unquestioned private-led instalment of surveillance systems

This element was only communicated during the introduction movie. However, one participant from the prototyping session mentioned how much Albert Heijn had surveillance cameras inside and was wondering what they were going to do with all that information.

6-4.
Reflections on the medium

6-4-1
The pros and cons of escape room as a medium for critical play

Pros: Players as the protagonist of their own story

Escape room style games as a medium of expression offer a great platform to tell intriguing stories that immerse the players into the world of the story. Compared to other storytelling methods such as books or movies, it is more immersive and makes the story relatable. This is because the players are the one performing the actions which influence events in the game's narrative. This makes them a protagonist and a hero of their own story which has a bigger emotional impact than just sitting back and watching a movie.

In Escape the Smart City this structure worked in favour of communicating the critical issue because the players were seeing the AI surveillance system Watcher as their common enemy. This made them feel as if they owned the problem in some way.

Pros: Inclusive and self-explanatory

The usage of familiar and popular game design made the whole game self-explanatory and easy to participate. Just guiding them into the room was enough to get the game started because a lot of people know that the rule of escape rooms are achieving a certain goal by solving puzzles within a limited amount of time. This helped immensely to immerse the players into the game world quickly because there was no need for sessions to explain the game rules beforehand. This solves a lot of problems in the field of experimental pervasive games where it usually requires the game master to explain the complex rules in a detailed manner because the game design is often unique and unheard of.

Pros: Membrane to wrap experimental game elements inside a bigger narrative.

The escape room structure worked as an membrane to wrap experimental game elements inside a bigger narrative to create one coherent game experience. Using the linear structure with time limits helped to make a convincing story of "hacking an evil system" and immerse participants into it. Each game elements such as the face-detection avoidance or the violence classification in the Escape the Smart city is rather weak if it was on its own as a standalone game. However, when used as a game element inside a bigger game narrative it suddenly makes sense and strengthens the entire experience. Using the escape room game structure allows the game designer to experiment on cutting-edge game elements while keeping the experience from falling apart.

Cons : Difficulty in terms of reaching wide audience

When trying to recruit participants, I realized that escape room s were difficult to use as a medium to address wide audiences. This was because the structure of escape rooms are extremely vulnerable to spoilers. This became a problem in recruiting players. Most of the participants want to know what they were going to do beforehand. However, saying anything to hint at the game experience will take away the surprise elements thus making the marketing extremely difficult. From this, I discovered that the use of the escape room as a medium for critical play has to overcome several marketing issues to reach a wider audience.

6-4-2

The game master as the smart city

A lot of the energy on this project was spent on trying to turn the city into an escape room by controlling it. It was extremely difficult to do so since the city in its essence is chaotic and open and is resilient to structures forced upon it. I had many play-tests cancelled due to a sudden rain destroying hidden clues or an unplanned gay parade adding a very drunk yet friendly participant into the session who destroys the whole game.

Due to its nature, all escape rooms have a strict linear structure which only aims for the escape at the end. It does not allow players to deviate from the rules but rather trap them in a structure. It was at the end of the project when I realised my nature as a game master was the smart city — always trying to make sure the players follow the rules by constant surveillance.

Eric Zimmerman says in the Rules of Play (2004) — “When play occurs, it can overflow and overwhelm the more rigid structure in which it is taking place, generating emergent, unpredictable results.” In this case, I was acting as the rigid structure to restrict the players from open-ended play.

6-5.

Reflections on game mechanics

6-5-1
Designing fictional motivations

During the prototyping phase, I realized I spent a huge amount of time trying to come up with a convincing motivation for the players to go through the experimental game elements that I had created.

This was especially true when trying to make a convincing motivation and reason for the players to enact extreme violence on the streets. To solve this I created a narrative that the Watcher system will enter an emergency state when detecting extreme violence on the street — which allows the players to hack into the system.

Designing these fictional motivations are what most of my time went into during the prototyping phase. I would sketch what sort of reasons would be believable and natural. Why would participants need to mask their face? Why would they need to delete Watcher? these questions were asked over and over again on the paper.

After some time I realized what I was doing was closer to a screenwriter and not a designer. The methods I was using was borrowed from professional screenwriters at Pixar explaining how to build a character-arc or the right way to reveal a plot.

In most live action role play (LARP) the participants themselves have to create a fictional motivation to follow and play the role. He or she may decide to be a vampire who is seeking revenge or a wizard trying to hide from humans. In any case, it demands a lot of effort from the participants to actively decide what their own fictional motivation is and enact the role.

I realized that Escape rooms as a LARP succeeded because it was good at designing these fictional motivations without demanding the user anything. “Get out of a locked room” immediately makes the participants into a protagonist of their own story with a clear strong motivation to achieve the goal. Through this experience, I learned that one of the most important things in creating a pervasive game is to design a convincing fictional motivation which turns the player into a protagonist in the story.

The key finding to design good fictional motivations for critical pervasive games is below.

1. Make the common enemy as believable as possible

2. Make the players wear a certain kind of uniform or a mask which separates them from the real world.

3. For each game task provide a clear reason why it needs to be done to avoid it becoming a “solving-a-puzzle” within a game.

4. At the start of the game, create a sensitizing phase where the players understand the narrative.

6-5-2

Controlling the game state in pervasive games

Controlling the game state was difficult in the open environment. In normal games, it is easy to control which game phase the players are in. However, in the game that takes place in the open environment, this is very difficult because the players can do anything they desire. During the prototyping phase, several ways were experimented to control and separate the game state. The most effective way to control each game phase was to restrict the amount of information the player has. Instead of giving players all the information at once, several props were used to reveal information gradually to control the range of actions the players can take.

1 Printers that print with the progression of the game

A printer that was connected via WiFi was used during the play-test. This printer printed out important information that included clues that were necessary to progress the game. The printed papers worked as a portable item that the players can carry around. Players reacted positively towards new messages coming out of the printer.

2 Terminal like UI which returns information

At the start of the game, players are provided with a smartphone with a unique Linux like UI. Players then activate hack.chat web that can communicate with multiple people on the web. This chat service was repurposed into a terminal UI for the game. Whenever the player solves a puzzle or aims to go to the next phase, they were prompted to enter commands into this terminal. This way of proceeding gave the players the feeling of control as well as feeling like an actual hacker.

3 Lock-box which contains info

The lock-box which contains clues was used in the outside environment. This way of hiding things in the city is commonly used in geocaching. In the game finding the location of the lock-box and figuring out what the key-lock was served to create 2 different game modes. The place to hide them needed to be visible as well as secure. In the end, it was attached to a pole.

6-6.

Reflections on critical message

6-6-1
Critical Pervasive Games

Combining practice from critical design and pervasive game were done throughout the project. In critical design practice, the design is used to mobilize debate and inquire into matters of concern (cf.,Dunne & Raby, 2013 ; Malpass, 2017). Many critical design projects take the form of design fiction which is the use of design to explore and criticise possible futures by creating speculative, and often provocative, scenarios narrated through designed artefacts(Bleecker, 2009).

My practice of combining these two led to critical pervasive games, which served as a platform to tell provocative stories that immerse the players into the world of the near future. When compared to other forms of showing critical design through exhibition or movies, it was more immersive and made the future scenario more relatable since it was a first-hand experience.

The nature of pervasive games to blend the ordinary world with the game world combined with the nature of design fiction to communicate speculative futures created a unique sphere where the participants were physically in the city space but mentally in a future alternative world. This was seen when the participants masked themselves and went out in the city in search for the SD card. One participant mentioned after seeing what the AI was capable of detecting from his facial expression, he felt protected by the mask even though there was no actual risk of being detected by Watcher. Although critical pervasive games leave a lot of room for further exploration, it might have the potential to address bigger socio-technical problems in ways that were not possible before.

Participants out in the city with the mask to avoid facial recognition

6-7.

Recommendations

Teaser video to spread the message

To enable the message to spread, communication before the game is an area to be improved and explored. Especially the communication through trailer videos might work as a way to generate a level of awareness. There can also be clues in the videos so that it expands the width of the game design as well as communicating the narrative.

Stand-alone non-site specific version

The future version of the game should take place completely outside in the city. Preferably it should not be location specific but something that can be hosted in several cities. This would require the game to work with only a tablet and a bag with items. Since the current version of Escape the Smart City works on Unity C# it would be fairly easy to create a stand-alone app that would work on a tablet.

Use AR for a richer experience

For further developments, the interactive experiences using computer vision has more room for experimentation with Augmented Reality. For example, mapping violence scores to people could be done by using ARCore that Apple recently announced.

Hosting this game in an Asian context

This game was done in the European context where it already has a strong culture of protecting privacy. What I would wish for as a next step is to host this game in Asian cities where massive amounts of money are spent on smart city developments. In cities such as Tokyo, Taipei and Shang-hai, there is not enough critical discussion about developments behind smart city infrastructure and surveillance done by an AI.

6-8.

Conclusion

This project started off by defining 8 problems of AI surveillance. During the prototyping phase, several ways of communicating these problems were explored. In the end, the game Escape the Smart City succeeded in creating a immersive experience of living in a city under surveillance by AI. Through the play-test, it was made clear that, by providing the players with interactive feedback on how AI surveillance would perceive them, the players were able to understand the black boxed nature of it and ask critical questions about their necessity and consequences. Also the in-situ experience outside created a heightened awareness of existing surveillance infrastructures.The opportunities for this project is to explore how to communicate about the game to a wider audience in and outside Europe.

Contributions of this project

1.Defined 8 problems of AI surveillance technology.

2.Trained a neural network to detect 2 types of surveillance cameras and made it open source.

3.Created an interactive play which uses an actual AI to classify if the image is violent or sexy.

4.Explored the possibility of critical pervasive games to raise awareness about a social concern.

5.Succeeded in creating a critical play to raise awareness about the problems of AI surveillance.

7-1. Afterword

Opening the black-boxed world

One of the best design decisions in the project that the supervisory team gave me was adding a plot twist at the end that all of what the participants were doing was for the AI to become smarter. This ending gave a depth to the whole project because it closely resembles our current relationship with AI.

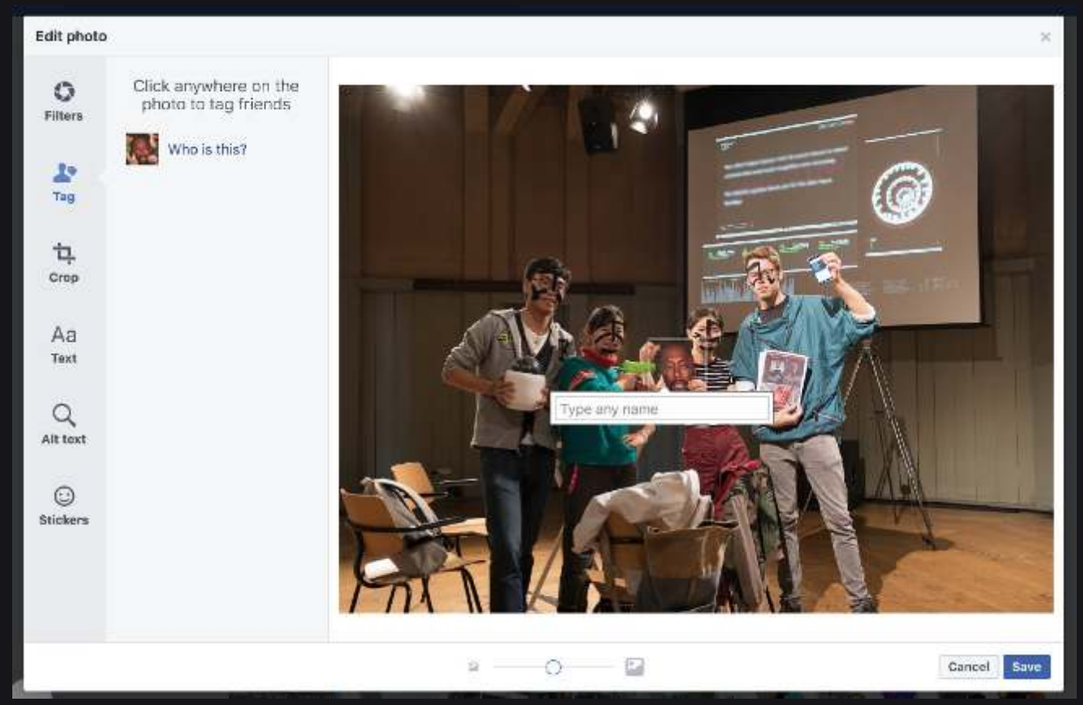
All the participants in the game were dumbfounded when they saw that what they have been doing was just a training program disguised as a mission. “We lost…But how do we win against an AI?” is an interesting question that the participants brought up in the post-interview.

We live in a black boxed world where the structures we have built to sustain ourselves are using all of us to train itself — in automated ways that we no longer are capable of understanding its full consequence. We understand less and less about the world as these ever learning complex structures assume more control over our lives.

Every time somebody watches a video on Youtube, it gets better at suggesting a video that the searcher will watch longer to increase ad revenue. It has no concerns if the video it is suggesting is a propaganda spreading violent ideologies or fake news.

So how do we “win against an AI?”. Obviously, there is no deleting Google or Facebook with a virus in a SD card. Nobody can delete the ever-growing amount of data it feeds upon. And it is not likely to stop using these systems since it has become part of our everyday life.

But what we can do is to try to understand the black-boxed society that we live in. Try to make sure where our data is going and how it was trained and how it will be used. Through this project, I hope I contributed to demystifying this black boxed world.



By tagging their face I can teach Facebook’s AI how to better detect a masked face…

7-2.

Acknowledgements

First of all I want to thank my supervisory team from TU Delft, Roy Bendor and Derek Lomas for giving me so much freedom, guidance, inspiration and support.

My big thanks to Roy for always providing me critical and constructive advices that really helped move the project forward. I got so much inspirations learning from you at TU Delft over the years (also ITD!) and you made me realise what kind of practice I want to do as a designer.

And also a big thanks to Derek for giving me inspirations every time we have a meeting and sending me cool literature by email. Your advice to think beyond school to reach wider audience helped me get my mind out of the box.

And to TU Delft for giving me the faculty scholarship to study two years here without having to worry about tuition.

Secondly I would like to thank my client coaches Tom Demeyer and Stefano Bocconi for having me at the Waag for a year.

Thanks to Tom for always making sure that I don't fly into crazy directions, making sure that I was on ground. I really appreciated your critical and witty comments when trying to decide the project direction.

Also thanks to Stefano for always kindly advising me even though I was chaotic and unorganised. Especially the sharp comments you gave on my writing really helped me understand and reflect on my topic.

And a big thanks all the fellow colleagues at Waag, especially at CODE for joining me on the play-tests and giving me advice.

Thanks to all the members of Studio Lab for giving me a nice working environment. Special thanks to Aadjan van der Helm for the arrangement. Also, thanks to all my friends for helping me keep happy and sane during the crazy TU Delft years.

And to all those in Japan.
Thanks to Daijiro-san for guiding me to the world of design. If not for you I would not be doing this.
And last and not the least, to my beloved family(in Japanese)
修士二年間支えてくれて本当にありがとう！これからも頑張るね！

7-3.
References

Bleecker, J. (2009). Design Fiction: A short essay on design, science, fact and fiction. Near Future Laboratory, 29.

Bogost, I. (2007). Persuasive games: the expressive power of videogames. Cambridge, MA: MIT Press.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Anderson, H. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. arXiv preprint arXiv:1802.07228.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability and Transparency (pp. 77-91).

Ditton, J. (2000).Crime and the City: Public Attitudes to CCTV in Glasgow. British Journal of Criminology, 40: 692-709.

Dunne, A., & Raby, F. (2013). Speculative everything: design, fiction, and social dreaming. MIT press.

Easterling, K. (2014). Extrastatecraft: the power of infrastructure space. Verso Books.

Flanagan, M. (2009). Critical play: radical game design. Cambridge, MA: MIT press

Gordon, E., & Walter, S. (2016). Meaningful inefficiencies: resisting the logic of technological efficiency in the design of civic systems. Civic Media: Technology, Design, Practice, 243-266.

Gredler, M. E. (2004). Games and simulations and their relationships to learning: Handbook of research on educational communications and technology, 2, 571-581

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Uhlmann, L. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology, 29(8), 1836-1842.

Hatcher, J. (2015, July 1). UK citizens unaware of 'smart cities' [News Article]. Retrieved from http://www.smartbuildingsmagazine.com/news/uk-citizens-unaware-of-smart-cities.

Hollands, R. G. (2008). Will the real smart city please stand up? Intelligent, progressive or entrepreneurial?. City, 12(3), 303-320.

Huizinga, Johan. (1955). Homo ludens; a study of the play-element in culture. Boston: Beacon Press

Leese, M. (2014). The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. Security Dialogue, 45(5), 494-511

Neirotti, P., De Marco, A., Cagliano, A. C., Mangano, G., & Scorrano, F. (2014). Current trends in Smart City initiatives: Some stylised facts. Cities, 38, 25-36.

Nisselson,E. (2017). 45 Billion Cameras by 2022 Fuel Business Opportunities. LDV Capital

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

Malpass, M. (2017). Critical design in context: History, theory, and practices. Bloomsbury Publishing.

Montola, M., Stenros, J., & Waern, A. (2009). Pervasive games: theory and design. CRC Press.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection.
In Proceedings of the IEEE conference on computer vision and pattern recognition, 779-788.

Sadowski, J.,& Pasquale, F.A.(2015). The Spectrum of Control: A Social Theory of the Smart City. U of Maryland Legal Studies Research Paper, 20,7

Stenros, J., & Montola, M. (2010). The Paradox of Nordic Larp Culture. NORDIC LARP,12 - 30.

Salen, K., Tekinbaş, K. S., & Zimmerman, E. (2004). Rules of play: Game design fundamentals. MIT press.

Townsend, A. M. (2013). Smart cities: Big data, civic hackers, and the quest for a new utopia. WW Norton & Company

Vincent, J (2018, Jan 23). ARTIFICIAL INTELLIGENCE IS GOING TO SUPERCHARGE SURVEILLANCE [News Article]. Retrieved from https://www.theverge.com/2018/1/23/16907238/artificial-intelligence-surveillance-cameras-security

Zabou. (2012, May). The CCTV Map. Retrieved from https://thecctvmap.wordpress.com/shoreditch/