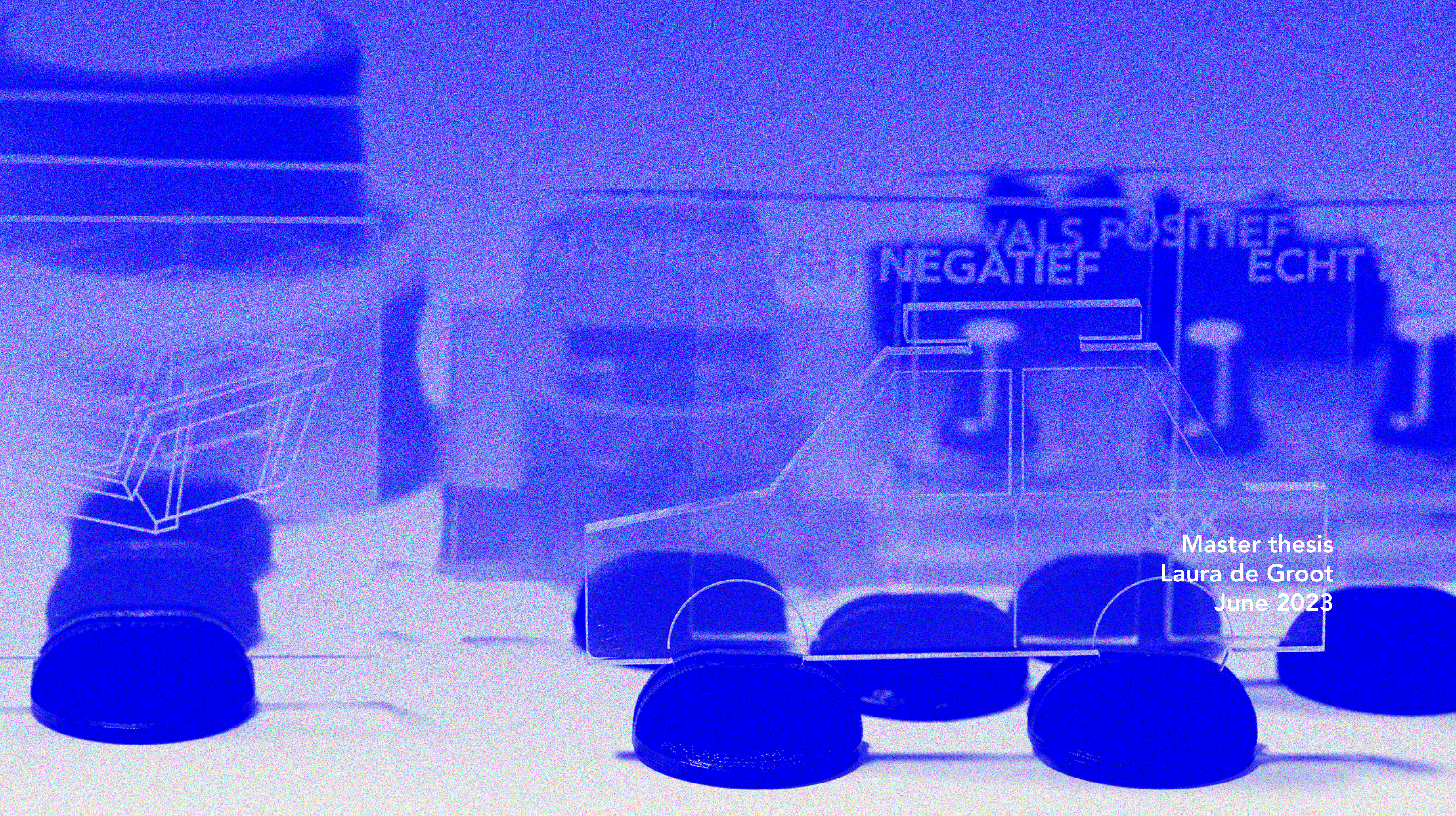
AGONISTIC MACHINE VISION DEVELOPMENT

A Tangible Approach to Involving Citizens in the Development Phase of Machine Vision Systems for Scan Cars



Master thesis
Laura de Groot
June 2023

Author
Laura de Groot

Master thesis
Delft University of Technology
Faculty of Industrial Design Engineering
Msc. Design for Interaction

Graduation Committee
Ianus Keller
Kars Alfrink

In collaboration with
Human Values for Smarter Cities

PREFACE

Dear reader,

This report is the confluence of the last few months, dedicated to this project. I have really enjoyed learning more about this topic, and bringing everything together in the end. So I hope you enjoy it too.

First I want to thank Kars and Ianus for their time, helpful feedback, and discussions throughout the project.

Thanks to the Human Values for Smarter Cities project team for giving a sneak peek into practice, and opening some interesting doors.

Thanks to everyone else who was somehow involved over the last few months. From the weekly openings and casual conversations at StudioLab to the many coffee breaks with my friends. Thanks to my friends, family, and participants for testing and discussing the project with me.

Enjoy!

Laura

SUMMARY

Citizens should be involved during the development of machine vision systems to ensure legitimacy. However, enabling this civic participation is a challenge due to a lack of awareness, knowledge, and understanding. This graduation project explores opening up the discussion about acceptability of a machine vision system, using the scan car development process in Amsterdam as a use case. Acceptability is explained on the basis of trade-offs made during the development phase. First, certain elements need to be understood before citizens are able to judge and critique. By providing a tangible approach to explaining the system, the design aims to improve the understanding of non-experts citizens about machine vision systems and to nurture a deliberate discussion.

After evaluating the prototype with pre-post knowledge tests and semi-structured interviews, the final prototype indicates improvement in subjective understanding of participants, enables them to form their own opinion about what is acceptable, and created a shared language to communicate with.

Ultimately, the design contributes to the field of participatory design approaches to public and responsible artificial intelligence (AI) by providing a practical example. This project suggests that merging explainable AI (XAI) and tangible user interfaces (TUI) can be a suitable approach for involving non-experts during development.

TABLE OF CONTENTS

1. INTRODUCTION

This chapter introduces the topic, scope and process of this graduation project.

INTRODUCTION

To address various urban challenges, cities are focusing on becoming 'smarter'. Using digital technologies as sensors, algorithms, and artificial intelligence has become more common in cities. Turned into 'smart cities', these new technologies, often seamlessly integrated into city context, can help monitor, regulate, and understand the use of the city. Benefits for a municipality can be lower costs and more efficiency. For example, scan cars, cars equipped with cameras that collect data by driving around the city, are commonly used for optimizing parking enforcement (see Figure 1).

The growing use of these systems in public context comes hand-in-hand with more concerns around its possible harms and ethical considerations are being researched (e.g., lack of fairness, transparency, legitimacy, and accountability). Without careful examination of these technologies, we might be designing cities ignoring the majority of civil society (Forlano, 2014). By involving civil society in the discussion around smart cities, we opt to inform decisions around deployment, use, or rejection that better represent civic perspectives.

Human intervention throughout the systems' life cycle can counteract and mitigate some of these possible harms. With participatory approaches towards urban artificial intelligence (AI) applications, these technologies could be aligned with civil society, fostering legitimacy and acceptance. However, enabling this civic participation in public AI is experienced as a challenge due to different knowledge positions of stakeholders.

Figure 1: *Example of a scan car in the city of Amsterdam*

The municipality of Amsterdam aims to develop ethically responsible, privacy-friendly, and safe machine vision systems for scan cars. By involving citizens throughout the development of these cars they aim to identify different perspectives and legitimize the choices made.

Citizens can be involved during different phases of scan car development, from ideating on applications and training the system to the appearance of the cars themselves. The focus of this graduation project is involving citizens in the development phase of the machine vision system, as presented in Figure 2. Due to a lack of awareness, knowledge, and understanding, citizens are not able to participate and discuss on equal footing (Alfrink et al., 2023). Unawareness of systems being used, their (in)direct impact, and a lack of technical expertise hinders this discussion, resulting in not legitimizing a system.

This leads this graduation project to explore and investigate how to involve citizens in participatory machine learning development, and how to improve their knowledge position to meaningfully take part in this co-construction. With this focus, the discussion around the acceptability of a machine vision system is explored by zooming in on trade-offs made and their context. By providing the necessary information in an interactive and tangible way, the proposed design aims to improve the knowledge positions of non-expert citizens, in order to make them full-fledged interlocutors in the co-construction of machine vision systems. Different interactions within a participatory discussion are explored with interactive sketches to inform the final design. This design is developed into a prototype to evaluate the design goals: (1) improving the knowledge positions of citizens of machine vision systems and (2) nurture a deliberate discussion around acceptability of a system.

The project contributes to the new field of participatory machine learning (ML), by providing a practical example aiming to facilitate contestability for citizens in the development process. The design introduces an example of tangible user interfaces (TUI) for explainable AI (XAI), and shows potential for a tangible approach to co-constructing public AI.

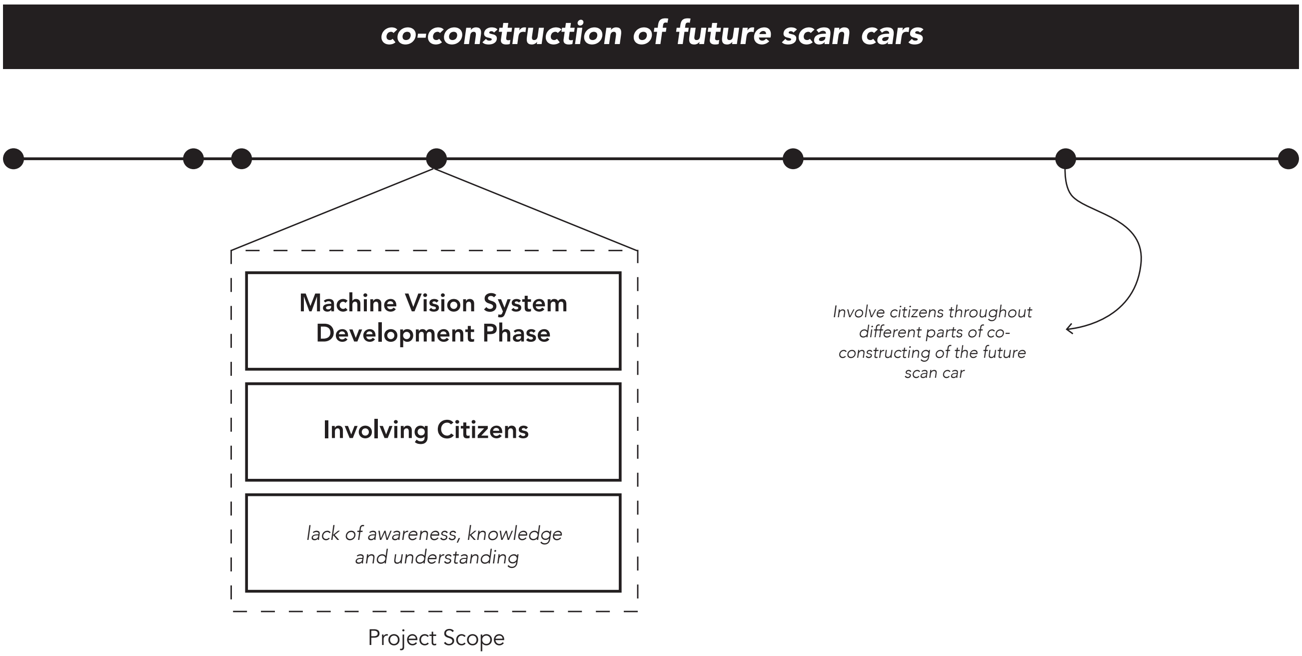


Figure 2: *Project focus within the scan car development process*

The graduation report is structured as follows: First, Chapter 2 stages the opportunity space by introducing the scan car development process in Amsterdam. Next, related work on the legitimacy of a machine vision system, participation, and understanding AI as non-experts are discussed in Chapter 3. These insights are combined to inform the design process. The resulting design is presented, evaluated, and discussed in respectively Chapter 4, 5 and 6. Finally, the report reflects on the design process and outcomes considering its contributions to participatory machine learning, public AI, and the development of scan cars in Amsterdam in Chapters 7 and 8. This is visualized in Figure 3.

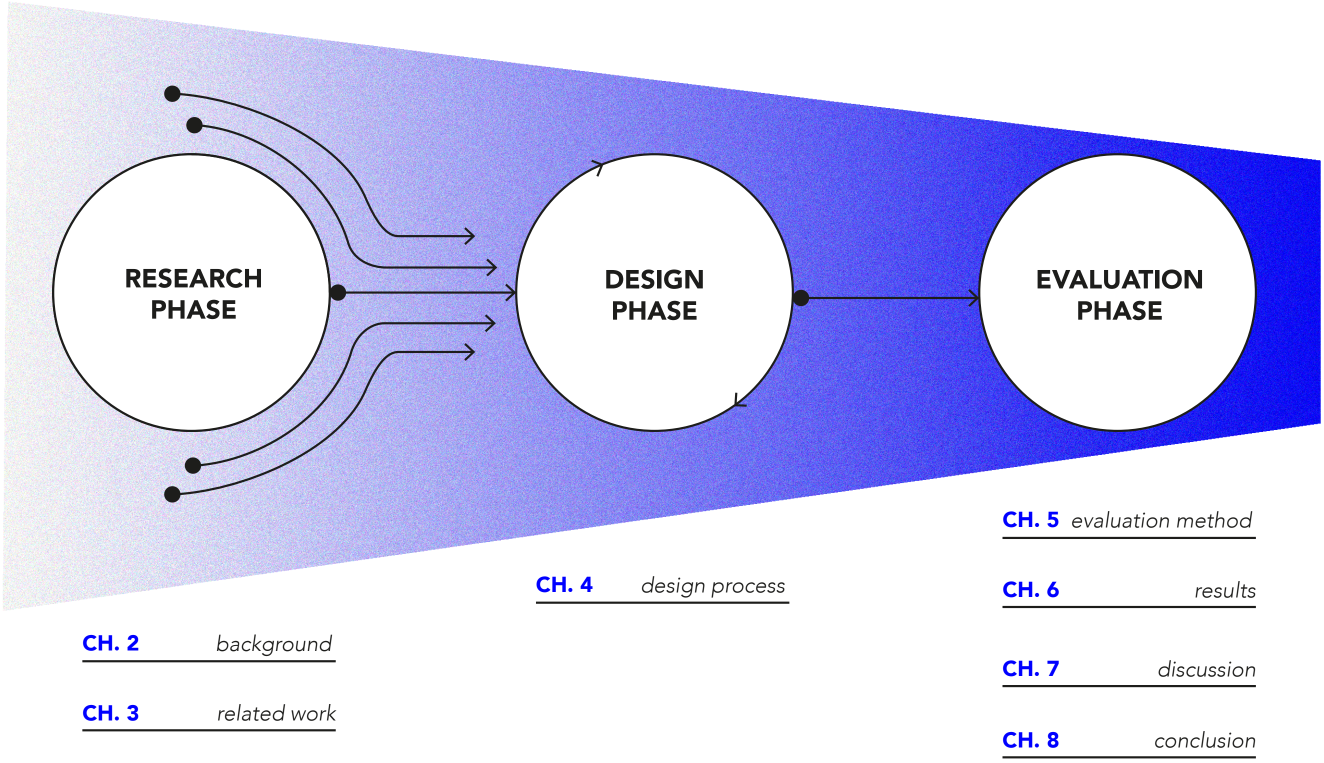


Figure 3: *Project outline with related chapters*

2. BACKGROUND: STAGING THE OPPORTUNITY

This chapter discusses relevant background information regarding smart city technology, specifically the development of scan cars in Amsterdam. It addresses the challenge of enabling civic participation due to different knowledge positions of stakeholders.

ALIGNING THE PUBLIC AND SMART TECHNOLOGY

Cities are redesigning their environment by adapting various technologies to become more efficient and safe. Applying these technologies in public context makes citizens part of the complete system. Without careful examination, untended consequences might arise, leading to designing cities for a specific group instead of the majority of civil society. Involving civil society in this examination can open new perspectives, informing responsible design of technologies in public context.

RIGHTS AS SMART CITIZEN

Wider involvement in making responsible urban technology is not only an idealistic perspective anymore. The adoption of regulations as GDPR, the European AI Act, and mandatory FRAIA by the Dutch Government emphasises the importance of careful considerations around the use of algorithms and artificial intelligence. Idealistic manifestos as TADA are being translated into more concrete regulations to protect data subjects. Data subjects, persons from whom data is collected, can be impacted by the decisions made by AI systems applied in public spaces.

As data subjects, citizens have the right to object, intervene, and transparent information (GDPR). Thus, the technologies should be understandable and open for critique by citizens to fulfill this right.

DATA-INWINNING PROJECT

Currently, the city of Amsterdam is developing new scan cars, with the objective to create an ethically responsible, privacy-friendly, and safe use of machine vision systems (see Figure 4). These cars with machine vision are mainly used for controlling parking fees, but other possible uses are being researched. Localizing waste, identifying construction containers on weak quays, tree maintenance or advertisement taxation are possible other applications for these future scan cars.

The municipality wants to involve citizens in the development of these cars to identify different interests, ensure quality and allow 'Amsterdammers' to participate in deciding on their own living environment. Some central questions to this project are:

- How can yet unknown wishes of citizens be incorporated into the design of the scanning vehicle and the set-up of the image recognition facility?
- Can the working and output of artificial intelligence use to be made transparent and 'contestable' for stakeholders?
- How can we work towards a scanning methodology with sufficient support from society?

Since not every citizen can participate, the municipality aims to compile a citizen panel to represent various interests. Citizens with negative, positive, and neutral attitudes towards scanning solutions in public space should be involved in this process. Seeing each other as equal partners, various

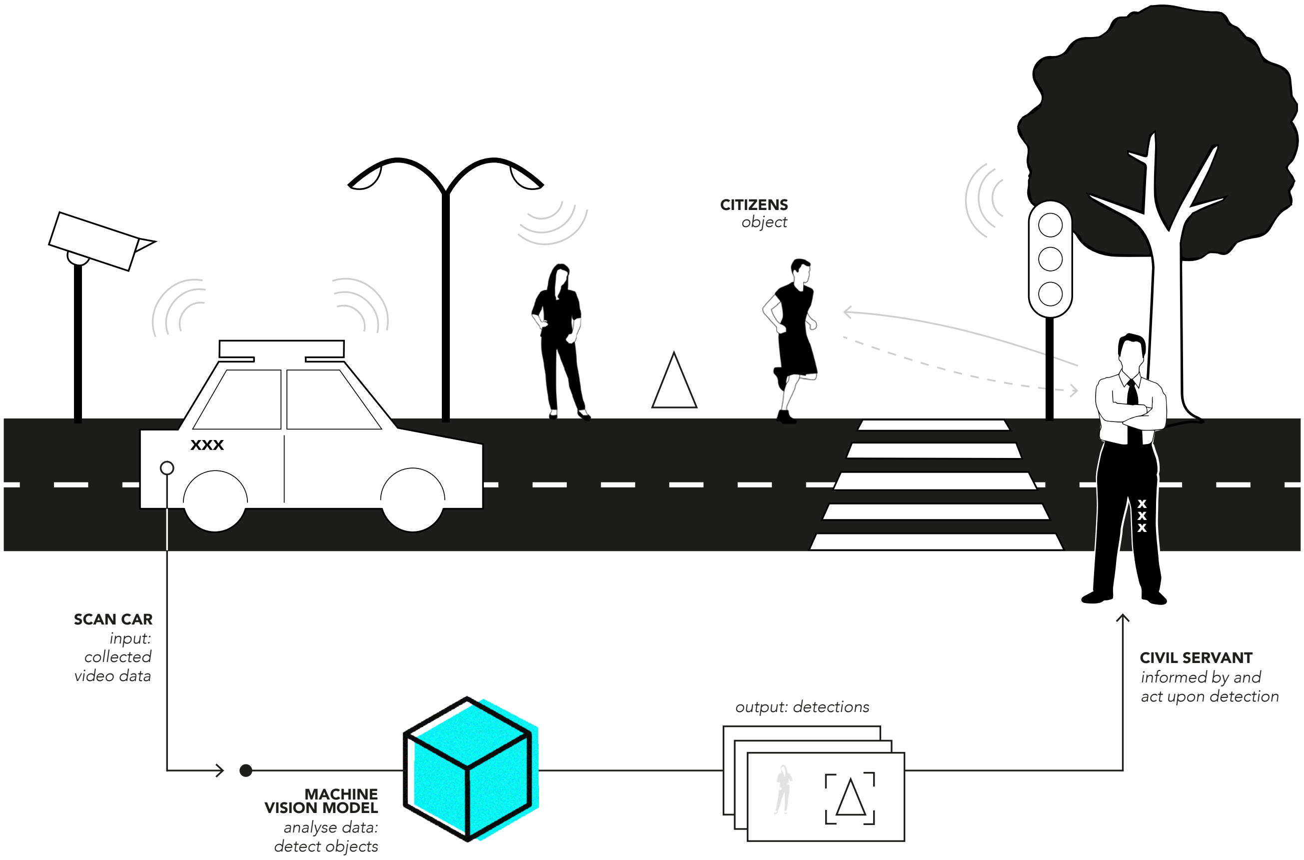


Figure 4: *Scan car system for object detection in public space*

stakeholders should discover and explore the perspectives, interests, and needs, to create an understanding of each other's views and legitimize choices made.

DEVELOPMENT PROCESS IN PRACTICE

The development of machine vision systems, operating on current scan cars, is outsourced. For future scan cars, the municipality intends to create such systems in-house.

Development at ARVOO

The machine vision systems for parking enforcement are currently developed by ARVOO. With their focus on efficiency, they train the model themselves, without including the stakeholder directly in the creation of the model. Data is collected, labeled, and used to train in-house. If after deployment certain mistakes occur often, the model can be updated by ARVOO upon these mistakes (Site visit ARVOO, personal communication, February 10, 2023).

Development at Municipality of Amsterdam

The municipality of Amsterdam has a Computer Vision team with developers to make similar models that will be in charge of models for future scan cars. Normally, after receiving an objective from another department, it is investigated if computer vision is a suitable tool. If so, a proof-of-concept model is developed, which later on can be scaled up for deployment.

During this process, there is a focus on creating a fair model, minimizing bias, and creating trust. While discussing expectations beforehand between departments, no clear agreements are made defining when the model is acceptable enough to be deployed. (Meeting Data-Inwinning project team, personal communication, February 24, 2023)

Impact Assessment informing Acceptability
Recently, a motion was passed in the 'Tweede Kamer', making it mandatory to perform an Impact Assessment Fundamental Rights and Algorithms (FRAIA) before algorithms are used to evaluate and make decisions about people. Even though scan cars are not directly used to make decisions about citizens, it can be assumed that similar assessments are done in advance by the municipality. Citizens could possibly be included in completing this assessment, discussing themes as accuracy, interpretability, and possible consequences, to define what is an acceptable impact before deployment.

CIVIC PARTICIPATION: A CHALLENGE

The Amsterdam Participation guideline notes the importance of careful and conscious consideration of how to involve Amsterdammers meaningfully. However, this civic participation, disregarding the topic, is often described as a struggle.

> *"Involving Amsterdammers in municipal processes is sometimes complex for all involved and requires us to keep up the conversation with each other, to be inquisitive and vulnerable, and - certainly as far as the municipality is concerned - to keep learning and not to shy away when things go wrong, but rather to strive to do it better and better."*
>
> *- Participatieparagraaf (2023)*

This struggle reflects in the process of citizens giving negative advice towards AI implementation for welfare benefits. Where from the municipality's perspective lots of effort was put into engaging people, it seemed that the provided information (in the forms of presentation and work cards) did not come across to inform the final advice. From the citizen's perspective, this involvement emphasizes the need for democratic control, transparency, call for privacy and recognizing the added value of such a system (Participatieraad, 2022). This indicates that improvement is needed to have a conversation, even if it will be complex and tense.

HUMAN VALUES FOR SMARTER CITIES

The main objective of using smart urban technologies often includes values as productivity, efficiency, innovation, and security. Forlano (2014) notes that involving civil society in the debate can introduce values beyond this objective, '*that might more accurately represent the everyday life of citizens*'.

As technology is not neutral, more attention is directed to implementing human and public values in technology. Various value sensitive design methods try to identify (in) direct stakeholders and elicit and represent their values (Friedman & Hendry, 2019). However, there is a lack of how to translate human values as privacy, transparency, and trust into concrete design guidelines for smart city technologies.

Need for New Approaches for Co-Designing
The Human Values for Smarter Cities is a research project that originated from this question for practical design knowledge for the alignment of public values in smart cities. Where both public officials and designers indicate that it is important to involve citizens and other 'end users' for the societal acceptance of smart city technologies, this requires a new approach to co-design. Designers need new methodologies to include citizens in the design of these complex technological services, including ways to understand and critique them.

The scan car and machine vision development process in Amsterdam is observed as a use case, to research ethical implementations of smart city technologies with multiple stakeholders and various values at play in practice. The different stakeholders are presented in Figure 5.

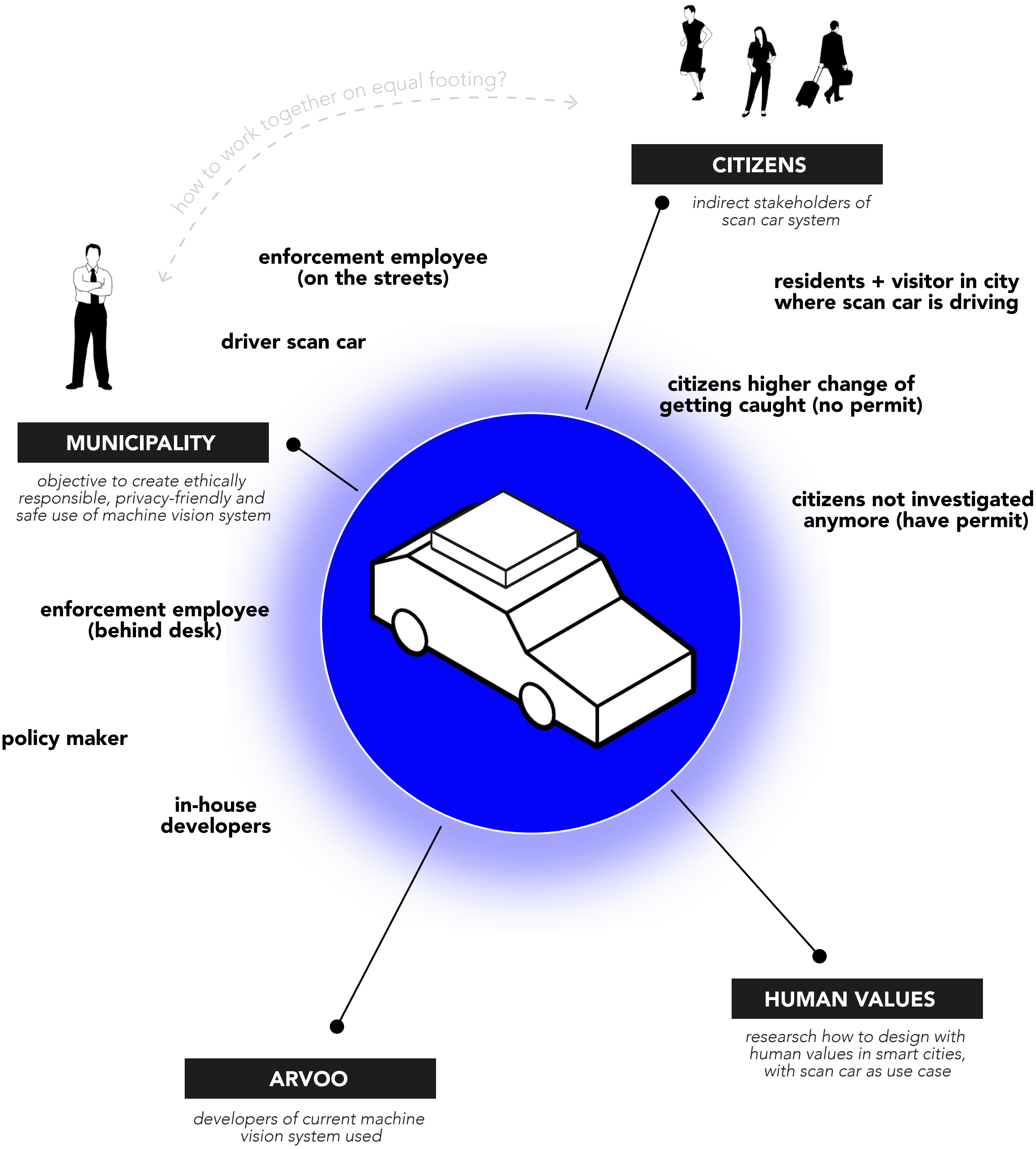


Figure 5: *Stakeholder map scan car development*

KNOWLEDGE POSITIONS OF STAKEHOLDERS

LACK OF KNOWLEDGE

From interviews with civil servants, Alfrink et al. (2023) summarized enabling civic participation as a challenge to creating contestable scan cars. Due to a lack of awareness, knowledge, and understanding citizens are not able to contest a system effectively, making discussion about the system difficult. The different knowledge positions between stakeholders as citizens, civil servants, and developers hinder this conversation. These knowledge positions do not only consist of technical knowledge but also awareness of a system being used or decision impacts on their own life as an (indirect) decision subject.

OPACITY OF MACHINE VISION SYSTEMS

With the intention to involve citizens in the co-construction of scan cars, the municipality of Amsterdam counters one possible cause of opacity of a machine vision system. Burell (2016) identified various possible reasons for opacity of machine learning algorithms, one of which is keeping the conversation about the technological systems used inside 'walls'. Opening up this conversation can provide more transparency.

Address Technical Illiteracy

To have this substantive conversation with citizens, their lack of awareness, knowledge, and understanding gets in the way. This lack due to technical illiteracy is a reason for opacity of a machine learning algorithm (Burell, 2016). Understanding complex and abstract concepts remains inaccessible to the majority of people. Burell notes the importance of *'making the public more knowledgeable about these mechanisms that impact their life opportunities and put them in a better position to directly evaluate and critique them'.*

DEFINING THE OPPORTUNITY

To involve citizens in the co-constructing of scan cars, next to providing space for conversation, they need more knowledge to be able to discuss and critique a system. This leads to the following design question:

"How to ***improve understanding*** *of* ***machine vision systems*** *to make stakeholders (citizens)* ***full-fledged interlocutors*** *in the co-construction of future scan cars?"*

This can be separated into different questions as how can we facilitate this conversation to open up and enable citizens to take part on equal footing? What should be discussed, what knowledge is needed for this, and how to improve the understanding of non-expert citizens? How can we facilitate this conversation to open up and enable citizens to take part on equal footing?

3. RELATED WORK

This chapter elaborates on related work from literature research, practice and other examples. This includes topics as legitimacy of scan cars, participatory machine learning and explaining AI to citizens. At the end of the chapter, conclusions are combined to inform the design opportunity with goals, requirements and an approach.

LEGITIMACY OF DESIGNING SCAN CARS

Citizens will be involved with the intention to legitimize choices made around the future scan cars. A way to neutralize possible harms, as a lack of fairness, legitimacy, and accountability, is by creating contestable AI systems. This means being 'responsive to human intervention throughout the systems lifecycle' (Alfrink et al., 2020). Civic intervention can take on different forms of contestation, see Figure 6. Individual decisions, i.e. given penalties, should be open to contestation. The development of the system being used or a policy can be contested through citizen participation before deployment (Alfrink et al., 2023). Consequently, civil society should engage in the debate over these smart city technologies ex-ante.

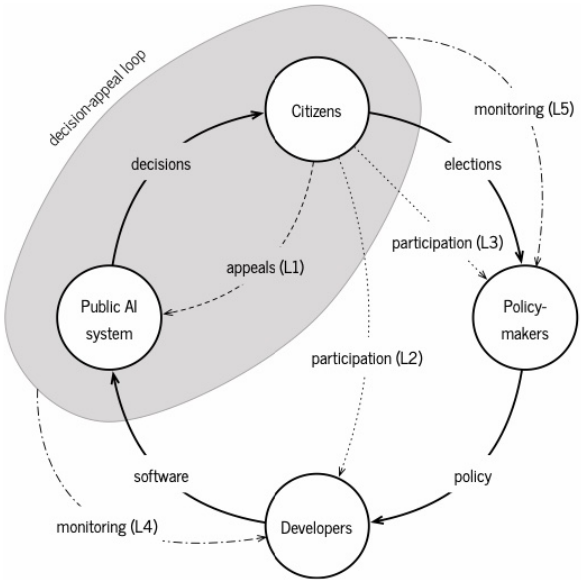


Figure 6: *Contestability loops (Alfrink et al., 2023)*

AGONISTIC ML DEVELOPMENT

Feedback from relevant stakeholder, directly and indirectly, should be included early on. Debating machine learning during the development phase is a practice to ensure the contestability of a system (Almada, 2019). At the beginning of the AI systems' life cycle agonistic approaches to ML development, as participatory design to include these stakeholders, can be fruitful to co-construct decision-making system.

DEFINING THE LEGITIMACY GAP

Within the development of Automated Decision-Making (ADM) systems König and Wenzelbuger (2021) identified a legitimacy gap. They advocate that there is a difference between ADM systems and Human Decision Makers in the public sector. Decisions made by a human can be guided by explicit values and rules, but at the same time can be influenced by implicit parameters as experience and intuition. These explicit and implicit parameters aggregate the distribution of decision outcomes. When a decision is automated through an algorithm, the decisions are informed by explicit decision parameters. By choosing a specific design, developers choose for specific parameters and a specific outcome distribution. This is visualized in Figure 7.

> *'With human decision-makers, abstract values and adequate procedure that safeguard a general orientation towards commonly accepted goals (such as impartiality, avoiding wrong classifications etc.) legitimize decision. With ADM systems, the epistemic basis is different, which introduces a different standard of evaluation. As it becomes possible to predictably produce different distributions of decision outcomes, this also calls for explicit value judgments about these outcomes. Hence, the legitimacy of the decision-making is no longer rooted in general values and an ethos that orient decision-making, and a legitimacy gap arises.'*
>
> *- Köning and Wenzelburger (2021)*

The involvement of those affected by decision-making can ensure the acceptance of a specific distribution of outcomes, and thus close the legitimacy gap. Specifically, direct inputs from the public can serve to secure the acceptance and legitimacy of the decision-making performance.

LEGITIMACY OF MACHINE VISION

Translating this to the development of a machine vision system used for scan cars, citizens and civil servants should be involved in this conversation to ensure legitimacy, as these stakeholders can have different perspectives of what is found to be acceptable. By identifying, hearing, and weighing these perspectives the model can be aligned more with civil society than just the systems' developers. This leads to the need for a discussion about what is acceptable model performance.

In the case of scan cars, the machine vision model does not make automated decisions. Even though the legitimacy gap focuses on ADM systems, this discussion is still relevant, because developers have the power over defining what a preferable outcome distribution is. Next, König and Wenzelburger (2021) focus on the impact the ADM systems have in high stake situations. In the case of object detection for scan cars, consequences are less severe and decisions can be reversed, making it a low stake application. However, the detections can be used to inform civil servants and result in legal follow-up steps, making it of higher interest and becoming more proportional. Lastly, lots of detections with such a model will be made, resulting in a high likelihood of the system being used. Therefore it is still important to debate the system, even if the following decision is not automated (Robinson, 2022).

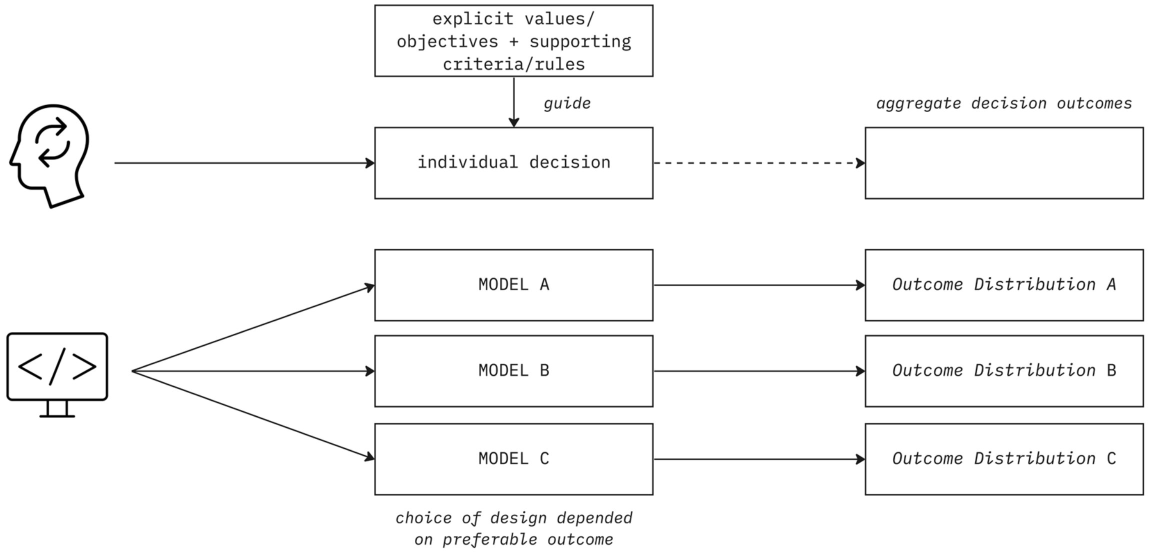


Figure 7: *Schematic visualization of legitimacy gap (adapted from Konig & Wenzelburger, 2021)*

OPENING UP FOR A DELIBERATE DISCUSSION

Following the definitions of Henin and Le Métayer (2022), to reach the legitimacy of a machine vision system, *'a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norm, values, beliefs and definitions'* needs to be formed.

To reach this generalized perception of what is acceptable, citizens first need to understand what the machine vision system is. This is followed by justification, convincing something is good, adequate or appropriate, and contestation, convincing something is bad, inadequate or inappropriate. Based on these elements, people can deliberately discuss what is appropriate and inappropriate for a model. This includes forming your own opinion, but also being able to understand other perspectives. Arguments should be informed and met with contrary arguments, weighing the different views out there.

CONCLUSION

Citizens should be involved to legitimize a machine vision model used by scan cars. They need to be involved in discussing the desired model outcome distribution and steering the model to acceptability. With an agonistic approach to ML development, there needs to be room for debate and different perspectives to exist.

To have a legitimate discussion, citizens need to judge what is appropriate or not. They need to be able to form their own opinion and understand others. Therefore they need to understand the machine vision model first to make informed arguments.

Explanation = goal is to make is possible for a human being to understand

Justification = goal is to convince that a decision is good, adequate or appropriate

Contestation = goal is to convince that the decision is bad, inadequate or inappropriate

Legitimacy = a generalized perception or a assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norm, values, beliefs and definitions

PARTICIPATORY MACHINE LEARNING

Engaging different stakeholders as active participants in the design process can be done in many ways. Participatory design approaches aim to involve stakeholders in collaboratively designing what they need, using their knowledge and experience. Almada (2019) notes participatory design as a way to contest ML development.

This section explores the upcoming themes around participatory design approaches regarding ML technologies, including the role of knowledge and participation from citizens' perspective.

PARTICIPATION TO LEGITIMIZE MACHINE LEARNING

Part of establishing the legitimacy of a participatory design process is letting participants know the purpose of the process and their influence on the outcomes. Different purposes match different strategies towards participation.

Crowd-searching initiatives for data labeling or collecting data are examples of present-day participation regarding ML. However, these kinds of involvements are not an easy way out to meaningful involvement. Emerging attention is directed to participatory machine learning, where participatory design approaches are combined with machine learning elements. Sloane et al. (2020) identified three forms of participation in relation to machine learning: participation as (1) work, (2) consultation, or (3) justification.

In order to legitimize the machine vision system, participatory activities should be seen as justification, meaning they must be long-term and genuine. The process should provide transparency and genuine knowledge sharing between stakeholders. Citizens should be involved in parts where their perspectives do care and have influence in facilitating co-construction instead of co-informing and preventing participation-washing.

AGONISTIC PARTICIPATORY DESIGN

Often participatory approaches are focused on consensus searching: aligning different stakeholders by bringing them together. However, the practice of an agonistic approach to ML development emphasizes the importance of dissensus, friction, and disagreement (Forlano et al., 2014). Different perspectives must be able to coexist because diverging views can be pushed back with the aim of reaching consensus. Within this space power structures need to be considered, because they can lead to silencing others (Kraff, 2019).

Giving room to diverging perspectives increases the chance of tensions. Therefore, agonistic participatory design approaches should allow for conflict and protect the participants at the same time.

DEALING WITH POWER AND KNOWLEDGE IMBALANCES

The dialogue between stakeholders in collaborative processes is influenced by their starting conditions beforehand. Power, knowledge, and resource asymmetries hinder participation on equal footing (Ansell & Gash, 2007). Power imbalances can bring stronger voices forwards, skewing the outcome of a process. Lacking skill and expertise can disengage stakeholders in discussions about technical problems. This hinders genuine knowledge sharing, which undermines the legitimizing intentions of participatory machine learning processes.

These imbalances result in stakeholders being unable to participate in a meaningful way. By improving these starting conditions beforehand, every stakeholder should be able to engage and be empowered to take part in the collaborative process and express their views.

PARTICIPATION FROM CITIZEN PERSPECTIVE

Parallel to top-down regulations as GDPR, bottom-up initiatives try to represent public opinion. While often initiated by authorities, these initiatives try to empower and give a voice to citizens. Where developers have technical knowledge, citizens have local knowledge. To manage power dynamics, one form of knowledge should not be privileged over the other (Schouten, 2022). Living Labs and platforms as Smart City Amsterdam are examples of bottom-up initiatives, where citizens are involved in sharing public opinions and urban decision-making.

Example: Play the City

Play the City is an example of including citizens in urban decision-making processes. Combining elements of game play, different stakeholders are put together in conversation, placing the problem in perspective. This is often done around a table, having a playing field to explore as shown in Figure 8.

CONCLUSION

Participation in the development of a machine vision system should be seen as justification to legitimize. This includes being transparent about intentions and the impact, as well as making room for genuine knowledge sharing. For this to happen, power and knowledge balances should be addressed to put everyone on equal footing.

Citizens can be empowered by improving their technical knowledge. Next, the session should give room to different views to exist and emerge. Game rules, different stakeholder roles, and exploration of the problem space together can help with this.

8: *Pictures from Play the City events*

TRADE-OFFS WITHIN MACHINE VISION DEVELOPMENT

To legitimize the machine vision system for a scan car, citizens should be involved in defining what is acceptable. To facilitate genuine knowledge sharing, the question arises what knowledge is needed to have this legitimate discussion. Therefore, we need to understand how a machine vision model is developed.

WHAT IS MACHINE VISION?

Machine vision is an application of computer vision, where image processing techniques are integrated into systems to detect elements. Videos and images are used by a model to 'make sense' or 'understand' the environment. What happens within the model actually is often described as a black box: the exact workings the model follows are unknown, due to its mathematical complexity and depth.

Scan cars use a specific subdomain of computer vision, called object detection: objects are detected, classified, and located within an image. Object detection models use supervised learning, where already classified data is used to train a model. The learning process takes the inputs and the desired outputs and updates its internal workings accordingly, so the produced output gets as close as possible to the desired output. By confronting the model with a newly labeled subset of images, the performance of different models can be compared in order to optimize. Object detection models use convolutional neural networks (CNNs) to make sense of the images, which even to experts have an opaque behavior. While humans label their surroundings unconsciously based on prior knowledge, machines 'see' each pixel as an individual number, which contains information. While meaningless to humans, these numbers are the input and communication to a computer. Translated through various layers of the (mathematical) models, each pixel is investigated.

DEVELOPING A MACHINE VISION SYSTEM

Developing machine vision systems is nowadays done by data engineers and scientists. A general overview of developing machine learning systems is shown in Figure 9. The first step consists of collecting data, which then is processed, labeled, and split in order to train a model. The actual model needs to be developed, which consists of selecting, training, tuning, and then evaluating its workings. Evaluating machine learning systems is commonly done via a confusion matrix, where the distribution outcomes are shown. This includes the number of false positives, true positives, false negatives, and true negatives. From this distribution performance rates as accuracy, precision, and recall can be calculated.

Developing and testing the algorithms is an iterative process where different variations are made, in order to see how a system behaves. During this process, issues can arise, which the model can be updated upon before it is deployed (Hapke & Nelson, 2020).

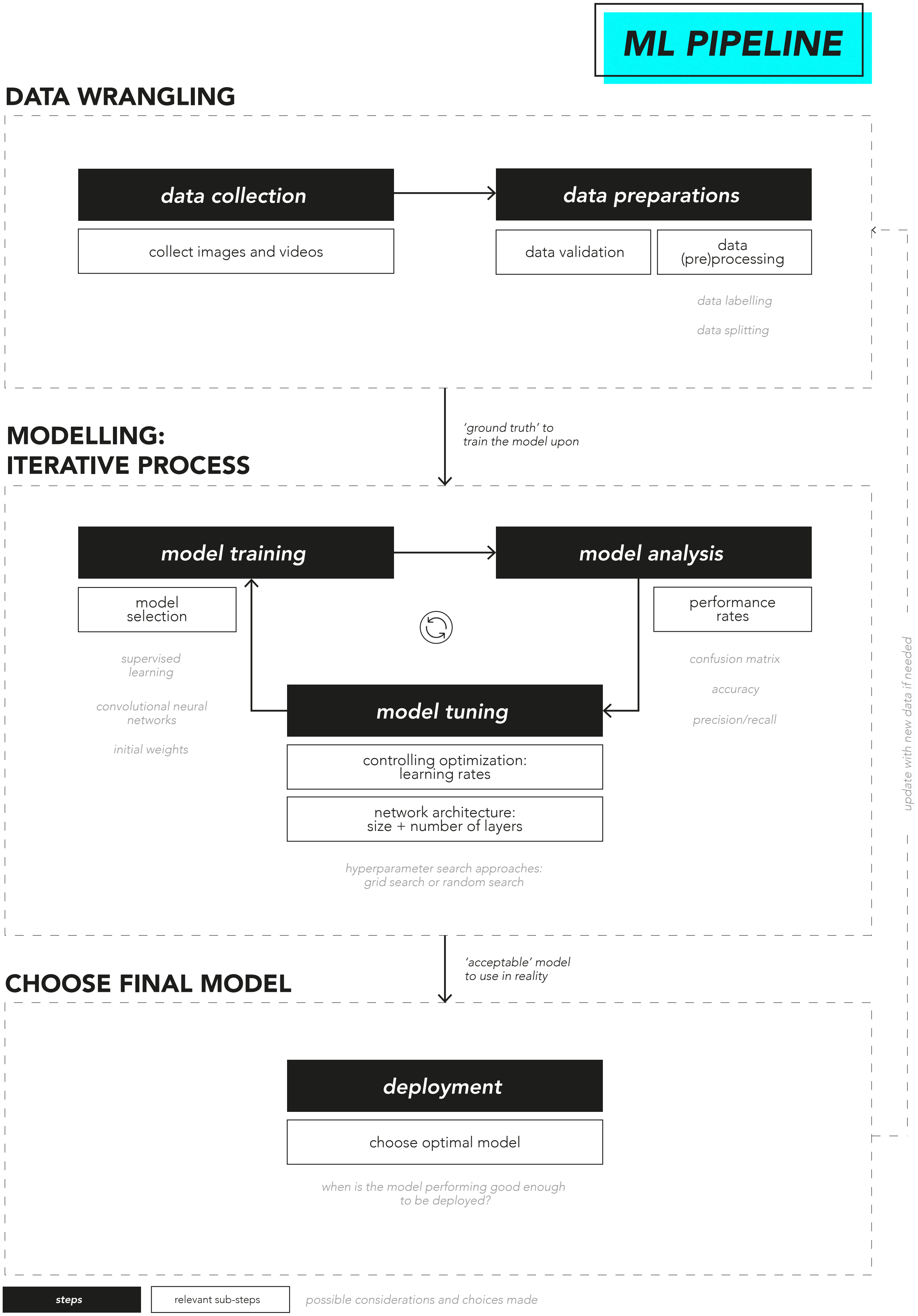


Figure 9: *Machine learning pipeline in steps*

TECHNICAL TRADE-OFFS DURING MODELING

In the course of developing a machine learning model, different decision-making points occur (i.e. problem formulation, technical approach, and evaluation metric). These can be seen as explicit and implicit decision-making points, that influence how the model functions after being deployed (Smith et al., 2022). Where model selection is often an explicit choice, due to technical possibilities, the actual model architecture is iteratively tested out, resulting in an implicit and hard to document design.

During this iterative process of optimizing the model, various trade-offs are made. A trade-off is '*a balance achieved between two desirable but incompatible features; a compromise*.' Smith et al. (2022) identified value tensions, where some values can be sacrificed in pursuit of others. A selection of trade-offs relevant to machine vision development identified is explained.

Accuracy vs. Interpretability

'*Accuracy concerns the ability of a model to make correct predictions, while interpretability concerns to what degree the model allows for human understanding. Models exhibiting the former property are many times more complex and opaque, while interpretable models may lack the necessary accuracy*.' (Johanson et al, 2011)

Accuracy vs. Efficiency

For a model to become more accurate, it is often made more complex to make better predictions, by adding layers or increasing the depth. This computational complexity takes more time to process, resulting in a longer execution time, which makes it less efficient.

Bias vs. Variance

During model training, the model is confronted with a training set. Variance is an error from sensitivity to shifts in this training set. High variance can lead to overfitting: the model is incorporating random noise or outliers in the patterns found. With the bias error models can be simplified, but miss important patterns in the data (underfitting). Generally, the chosen model should accurately capture patterns in the training set, but should also be able to generalize well to unseen data.

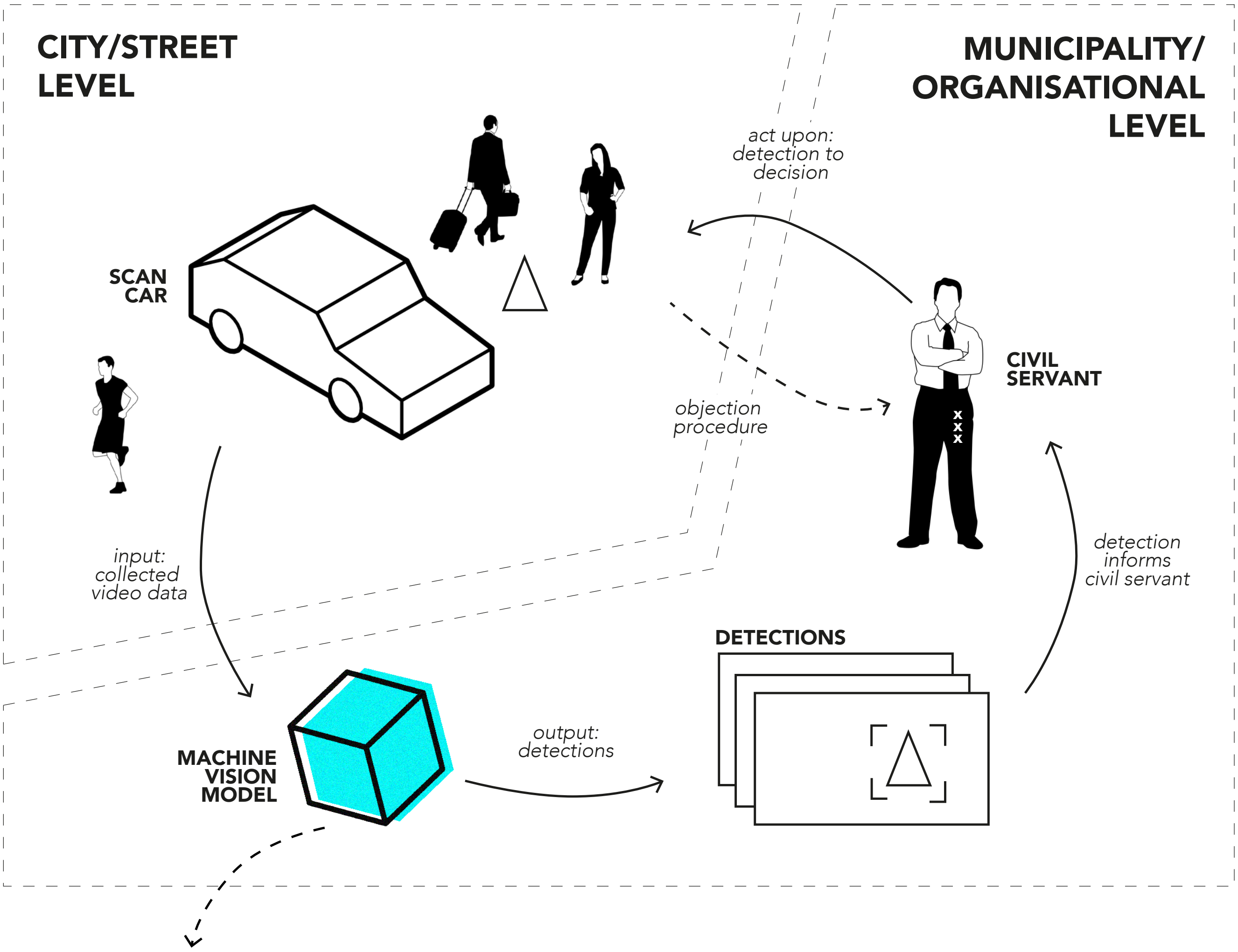
Precision vs. Recall

Next to accuracy, precision and recall are evaluation metrics. With precision, the amount of actual objects within the made detections is evaluated. Thus, how precise are detections made? Recall refers to of all the detections the model should have made, how many did it make? During the development, the model can be optimized to minimize a certain error, and thus optimize for the precision or recall rate. By optimizing for one type of error, more errors of the other will occur.

SEEING TRADE-OFFS IN CONTEXT TO EVALUATE ACCEPTABILITY

According to König and Wenzelburger, to establish the legitimacy of ADM systems, these trade-offs should be discussed with stakeholders. In order to realize this desire, first the decision situation, stakes, and values associated with possible decision outcomes should be clarified. Secondly, the possible trade-offs need to be clarified and evaluated, to in the end aggregate preferences and come to a design choice.

Translating these steps into debating the acceptance of a machine vision system, its context and consequences need to be clarified to all stakeholders. While the isolated trade-offs might be too abstract, placing them into context opens up more comprehensible questions. Seeing the trade-offs in context are visualized in Figure 10.



MACHINE LEARNING PIPELINE

data wrangling → **modelling** → **choose model**

data wrangling
- data validation
- data preparations

what is the 'ground thruth' we are training upon?

modelling
- model training
- model analysis
- model tuning

when is the model working acceptable enough?

choose model
- optimal model
- deployment

what is the optimal model perfromance?

TRADE-OFFS

1 accuracy vs. efficiency

How accurate can the model be, but still work efficiently in the context? I.e., how long can the model take to accurately detect an object, versus when is this not efficient anymore? Is this loop efficient to be deployed?

2 accuracy vs. interpretability

If the model becomes more complex, it perfoms more accurately, but it becomes less clear how a certain detection came to be: the model becomes less interpretable. Can we understand how the input became the output? Is this important?

3 bias vs. variance

During training the model, we need to find a balance between how good the model recognizes objects from the data set, but also is able to recognize objects it has never seen before. With what threshold should the model detect?

4 precision vs. recall

Should we optimise to have less false positive detections, or false negative detections? What are the costs and consequences of these mistakes, and what should we optimise for?

Figure 10: *Seeing trade-offs in context*

CONCLUSION

Citizens need to discuss the acceptability of certain trade-offs within machine vision development in order to ensure legitimacy. These trade-offs need to be placed in context to open up comprehensible questions to discuss. Therefore stakeholders need to understand the following elements:

- Trade-offs: what trade-offs are made during modeling?
- Modeling: what does the modeling phase entail?
- Evaluation Metrics: how is a model evaluated?
- Detection: what is a detection?
- Decision Situation: what are the consequences of a detection? How are they translated into a decision?

By recognizing how these elements interact with each other, stakeholders can inform their own perspectives to take part in the discussion.

IMPROVING AI KNOWLEDGE FOR NON-EXPERTS

When talking about AI, often abstract and complex information are used which can be hard to grasp for non-experts, which hinders communication. Providing the information to participants does not necessarily lead to them understanding and using this information. Working with different knowledge positions a suitable mode of presentation of information is needed to improve understanding and transfer knowledge.

IMPROVE UNDERSTANDING OF AI

A term associated with improving the understanding of AI concepts is AI literacy. Long and Magerko (2020) define AI literacy as '*a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace*'. Competencies do not solely focus on understanding AI but include being able to interact and form opinions around it. The various knowledge positions of citizens towards AI can be seen as different competency configurations.

In order to be a full-fledged interlocutor that can critically evaluate the trade-offs during the development phase, desired competencies include:

- *Decision-Making*: Recognize how computers reason and make decisions
- *Programmability:* Understand that agents are programmable
- *Human Role in AI:* Recognize that humans have an important role in programming, choosing, and fine-tuning AI systems
- *Interpreting Data Critically:* Understand that data cannot be taken at face-value and requires interpretation.
- *Action & Reaction:* Understand that some AI systems have the ability to physically act on the world
- *Imagine Future AI:* Imagine possible future applications of AI and consider the effects of such applications on the world. (Long & Magerko, 2020)

Building on these competencies, stakeholders can be empowered to critically discuss the machine vision model. To improve these competencies Long and Magerko identified various design considerations for creating learned-centered AI activities. With this focus on learning, the design considerations look into fostering AI literacy among audiences without technical backgrounds, matching similar knowledge positions as citizens. Nine possibly interesting design considerations are illustrated in Figure 11.

embodied interactions

Consider making interventions which puts user 'in the agent's shoes' as sensemaking of reasoning process; i.e. embodied simulations and/or hands-on physical experimentation with AI technology.

unveil gradually

To prevent cognitive overload, consider the option to inspect and learn about different system components; explain few at once; introduce scaffolding that fades as user learns more.

critical thinking

Encourage learners to be critical consumers of AI technologies by questioning their intelligence and trustworthiness.

lower barrier to entry

Consider how to communicate AI concepts to learners without extensive backgrounds in math or CS (e.g. reducing required prerequisite knowledge/skills, relating AI to prior knowledge, addressing learner insecurities about math/CS ability).

leverage learner interests

Consider leveraging learners' interests (e.g. current issues, everyday experiences, or common pastimes like games or music) when designing AI literacy interventions.

acknowledge preconceptions

Acknowledge that learners may have politicized/sensationalized preconceptions of AI from popular media and consider how to address, use, and expand on these ideas in learning interventions.

new perspectives

Consider introducing perspectives in learning interventions that are not as well-represented in popular media (e.g. less-publicized AI subfields, balanced discussion of the dangers/benefits of AI).

social interaction

Consider designing AI learning experiences that foster social interaction and collaboration.

explanability

Consider including graphical visualizations, simulations, explanations of agent decision-making processes, or interactive demonstrations in order to aid in learners' understanding of AI.

Figure 11: *Selected design consideration (adapted from Long & Magerko, 2020)*

EXPLAINING AI TO DIFFERENT STAKEHOLDERS

Within research, a lot of focus is directed to the explainability of AI, also known as XAI. This can be seen as a way to open up the 'black box', provide more transparency and increase users' understanding of AI. Nowadays XAI research is mainly focused on experts, explaining how complex models work ex-ante, and non-expert end users, why a model came to a certain decision or advice post hoc to an individual. Generalizing insights from experts to non-experts might not be feasible, because the different receivers have different needs and abilities to process explanations (Jiang et al., 2022). However, this research could possibly inform how non-experts could be informed ex-ante.

Next to different receivers of explanations, the explanations themselves can differ. Where one focuses on technical elements, others focus on context and relations. Jiang, Kahai and Yang (2022) summarized these different explanations into three dimensions, as shown in Figure 12. One model can have different dimensions of explanations to become interpretable and understandable to its users throughout the system.

Contextualizing Explanations

Various examples of explanations on human-ground level try to turn abstract concepts into context-specific concepts. Shen et al. (2020) found that contextualizing field specific terminologies can support non-expert public understanding of model performance. Cai et al. (2019) note that visual examples can improve subjective understanding of machine learning algorithms. Similar effects of contextualizing were found by Wolf (2019), introducing explainability scenarios to help envision possible use.

Interactive Interfaces

A commonly researched form of explanations is using interfaces, for both expert and non-expert explanations. Cheng et al. (2019) found that interactive explanations can improve non-expert users' comprehension of decision-making algorithms. However, this improvement comes with the trade-off of taking more time.

Opposing Effects of XAI

While explanations have the objective to clarify an AI system, Jiang et al. (2022) note various opposing effects of XAI.

For example, providing a detailed explanation of the mathematical workings within an AI model might overwhelm non-expert users. Jiang et al. (2022) therefore note that to tailor explanations to the receivers, human factors as reasoning processes and knowledge need to be considered. Secondly, not every moment is in need of an explanation. When providing one, this should be carefully considered to prevent overload. Thirdly, users of a system can be negatively influenced by

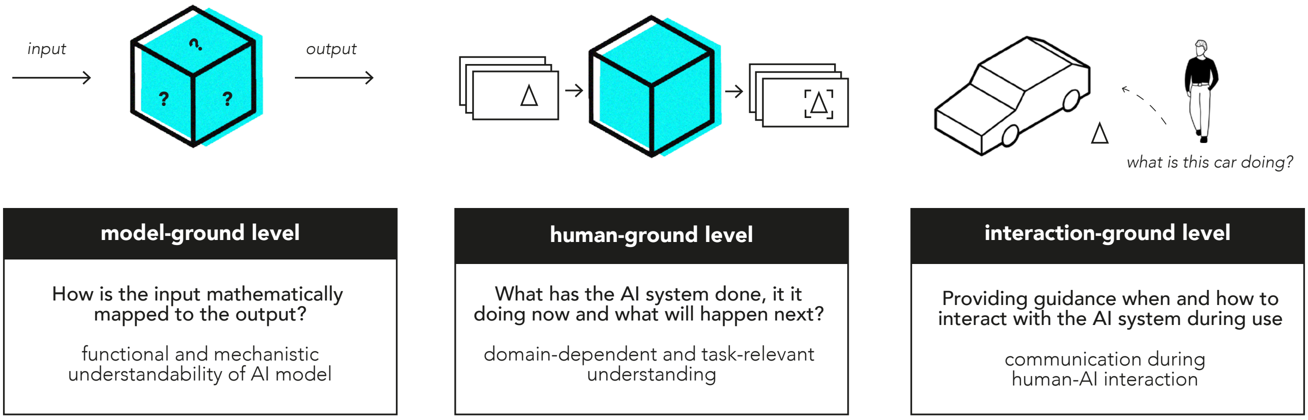


Figure 12: *Dimensions of explanations (adapted from Jiang et al., 2022)*

their epistemic uncertainty (ambiguity due to a lack of knowledge, i.e. about the system or its context). Due to this uncertainty, mistrust, overconfidence or confusion can occur. Not understanding and being aware of not knowing can have an adverse effect. Thus understanding not only is objective but also has a subjective element to it.

TANGIBILITY: MATERIALIZING THE ABSTRACT

Moving from an interactive interface to a more spatial approach, tangible experiences can be helpful for improving the understanding of AI. By including several design considerations, as lowering the barrier to entry, social interaction, and embodied interaction, this can be a fruitful direction for improving the understanding of non-experts.

Materializing the Abstract

Tangible User Interfaces (TUI) allow users of computer systems to interact with the content by interacting with physical objects. These physical tokens take advantage of the implicit or intuitive knowledge of humans in their everyday life.

This tangible approach contains properties, that promise to foster learning (Schneider et al., 2011), as enabling an enactive mode of reasoning activities. With physical tokens, the users are able to reason and explore concrete situations in 'ordinary language'. Putting these objects in structures as slots limits the movement of a token. This combination of token and constraint can represent a certain interaction syntax or relation (Ullmer & Ishii, 2005). One physical element can have different roles, as presented in Figure 13.

TUIs support collaboration between people, by allowing multiple users to interact simultaneously with the systems and keeping room for interaction with each other. They can promote and enhance social learning processes, due to shared representations that can facilitate interaction and reduce cognitive overload (Schneider et al., 2011).

Merging XAI and TUI

Colley et al. (2022) propose a conceptual framework where XAI merges with the field of TUI to provide intuitive and interactive interfaces. The tangible approach has been applied to different fields, as sound design and urban decision-making, but the field of AI is new. Tangible XAI (TangXAI) just started emerging, and practical examples are still limited.

The initial framework by Colley et al. (2022) helps to create an understanding between technical explanations on model-ground level and the viewpoint of a user in an embodied manner (on human-ground level). Combining these two levels seems as a helpful direction to improve understanding of citizens. The framework is shown in Figure 14.

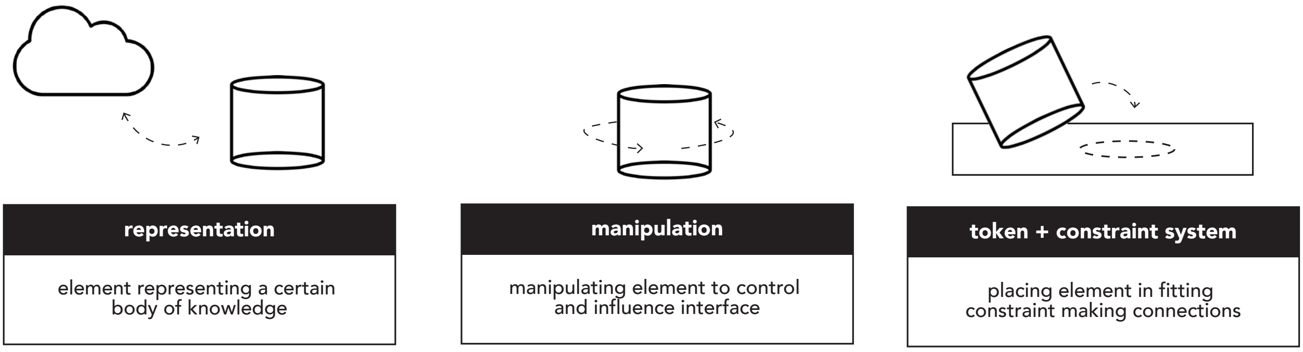


Figure 13: *Possible interactions with a tangible component*

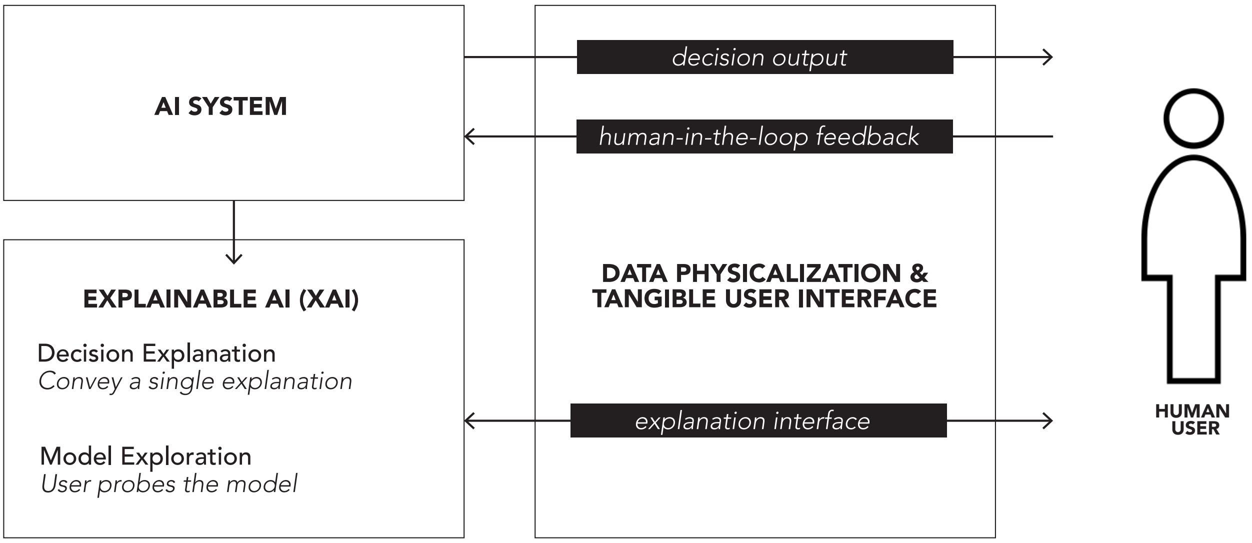


Figure 14: *Conceptual framework for TangXAI (adapted from Colley et al., 2022)*

Example: The Machine Learning Machine
An example of a TUI for teaching machine learning is the Machine Learning Machine made by Kaspersen et al. (2021). Young learners engaged in creating data, using a 'Trainer' to train a model and an 'Evaluator' to evaluate and test the model they have created. By influencing the performance, they were able to explore and reflect upon the Machine Learning model. Kaspersen mentioned how these iterations were triggered by reflecting on previous versions of the model, indicating that this improved their students understanding.

Figure 15: *The Machine Learning Machine (Kaspersen et al., 2021)*

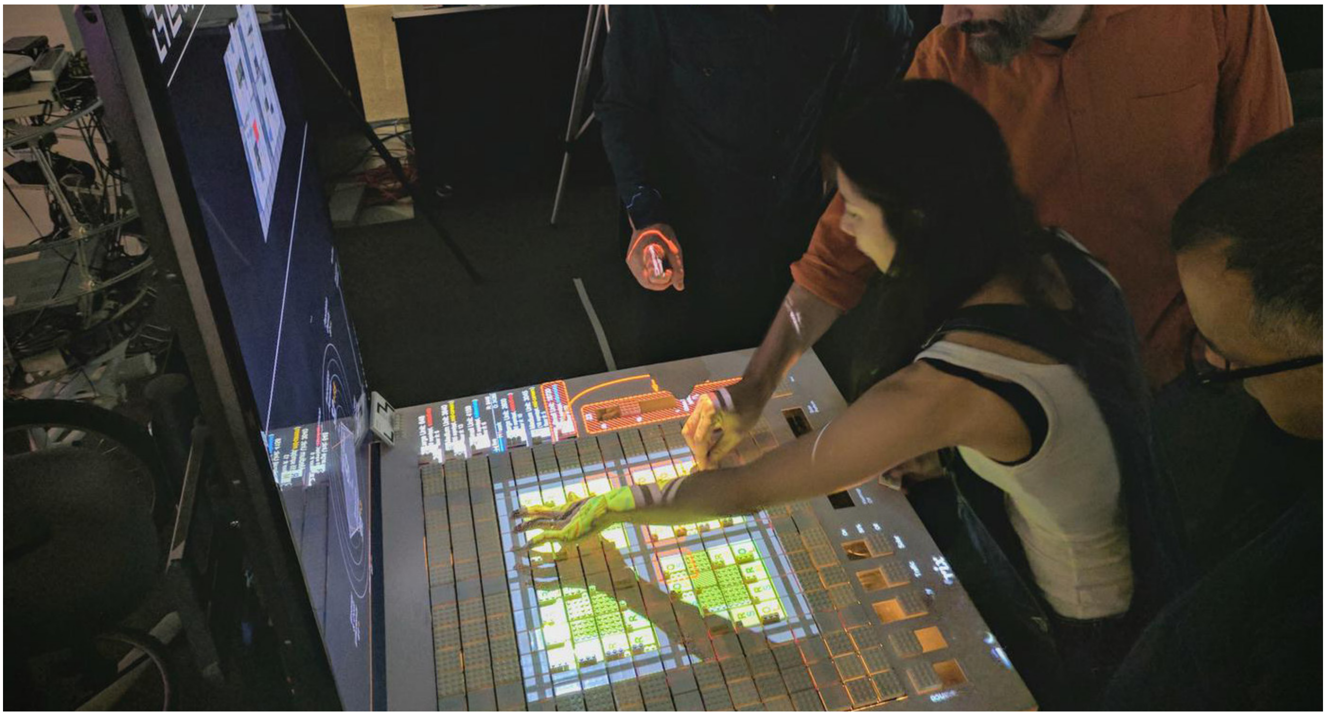


Figure 16: *CityMatrix (Zhang, 2017)*

Example: CityMatrix
CityMatrix is TUI for collective urban decision-making (see Figure 16). This tool helps urban professionals and non-professionals to *'understand the city better to make more collaborative and better-informed decisions'* (Zhang, 2017). By using tangible elements the decision-making process becomes more accessible. Providing real-time feedback and actual data helps the evaluation, which enables different explorations informing users of their own perspectives.

Example: Model Sketching
Lam et al., (2023) propose an interface based technical workflow where human-understandable concepts are presented to iteratively sketch the logic of a model (see Figure 17). This example combines XAI on human-ground level and iteratively exploration in the early stages of ML model design. Focusing on using concepts, stakeholders who lack technical expertise could be enabled to explore prototyping ML models.

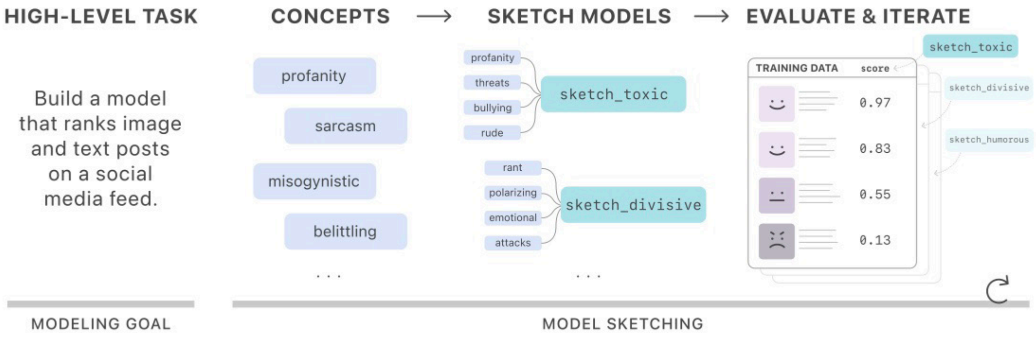


Figure 17: *Model Sketching workflow (from Lam et al., 2023)*

EVALUATING UNDERSTANDING OF AI

To understand if non-expert understanding of AI systems improved, various quantitative and qualitative evaluation methods are used. D.T.K Ng et al. (2021) extracted 3 main tools for assessing AI literacy.

Tools for Assessing AI Literacy

Firstly AI knowledge can be assessed via a knowledge test, where participants answer questions regarding AI concepts. Comparing pre-post tests can show improvement in cognitive knowledge of participants. Secondly, questionnaires are applied to assess perceptions or perceived abilities, using scales and open-ended questions. Combining these two elements, can give more insight into why something is perceived in a certain way. Cheng et al. (2019) used scales to identify participants' subjective understanding and knowledge of relevant topics.

Lastly, follow-up interviews are used to assess AI literacy. For example, Kaspersen et al. (2021) evaluated the Machine Learning Machine by interviewing participants after interacting with the artifact. This can be insightful to extract how participants can use knowledge and what remains unclear. Follow-up interviews give the space to go in depth around an answer in comparison with questionnaires.

Mixed-Method Approach

Cheng et al. (2019) used a mixed-method approach to investigate non-expert users understanding. With this approach quantitative an qualitative data are combined in one study, to provide more in-depth findings. This can include on one side quantitative questionnaire, combined with interviews.

CONCLUSION

In order to improve the understanding of AI of non-experts, information should be presented in a suitable and comprehensible way for its receivers. Explanations need to be carefully put together regarding how, when, and to whom they are presented to stay away from opposing effects. Understanding does not only consist of objective understanding but also has a subjective element. The feeling of not knowing can negatively influence their level of understanding of non-experts.

Contextualized explanations and interactive interfaces can improve comprehension of individual non-experts of machine learning algorithms. Adding tangible elements to this interface can improve shared understanding within groups, lowering the cognitive load and having a shared visible representation. The field of XAI and TUI merging could possibly help improve understanding of non-experts in a group setting. Having an iterative element can help reflection among participants, improving their understanding of a ML system.

Knowledge tests, questionnaires, and follow-up interviews can be used to evaluate the objective (knowledge test) and subjective understanding (questionnaire/interview) of AI concepts.

DESIGN METHODS FOR INTERACTIVE SYSTEMS

Where participatory design approaches involve stakeholders in the design process, other design methods can help to find fitting designs within this setting. As information should be presented in a comprehensible way for citizens, what is working depends on the settings and the citizens.

EXPLORING INTERACTIONS

Interactions happen everywhere, including how people use, understand, and experience products and situations. By exploring and analyzing these interactions, knowledge can be collected to inform the design. This can be seen as a search for the right thing: '*a product that transforms the world from its current state to a preferred state*' (Zimmerman, 2003). This holistic approach to design looks at the user, system, and context as connected (see Figure 18).

An interactive system consists of the different components, that together define the system a user interacts with. This includes content, structure, behavior, and appearance. Barfield et al. (1994) describe these elements as follows:

- *Content:* the functionality and information that is accessible to users.
- *Structure:* the apparent organization of information and functionality.
- *Behavior:* all potential system behaviors (manifest behavior), and all potential user actions (evoked behavior), together defining the potential for interaction.
- *Appearance:* all parts of the system that the user can perceive

INTERACTIVE SKETCHING

To find what is the right thing, quick & low effort artifacts can be used in the beginning to explore, question, or propose something before committing to a concept. Buxton (2011) describes this as sketches that during time can evolve into prototypes, being more specific and testable. By giving impressions of certain interactions, users can experience the interactions before investing in developing the actual human-computer interaction, giving room to multiple iterations. With 'Wizard-of-Oz' prototyping, different interactive concepts can be explored quickly, narrowing down the design space. These sketches can be critiqued internally by the designer and externally by putting them in front of possible users.

CONCLUSION

To design for certain interactive interactions between humans and human-computer, the contest of use needs to be explored to inform the design. By iteratively sketching user experiences and critique in context can inform finding the 'right' system.

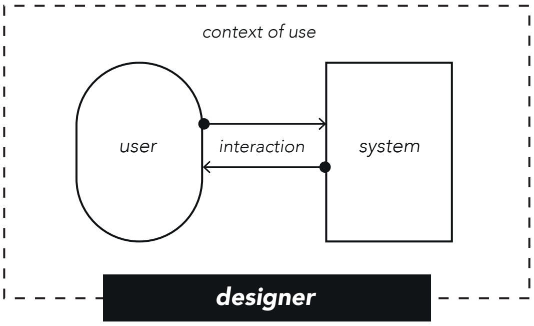


Figure 18: *Central model for interaction design (adapted from Barfield et al., 1994)*

CONCLUSION

This section combines the conclusions from the background research and related work. First, the design questions are complemented by design goals. This is followed by stating the design space and the compiled design requirements. Lastly, the design approach is summarized.

DESIGN GOALS

To legitimize the machine vision system being used by scar cars, citizens should be involved during the development process. To prevent participation-washing, this involvement should be transparent and facilitate genuine knowledge sharing. They should be able to discuss the acceptability of such a system. However, citizens lack awareness, knowledge, and understanding of relevant information to judge such a system.

Therefore the design should improve understanding of machine vision systems. The design goals indicate the desired effects of the design. This resulted in the formulation of Design Goal 1, divided into 3 sub-design goals. Next, the design should nurture a deliberate discussion within a participatory session. This is separated in the design triggering a discussion, but also enabling participants to form their own perspectives to take part. This objective is reflected in Design Goal 2, including 3 sub-design goals.

1: The design should improve the understanding of citizens about machine vision system

1A: The design should improve the understanding of citizens about what the scan car 'sees' and 'detects'
1B: The design should improve the understanding of citizens about the decision situation of the system (what happens with a detection)
1C: The design should improve the understanding of citizens about trade-offs within model training, tuning, and validation

2: The design should nurture a deliberate discussion between all participants (expert & non-expert)

2A: The design should trigger discussion about the machine vision system and its implications
2B: The design should empower citizens to take part in the discussion about when a machine vision system is acceptable
2C: The design should enable citizens to form their own perspectives about what is acceptable, and understand others

DESIGN REQUIREMENTS

Where the design goals indicate the desired effects of the design, research revealed several requirements. First, the use case of scan car development in Amsterdam defines the design space. To prevent participation-washing, citizen involvement should be transparent and provide genuine knowledge sharing. Analyzing the current development process, citizens should be involved between the creation of the proof-of-concept model and the actual model to actually have influence. This proposition, and envisioned context of use of the design is visualized in Figure 19.

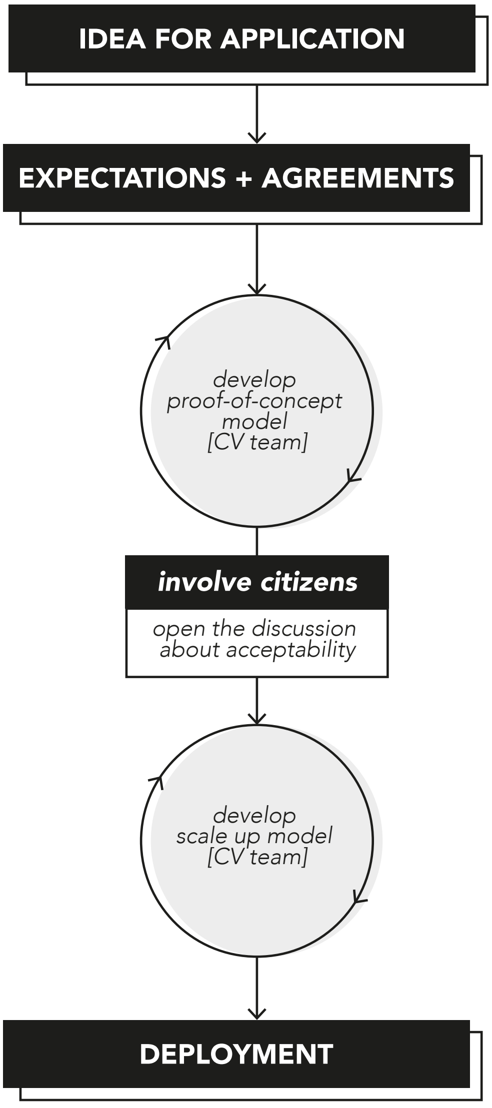


Figure 19: *Placement of design intervention*

To address the legitimacy of the machine vision model at this identified point, a discussion about the acceptability of the model should take place. The requirements for this discussion contain of various elements that are concluded in the following sections: (1) content, (2) structure, (3) behavior, and (4) materialization.

(1) Content

The content of the discussion should be the acceptability of a machine vision model. Acceptability is influenced by the trade-offs made during the development process. Therefore, these trade-offs need to be up for discussion to ensure legitimacy. They need to be placed in context to open up comprehensible questions.

Specifying this to the use case results in focusing on *Accuracy versus Interpretability* and *Precision versus Recall*. These trade-offs can evoke value tensions and friction, due to different interests of stakeholders. Next, due to the placement of the design intervention, is it difficult to for example judge efficiency of a proof-of-concept system that still needs to be scaled up. Therefore, the focus is directed to the two trade-offs, because these can inform and steer further development processes. To participate in this discussion certain basic knowledge is needed. So the design must contain and articulate the following information:

- Trade-offs: What is *Accuracy versus Interpretability* and *Precision versus Recall?*
- Modeling: what does the modeling phase entail?
- Evaluation Metrics: how is a model evaluated?
- Detection: what is a detection?
- Decision Situation: what are the consequences of a detection? How are they translated into a decision?

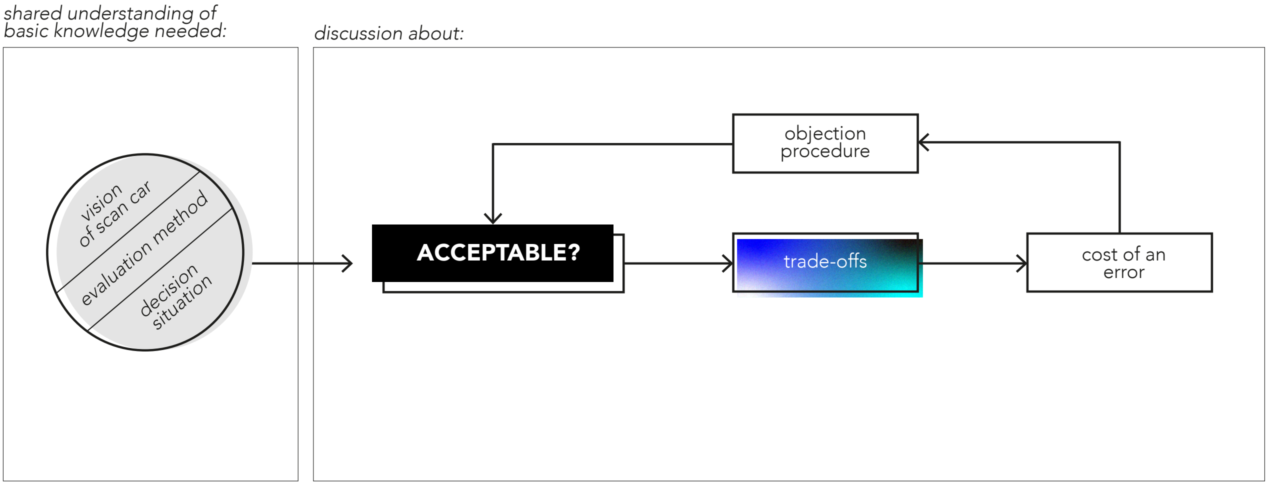


Figure 20: *Structure of needed information*

(2) Structure
This information (content) and its function needs to be ordered in a way comprehensible for the receivers. The relationship between the content elements and their role in the discussion is visualized in Figure 20. Some elements are needed to inform the discussion, while other elements are more directly connected to the discussion.

To prevent confusion from cognitive overload, the information needs to be unveiled gradually.

(3) Behavior
Placing this system in the context, various interactions should be supported, as visualized in Figure 21. First, the interaction between the user and the system should transfer the information to improve knowledge. This means that the way information is presented should match the non-expert receivers.

Next, interactions between the user and system (participant and design) should include technology. This interaction should be interactive and iterative, as this can trigger reflective thinking to improve understanding. This could include probing or manipulating something in the system, and receiving something back.

The design will be used within a participatory setting. Six to eight people, as a citizen panel, will be present. The design should enable possible interaction between these people, in the form of discussion or creating shared understanding. The design should enable the participants to form their own opinion about what is acceptable and discuss this with each other. Therefore the information should be presented in a way to facilitate this opinion-forming and promote critical thinking to facilitate deliberation.

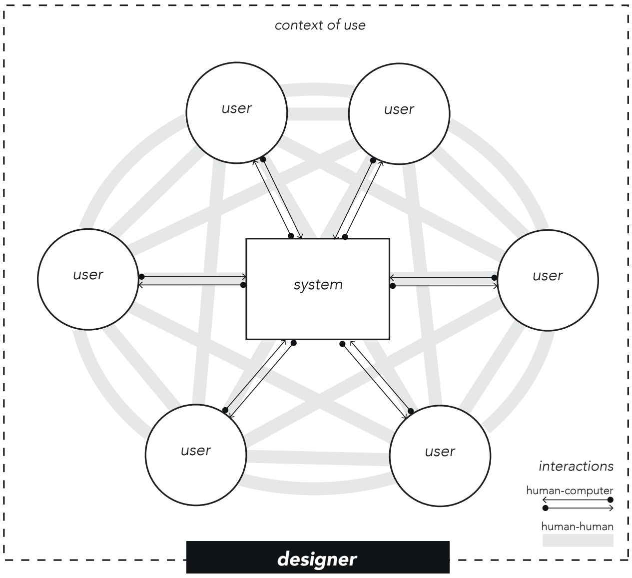


Figure 21: *Adjusted model for (group) interaction design*

To manage power imbalances between participants, all information should be available and comprehensible for everyone.

(4) Appearance
A system providing direct feedback can help improve understanding by triggering reflections.

Explanations should happen on human-ground level to match graspable concepts for non-experts. This can include contextualizing, giving examples, and placing theory in context. This opens up a space for materializing abstract concepts and information into tangible representations. As this can help individuals to better understand, it can also create a shared understanding within a group setting enhancing the interactions.

As the explanation and representations provided should match the receivers, what works will be iteratively explored.

DESIGN APPROACH

A holistic approach should help to explore the different interactions through interactive sketches. With the focus on quick and easy sketches, an iterative process can find suitable modes of explaining and discussing.

4. PROPOSED DESIGN

This chapter presents the conceptualization process of the design. Starting from an 'empty discussion', artifacts were added to generate responses and insights. Through an iterative process, the sketches evolved into a final concept. The proposed design is showcased in depth at the end of the chapter, including the structure, components, their interaction, and a journey of the design intervention.

DESIGN PROCESS

The design process had as a main point of view, as underlying perspective, 'trying something out to see if it works'. The combination of a complex topic yet a clear situation where the design should operate led to the focus of iteratively developing concepts fitting with the session and the participants.

STARTING POINT

The design process kicked off with a discussion about what is acceptable for a machine vision system in a participatory design setting. Without additional materials, this discussion functioned as the first iteration, the starting point, of the design. Based on the description of Zimmerman (2003) of designing the right thing, *'a product that transforms the world from its current state to a preferred state'*, led to the question: How can this session be changed and improved to reach the design goals? What can nurture this discussion and how?

Various iterations were made upon this discussion to address the design goals. By adding different artifacts 'to the table', it was explored what helps to clarify or question certain aspects of a machine vision system to improve the discussion. This process is visualized in Figure 22.

Figure 22: *Design discussion set-up*

DESIGN ACTIVITIES

The design process consisted of several design activities. Most activities can be characterized by an iterative design cycle, where a design is created, tested, and analyzed to inform the design again (see Figure 23). The created artifacts, evolving from sketches to prototypes, were first reviewed by internal critique, before presenting to peers for external critique. These group discussions informed the design regarding what is needed, working or not working, to revisit the design again.

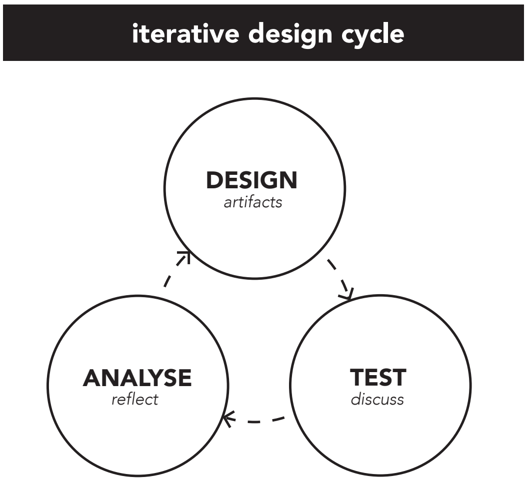


Figure 23: *Iterative design cycle within design activities*

The design activities can be placed on different exploration themes before converging into the final design (see Figure 24). Firstly the discussion in a participatory setting was explored to find and verify the need for certain information.

This information was translated into different artifacts to present 'on the table'. These artifacts can be described as 3D or interactive sketches. Characteristics of these artifacts were the focus on a low effort to create (use existing objects, paper objects) and usage to communicate. After getting a hold on suitable components, and what the artifacts should be, iteratively combining the elements together explore the interactions between and overall structure. This part of the design can be seen as a storyline or narrative addressing how to order the artifacts.

From these ordered artifacts, the final design stems. Moving away from paper objects, the artifacts are transferred into a more formalized design to appear more realistic, take less imaginary effort, and be taken more seriously by participants. The result of this last iteration is presented in the next chapter.

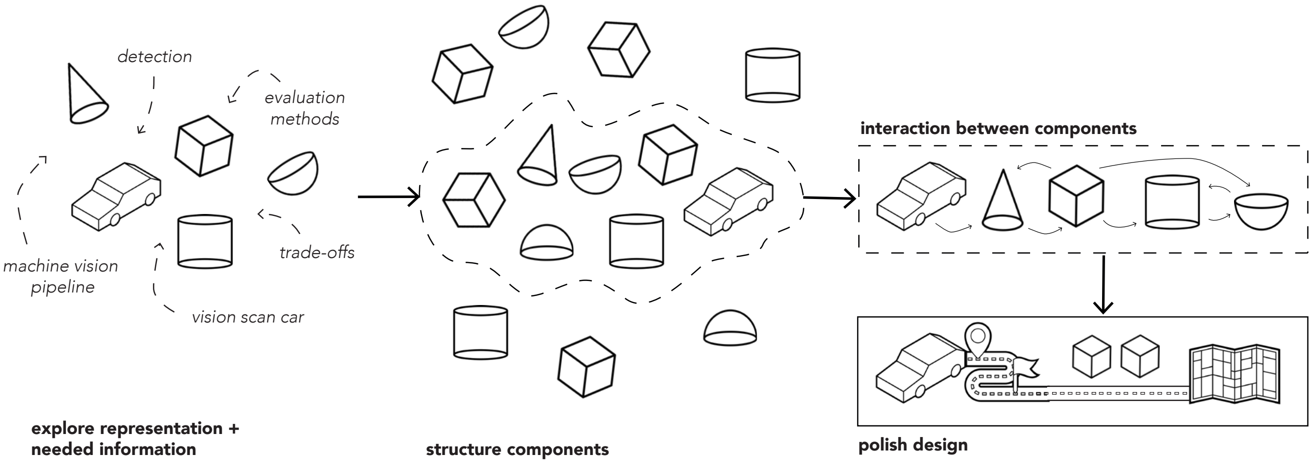


Figure 24: *Themes of design activities*

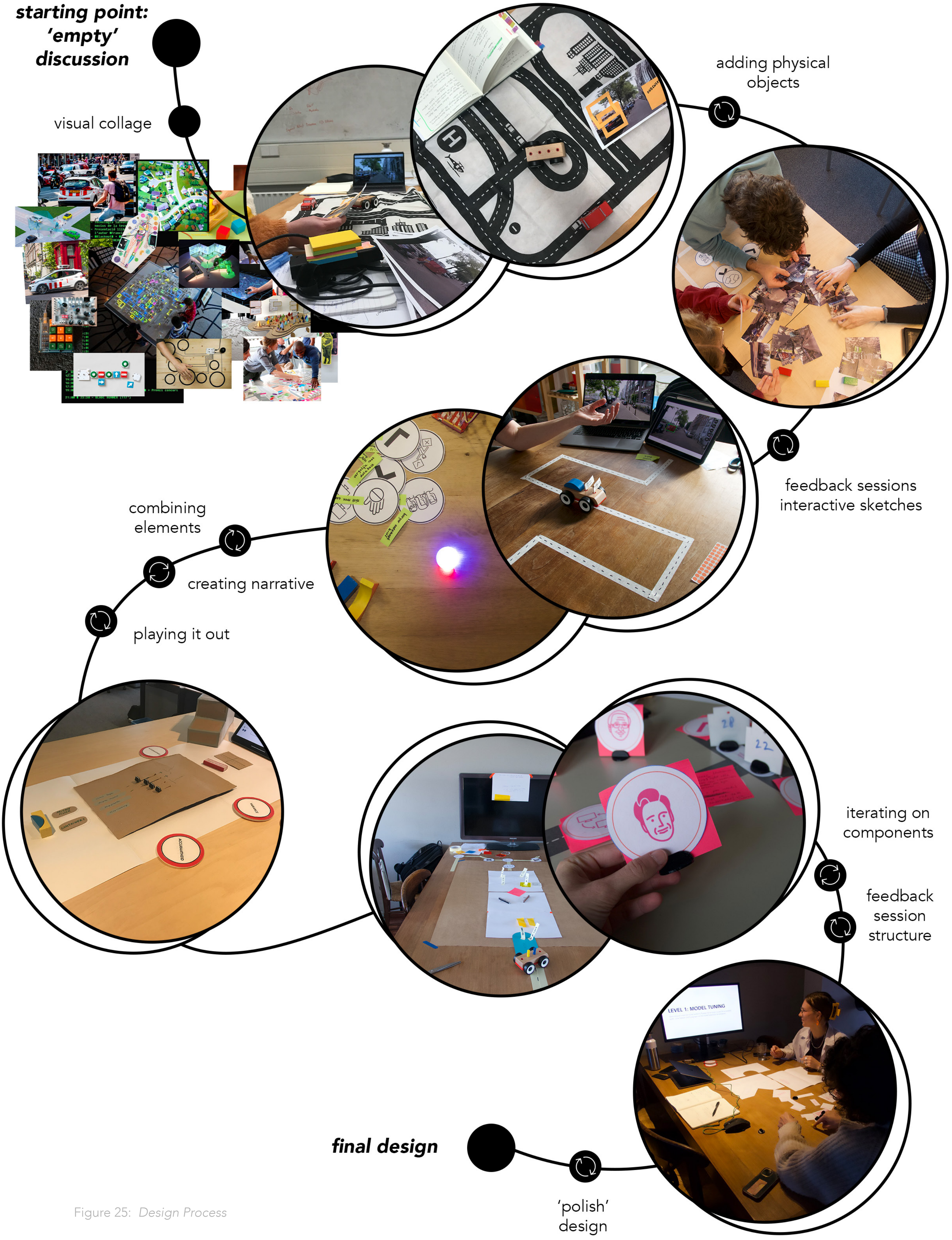


Figure 25: *Design Process*

KEY INSIGHTS FROM DESIGN ACTIVITIES

The different design activities led to various insights that informed the final design. These takeaways contain reasoning for certain design decisions and helped the design evolve. The key insights that led to the final design are summarized in the following pages.

EMPTY DISCUSSION

During the first 'empty' session it came forward that the participants need to understand the application and context in order to form an opinion. They need to understand what the system 'sees' because it is hard to envision.

- *Validation of design goals*
- *Show video clips and images to show vision*

TRANSLATING CONTEXT INTO PHYSICAL OBJECTS

Placing toy cars, printed images, and wooden blocks representing different abstract concepts helped to clarify the context. However, printed images of detections were not visible to everyone around the table. Next, participants need a specific scenario, case, and application to form their opinion, because this is hard to envision. Playing out the decision situation on the table, and physically going through this system is found to be very clear..

- *Use a central screen for certain information provisions to make what we are talking about visible to everyone at the same time*
- *Specify for collecting containers*
- *Play out the decision situation*

INTERACTIVE SKETCHES

Don't put everything on the same table at once, because participants will focus on different things, and experience an overload of stimuli.

- *Present components step-by-step*

Having constraints makes clear what object fits with what object, i.e. placing the collected 'data' of the scan car into the 'model', clarifies the connection. Currently, I am talking the participants through the interactive session, which puts me in the facilitating chair (not the desired power balance).

- *Use storyline or other instructions to prevent becoming the facilitator who talks*

Giving participants the opportunity to 'tune' the model makes them more critical: seeing the model perform less makes them critical towards the role of the developer. Reflecting on these sketches led to the insight that trade-offs were not mentioned, thus these need to be added somehow.

- *Let participants tune the model, even if it performs less*

CREATING NARRATIVE

Different artifacts have different connections with each other. These components need to be structured in a way the relations become clear. First, the components were presented from the context, tuning the model, and later on the decision situation. However, this lead to getting acquainted with the trade-offs and acceptability of the model in the end, while being the desired center point of discussion:

they need to get introduced earlier on, with clear definitions before using these terms.

- *Introduce trade-offs and other relevant terms earlier on, before using them*

In this narrative, tuning happened before having an understanding of the consequences of the model. This lead to being uncertain about where to tune the model towards, and decide what is acceptable.

- *Make sure people understand the consequences of the model before tuning*

Where some elements were thought to provoke discussion around a certain topic, this did not always happen. Thus, participants need an extra push/stimulus to start this discussion. This is in line with people needing to have a goal or clear objective of why they are tuning this model.

- *Ask clear and specific questions if you want participants to think about something*
- *Give clear instructions on what to do: introduce game elements to have a clear objective of the session.*

Because it became clear that different stakeholders would not be present at the table, the choice was made to introduce different perspectives. During this introduction, the participants took these perspectives as their player role. This resulted in overacting, and not thinking about what they themselves find important

- *Present perspectives not as a role but as another opinion that is out there*

Let people try out something iteratively to get more familiar and explore different consequences

- *Iteratively tune the model to become aware of what is more acceptable*

FINAL CONCEPT

The final concept contains similar components as previous iterations. The final concept can be seen as an intervention that tries to provide information, and opens up the process of machine vision development, by including users to think about certain decisions made. This includes elements about the whole system, the consequences, and the making of the model.

The design translates abstract concepts as machine vision and model tuning into tangible components to explore the whole system together in a group session.

The intervention can be approached as a game, searching for what it means to be an acceptable model. During this game, players are exploring different aspects regarding what is acceptable.

In this variant, the game is focused on making a machine vision model for detecting containers in Amsterdam. The relevant trade-offs for this use case are *Accuracy versus Interpretability* and *Precision versus Recall*.

GAME STRUCTURE

Due to the complexity of the content, and experience and knowledge differences of the players, the structure of the design will go in depth gradually as visualized in Figure 26. Inspired by the steps identified to address the legitimacy gap, the game has different phases.

Introduction

During the introduction phase, players will be introduced to the machine vision system and the relevant trade-offs. In order to understand these trade-offs, players will be provided with information regarding what is a detection, relevant terms, and evaluating model outcomes (for this model the confusion matrix). Going through this introduction together is done in order to create a shared understanding of elements.

Level 1: Explore Consequences

In order to form an opinion about the trade-offs and their acceptability, the players discover the consequences by exploring the model outcomes. This is done by evaluating the model in the context of the whole system (level 1).

Level 2: Model Tuning

After exploring the model in context, the players will iteratively explore tuning the model to evaluate and aggregate their preferences. During level 2 the players are introduced to other perspectives of citizens, in order to stimulate deliberative consideration. After this introduction, the players have the chance to update their model if desired.

The end of the game is reached when the players have used all their attempts to tune the model.

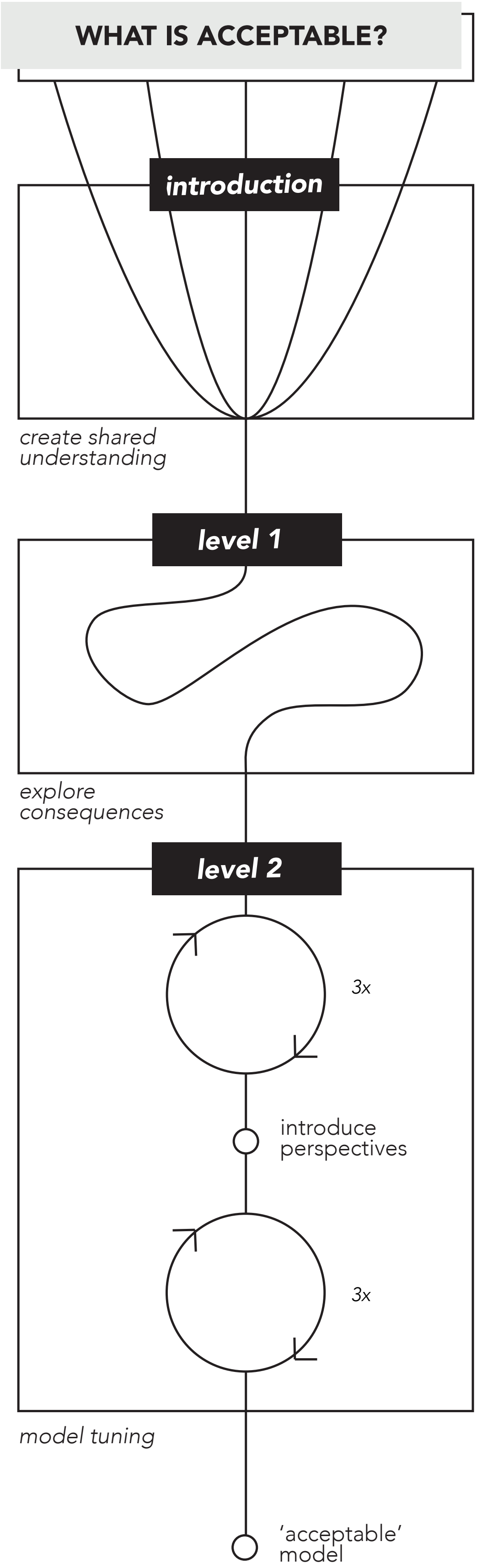


Figure 26: *Game structure*

COMPONENTS OF THE GAME

The design intervention consists of various components to play with. It is a combination of physical playing pieces on the table and an interactive interface at the head of the table. The physical elements and interface interact with each other in various ways. These interactions are explained in the journey.

The interface, made in Figma, is shown on a screen, and controlled with a wireless mouse by the players. As a central point, visible to everyone, the screen is used to clarify the operating context and certain elements of information. Showing elements as the 'vision' of the scan car, and what is 'sees' as a detection is more clear than communicating these elements with words. Next to that, the screen shows the confusion matrix and its explanations, trade-offs, and the changing model performances of the players.

The machine vision model phases during development are represented by various parts, functioning as a guiding *playing field*:

- Input
- Model Tuning
- Evaluate
- Decision Situation

The *playing pieces* that interact with the playing field and interface are:

- Trade-off Tokens
- Detections Pawns
- System Icons
- Citizen Perspectives
- Tuning Elements

These elements are shown on the following pages. This is followed by an 'experience journey' of the game, where the steps of the design intervention are played out step by step.

guiding placeholders playing field

PLAYING PIECES

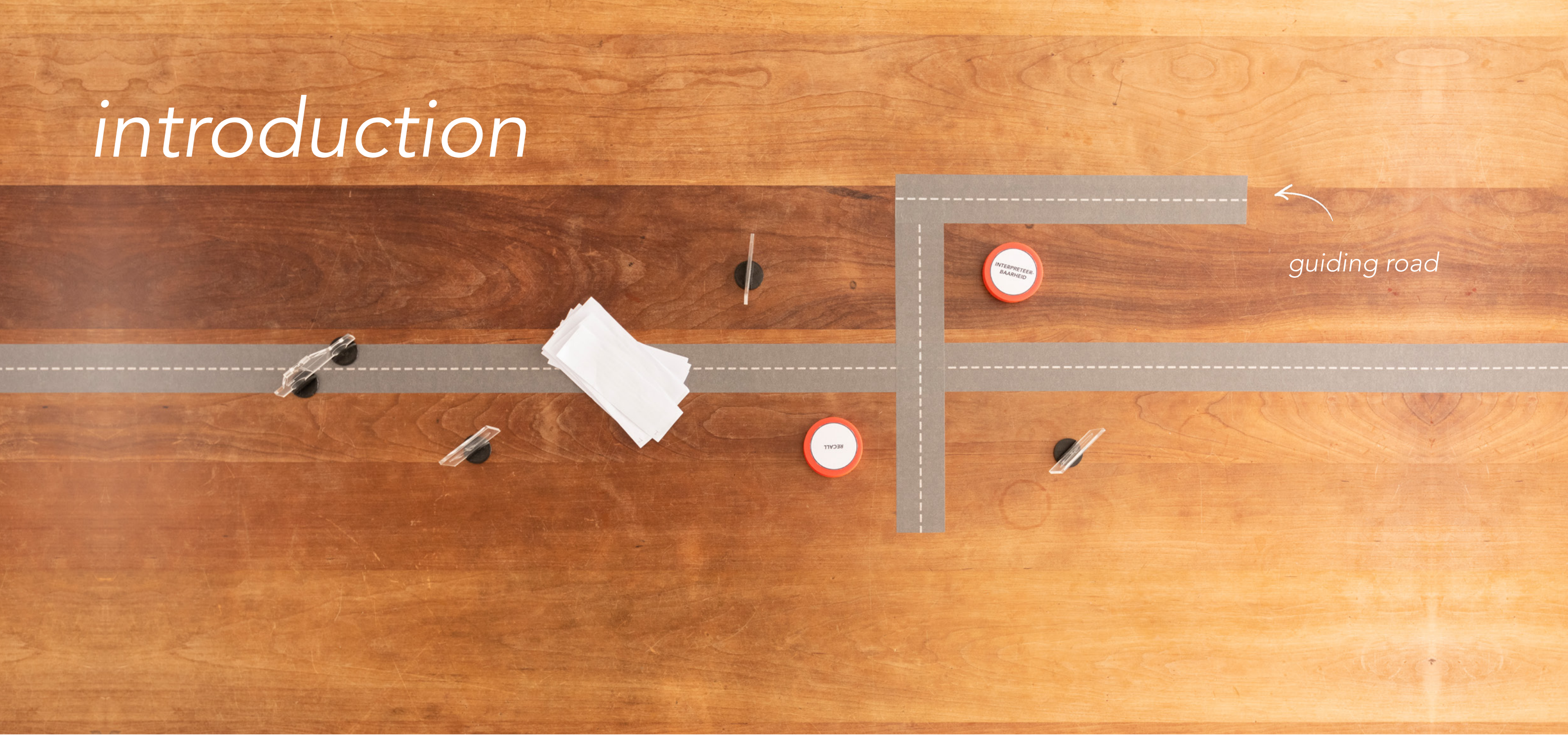
The playing pieces are movable tangible playing elements that represent certain parts of information. These elements are introduced during the different phases to complete the machine vision system.

The playing pieces are made out of various colors of PMMA with line and surface engravings. Some of these engravings are filled in with paint in order to emphasize what is engraved. Being made out of this material, we are able to use whiteboard markers on the material. Pieces as the trade-offs, citizen perspectives, and system icons are supplemented with informational stickers on the other side.

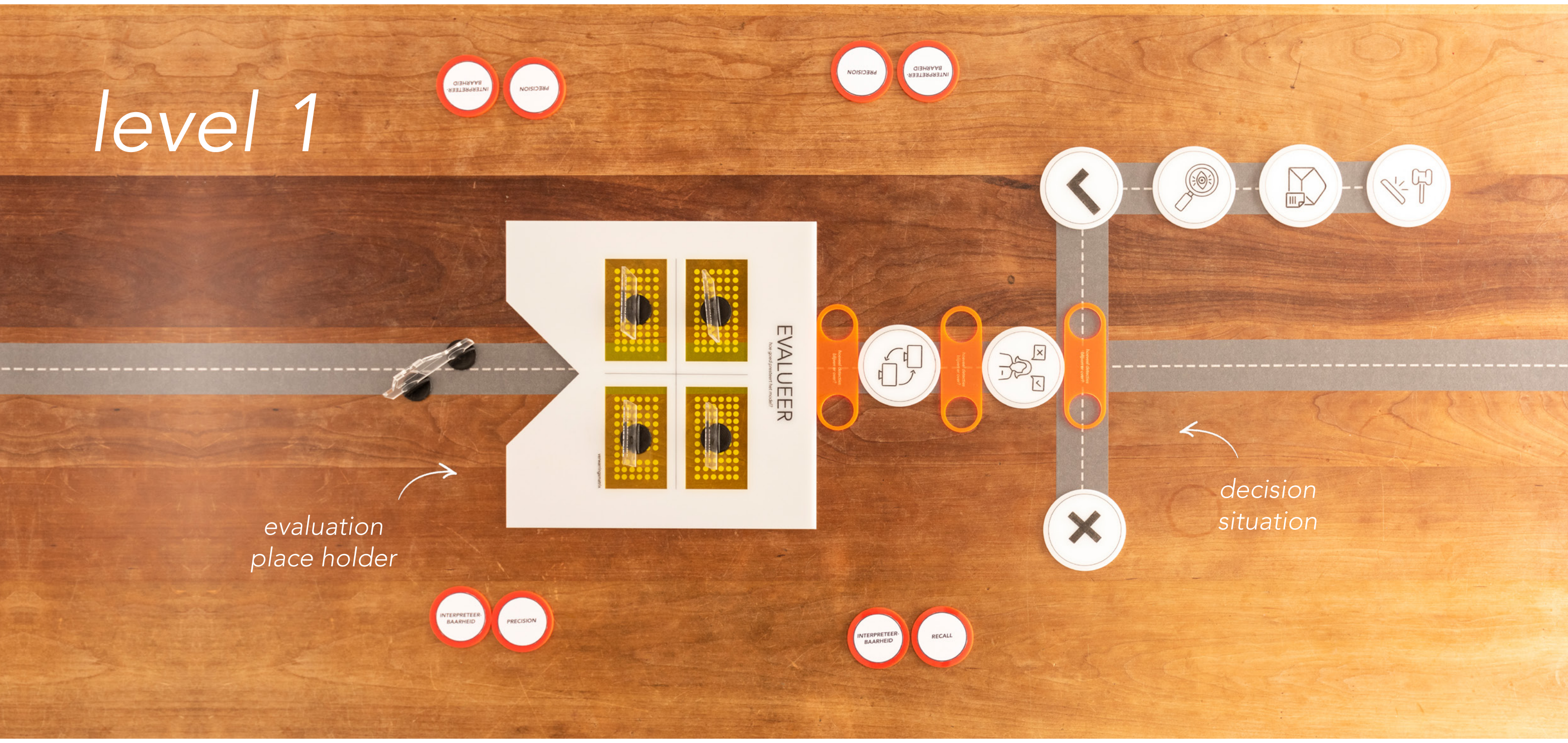
In order to make the playing field more interactive, actual buttons, sliders, and rotary knobs are added to play with. Transparent elements as the scan car and containers are kept upright by little holders. The black sliders and pawn holders are 3D printed with black PETG.

The next page shows an overview of the table setting as the playing field, throughout the different phases. This includes the playing pieces in their place.

introduction



level 1

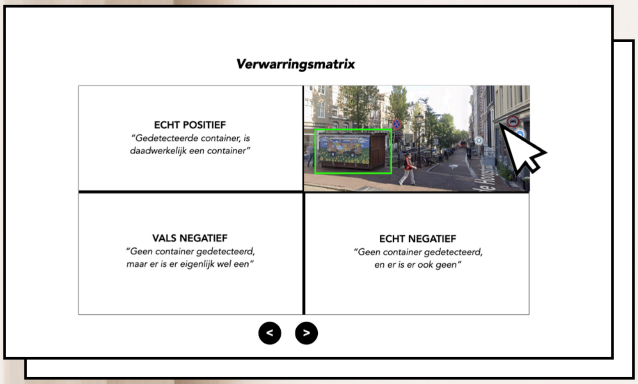


level 2

INTERFACE ON SCREEN

The playing field is combined with an interface shown on a screen. This interface is made within Figma to add various interactive elements to it.

An overview of all screens and a link to the Figma project can be found in Appendix A.



Verwarringsmatrix

ECHT POSITIEF
"Gedetecteerde container, is daadwerkelijk een container"

VALS NEGATIEF
"Geen container gedetecteerd, maar er is er eigenlijk wel een"

ECHT NEGATIEF
"Geen container gedetecteerd, en er is er ook geen"

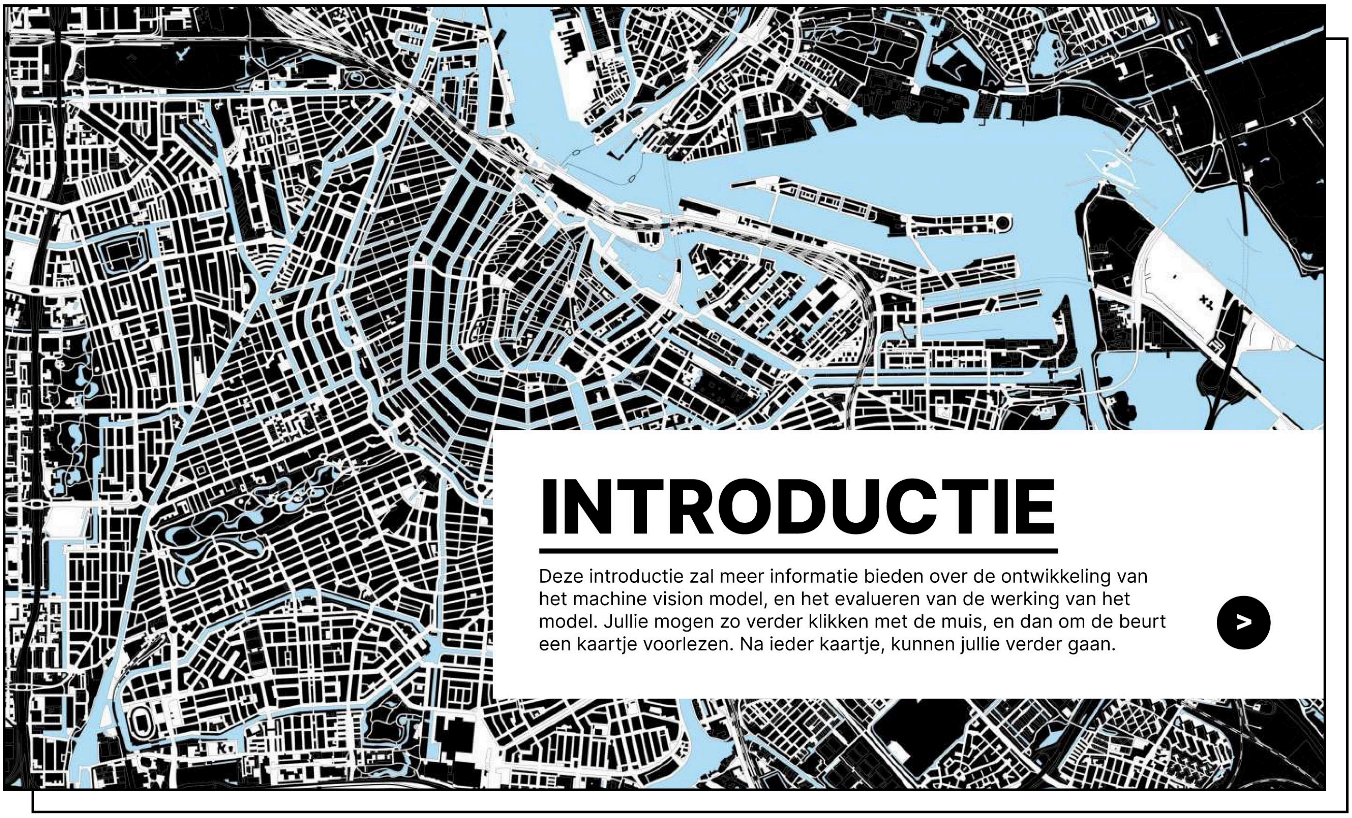
TRADE-OFF

Accuraatheid versus Interpreteerbaarheid

Is het belangrijker om heel accuraat te zijn, of is het belangrijker om te begrijpen hoe de detecties precies tot stand zijn gekomen?

Leg boven wat jij belangrijker vind

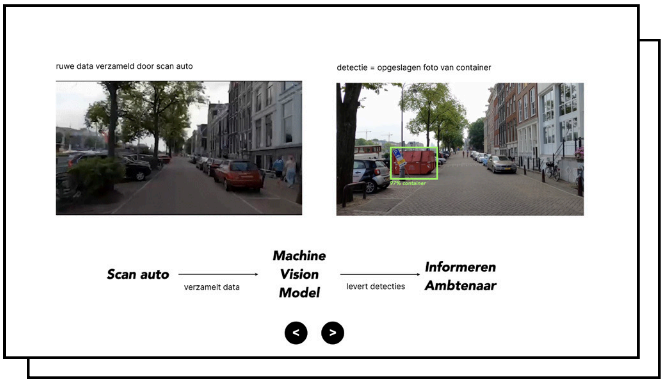
some of the screens



INTRODUCTIE

Deze introductie zal meer informatie bieden over de ontwikkeling van het machine vision model, en het evalueren van de werking van het model. Jullie mogen zo verder klikken met de muis, en dan om de beurt een kaartje voorlezen. Na ieder kaartje, kunnen jullie verder gaan.

>



ruwe data verzameld door scan auto

detectie = opgeslagen foto van container

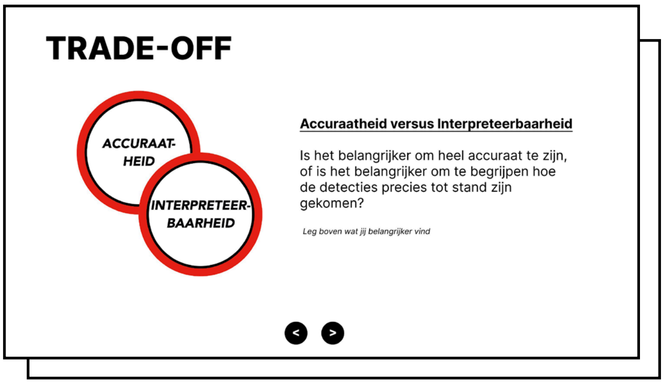
Scan auto → verzamelt data → **Machine Vision Model** → levert detecties → **Informeren Ambtenaar**

< >

Verwarringsmatrix

ECHT POSITIEF *"Gedetecteerde container, is daadwerkelijk een container"*	
VALS NEGATIEF *"Geen container gedetecteerd, maar er is er eigenlijk wel een"*	**ECHT NEGATIEF** *"Geen container gedetecteerd, en er is er ook geen"*

< >



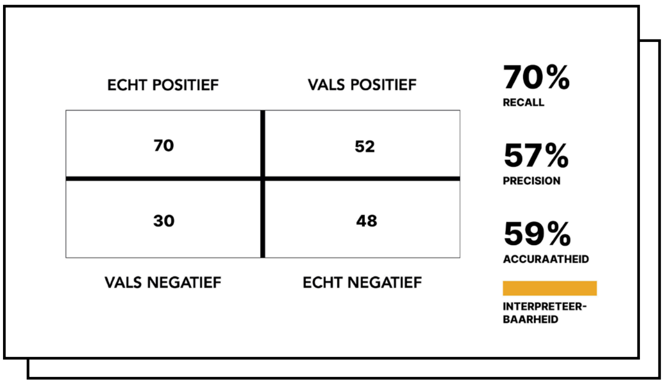
TRADE-OFF

Accuraatheid versus Interpreteerbaarheid

Is het belangrijker om heel accuraat te zijn, of is het belangrijker om te begrijpen hoe de detecties precies tot stand zijn gekomen?

Leg boven wat jij belangrijker vind

< >



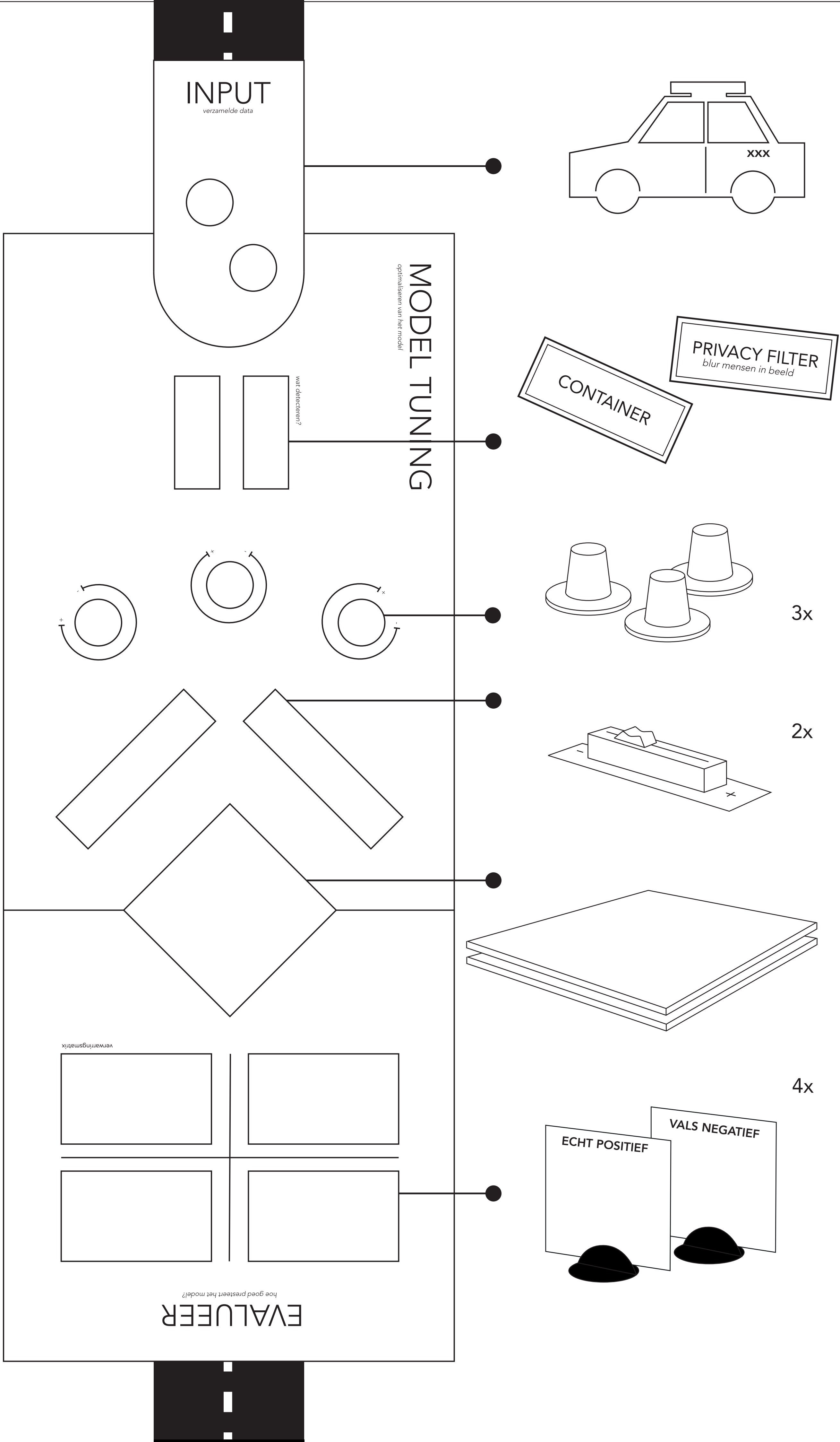
ECHT POSITIEF	VALS POSITIEF
70	52
30	48
VALS NEGATIEF	ECHT NEGATIEF

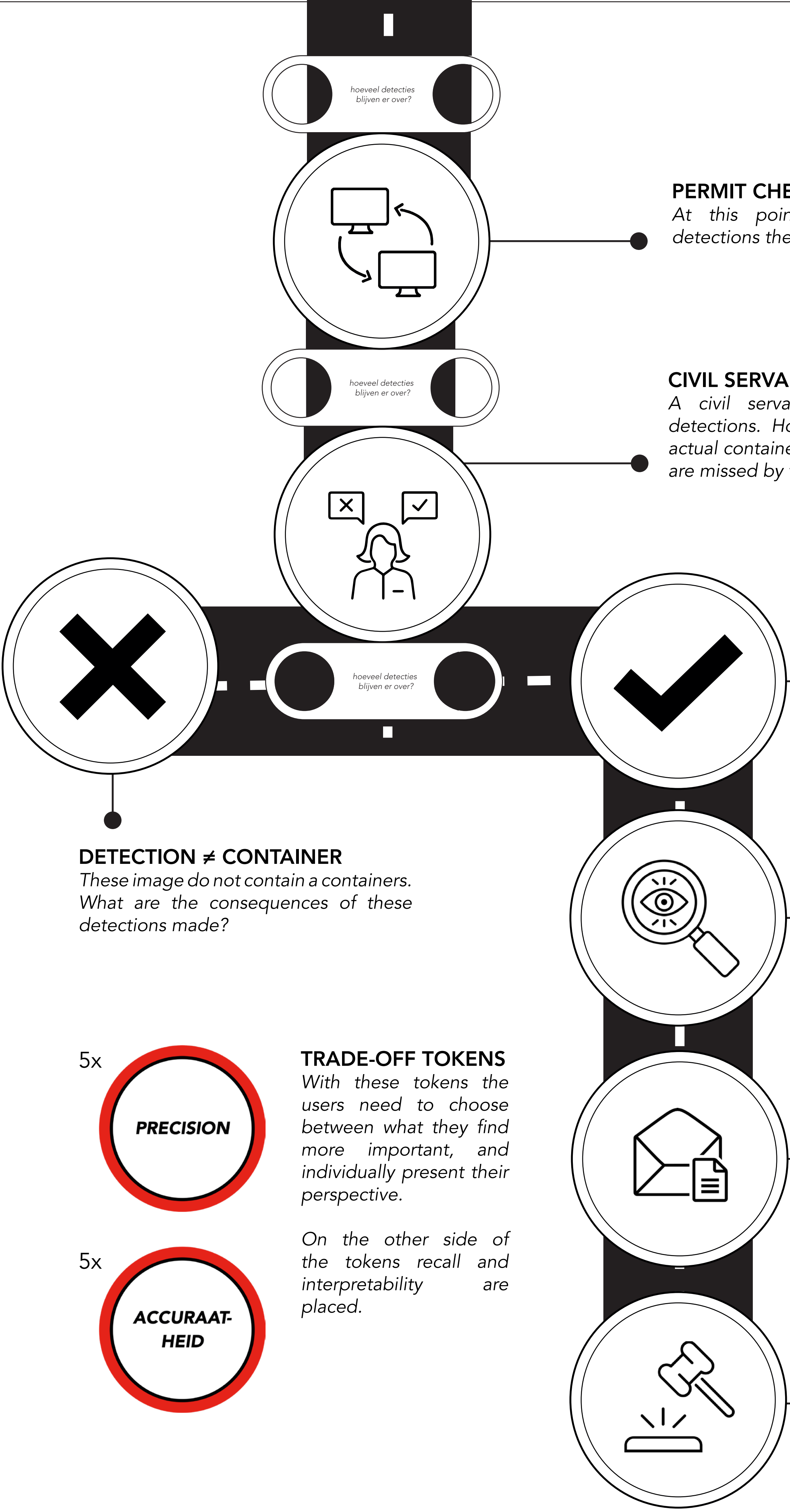
70%
RECALL

57%
PRECISION

59%
ACCURAATHEID

INTERPRETEER-BAARHEID


INPUT
verzamelde data
MODEL TUNING
optimaliseren van het model
wat detecteren?
xxx
PRIVACY FILTER
blur mensen in beeld
CONTAINER
3x
2x
4x
verwarringsmatrix
ECHT POSITIEF
VALS NEGATIEF
EVALUEER
hoe goed presteert het model?



PERMIT CHECK
At this point, check for how many detections there is a permit registered.

CIVIL SERVANT
A civil servant checks the remaining detections. How many detections are of actual containers? What if some containers are missed by the model?

DETECTION ≠ CONTAINER
These image do not contain a containers. What are the consequences of these detections made?

DETECTION = CONTAINER
These images are of containers without a permit.

INVESTIGATE
Civil servant goes to the location to decide what needs to happen next.

RECEIVE FINE
Owner receives a fine or letter. What if someone else does not get this, because the container was missed?

OBJECTION
If you disagree, you can object. How interpretable should the model be? What if other objects are missed by the model, and we know of the error rate?

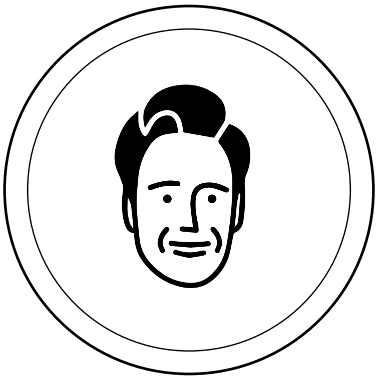
5x **PRECISION**

5x **ACCURAAT-HEID**

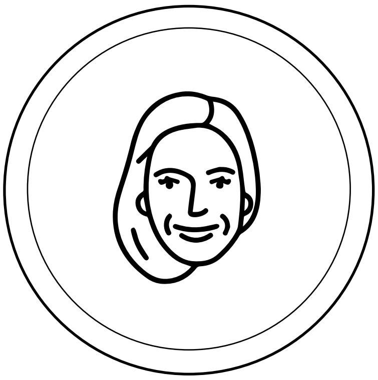
TRADE-OFF TOKENS
With these tokens the users need to choose between what they find more important, and individually present their perspective.

On the other side of the tokens recall and interpretability are placed.

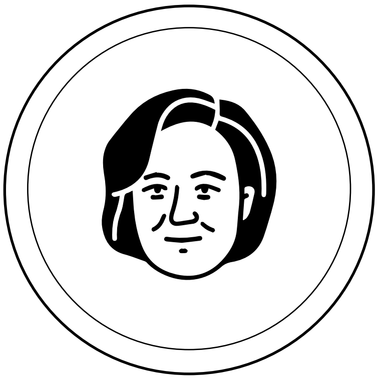
RALPH, 41
A civil servant who checks the images and investigates the locations. Wants accurate system, otherwise why would you want it?



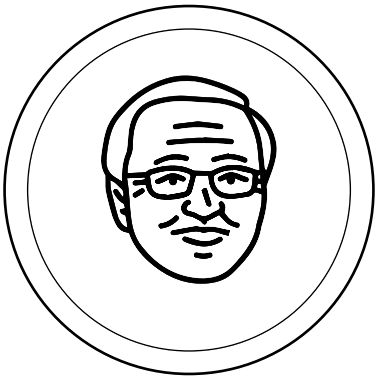
JORIS, 37
A citizen working in construction, thus often in need of a container. If I get a fine, everyone should get one. But I rather do not have a system like this.



CAMILLE, 31
A citizen living in the city center. If containers are illegally occupying parking spots or places on the pavements, there need to be strict enforcement



AMANDA, 54
A civil servant wanting to have interpretable systems if we use them in connection with citizens, even though they become less efficient.



MICHEAL, 61
A citizen, not fond of using technologies like scan cars. Wants it as humane as possible: like a person model needs to be able to explain its decisions.

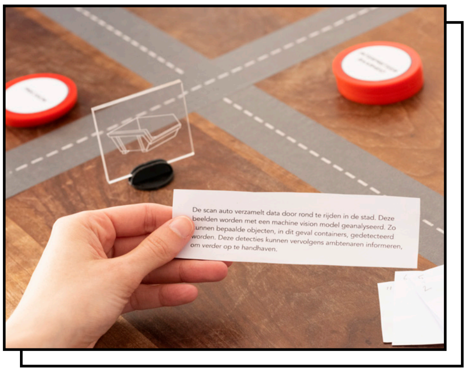
introduction

START OF INTERVENTION

The intervention starts of with following ‘playing field’: containers, a scan car, a road, trade-off tokens, instruction cards and the interface. The arrangement will slowly be reassembled into a more complex playing field of the complete machine vision system.

READ ONE BY ONE

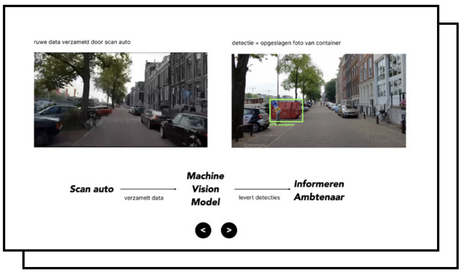
The users are instructed to one by one read a card from the table. These cards contain more information about the topic. After each card is read aloud, the users continue by clicking on the screen with their mouse. This procedure continues throughout the whole introduction.



De scan auto verzamelt data door rond te rijden in de stad. Deze beelden worden met een machine vision model geanalyseerd. Zo kunnen bepaalde objecten, in dit geval containers, gedetecteerd worden. Deze detecties kunnen vervolgens ambtenaren informeren, om verder op te handhaven.

VISION OF SCAN CAR

What does the scan car see? How is a machine vision model used in order to analyze the data collected by the scan car? What is left over as a detection?



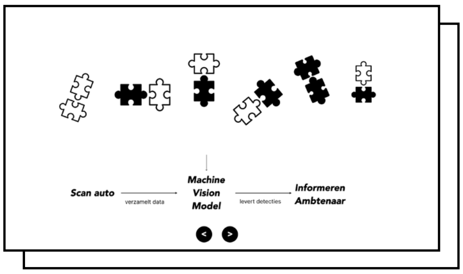
ruwe data verzameld door scan auto

detectie = opgeslagen foto van container

Scan auto → verzamelt data → **Machine Vision Model** → levert detecties → **Informeren Ambtenaar**

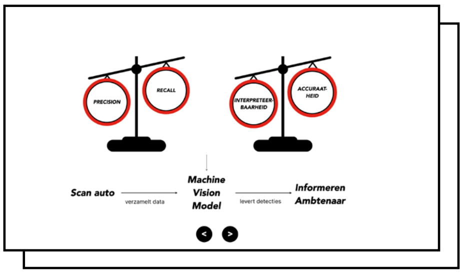
OPTIMIZATION OF THE MODEL

We want this model to detect containers as good as possible. This process of improving the model is known as the optimization of it. However, what the 'best performing model' is, can be different for different people. Therefore the citizens are involved, and are reminded here of their purpose.



INTRODUCE TRADE-OFFS

At this point the two trade-offs are introduced. The focus of this intervention is accuracy versus interpretability, and precision versus recall.

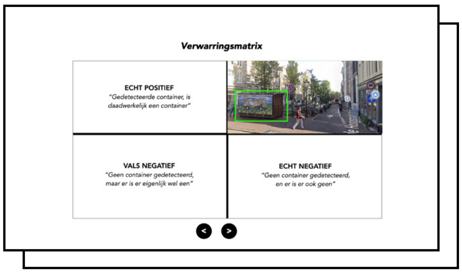


DIVIDE TOKENS & MAKE A CHOICE

The players are now instructed to divide the trade-off tokens, so that everyone has two in front of them. Without an explanation of the terms, the players need to choose what they think is more important. By turning the tokens, they lay the more important elements above.

INTRODUCE CONFUSION MATRIX

In order to learn more about the terms on the tokens, an explanation of evaluating the model and its performance is given. This contains an interactive confusion matrix, where definitions in context and example images of certain detections are combined. If the four possibilities are clear, the players can continue.



Verwarringsmatrix

ECHT POSITIEF
"Gedetecteerde container, is daadwerkelijk een container"

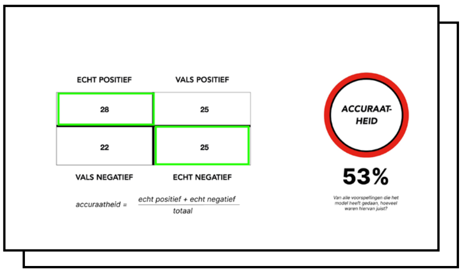
VALS NEGATIEF
"Geen container gedetecteerd, maar er is er eigenlijk wel een"

ECHT NEGATIEF
"Geen container gedetecteerd, en er is er ook geen"

< >

DEFINING ACCURACY

The players receive an explanation of what accuracy means within evaluation a machine vision model. With this explanation the confusion matrix can show how this percentage comes about.



ECHT POSITIEF	VALS POSITIEF
28	25
22	25
VALS NEGATIEF	ECHT NEGATIEF

$$\text{accuraatheid} = \frac{\text{echt positief + echt negatief}}{\text{totaal}}$$

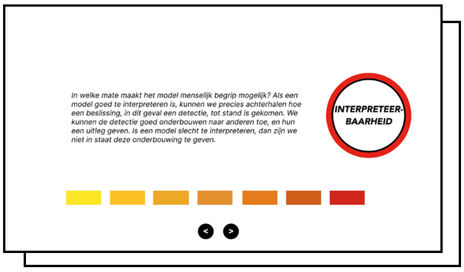
ACCURAAT-HEID

53%

Van alle voorspellingen die het model heeft gedaan, hoeveel waren hiervan juist?

DEFINING INTERPRETABILITY

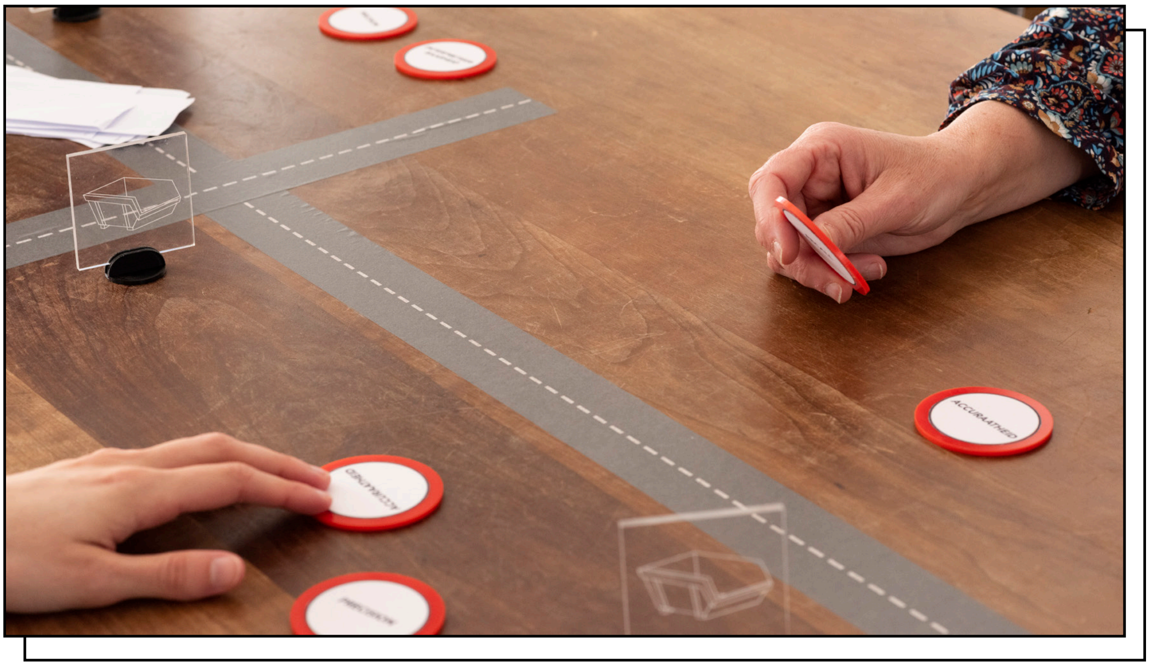
The other side of the trade-off is introduced by clicking on the token on the screen. Then the term interpretability is introduced with an explanation and a color coding.



In welke mate maakt het model menselijk begrip mogelijk? Als een model goed te interpreteren is, kunnen we precies achterhalen hoe een beslissing, in dit geval een detectie, tot stand is gekomen. We kunnen de detectie goed onderbouwen naar anderen toe, en hun een uitleg geven. Is een model slecht te interpreteren, dan zijn we niet in staat deze onderbouwing te geven.

INTERPRETEER-BAARHEID

< >

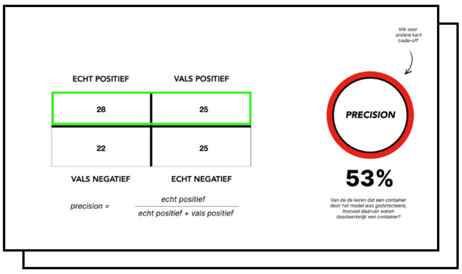


CHOOSE BETWEEN TRADE-OFF

The players are asked to choose what they think is more important based on the information that is available now. This is done with the aim to trigger discussion, and create shared understanding about what people think the terms imply.

DEFINING PRECISION

The players introduce each other to the fact that we should not only focus on what the model does right, but also should look into the mistakes that the model makes. With precision we look into, of all the detections of containers that the model made, how many images do actually contain a container?



klik voor
andere kant
trade-off

ECHT POSITIEF	VALS POSITIEF
28	25
22	25
VALS NEGATIEF	ECHT NEGATIEF

$$precision = \frac{echt\ positief}{echt\ positief + vals\ positief}$$

53%

Van de de keren dat een container door het model was gedetecteerd, hoeveel daarvan waren daadwerkelijk een container?

DEFINING RECALL

Another mistake the model can make is missing containers, by not making a detection while the model should have made one. This is evaluated with the term recall. So how many false negative instances do we allow in the model performance?

ECHT POSITIEF	VALS POSITIEF
28	25
22	25
VALS NEGATIEF	ECHT NEGATIEF

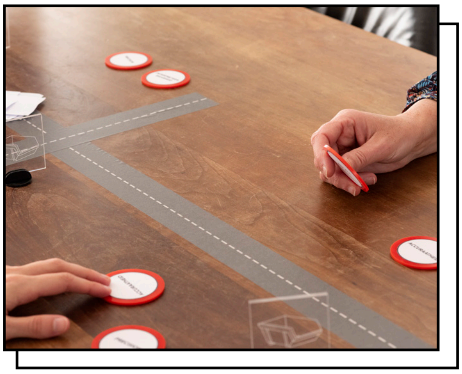
$$recall = \frac{echt\ positief}{echt\ positief + vals\ negatief}$$

56%

Van alle containers die gedetecteerd zouden moeten worden, hoeveel waren er daadwerkelijk gedetecteerd?

CHOOSE BETWEEN TRADE-OFF

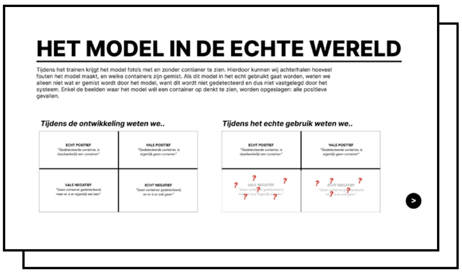
If we want to minimize one of these mistakes, in practice it will lead to more mistakes of the other sort. Thus the players are asked to choose what they find more important to optimize for. This is again done with the aim to trigger discussion, and create shared understanding about what people think the terms imply.



level 1

EXPLORING THE DECISION SITUATION
In order to form an opinion about what is important, the players need to see the model in context of the whole system. The model makes detections, but the system around it can translate these detections into a decision. During level 1, players explore what steps there are in this complete machine vision system used for enforcing containers. The playing field is expanded: a physical confusion matrix, detection pawns, placeholders and system icons are added to the road. These system icons contain information about the process, and questions to trigger discussion about certain consequences.

FROM TRAINING TO DEPLOYMENT
Before entering the playing field of the table, the players are explained that, if the model is used in reality we are not aware of the 'negative detections'. These aren't made or saved. Only during training the model, we have an idea of this percentage. Thus we need to keep in mind that in reality a similar percentage of containers will be missed.



HET MODEL IN DE ECHTE WERELD

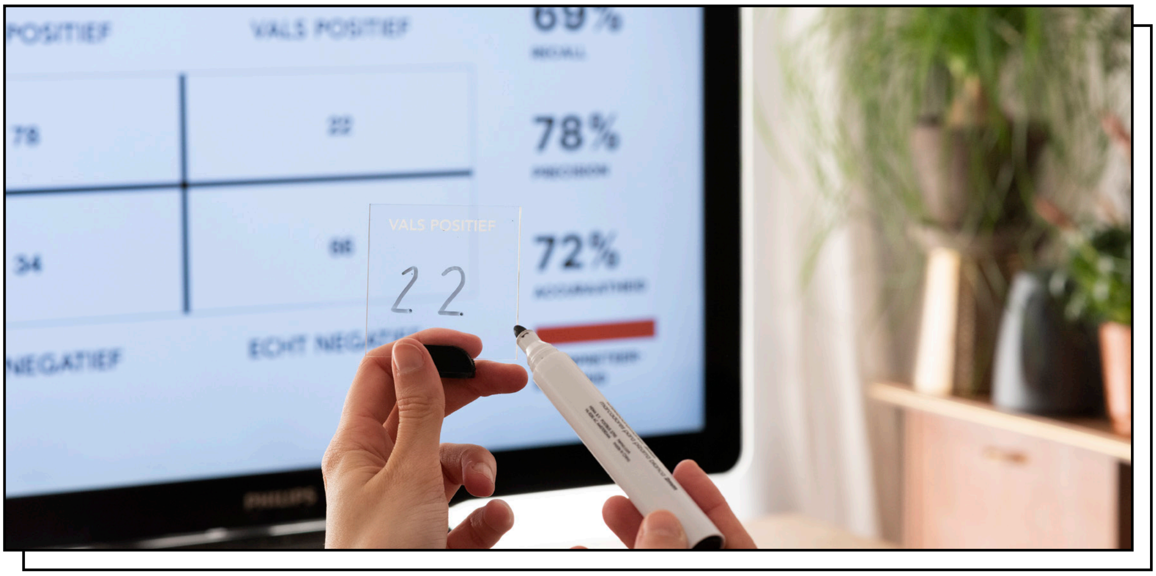
Tijdens het trainen krijgt het model foto's met en zonder contianer te zien. Hierdoor kunnen wij achterhalen hoeveel fouten het model maakt, en welke containers zijn gemist. Als dit model in het echt gebruikt gaat worden, weten we alleen niet wat er gemist wordt door het model, want dit wordt niet gedetecteerd en dus niet vastgelegd door het systeem. Enkel de beelden waar het model wél een container op denkt te zien, worden opgeslagen: alle positieve gevallen.

Tijdens de ontwikkeling weten we..

ECHT POSITIEF "Gedetecteerde container, is daadwerkelijk een container"	VALS POSITIEF "Gedetecteerde container, is eigenlijk geen container"
VALS NEGATIEF "Geen container gedetecteerd, maar er is er eigenlijk wel een"	ECHT NEGATIEF "Geen container gedetecteerd, en er is er ook geen"

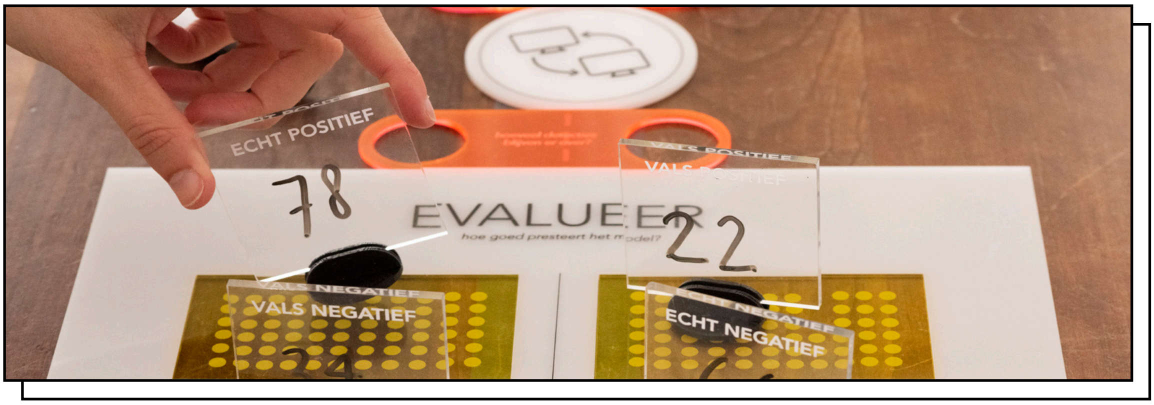
Tijdens het echte gebruik weten we..

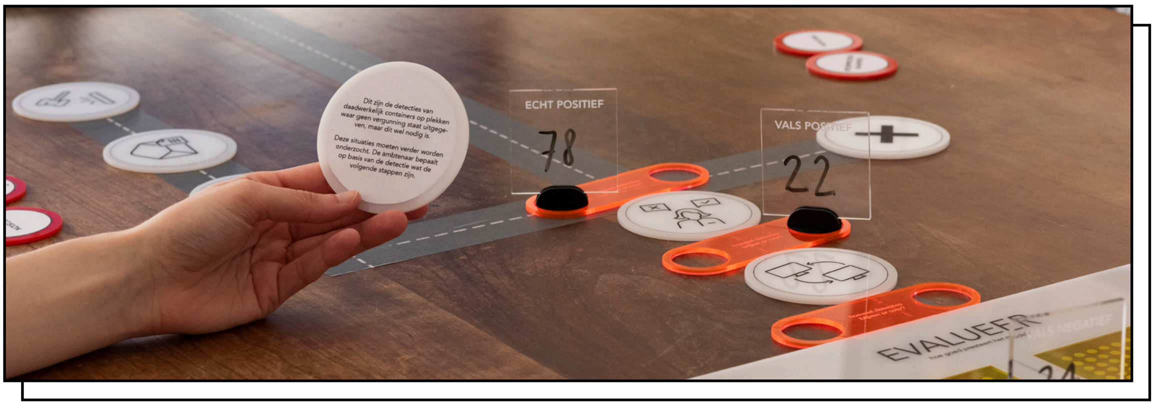
ECHT POSITIEF "Gedetecteerde container, is daadwerkelijk een container"	VALS POSITIEF "Gedetecteerde container, is eigenlijk geen container"
? ? ? VALS NEGATIEF ? "Geen container gedetecteerd, maar er is er eigenlijk wel een" ? ?	? ? ? ECHT NEGATIEF ? "Geen container gedetecteerd, en er is er ook geen" ?



UPDATE DETECTION PAWNS

A certain model composition is shown on the screen and the players are instructed to copy the numbers to their pawns on the playing field. With a whiteboard marker these numbers can be updated during exploration of the system.





DECISION SYSTEMS

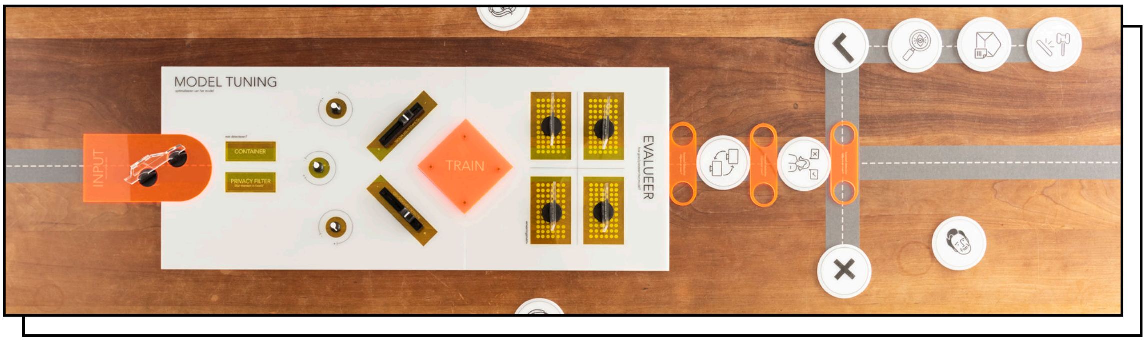
Like a board game the players can explore what happens to these detections by going through the system. After this process, the players hopefully have a better understanding of what certain consequences of certain mistakes can be. Based on their insights they can adjust their trade-off tokens, or not.

level 2

Jullie krijgen de optie om weer terug naar de tekentafel te gaan, en om het model aan te passen. Wanneer werkt het model acceptabel genoeg om in de echte wereld gebruikt te worden?

TUNING THE MODEL ACCEPTABLE

During level 2 the playing field is expanded with elements to tune the machine vision model: representation of input, the model, sliders, rotary knobs and a training button are added. After adjusting the sliders, the players can train the model. This will result in different performances of the model. With this the players can explore what they think is acceptable for the model to perform, and get a sense of the development of the model and the human role within.

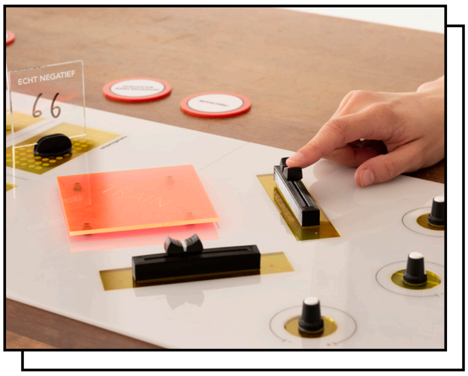


ITERATIVE TUNING

The players have 3 attempts to tune, train and evaluate the model. This iterative element is added, so they can reflect on what has happened, in order to identify links and critically think about what they find better. By doing this together, it will hopefully trigger discussion.

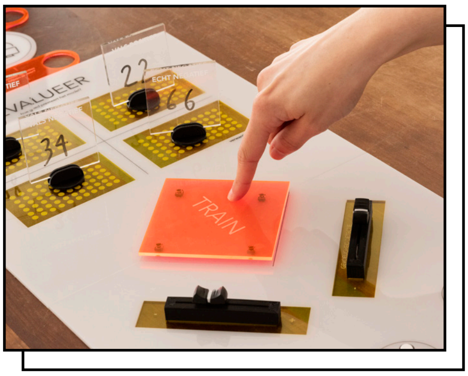
TUNING COMPONENTS

The model consists of two forms of tuning components: *rotary knobs* representing hyperparameters regarding the tuning process and *sliders* representing complexity and sensitivity of the model. Adjusting these elements will influence the model performance.



TRAINING BUTTON

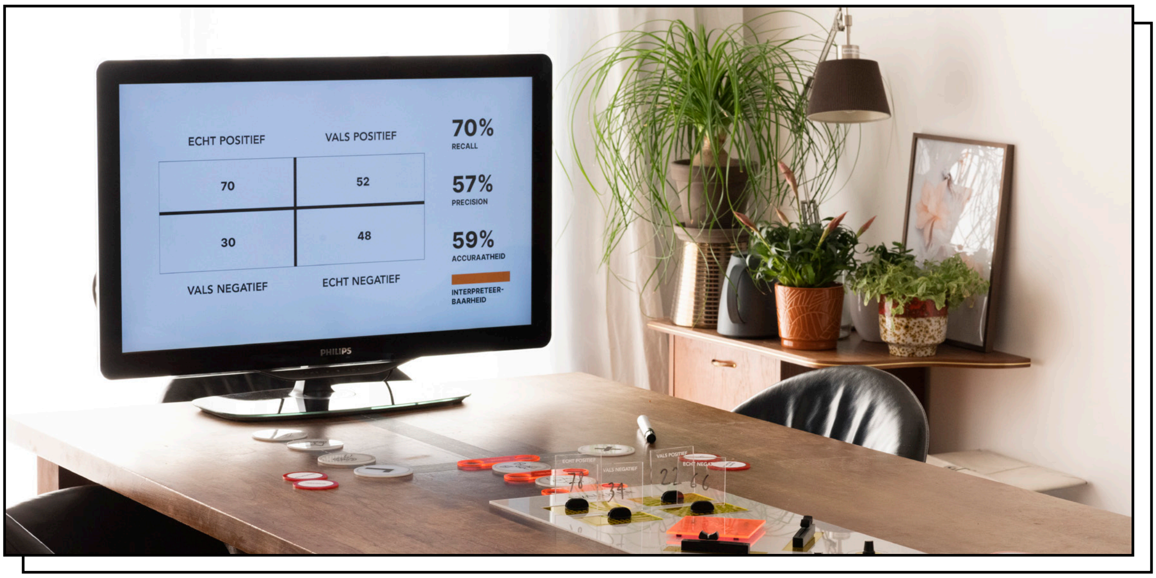
By clicking on the button in between model tuning and evaluating, the players can train the model. If the players want to see how a certain combination of parameters will perform, they can send this to the screen by clicking on the button.



'LOADING'

This interaction will result in a loading symbol on the screen. The model is now training, and it gives time to connect the composition of variables on the table to a certain model performance outcome (read more in *managing the system*).

repeat 3x



REVEAL PERFORMANCE

On the screen the model performance will be shown, where the participants can reflect upon (and tune the model again).

PERSPECTIVES OF OTHER PEOPLE
Within a deliberate discussion it is important to be aware of other opinions, even if they are not your own. Therefore the perspectives of other citizens, not present around the table, are introduced. The aim of this introduction is that the players will be more aware of other perspectives, and possibly consider and weigh in these with what is found to be acceptable. Next to that, the perspectives are introduced in order to make them aware that people want different things, and it is not possible to satisfy everyone completely. Each perspective is read out loud by a player. The players do not slip in to the role of this other person, but will only share the perspective.

TUNE AGAIN IF DESIRED: 3 MORE ATTEMPTS
The players have the chance to go back and update the model again, if they desire to. When they think the model is performing acceptable enough or have used all their attempts, the game will come to an end.

is the model performing acceptable for you?

finish: the end

Managing the Session

Due to earlier iterations, the decision was made to step back as a facilitator during the design intervention, to prevent interfering with the process. Other advantages of this approach are being able to observe the interactions and providing players with the same content. Due to this, instruction cards and instructions on the screen were added to guide the participants through the different phases. A wireless mouse was given to control the screen, and made players continue through the Figma interface by themselves.

In order to let players tune the machine vision model iteratively, this process was limited to tuning the complexity and the sensitivity of the model. For these variables, different variations of the model performance are created. With a Wizard-of-Oz approach, this process was managed. After players decided on the composition of variables, they click on the ‘train’ button. When this happens, I continue the Figma to a loading icon, which represents the waiting time to train a model, and gives me time to look at the chosen variables and continue to the fitting screen, each connected to another key (see Figure 27).

The details of these screens can be found in Appendix A.

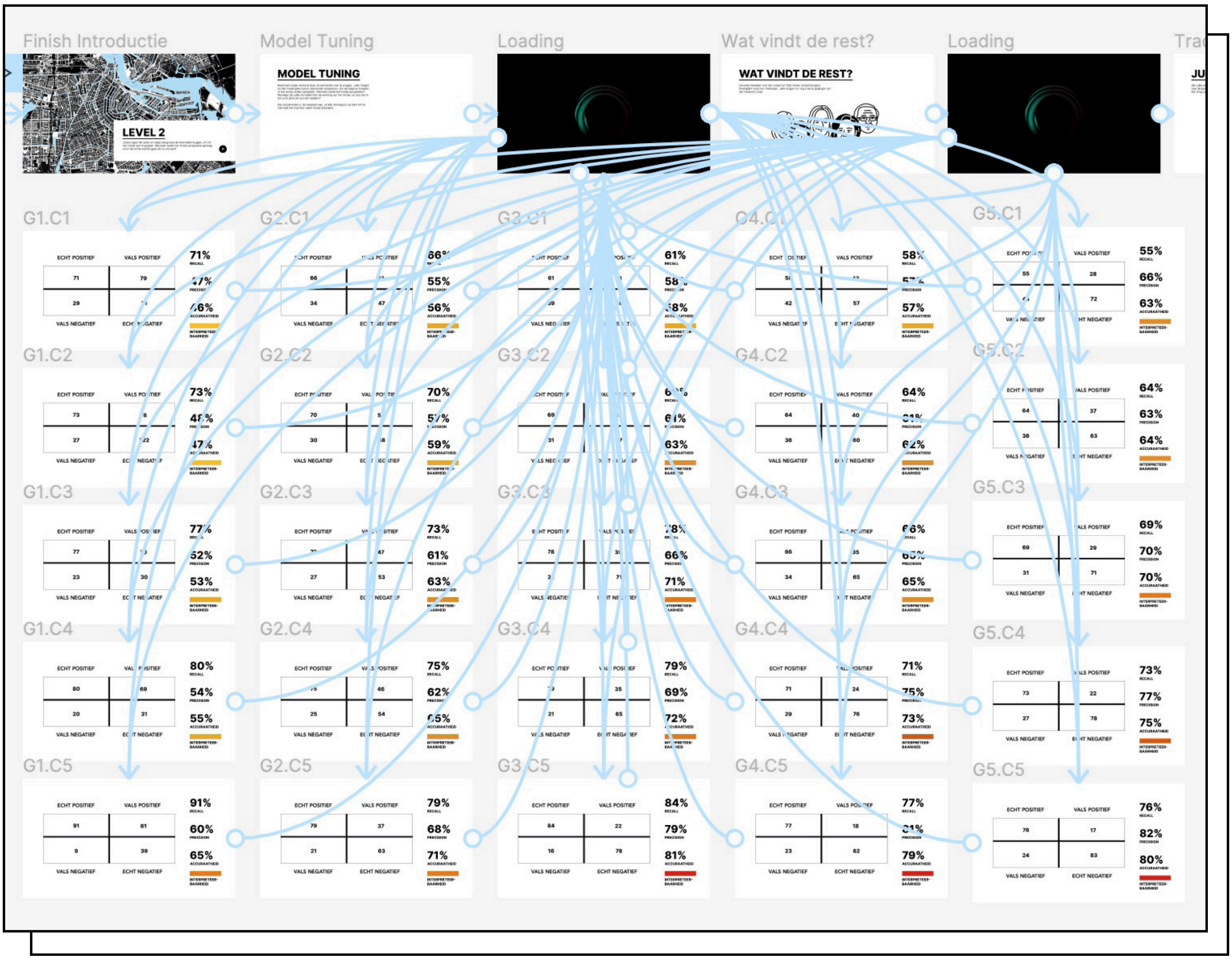


Figure 27: *Figma interface on the background*

DESIGN RATIONALE

Within the design process, a lot of choices were made. In this section, a few important ones are explained.

Container Context
To become more comprehensible, the design needed a specific application field. The choice of containers resulted from it being a clear and recognizable object, the municipality actually considering this option and a relatively clear decision situation.

Trade-off Tokens
Participants are asked to choose what they think is important by turning their tokens, to focus on these elements. This is done several times to familiarize them with the terms and main discussion points. First, they are asked to choose without being introduced to the terms. This might be uncomfortable but also opens up the space to talk and learn. Throughout, participants are able to turn and thus change what they find important.

Adding Perspectives
Because it became clear that no actual other stakeholders were going to participate, perspective tokens were added to represent their interests.

Material Choice
The final prototype was made out of PMMA to present a more finalized concept during the final evaluation, so participants would take the concept more seriously.

Rotary Knobs
As the sliders influence the shown model performance, the rotary knobs do not influence this. This choice was made due to the complexity of 'wizard-of-oz' the influence of multiple parameters. It was found more important to have fewer parameters and a more accurate influence on the performance than the other way around, so participants were able to identify relationships.

Language Barrier
As the design is aimed to introduce non-experts to complex topics, the decision was made to hold the sessions in Dutch. This led to translating some terminology into Dutch equivalents, i.e. Accuracy became Accuraatheid. The decision was made to keep the terms precision and recall in English, as no clear translation was found for recall that sparked clarification.

Game Control
Throughout the design participants control and continue through the interface themselves. As described earlier, this decision was made to let participants independently go through the playing field. As there is a lot of information, some structure, in this case, game rules, was added. Presenting the design intervention as a game was inspired by Play the City, as it can help to engage people in the game, deal with power balances, and opens up room to explore and make mistakes. Next, a game approach matched with dividing the information into different levels to prevent cognitive overload.

Within a game, the players need a clear objective. Therefore this was added, making an acceptable model and identifying where your perspectives differ.

5. EVALUATION METHOD

This chapter describes the evaluation of the design presented in the previous chapter. It contains the goal of the evaluation and explains the chosen method.

EVALUATING THE DESIGN

GOAL OF EVALUATION

The main goal of this evaluation is to investigate how the design performs regarding the design goals compiled in Chapter 3, within the context of a participatory session. The main intention of the evaluation sessions is generating data which will be used to evaluate the performance of the design intervention, and possibly bring limitations (and thus points of improvement) to the surface.

EVALUATION SESSIONS SETUP

The design was evaluated within multiple workshop sessions, imitating the context of a citizen panel assembled by the municipality of Amsterdam. This entails having a room with a table and a big screen to show things on. The participant should have different knowledge positions, as a mix of expert and non-experts. In order to get various participants together, the decision was made to relocate the prototype and move the session to a place easy for participants to get to.

> *"How to* ***improve the understanding*** *of* ***machine vision systems*** *to make stakeholders (citizens)* ***full-fledged interlocutors*** *in the co-construction of future scan cars?"*

The setup for the session contains of the various phases: (0) *pre-session*, (1) *introduction to the session*, (2) *introduction to topic*, (3) *level 1*, (4) *break*, (5) *level 2* and (6) *after intervention* as visualized in figure 28.

The complete evaluation set-up plan can be found in Appendix B.

EVALUATION SET-UP

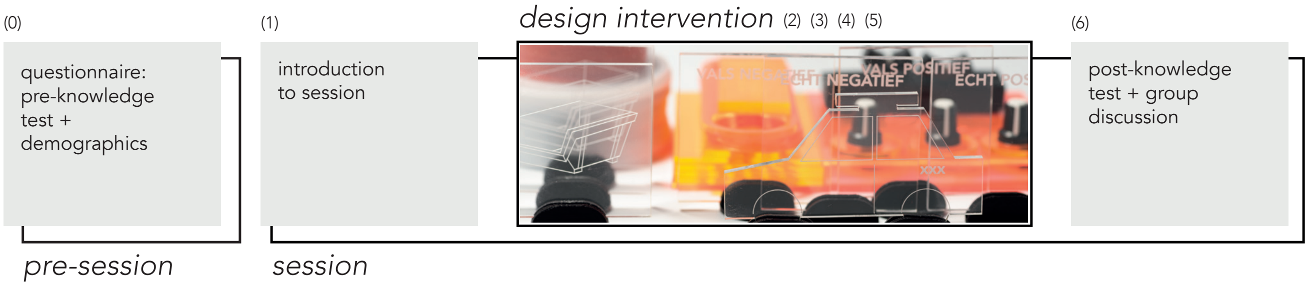


Figure 28: *Evaluation Set-up*

Formal Evaluation

Participants for the evaluation sessions were recruited through personal connections. After agreeing to take part in a session, participants received a pre-session questionnaire to fill in. This contained a variety of questions regarding demographic information, self-reported knowledge and objective understanding questions matching the topics. This information can be found in Appendix D. At the beginning of an evaluation session, participants were asked to sit down at the table. After an introduction of the session and the goal of today, they received the instructions to start interacting with the design. From this point on, instructions were given via the screen. After the various parts of the design and a break, the participants received a post knowledge test to fill in. The session was closed with a semi-structured group interview in order to evaluate together.

Participants

In total 14 participants took part in 3 final evaluation sessions. An overview of the participants per session is shown in Figure 30.

The participant are mapped per session based on their attitude towards smart city technologies and their self-reported knowledge, as visualized in Figure 29.

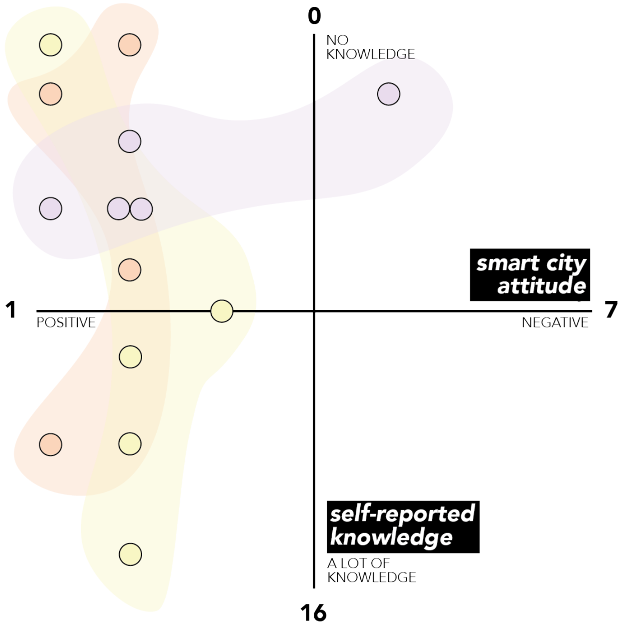


Figure 29: *Participants self-reported information distributed per session*

session #1

- ● **P1:** *women, 18-24, wo,* student engineering
- ● **P2:** *men, 25-34, wo,* student architecture
- ● **P3:** *women, 55-64, hbo,* photographer
- ● **P4:** *men, 18-24, wo,* student engineering

session #2

- ● **P5:** *men, 25-34, hbo,* student maritime
- ● **P6:** *women, 18-24, wo,* student medicine
- ● **P7:** *men, 25-34, hbo,* meal delivery
- ● **P8:** *women, 18-24, wo,* student philosophy
- ● **P9:** *women, 18-24, wo,* student bioscience

session #1

- ● **P10:** *men, 18-24, vwo,* student mathematics
- ● **P11:** *men, 55-64, hbo,* IT developer
- ● **P12:** *men, 25-34, wo,* robotics engineer
- ● **P13:** *women, 35-44, mbo,* teacher special education
- ● **P14:** *women, 25-34, wo,* e-health consultant

Figure 30: *Demographics of participants per session*

Data Collection

For the evaluation a mixed-method approach is used, inspired by Cheng et al., (2019). This approach aims to collect and combine both quantitative and qualitative data, as subjective and objective data. This data is collected through different ways before, during and after the design intervention. An overview of data collection is shown in Figure 31.

Demographic Information

Participants are asked to report their gender, age, level of education and current occupation.

Self-reported Attitude Smart City

With a 7-point Likert scale participants are asked to report their attitude towards using scan cars in general. Because the scale alone does not explain why people have this attitude, an open question was added to understand their underlying reasoning.

Self-reported Knowledge

To evaluate the level of expertise participants are asked to report their knowledge regarding relevant fields as programming, algorithms, machine learning and machine vision.

Objective Understanding

In order to evaluate if the design intervention improved understanding of the machine vision system, a pre-post knowledge test will be conducted. With quiz questions participants' objective understanding of the system will be assessed. These categories match the sub-design goals (1A, 1B, C). Per theme, question types were set up. In order to compare the pre- and post-test answers, the questions asked where similar, yet different. I.e. '*What is false positive in this situation? What is false negative in this situation?*' In total there are 20 objective understanding questions, which will be divided into the two tests. The complete list of questions can be found in Appendix E. The complete pre-session questionnaire can be found in Appendix D.

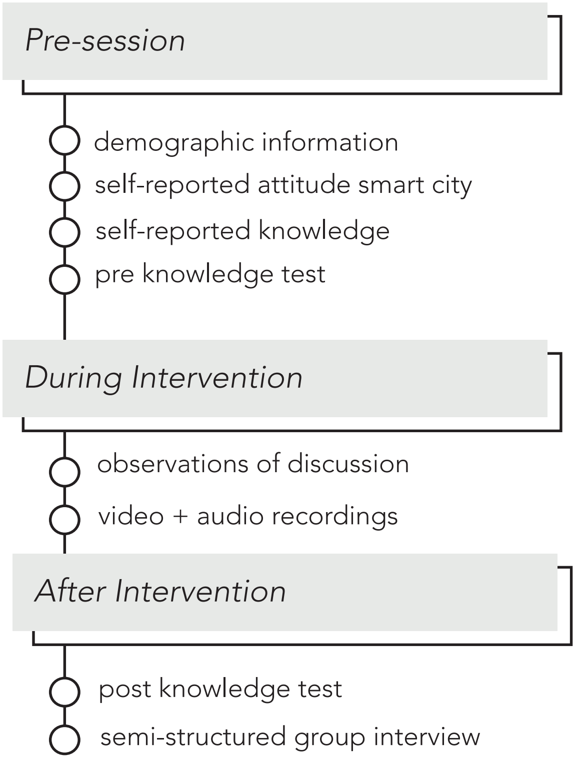


Figure 31: *Data collection*

Observations of Discussion

In order to evaluate if the design intervention triggers discussion, this is observed during the session. The sessions are filmed and audio is recorded. During the session, timestamps of discussion that occurred were noted down. A discussion is defined as a '*consideration of a question in open and usually informal debate*'.

Semi-structured Group Interview

The design intervention is followed by a semi-structured group interview, in order to investigate the subjective experience and understanding of participants. The interview questions can be found in Appendix F.

Analysis Plan

The collected data will be analyzed in separate ways, before combining objective and subjective elements regarding the same themes, see Figure 32. All data will be analyzed in a qualitative manner.

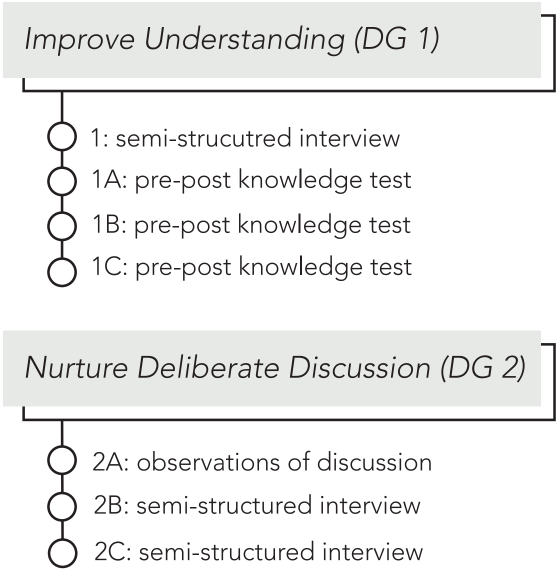


Figure 32: *Analysis plan*

Analysis DG1: Improve Understanding

The objective data from the pre-post knowledge test will be imported to an excel file, in order to calculate the individual and group performance grades. The subjective data from the semi-structured interview will we translated from handwritten notes to online notes, and clustered around similar themes. Interesting quotes will be noted down during the interview.

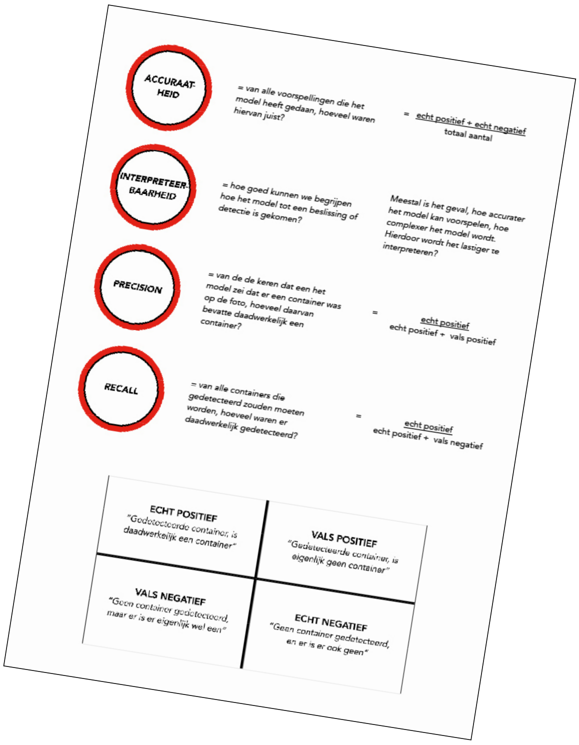
Analysis DG2: Nurture Discussion

The observation notes of the session and its discussions will be translated into a timeline where blocks of discussions and their length (from the timestamps) will be visualized. These blocks of discussion will be linked to the part of the session, where the participants were in the process. This will link certain discussions, topics and artifacts together. From the semi-structured interview notes will be combined in order to see if participants felt able to take part in the discussion, and what could be of influence to this.

Alterations after Session 1

After the first session a few small adjustments were made. These adjustments include unclear sentences, missing or unclear actions/ assignments, spelling mistakes, connections in Figma that didn't work and needed my help on the background during the session.

The biggest adjustment was made by providing the two next sessions with a small list of definitions that are explained during the introduction phase. The participants really felt that they needed this, in order to reason around the trade-offs. Therefore I decided to print a small list, 'a cheat sheet', to not make this the bottle neck of the session.



ACCURAAT-HEID = van alle voorspellingen die het model heeft gedaan, hoeveel waren hiervan juist? $= \frac{\text{echt positief + echt negatief}}{\text{totaal aantal}}$

INTERPRETEER-BAARHEID = hoe goed kunnen we begrijpen hoe het model tot een beslissing of detectie is gekomen?

Meestal is het geval, hoe accurater het model kan voorspelen, hoe complexer het model wordt. Hierdoor wordt het lastiger te interpreteren?

PRECISION = van de de keren dat een het model zei dat er een container was op de foto, hoeveel daarvan bevatte daadwerkelijk een container? $= \frac{\text{echt positief}}{\text{echt positief + vals positief}}$

RECALL = van alle containers die gedetecteerd zouden moeten worden, hoeveel waren er daadwerkelijk gedetecteerd? $= \frac{\text{echt positief}}{\text{echt positief + vals negatief}}$

ECHT POSITIEF "Gedetecteerde container, is daadwerkelijk een container"	**VALS POSITIEF** "Gedetecteerde container, is eigenlijk geen container"
VALS NEGATIEF "Geen container gedetecteerd, maar er is er eigenlijk wel een"	**ECHT NEGATIEF** "Geen container gedetecteerd, en er is er ook geen"

Figure 33: *Additional cheat-sheet*


HET MODEL IN DE ECHTE WERELD
Tijdens het echte gebruik weten we..
PHILIPS

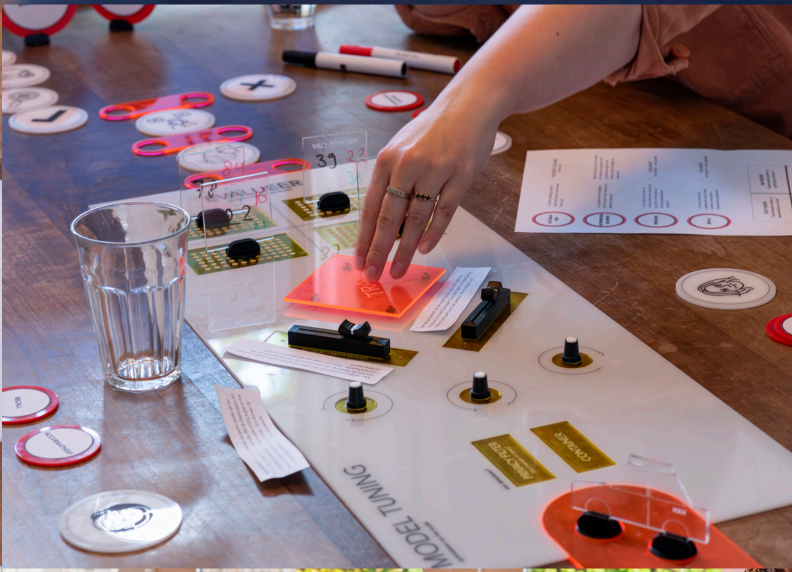

EVALUEER

SESSION 2

SESSION 3


ECHT POSITIEF


TRAIN


MODEL TUNING
VALIDEER


VALS POSITIEF
ECHT POSITIEF
22
58


ECHT POSITIEF
VALS POSITIEF
78
22
34
66
VALS NEGATIEF
ECHT NEGATIEF
69%
RECALL
78%
PRECISION
72%
ACCURAATHEID
INTERPRETEER-BAARHEID
PHILIPS
INTERPRETEERBAARHEID

6. RESULTS

This section presents the results of the formal evaluation of the prototype. The results of the evaluation are presented in more detail in the following sections:

- Results: Improving Understanding
- Results: Nurture Deliberate Discussion
- Additional Insights

RESULTS: *IMPROVE UNDERSTANDING*

This section presents the results regarding the first design goal: improve understanding of the machine vision system. The 3 sub-design goals are the *vision of the scan car* (1A), *decision situation* (1B), and *trade-offs within model development* (1C).

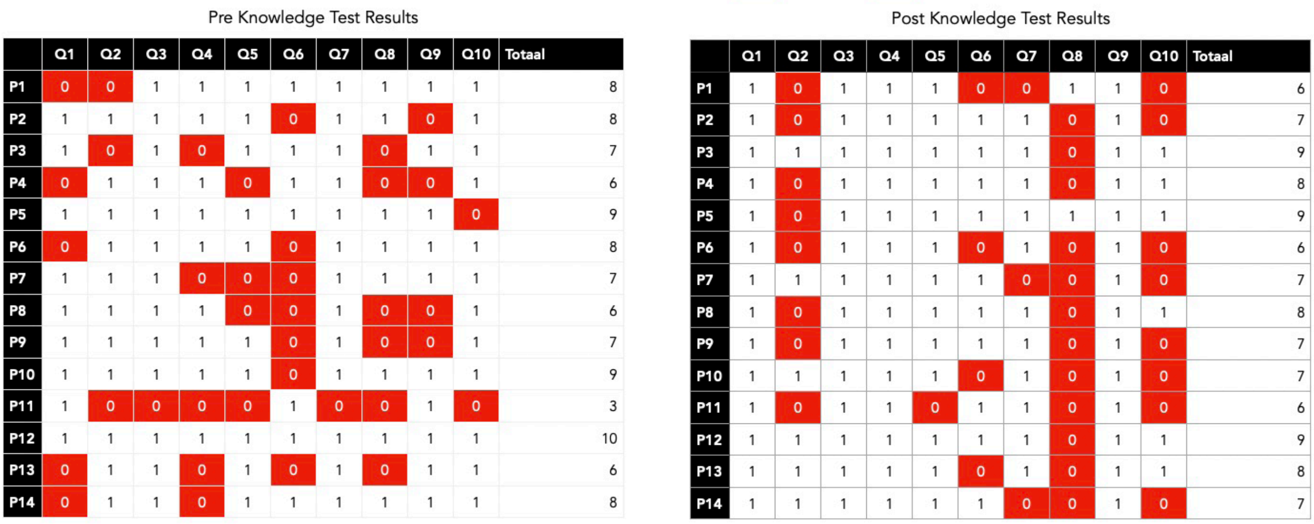
OBJECTIVE UNDERSTANDING

Analyzing the pre-post knowledge test resulted in the following scores: the average score on the test before the session was 7.29 out of 10. This means on average 7 questions were answered correctly. The lowest score of a participant was 3, whereas the highest score was 10 (a participant who answered each question correctly).

The average score of the test made after the session was 7.43 out of 10. This results in an average improvement of 0.14 points. The lowest score by a participant did improve from a 3 to a 6. No one was able to finish the test without mistakes. Looking at the individual performances, not everyone improved: 5 participants improved, 3 participants scored the same and 6 participants scored lower than before.

Analyzing the answers per question in depth gives the following overview shown in Figure 34. In this overview a different distribution of mistakes becomes visible. Where in the pre-test the mistakes are more diffused around the questions, in the post-test there are 3 questions where more people answered the question wrongly than correctly.

This can indicate several things:
- The topic of the question is still unclear or not understood after the session
- The question is formulated as unclear
- The question is too difficult



Pre Knowledge Test Results

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Totaal
P1	0	0	1	1	1	1	1	1	1	1	8
P2	1	1	1	1	1	0	1	1	0	1	8
P3	1	0	1	0	1	1	1	0	1	1	7
P4	0	1	1	1	0	1	1	0	0	1	6
P5	1	1	1	1	1	1	1	1	1	0	9
P6	0	1	1	1	1	0	1	1	1	1	8
P7	1	1	1	0	0	0	1	1	1	1	7
P8	1	1	1	1	0	0	1	0	0	1	6
P9	1	1	1	1	1	0	1	0	0	1	7
P10	1	1	1	1	1	0	1	1	1	1	9
P11	1	0	0	0	0	1	0	0	1	0	3
P12	1	1	1	1	1	1	1	1	1	1	10
P13	0	1	1	0	1	0	1	0	1	1	6
P14	0	1	1	0	1	1	1	1	1	1	8

Post Knowledge Test Results

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Totaal
P1	1	0	1	1	1	0	0	1	1	0	6
P2	1	0	1	1	1	1	1	0	1	0	7
P3	1	1	1	1	1	1	1	0	1	1	9
P4	1	0	1	1	1	1	1	0	1	1	8
P5	1	0	1	1	1	1	1	1	1	1	9
P6	1	0	1	1	1	0	1	0	1	0	6
P7	1	1	1	1	1	1	0	0	1	0	7
P8	1	0	1	1	1	1	1	0	1	1	8
P9	1	0	1	1	1	1	1	0	1	0	7
P10	1	1	1	1	1	0	1	0	1	0	7
P11	1	0	1	1	0	1	1	0	1	0	6
P12	1	1	1	1	1	1	1	0	1	1	9
P13	1	1	1	1	1	0	1	0	1	1	8
P14	1	1	1	1	1	1	0	0	1	0	7

Figure 34: *Results pre-post knowledge tests*

Analyzing the answers to the test, every participant perceived the role of the scan car correctly: to collect data. The second question showed that the process of identifying people to blur them was not clear. The questions regarding the decision situation (DG 1B) were answered correctly by every participant in the post-test, while some mistakes were made in the pre-test. This could indicate an improved understanding of the consequences of the system and possible mistakes that are made.

The following six questions regarding trade-offs within the machine vision development phase showed a less distinct improvement in understanding. Especially Q8 and Q10 showed that there is not a sufficient understanding of what interpretability and recall exactly entail. Reflecting on these questions and the content of the session, I think the session introduced the model as something that can be completely interpreted, while in reality, the use of a convolutional neural network is inherently not interpretable due to structure. By simplifying the model, and not introducing the actual structure, this information might have gotten lost. For Q10, the term recall was addressed multiple times during the session. Therefore I believe that the participants did not really understand the definition and what it entails.

Concluding, the tests show that objectively the understanding of the participants did not improve during this session.

SUBJECTIVE UNDERSTANDING

The participants were able to identify different elements of a machine vision system afterward: from the scan car that collects videos as input, the model that uses this video and makes detections, to the consequences the model can have in the real world. By summing up these elements, participants referred back to the components on the table. Participants also noted that there are trade-offs during this process, referring back to their tokens. The concept of a trade-off appeared to be clear:

> *'That I choose for one side doesn't mean that I don't find this important, but that's why it is a trade-off' - P14.*

Especially level 1 was experienced as helpful by several participants, in understanding what they found to be important: it facilitated insight into what happens to a detection. The realization that there is a human checking the detections was mentioned as a big influence on their decisions. Some participants mentioned that their opinions would be different if the whole process was automated. From this, it appears that participants did understand the relation between the machine vision model, and the decision situation (DG 1B).

> *'If the context changed, you would make different judgments. So the results we make here practically only apply to containers' - P7*

Lack of Understanding

The relations between some other elements appeared to remain unclear. For example, the term trade-off seemed to be understood, but P6 referred to every slider as a single trade-off and stated that in reality there probably would be hundreds of trade-offs. It wasn't understood that the two trade-offs of today could be influenced by multiple variables, and not only a specific slider or parameter. Next to that, the tuning elements in general provoked a lot of confusion. Where the sliders did have an impact on the model performance, the rotary knobs did not have a direct influence. This was mentioned as frustrating or confusing after each session by the participants and even suggested to be let out by a few participants.

Also, the effect, or expected effect, of the sliders was confusing to participants. At times participants expected a certain number or percentage to go down, but sometimes this did not happen after updating the model.

They found it difficult to find connections between or explanations for this behavior of the model. For example, some participants also questioned if the complexity slider always had the same influence, because they did not know this for sure. It appeared that it was clear that a model can be tuned, but what this exactly entails remained unclear to most participants (DG 1C).

The relation between tuning the model and deploying the model in the actual world also appeared to be misunderstood by some participants, especially during session 1. During this session the participants made statements like *'the civil servant can also check the false negatives, then we don't miss anything'* (P1). Reflecting on these statements afterward, confirmed that it was not clear that, in reality, we don't know these negatives, because they aren't detected. When discussing this afterward, it appeared that this difference was not clear to everyone, while it is a very essential element within the system. During the other two sessions, this misunderstanding did not occur.

Complexity of Language
All participants acknowledged the complexity and difficulty of the topics that they were introduced to. Most of them reflected on the session as complex, but 'followable'.

> *'I normally stand in the doll corner throughout the day, or am finger painting, so this is all new to me.' - P13*

While terms such as false positive and accuracy seemed clear, the terms interpretability and recall remained puzzling for most participants. During two interviews, a participant noted that a more 'Jip-en-Janneke' term could be helpful for example recall. Where accuracy and precision involve elements participants were able to imagine, recall was a completely new word. Even though almost every participant found recall more important to optimize for, they were not able to explain what it means clearly during the interview. This is in line with the post-knowledge test, where only 2 participants had a correct answer to the question about what recall is. Interpretability was also experienced as difficult: it was difficult to imagine what this exactly entails. Most participants felt that if there is a picture of a container, what is there more to interpret? This indicates that this term was not clear.

> *'When I had to choose between the tokens the first time, it meant nothing to me. Now I thought I did have a better understanding. Once the questionnaire came afterward, I still the idea that I didn't understand it so well, because then it wasn't quite clear after all, even though I felt it was clear to me.' - P11*

Another possible indicator of participants understanding of the machine vision system is that there was a demand for more knowledge and more information on certain topics. Group 1 indicated they would like to know more about how much data is stored, who gets to see all this, how long a civil servant exactly working on this, etc. Groups 2 & 3 indicated that they would like to know more about what we are actually tuning, or how many containers are detected (the scale). On the one hand, this could be an indication that elements are still unclear, but on the other hand, it could also indicate that people already have a better grip on the system, and now want to know more details about it. Based on this, I don't think wanting to ask more questions is a negative thing, but more an indicator of clarifying and being critical.

CONCLUSION
To conclude, the design intervention improved subjective understanding of some elements such as the decision situation and vision of a scan car. However, terms such as recall and interpretability and relations between elements remained unclear.

RESULTS: *NURTURE DISCUSSION*

OBSERVING DISCUSSION

This section presents the results regarding the second design goal: nurture a deliberate discussion about what is acceptable for the model. The 3 sub-design goals are *trigger discussion* (2A), *empower to take part in the discussion* (2B), and *enable to form their own perspective and understand others* (2C). Each sub-goal is evaluated separately.

TRIGGER DISCUSSION

During the session, discussions were observed. Each discussion, *a consideration of a question in an open and usually informal debate*, was noted down with timestamps. These observations are visualized by black boxes in Figure 35. Afterwards, the recorded footage was revisited at these timestamps, to see what happened beforehand. This resulted in adding a second row of artifacts to each discussion timeline. They are connected to the topic of the conversation. The levels are visualized per with different colors, and the model tuning iterations are split up.

From this, it appears to be that some specific components initiated the discussion. For example, making a choice concerning the trade-off resulted in a discussion among the participants about why they preferred one element over the other. During level 1, the questions asked on the back of the system icons, made participants discuss them. During level 2, it appears that participants were iteratively discussing how to adjust the settings, and reflecting on the outcomes seen on the screen. Each session ended with a discussion about what their final model should be, and what is ‘their best model’.

While some elements triggered discussion within all three sessions, other elements didn’t trigger discussion consistently. For example, not all perspectives of people were discussed in depth. During the second session at the end of level 1, choosing a side of a token did not trigger discussion, whilst participants had different preferences at this point.

After the session, during the interview, the participants did not reflect on what exactly caused their conversations and discussions during the session. The term discussion was mentioned while giving other answers. P2 said that the ‘session sparked discussion in various ways’, but he was not able to explain more in-depth. During sessions 1 and 2 the participants agreed that there would be more discussion if the participants had more distinct and varying opinions about containers. During session 3 the participants reflected on the role of the design as giving structure to the conversation, but not triggering certain conversations.

Thus analyzing the observations and interviews about the sessions, it appeared that there are components that can trigger discussion, but this role of the design was not consciously named by the participants.

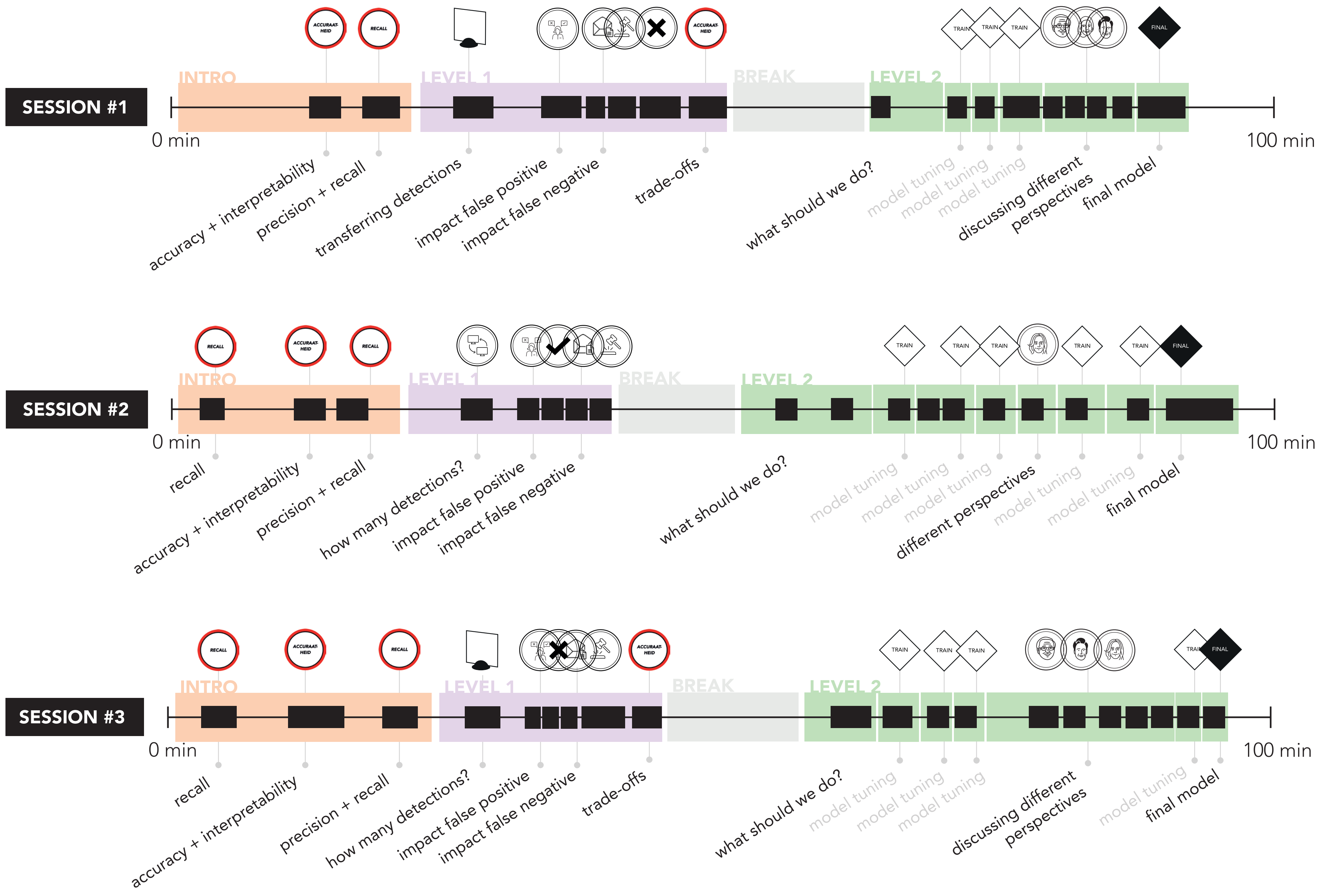


Figure 35: *Observations of discussions per session*

EMPOWER TO TAKE PART IN DISCUSSION

All participants said that they felt able to take part in the group discussion if they wanted to. '*Everyone had their say*', but some voices were heard more often than others. While P7 stated that '*everyone contributed to the decisions made (during model tuning)*', P8 responded with an opposing sound that '*not everyone always contributed to the decision, because some people took the lead*'. P8 explained that she and P6 were able to participate in the discussion, but sometimes stayed more in the background of the conversation during making a decision. This resulted in a reflection of P5 on his role during the session, *'did I scream too loud?*'. This illustrated something that wasn't reflected on during other sessions. Everyone participated in the conversation, but some participants spoke more often than others and were more dominant.

Knowing over Knowledge

When reflecting on what made participants able to take part in the discussion, all groups mentioned that knowing each other, even though the connection was thin, helped them to share opinions and ask questions. None of the participants reflected on the role that information could have in this process. The participants did not note newly obtained knowledge as an element that made the discussion possible. But when asked if they had enough information to take part in the conversation, every participant did think they had enough information. However, there were elements still unclear, or new questions arose as described in the previous results.

Something to Talk About

Various participants noted that they did need something in front of them to talk about. One of the participants said that '*if it only had been the slides, a lot less would come out of this session, because now you are more engaged with it*' (P4). P8 made a similar comment sayings *'this holds more attention: okay, what exactly are we doing and talking about now?'* Thus it appeared that the physical components in front of them made it easier for participants to keep track of the conversation topic, thus making it easier to take part in the conversation. It appeared that it was easier to connect certain information to certain tangible elements, which made it easier to go back and forth between them. For example, P10 said the slider eased the conversation about the sensitivity of the detection from 'trying to explain that you want the sensitivity lower with unfamiliar terms, to just lowering the slider. It's a lot clearer for everyone what you are talking about and what you mean.' Thus making it tangible seems to have a positive influence on the discussion.

'You need something like this to talk about with each other, otherwise it is very conceptual' - P1

'It's also nice that you can look at something, instead of just listening to each other' - P9

'Because something in front of us, easier to talk about, because pointing at something is easier than abstract concept' - P11

Starting to Talk

One group reflected on reading instructions out loud in turns during the introduction. For them it lowered the barrier to take part in the conversation, so made it easier to discuss or ask questions later on.

Overall, the participants did feel that they could take part in the discussion. The main reasons for this are being familiar with other participants and having something in front of you to talk about. However, the intensity of taking part varied per participant, which could be due to various personalities, or being less certain about the topic. The new information they were confronted with, was not connected by participants to empowering them in their discussions.

ENABLE TO FORM PERSPECTIVE

In each session, the participants ended with a variation of the model that they all found to be acceptable. When reflecting on what entails an acceptable machine vision model, the participants first stated how they ended with a specific model. This was experienced as a trial-and-error process, that P14 described as *'looking at the results (on the screen), our preferences and eventually changing (the model) towards these preferences.'* When trying to explain why this was 'the best' version, each group noted that they needed to find a balance. Because of working together in a group, the end model felt like a group accomplishment (P14).

Changing your Opinion

All participants felt like they could form an opinion about the trade-offs and what they found important better than before the session. During session 3 choosing a side of a trade-off was experienced as difficult the first time, because they had no idea what certain terms mean. This changed over time into having a better feeling of what they find important. P13, with agreeing on sounds of the group, described this as *'by experiencing it yourself, you learn more about it. So having it found out for yourself, you understand better where your opinion comes from.'* P5 noted that his opinion changed after seeing what happened to the detection in real life, which made him adjust his token.

However, it was also noted that there is still some uncertainty in forming their own perspective. One participant noted that she did not really have direct experience with placing containers in Amsterdam, making it difficult to form an opinion about this. Next to that, there are still some doubts about assumptions made during the session. Reflecting on what the session would be if an expert, a developer of this system, was at the table made P2 note that *'if someone with more knowledge shared their opinion, I would be more likely to follow them.'* P14 had a similar feeling and would see this expert role more as a fact-checker to confirm her assumptions, than someone with her own opinion because she would be influenced.

Room for Other Opinions

As part of a deliberate discussion, it is not only about forming your own opinion but also about understanding and exploring other perspectives. During each interview, the perspectives of citizens introduced were mentioned but reflected on the influence differently. P6 noted that the civil servant perspective made her think more about the consequences of false positives, while another participant said that the different perspectives directly went to the background. It appeared that participants were able to empathize with different perspectives.

> *'As a developer, I can imagine you want complexity very high because then it performs better. But you have to deal with people, hear (their) opinions. The developer's (version) is not always the right version. ' - P14*

Each group acknowledged the importance of hearing these different perspectives but noted that it would be better to actually hear these perspectives in person. It also appeared that people noticed that not all perspectives can get aligned because there are different opinions about what is acceptable.

> *'We are dealing with something in public space, with people, so you need to give them voice.' - P11*

> *'If Ralph had been at the table here, it would have been very different. Then you really get into a discussion, because someone really has that opinion. It is easy*

to refute such a perspective (on a token), whereas if someone is sitting in front of you, it becomes much harder and more complicated. You don't refute a person, but you can refute a card.' - P3

'You can't do something with all opinions, but then you can weigh in the opinion of Joris or Ralph.' - P6

To conclude, participants were more able to form an opinion about what they found acceptable during the session. However, there are still some doubts. They appeared to be able to understand other perspectives but sometimes disagreed. These other perspectives were not included in deciding on the final acceptable model.

CONCLUSION

The design intervention appears to trigger discussion among participants due to its tangible representation of the machine vision system. The question remains if it was always the intended discussion point. Taking part in the discussion was namely facilitated by knowing each other. It remains unclear what the role of new information exactly was. Participants were able to form their perspectives better than beforehand. Still, questions about their assumptions remain and the perspectives of people around the table are easily dismissed.

ADDITIONAL INSIGHTS

Next to reflecting on the design goals, the sessions unveiled other points worthy of reflecting upon.

Positive Reaction to Game Approach
After the sessions, multiple participants noted the feeling of playing a game in a positive way in their first reaction to the session. They noted that this approach kept their attention during the session.

'It's like playing a game, but one no one has played before. Together, we take the game out of the plastic.' - P2

Consensus versus Dissensus
During all sessions the participants ended with one final acceptable model: they found a consensus. This could be due to the similar attitudes of the participants towards smart city technologies, but could also be due to the assignment to find a model acceptable for everyone. This assignment also contained identifying elements that you are not agreeing upon. However, no group ended with concrete points of friction.

Exploring versus Guidance
A few participants noted that they would have liked a bit more guidance throughout the design intervention. Where one participants suggested to *'be taken by the hand of an expert*' during the introduction, another participant wanted more guidance throughout the decision system and model tuning. He noted that they were figuring things out themselves, but it would be more helpful is they got some help at these points, if struggling. However, I believe that this struggle, finding something out yourself adds to the learning experience and understanding what is exactly happening.

Better or Best
After the first session, the participants asked if they ended up with a 'good' model. After sessions 2 and 3 participants were asked if they made 'the best model' or 'a better model than the previous groups'. While it matches the game elements from the session, this could be an indicator that the subjective element of the 'best model' was not conceived clearly. It appeared that the participants still felt the need for confirmation of someone with more knowledge to validate if they actually made a good model (even though they are the judges themselves).

7. DISCUSSION

This chapter zooms out to reflect on the design in relation to the design field, academic field and the context of design. Recommendations, opportunities and limitations are discussed regarding to these topics.

DISCUSSION

The following sections reflect on the overall project and try to put the project in perspective of academic literature and the design field. This is followed by recommendations for the municipality on involving citizens in the co-construction of responsible scan cars and the limitations of the study.

> *"How to **improve the understanding** of **machine vision systems** to make stakeholders (citizens) **full-fledged interlocutors** in the co-construction of future scan cars?"*

SUMMARY OF KEY FINDINGS

Improve Understanding

This project aimed to explore how to (1) improve the understanding of machine vision systems by citizens and (2) open up the discussion about its acceptability during development. The design intervention created appears to improve the subjective understanding of participants but did not improve objective understanding. The intervention appeared to make participants more familiar with necessary components such as trade-offs and the decision situation. However, knowing more also raised more questions, such as being aware of not knowing everything. This epistemic uncertainty can lead to a critical attitude.

Nurture Deliberative Discussion

Next, the design intervention aimed to nurture a deliberate discussion between participants. This includes the design to trigger discussion, empower participants to take part in the discussion, and enable them to form their own perspective. Having 'something', in this case, the prototype, in front of the participants appeared to start the discussion and create a shared understanding. While exploring the design intervention, participants were able to form their own opinion based on the system but still found it difficult to deal with other perspectives.

IMPLICATIONS FOR PARTICIPATORY MACHINE LEARNING

As Sloane et al. (2020) elaborated, participation in public AI can take on different forms. Within this use case of scan car development, and the aim of the municipality to involve citizens to legitimize certain decisions, participation should be seen as justification. Becoming transparent and having genuine knowledge sharing is essential to prevent participation-washing. Transparent refers to being clear about the objective of involvement. For genuine knowledge sharing, a shared understanding of the context and terminology is desired to communicate between parties.

To, in this use case, transfer knowledge around the acceptability of an ADM system, König and Wenzelburger (2021) identified various steps as setting the stage, creating a shared understanding, and aggregating preferences. This project is a practical example of this envisioned stakeholder involvement to close the legitimacy gap. It is an attempt to facilitate contestability between citizens and developers, by opening up the machine vision development phase, showing potential for new participatory approaches to public AI. Translating generic steps and knowledge from literature into a context-specific example can help to inform practical design knowledge for public AI.

This proposed involvement differs from other examples, as stakeholders are included during the development phase instead of before or after. Participatory machine learning at this moment might not entail tuning the actual model together due to the complexity and length of the process. Involving stakeholders would more likely take on the role of steering and informing the model.

Merging TUI and XAI

As interactive interfaces improve comprehension of decision-making by non-expert stakeholders, adding tangible elements appears to retain this benefit (Cheng et al. 2019). The proposed benefits of merging these fields, in this project, seem to manifest themselves: the design intervention showed to enable an enactive mode of reasoning, making users able to explore the complex and abstract topic in a comprehensible manner, and facilitate social interaction. Participants reflected on the design needing to have 'something' to talk about to start the conversation. This tangible approach appears to engage in and open up discussion about complex and abstract technologies for citizens.

Where Colley et al. (2022) propose a conceptual framework for merging XAI with TUI, practical examples lack. This project is an example of using a tangible approach to machine vision development, to address this design gap. Reflecting on the framework, the design differs from not interacting with an actual AI system. By tuning the model, the users can explore the model that gives feedback after clicking on the train button. Future work could explore integrating a working machine vision system and iterate on the received feedback. This could include focusing on the explanatory interface on the screen with more visual examples or integrating feedback into the tangible components.

IMPLICATIONS FOR IMPROVING NON-EXPERT AI LITERACY

Where a lot of research is focused on ex-ante expert or non-expert post hoc understanding, this project looked into non-expert understanding ex-ante deployment. Improving the knowledge positions of non-experts should happen on the human-ground level. Making abstract concepts into contextual concepts helps to improve understanding. Similar results were found by Shen et al. (2020), as terminology remains difficult and confusing, but giving a contextual explanation or visualization makes it more clear. For example, showing a visual example of a false negative detection and description helped participants to clarify.

Next to explainability, the proposed design combines some design considerations proposed by Long and Magerko (2020). Social interaction, lowering the barrier to entry, and unveiling gradually seem to improve the understanding of AI of non-experts. These elements did not only emerge from literature but were also identified and emphasized during the design process. Uncovering the complex subject together was experienced positively. It is difficult to say how introducing new perspectives improved understanding, as these were easily refuted by participants.

To analyze improved understanding, from this project, it seems important to include subjective understanding of participants. Discussing what was clear and unclear revealed points where the design could improve and be more clear.

IMPLICATIONS FOR DESIGN METHODS FOR PUBLIC AI

This project was started with a holistic and explorative design approach. On one side this created space for many different explorations, but on the other side made it harder to reflect on the process. Some relevant insights, while out of scope, might have gone lost during the process. However, it seemed to be effective to directly start exploring interactions in a social context, resulting in insights regarding group engagement instead of individual interactions. Where a lot of XAI research is focused on individual experiences, viewing explanations as something situated through this approach helped to in a relatively short time design a group intervention. This is in line with Jiang et al. (2022), who notes that explanations should match the receivers and should be carefully provided because not everything needs to be explained.

Dividing the requirements into content, structure, behavior, and appearance helped to manage the complexity of the desired system. As appearance was partly left open for exploration, it was difficult to reflect on the requirements. Further research could improve defining and substantiating the needed look and feel of the design.

Where in the end no actual machine vision technology was used within the design, a 'wizard-of-oz' approach helped participants to envision training an actual model. This low effort approach seems to be an effective way to help find the desired interactions.

RECOMMENDATIONS FOR SCAN CAR DEVELOPMENT AMSTERDAM

Using the scan car development trajectory in Amsterdam as a use case helped to evolve the design intervention to be more realistic and concrete. It appears to be possible to involve citizens during the machine vision development phase. If comprehensible questions and consequences are clarified, citizens could be able to inform the design and help legitimize decisions made and become more transparent.

However, this civic participation will not be easy. There are several recommendations from this project, of which the first one is to acknowledge the preconceptions participants have beforehand. As most people are not aware of what a scan car is seeing or doing, this should be discussed with each other. From here, what is happening can be clarified.

Next, to involve citizens in complex topics such as public AI, it is advised to take time to create a shared understanding of the topic and relevant terminology. First, citizens need to understand the system, before they can identify their own perspective. Translating abstract and complex terms to something tangible and giving concrete examples can help clarify and communicate between these different parties.

Lastly, introducing perspectives of 'other stakeholders' in this project were easily dismissed and not included in the judgment of participants. Therefore it is advised to get these different perspectives around the table, from civil servants to people against scanning applications.

LIMITATIONS

The first set of limitations of this project is due to the composition of participants during the evaluation. While they were not familiar with the scan car project, they were not a fair representation of civil society, due to mostly being higher educated. Not all participants lived in a city where similar cars go around or had experience with receiving fines and permits from a municipality. This made it harder to envision and form an opinion. Reflecting on the attitudes towards smart city technologies, most participants were relatively positive towards using for example scan cars. Their main reason was efficiency. While this opinion should be represented, other perspectives should be heard as well. Therefore more research is needed to see how the design behaves in an environment with different attitudes towards AI and educational levels (and the possible tensions and frictions that it leads to).

While the project intended to improve the discussion between expert and non-expert stakeholders, in the end, the project focused on the knowledge positions of non-experts. During the evaluation session, interaction between non-experts was possible. However, the interaction between experts and non-experts should be researched more. Possible points for further research could be how power and knowledge balances play out in reality and where to involve experts (civil servants and developers). For example, should experts be involved during the intervention, or should, they be involved after the intervention in facilitated a discussion?

Another limitation of this project includes the research set-up for objective understanding. As the objective understanding of the participants did not improve, while participants subjectively indicated having a better understanding, the evaluation method needs more attention. For example, questions in the future could be validated and possibly be mixed between participants. The evaluation session lasted around two hours. After this time, participants were asked to fill in the post-knowledge test and take part in the follow-up interview. Some indicated that they were quite tired, which might have influenced their performance.

FUTURE WORK

As this concept is still a conceptual prototype, future work could look into other possible practical implementations of the contestability loop between citizens and developers. Next, future work could look into creating a tangible language to communicate around public AI to citizens and facilitate more public knowledge. Lastly, researching the role of trust and distrust in a machine vision system might open up other relevant perspectives on what stakeholders find acceptable or not. Starting from something unacceptable might lead to other perspectives of what is acceptable, including how values play a role.

8. CONCLUSION

This chapter concludes on the project by providing a general conclusion and personal reflection on the project.

CONCLUSION

As more smart city technologies are being developed, more attention should be directed on how to include citizens to align such technology with public values and facilitate contestability. This involvement can take place during different phases. To ensure the legitimacy of a machine vision system, citizens should be involved to discuss when a model is performing acceptably. However, enabling this civic participation is a challenge due to a lack of awareness, knowledge, and understanding.

This graduation project explores opening up the discussion around acceptability of a machine vision system, using the scan car development process in Amsterdam as a use case. This discussion clarified acceptability based on trade-offs made during the development phase. To open up this discussion, citizens first need to be able to understand before they can judge such a system and its trade-offs. This resulted in creating a tangible approach to machine vision development, merging elements from fields as TUI and XAI.

The final design was evaluated within three sessions. The results suggest that providing a tangible representation and context-specific examples improved the subjective understanding of participants. The design enabled participants to form and articulate their own opinion about what is acceptable and take part in this discussion. The findings suggest that a tangible approach to participatory machine learning could involve citizens and other non-expert stakeholders in informing and steering a machine vision model during the development phase to close the legitimacy gap.

Facilitating this communication, the project shows a possible contestability loop between citizens and developers in the co-construction of public AI. Ultimately, the design contributes to the field of participatory design approaches to public and responsible AI by providing a practical example.

PERSONAL REFLECTION

Looking back at this project and the last few months, it is time to put, for now, a dot behind this project.

Exploring the complexity of this topic was a challenge, but one I really enjoyed diving into. Sometimes a bit too much, but it opened an interest into smart technologies and how this should or could be designed in the future. As there is still a lot to explore here, it is something I would love to learn more about.

The most valuable thing that I will take from this project is that I can be a bit more sure about my own capabilities, and doubt myself a bit less. While I reflect on most projects with a list of things that could have been better, I am really proud and content with the work I delivered this time.

The combination of researching literature, having conversations with different people, and just trying things out was a balance that really worked for me. Reflecting on myself as a designer, I think I would say I am someone with an analytical eye that likes to dive into complexity, but also someone who likes to keep in touch with practice. And this combination is hopefully something I will do in the future.

Laura

REFERENCES

Alfrink, K., Keller, I., Doorn, N., & Kortuem, G. (2023). Contestable camera cars: A speculative design exploration of public AI that is open and responsive to dispute. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3544548.3580984

Alfrink, K., Keller, I., Kortuem, G., & Doorn, N. (2022). Contestable AI by design: Towards a framework. *Minds and Machines.* https://doi.org/10.1007/s11023-022-09611-z

Almada, M. (2019). Human intervention in automated decision-making. *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law.* https://doi.org/10.1145/3322640.3326699

Ansell, C., & Gash, A. (2007). Collaborative governance in theory and Practice. *Journal of Public Administration Research and Theory,* 18(4), 543–571. https://doi.org/10.1093/jopart/mum032

Barfield, L., van Burgsteden, W., Lanfermeijer, R., Mulder, B., Ossewold, J., Rijken, D., & Wegner, P. (1994). *Interaction design at the Utrecht School of the Arts. ACM SIGCHI Bulletin*, 26(3), 49–86. https://doi.org/10.1145/181518.181525

Bryson, J. M., Quick, K. S., Slotterback, C. S., & Crosby, B. C. (2012). Designing public participation processes. *Public Administration Review,* 73(1), 23–34. https://doi.org/10.1111/j.1540-6210.2012.02678.x

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 205395171562251. https://doi.org/10.1177/2053951715622512

Buxton, B. (2011). *Sketching user experiences: Getting the design right and the right design*. Morgan Kaufmann.

Cai, C. J., Jongejan, J., & Holbrook, J. (2019). The effects of example-based explanations in a machine learning interface. *Proceedings of the 24th International Conference on Intelligent User Interfaces.* https://doi.org/10.1145/3301275.3302289

Chapter 3 – Rights of the Data Subject. General Data Protection Regulation (GDPR). (2018, October 5). https://gdpr-info.eu/chapter-3/

Cheng, H.-F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019). Explaining decision-making algorithms through UI. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3290605.3300789

Colley, A., Väänänen, K., & Häkkilä, J. (2022). Tangible explainable AI - an initial conceptual framework. *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*. https://doi.org/10.1145/3568444.3568456

European AI Act. The Artificial Intelligence Act. (2021, September 9). https://artificialintelligenceact.eu/the-act/

Forlano, L., & Mathew, A. (2014). From design fiction to design friction: Speculative and participatory design of values-embedded urban technology. *Journal of Urban Technology*, 21(4), 7–24. https://doi.org/10.1080/10630732.2014.971525

Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination.* MIT Press.

Hapke, H., & Nelson, C. (2020). *Building Machine Learning Pipelines: Automating model life cycles with tensorflow.* O'Reilly.

Henin, C., & Le Métayer, D. (2021). Beyond explainability: Justifiability and contestability of Algorithmic Decision Systems. *AI & SOCIETY*, 37(4), 1397–1410. https://doi.org/10.1007/s00146-021-01251-8

Ishii, H. (2003). Tangible bits. *Proceedings of the 8th International Conference on Intelligent User Interfaces*. https://doi.org/10.1145/604045.604048

Jiang, J., Kahai, S., & Yang, M. (2022). Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. *International Journal of Human-Computer Studies*, 165, 102839. https://doi.org/10.1016/j.ijhcs.2022.102839

Johansson, U., Sönströd, C., Norinder, U., & Boström, H. (2011). Trade-off between accuracy and interpretability for predictive in silico modeling. *Future Medicinal Chemistry,* 3(6), 647–663. https://doi.org/10.4155/fmc.11.23

Kaspersen, M. H., Bilstrup, K.-E. K., & Petersen, M. G. (2021). The Machine Learning Machine: A tangible user interface for teaching machine learning. *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction.* https://doi.org/10.1145/3430524.3440638

König, P. D., & Wenzelburger, G. (2021). The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it. *Technology in Society,* 67, 101688. https://doi.org/10.1016/j.techsoc.2021.101688

Lam, M. S., Ma, Z., Li, A., Freitas, I., Wang, D., Landay, J. A., & Bernstein, M. S. (2023). Model sketching: Centering Concepts in early-stage machine learning model design. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3544548.3581290

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3313831.3376727

Ministerie van Algemene Zaken. (2023, March 3). *Impact assessment mensenrechten en algoritmes. Rapport | Rijksoverheid.nl.* https://www.rijksoverheid.nl/documenten/rapporten/2021/02/25/impact-assessment-mensenrechten-en-algoritmes

Ng, D. T., Leung, J. K., Chu, S. K., & Qiao, M. S. (2021). Conceptualizing ai literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Parkeercontrole. Amsterdam Algoritmeregister. (n.d.). https://algoritmeregister.amsterdam.nl/parkeercontrole/

Participatie - Gemeente amsterdam. (n.d.). https://www.amsterdam.nl/bestuur-organisatie/invloed/participatie/

Participatieparagraaf. Rekenkamer Metropool Amsterdam. (2023, January 16). https://www.rekenkamer.amsterdam.nl/onderzoek/participatieparagraaf/

Participatieraad. (2022, March 16). *Gevraagd Advies Pilot Artificial Intelligence (AI) Bij de Aanvraag van een (bijstands) uitkering.* Participatieraad Amsterdam https://participatieraadamsterdam.nl/gevraagd-advies-pilot-artificial-intelligence-ai-bij-de-aanvraag-van-een-bijstandsuitkering-2/

Robinson, D. G. (2022). *Voices in the code a story about people, their values, and the algorithm they made.* Russell Sage Foundation.

Schneider, B., Jermann, P., Zufferey, G., & Dillenbourg, P. (2011). Benefits of a tangible interface for collaborative learning and interaction. *IEEE Transactions on Learning Technologies, 4*(3), 222–232. https://doi.org/10.1109/tlt.2010.36

Schouten, B. (2022). *The Civic Empowerment Toolbox.* Jap Sam Books.

Shen, H., Jin, H., Cabrera, Á. A., Perer, A., Zhu, H., & Hong, J. I. (2020). Designing alternative representations of confusion matrices to support non-expert public understanding of algorithm performance. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–22. https://doi.org/10.1145/3415224

Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2020). Participation is not a design fix for Machine Learning. *Equity and Access in Algorithms, Mechanisms, and Optimization.* https://doi.org/10.1145/3551624.3555285

Smith, J. J., Amershi, S., Barocas, S., Wallach, H., & Wortman Vaughan, J. (2022). REAL ML: Recognizing, Exploring, and Articulating Limitations of Machine Learning Research. *In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022*, https://doi.org/10.1145/3531146.3533122

Stolterman, E., & Wiberg, M. (2010). Concept-driven interaction design research. *Human–Computer Interaction,* 25(2), 95–118. https://doi.org/10.1080/07370020903586696

Tan, E. (2017). *Play the city: Games informing the Urban Development.* Jap Sam.

van Turnhout, K., Bennis, A., Craenmehr, S., Holwerda, R., Jacobs, M., Niels, R., Zaad, L., Hoppenbrouwers, S., Lenior, D., & Bakker, R. (2014). *Design patterns for mixed-method research in HCI. Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational.* https://doi.org/10.1145/2639189.2639220

Wolf, C. T. (2019). Explainability scenarios. *Proceedings of the 24th International Conference on Intelligent User Interfaces.* https://doi.org/10.1145/3301275.3302317

GRADUATION BRIEF

IDE Master Graduation

Project team, Procedural checks and personal Project brief

This document contains the agreements made between student and supervisory team about the student's IDE Master Graduation Project. This document can also include the involvement of an external organisation, however, it does not cover any legal employment relationship that the student and the client (might) agree upon. Next to that, this document facilitates the required procedural checks. In this document:

- The student defines the team, what he/she is going to do/deliver and how that will come about.
- SSC E&SA (Shared Service Center, Education & Student Affairs) reports on the student's registration and study progress.
- IDE's Board of Examiners confirms if the student is allowed to start the Graduation Project.

USE ADOBE ACROBAT READER TO OPEN, EDIT AND SAVE THIS DOCUMENT

Download again and reopen in case you tried other software, such as Preview (Mac) or a webbrowser.

STUDENT DATA & MASTER PROGRAMME

Save this form according the format "IDE Master Graduation Project Brief_familyname_firstname_studentnumber_dd-mm-yyyy". Complete all blue parts of the form and include the approved Project Brief in your Graduation Report as Appendix 1 !

family name de Groot
initials L.M. given name Laura
student number 4648382
street & no.
zipcode & city
country
phone
email

Your master programme (only select the options that apply to you):
IDE master(s): ○ IPD ● DfI ○ SPD
2nd non-IDE master:
individual programme: - - (give date of approval)
honours programme: ○ Honours Programme Master
specialisation / annotation: ○ Medisign
○ Tech. in Sustainable Design
○ Entrepeneurship

SUPERVISORY TEAM **

Fill in the required data for the supervisory team members. Please check the instructions on the right !

** chair Ianus Keller dept. / section: HCD
** mentor Kars Alfrink dept. / section: SDE
2nd mentor
organisation:
city: country:

comments (optional) Working relationship with Human Values for Smarter Cities project

Chair should request the IDE Board of Examiners for approval of a non-IDE mentor, including a motivation letter and c.v..

Second mentor only applies in case the assignment is hosted by an external organisation.

Ensure a heterogeneous team. In case you wish to include two team members from the same section, please explain why.

Procedural Checks - IDE Master Graduation

APPROVAL PROJECT BRIEF

To be filled in by the chair of the supervisory team.

chair Ianus Keller date - - signature

CHECK STUDY PROGRESS

To be filled in by the SSC E&SA (Shared Service Center, Education & Student Affairs), after approval of the project brief by the Chair.
The study progress will be checked for a 2nd time just before the green light meeting.

Master electives no. of EC accumulated in total: ______ EC

Of which, taking the conditional requirements into account, can be part of the exam programme ______ EC

List of electives obtained before the third semester without approval of the BoE

○ YES all 1st year master courses passed

○ NO missing 1st year master courses are:

name date - - signature

FORMAL APPROVAL GRADUATION PROJECT

To be filled in by the Board of Examiners of IDE TU Delft. Please check the supervisory team and study the parts of the brief marked **.
Next, please assess, (dis)approve and sign this Project Brief, by using the criteria below.

- Does the project fit within the (MSc)-programme of the student (taking into account, if described, the activities done next to the obligatory MSc specific courses)?
- Is the level of the project challenging enough for a MSc IDE graduating student?
- Is the project expected to be doable within 100 working days/20 weeks ?
- Does the composition of the supervisory team comply with the regulations and fit the assignment ?

Content: ○ APPROVED ○ NOT APPROVED

Procedure: ○ APPROVED ○ NOT APPROVED

comments

name date - - signature

Agonistic Machine Vision Development project title

Please state the title of your graduation project (above) and the start date and end date (below). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

start date 09 - 01 - 2023

02 - 06 - 2023 end date

INTRODUCTION **

Please describe, the context of your project, and address the main stakeholders (interests) within this context in a concise yet complete manner. Who are involved, what do they value and how do they currently operate within the given context? What are the main opportunities and limitations you are currently aware of (cultural- and social norms, resources (time, money,...), technology, ...).

Within smart cities, there is a growing development of using sensors and automated decision-making systems (ADM) around us. Using these new technologies, often seamlessly integrated with the city context, can help monitor, regulate and understand the use of the city. Benefits for a municipality can be lower costs and more efficiency. For example, scan cars with machine vision are already going around in multiple cities (image 1). They are mainly used for controlling parking fees, but other possible uses are being researched. The city of Amsterdam is planning to develop new scan cars and applications.

With the use of these systems in a public context, more concerns around its possible harms and ethical considerations are being researched (e.g., lack of fairness, transparency, legitimacy, and accountability). Technology is not neutral: values are embedded in technology through design. The city of Amsterdam wants to design future scan cars using understandable, transparent, and ethical machine vision systems (Human Values for Smarter Cities). Stakeholder participation can help carefully consider how values, such as fairness, trust, and privacy, should be embedded to create a responsible system.

For a democratic society to function properly, people must be able to understand and criticize how their cities are run (Human Values for Smarter Cities). Within the context of scan cars, this implies that people must be able to understand and criticize the system that is going to be used. Being 'responsive to human intervention throughout the system's life cycle can facilitate this desired contestation or relationship (Alfrink et al., 2020). A current challenge identified around contestability in public AI is various knowledge positions: citizens have insufficient awareness, knowledge, and understanding of systems to contest (Alfrink et al., 2022). To prevent 'participation-washing,' stakeholders need to know why what, and how to contest a system.

At the beginning of the AI systems' lifecycle, agonistic approaches to ML development, such as participatory design, can enable stakeholders to "co-construct the decision-making process" (Vaccaro et al., 2019). Having different perspectives and knowledge positions makes it hard to have practical discussions The debate around future scan cars is complicated by personal views and technical elements (technical illiteracy). Hence, machine vision systems must be better explained to citizens to improve the debate around the design of the future scan car and make (all) stakeholders full-fledged interlocutors.

This graduation project will focus on citizens' current lack of awareness, understanding, and knowledge (various knowledge positions). It will explore how to provide an 'understanding' of machine vision within a participatory session to make all stakeholders full-fledged interlocutors. The gained insights will contribute to the development of new scan cars responsibly and design methods regarding creating public AI.

space available for images / figures on next page

Initials & Name L.M. de Groot Student number 4648382

Title of Project Agonistic Machine Vision Development introduction (continued): space for images

image / figure 1: Camera Car in Amsterdam

image / figure 2: Participation Amsterdam

Initials & Name L.M. de Groot Student number 4648382

Title of Project Agonistic Machine Vision Development

PROBLEM DEFINITION **

Limit and define the scope and solution space of your project to one that is manageable within one Master Graduation Project of 30 EC (= 20 full time weeks or 100 working days) and clearly indicate what issue(s) should be addressed in this project.

The city of Amsterdam wants to develop more understandable, transparent, and ethical machine vision systems for the scan cars of the future. The city will use participatory processes to create designs responsibly. Involving direct and indirect stakeholders in the early stages of the development of machine learning systems can address different perspectives and prevent potential harm in the future. While they are part of the interplay between technology and society, stakeholders have different knowledge positions regarding awareness, knowledge, and understanding of the topics (Alfrink et al., 2022). These knowledge positions can limit the discussion between stakeholders in participatory design sessions and need to be improved to make everyone full-fledged interlocutors.

The current challenge is understanding aspects of machine vision systems to co-construct contestable machine learning systems. This challenge results in the main research question:

"How to improve the understanding of machine vision systems to make stakeholders full-fledged interlocutors in the co-construction of future scan cars?"

This research is relevant for the responsible design of public AI for new design methods around how to design public AI responsibly. It will add new insights for creating contestable AI systems in the city context and add to the development of new scan cars of the future.

ASSIGNMENT **

State in 2 or 3 sentences what you are going to research, design, create and / or generate, that will solve (part of) the issue(s) pointed out in "problem definition". Then illustrate this assignment by indicating what kind of solution you expect and / or aim to deliver, for instance: a product, a product-service combination, a strategy illustrated through product or product-service combination ideas, ... In case of a Specialisation and/or Annotation, make sure the assignment reflects this/these.

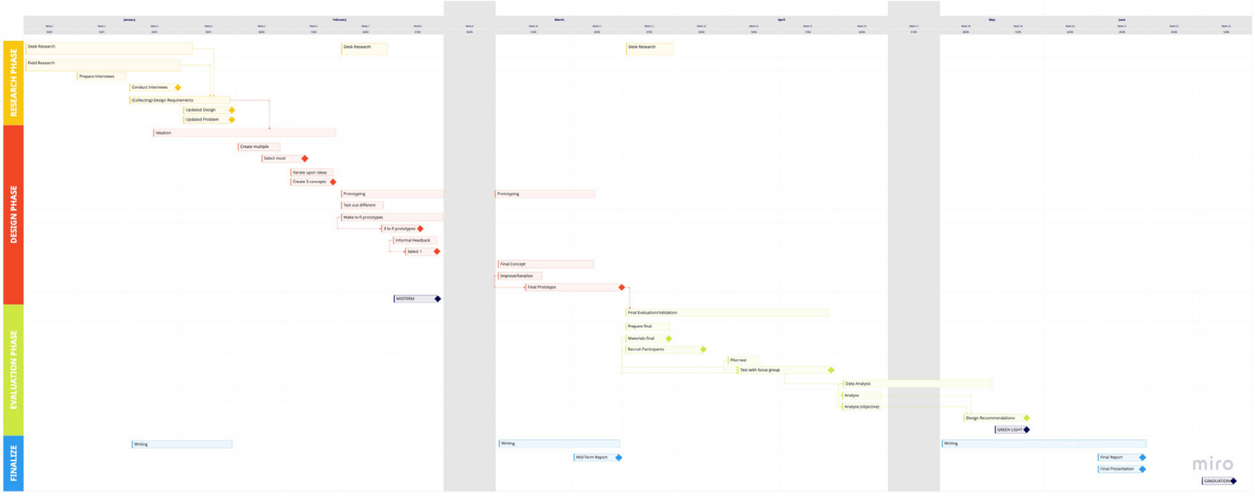
The desired outcome of this project is a (high fidelity) prototype that provides hands-on interaction with machine vision. Citizens can use it during the early design phase of future scan cars to understand the technical aspects of machine vision to nurture discussion about the design implementation.

PLANNING AND APPROACH **

Include a Gantt Chart (replace the example below - more examples can be found in Manual 2) that shows the different phases of your project, deliverables you have in mind, meetings, and how you plan to spend your time. Please note that all activities should fit within the given net time of 30 EC = 20 full time weeks or 100 working days, and your planning should include a kick-off meeting, mid-term meeting, green light meeting and graduation ceremony. Illustrate your Gantt Chart by, for instance, explaining your approach, and please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any, for instance because of holidays or parallel activities.

start date 9 - 1 - 2023

2 - 6 - 2023 end date



The project's first phase will focus on researching and exploring the current challenge regarding knowledge positions and participatory design (in Gemeente Amsterdam). The research will combine desk research with field research (in the shape of semi-structured interviews). Topics to research and take inspiration from are contestable/public AI, value-sensitive design, participatory design, and current applications of ADM systems in a socio-technical context, but also evaluation and validation methods. At the end of this phase, the problem definition and design direction, with its design requirements from research, can be updated.

The second phase will involve iteratively developing and prototyping concepts that help understand technical aspects of machine vision and the discussion of future scan cars. After ideating and selecting a few ideas, I will make lo-fi prototypes of a few concepts to informally research and get feedback on the designs. After this, one final iteration will provide the final design and a final hi-fi prototype.

The third phase will focus on formally evaluating the final design. The exact method for evaluation depends partly on the research and design phase. However, it will include testing the design intervention and its influence with a focus group. This evaluation will provide different results to answer the research question. I will question this focus group on their knowledge position before and after the design intervention to test (objectively) their level of understanding. Small discussions/interviews can show the subjective experience. Being full-fledged interlocutors also can be investigated from an objective (analyzing and transcribing discussion) and subjective (interviewing if stakeholders felt heard or important) perspective.

The last cycle focuses on analyzing the results and translating them into recommendations. During each phase, I will dedicate time to capturing insights and progress in writing during the final two weeks. These drafts will provide a base for writing the report and presentation graduation.

MOTIVATION AND PERSONAL AMBITIONS

Explain why you set up this project, what competences you want to prove and learn. For example: acquired competences from your MSc programme, the elective semester, extra-curricular activities (etc.) and point out the competences you have yet developed. Optionally, describe which personal learning ambitions you explicitly want to address in this project, on top of the learning objectives of the Graduation Project, such as: in depth knowledge a on specific subject, broadening your competences or experimenting with a specific tool and/or methodology, ... Stick to no more than five ambitions.

During my electives, I explored various topics around new human-AI interactions emerging in various fields around us. After the elective AI & Society, I found the combination of these new technologies and how they can influence how we shape our lives very interesting. Next, I tried to familiarize myself with different design angles regarding human-AI interaction by following courses such as Advanced Machine Learning for Design, Designing Intelligence, and More-than-human Design. These courses resulted in going back to my interest in the interplay of applying/using new technologies in social environments/contexts.

This graduation project is an opportunity to explore the topic in an applied use case instead of focusing on only theory and knowledge. It will be the first time combining design principles with knowledge of AI, which will be very interesting.

So for this project, I have two main ambitions:
(1) Experience implementing AI/ML technology within a design instead of ‘wizard-of-oz’-ing these aspects.
(2) Learn how to work with various stakeholders, each with their own interests.

FINAL COMMENTS

In case your project brief needs final comments, please add any information you think is relevant.