

APPENDIX

APPENDIX A: FIGMA

Link to Interface Prototype:

https://www.figma.com/file/f3riChPSPf6B9Lv0fbmxtR/Prototype_Final?type=design&node-id=0%3A1&t=WYapNG5Tanpucouc-1

APPENDIX B: SESSION PLAN

Session Plan

Introduction to Session

<ul style="list-style-type: none">• Welkom allemaal, top dat jullie hier vandaag konden zijn, en alvast bedankt voor het meedoen. Is het goed als ik foto's maak + opneem wat jullie	<i>Op scherm: slide 1</i>
<ul style="list-style-type: none">• Zoals volgens mij de meesten wel hebben meegekregen, ben ik afgelopen tijd bezig geweest met mijn afstudeerproject. Hierbij ben ik gaan onderzoeken hoe je burger zou kunnen betrekken tijdens de ontwikkeling van machine vision systemen. Uiteindelijk is hier een bepaald prototype uit ontstaan. Daar gaan jullie zo mee aan de gang, en het doel van deze evaluatie is dan ook dit ontwerp evalueren, en voor mij om te kijken of het	<i>Op scherm: slide 2</i>
<ul style="list-style-type: none">• Dan nu over vandaag, iets meer context: Op dit moment is de gemeente Amsterdam bezig met het ontwikkelen van verschillende technologieën om te gebruiken in de stad, om zo efficiënter te werk te gaan, maar ook om het leven in de stad fijner en veiliger te maken. Een project hiervan is de ontwikkeling van scan auto's. Deze komen jullie misschien bekend voor [<i>plaatje scan auto toevoegen</i>]: op dit moment worden deze scan auto's ingezet om parkeercontroles uit te voeren. Boven op de auto zit een	<i>Op scherm: slide 3</i>
<ul style="list-style-type: none">• Een daarvan is het detecteren van puin containers in de stad [<i>foto puincontainer</i>]. De gemeente wilt de technologie hiervoor inzetten. Voor het plaatsen van zulke containers is een vergunning nodig. Maar vaak is deze vergunning er niet, waardoor zulke puin containers illegaal door de stad verspreid staan. Nadelen hiervan zijn dat zulke containers ruimte op straat, op de stoep maar ook in parkeervlakken inneemt. Daarnaast zijn volle containers vaak zwaar, en is er op dit moment een probleem met zwakke kades in de stad. Hier zouden helemaal geen zware voorwerpen geplaatst mogen worden, aangezien een verzakking van de kade veel schade kan aanrichten. Om dus de plaatsing van containers in de stad beter en efficiënter te kunnen controleren, wilt de gemeente hier de scan auto met een machine vision model voor gebruiken.	<i>Op scherm: slide 4</i>
<ul style="list-style-type: none">• Jullie doel van vandaag is om aan het eind van de sessie een samenstelling van machine vision model te hebben, wat jullie allemaal goed genoeg en acceptabel vinden, waar jullie allemaal mee zouden kunnen leven. Aangezien hier een aantal subjectieve elementen in verworven zitten, is er niet één consensus, een juiste beste model, of een model van iedereen 100% tevreden mee moet of kan zijn. Daarom vraag ik jullie ook om bij de uiteindelijke samenstelling, juist deze verschillende pijnpunten of	<i>Op scherm: slide 5</i>

<ul style="list-style-type: none"> • Vanaf nu zal ik naar de achtergrond verdwijnen en gaan observeren. Als er dingen onduidelijk zijn of er zijn vragen, stel ze aan elkaar, of onthoud ze voor achteraf. De verdere instructies zullen vanaf nu op het scherm verschijnen. Dus ik geef nu de muis aan een van jullie, daarmee kunnen 	Op scherm: slide 6
---	--------------------

Introduction to Trade-offs

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	Op scherm: slide 7
Het model dat hier voor zal worden gebruikt, wordt op dit moment ontwikkeld door de gemeente. Dit is is een iteratief proces, waarbij steeds gekeken wordt naar hoe het model 'beter' kan detecteren. Dit optimaliseren kan vergeleken worden met een zoektocht naar het allerlaagste punt van een vallei: dat is de beste variatie van het model. Alleen is er in dit geval niet één antwoord over wat dat beste model precies is, aangezien dit subjectieve	Op scherm: slide 8
Tijdens deze zoektocht worden verschillende afwegingen gemaakt, die meespelen in wat het gewenste model is. Bij deze afwegingen, vanaf nu trade-offs genoemd, moet een compromis tussen twee elementen gesloten worden, een balans gevonden. <u>Klik op de weegschaal om de trade-offs te bekijken</u> . Vandaag is de focus op de volgende trade-offs: accuraatheid versus interpreteerbaarheid, en precision versus recall. Op tafel liggen rode tokens met deze trade-offs; <u>Verdeel de muntjes op tafel zodat iedereen er twee</u>	Op scherm: slide 9
Om een mening over deze trade-offs te kunnen vormen is het van belang om te weten waar deze vandaan komen. Tijdens het tunen, het optimaliseren, van het model wordt steeds geëvalueerd hoe goed het model kan voorspellen of er een container te zien is. Tijdens dit proces geven we het	Op scherm: slide 10
Er zijn dan vier opties mogelijk, die worden weergegeven in een verwarringsmatrix. Dit zijn: <ul style="list-style-type: none"> - Echte positief: er is een container gedetecteerd, en er is daadwerkelijk een container - Echt negatief: er is geen container gedetecteerd, en er is er ook geen - Vals positief: er is een container gedetecteerd, maar er is eigenlijk geen 	Op scherm: slide 11
Stel je voor, om te kijken hoe goed het model presteert, laten we het model 50 foto's met en 50 foto's zonder containers analyseren. Hoeveel goede voorspellingen maakt het model nu? <u>Klik op evalueer</u> . Op basis van deze resultaten kunnen we evalueren hoe accuraat het model is. Accuraatheid van een model is het percentage van de voorspellingen dat het model goed had. In dit geval is het 53%. <u>Klik op het percentage om te zien waar dit getal</u>	Op scherm: slide 12/13

Interpreteerbaarheid is de mate waarin het model menselijk begrip mogelijk maakt. Dus, hoe goed zijn we in staat te begrijpen hoe het model tot een bepaalde beslissing of detectie is gekomen. In een ideale wereld zouden we het liefst een heel accuraat en een heel interpreteerbaar model hebben. Alleen kan dit in de praktijk niet. Om een model accuraat te maken, maken	Op scherm: slide 14
Dit leidt tot de eerste 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? <u>Leg allemaal de</u>	Op scherm: slide 15 Observeer: Discussie?
Maar accuraatheid is niet de enige manier om te evalueren hoe het model presteert. De focus moet niet enkel zijn op wat het model goed doet, maar we moeten ook kijken naar de fouten die worden gemaakt. En hoe erg is het dat het model deze fouten maakt? Met precision kijken we van de keren dat	Op scherm: slide 16
Met recall kijken we naar van alle containers die gedetecteerd zouden moeten worden, hoeveel zijn er daadwerkelijk gedetecteerd? Dus hoeveel containers heeft het model gemist? En hoeveel containers zal het model dus in werkelijkheid kunnen missen? Als dit willen minderen, dan zullen er meer	Op scherm: slide 17
Dit is de tweede trade-off van vandaag: precision versus recall. Als we de ene soort fout willen minderen, dan leidt dit tot meer fouten van de andere soort. Dus de vraag is, is het belangrijker om zoveel mogelijk werkelijke containers te detecteren, of is het belangrijker dat alle voorspellingen, detecties, die we doen, daadwerkelijk containers zijn? <u>Leg allemaal de munt neer. Waarom</u>	Op scherm: slide 18 Observeer: Discussie?

(aanpassen wat op tafel ligt)

- Evaluer element met pionnen
- Whiteboard marker
- Systeem iconen met straat
- Tokens trade-offs

Level 1: Decision Situation

Start scherm level 1: instructies: gevolgen van detecties bekijken	Op scherm: slide 19
Kies een model + neem deze gegevens over op tafel	Op scherm: slide 20
Ga langs iconen op tafel	Op scherm: slide 21 Observeer: Discussie aan de
Alles gehad? Tokens omdraaien naar wat jij belangrijk vind	Op scherm: slide 22

Pauze

(aanpassen wat op tafel ligt)

- Model tuning toevoegen

Level 2: Model Tuning

Start scherm level 2: instructies maak het model zodat jullie kunnen tunen. Pas op, het is wel een versimpelde weergave van hoe het in het echt gaat, focus is meer van het begrijpen van verschillende elementen. Jullie krijgen 3	Op scherm: slide 23
Onderdelen in het model doen	Op scherm: slide 23
Train button: klik	Op scherm: slide 24
Performance overview op scherm	Op scherm: slide 24
Train button: klik	Op scherm: slide 24
Performance overview op scherm	Op scherm: slide 24
Train button: klik	Op scherm: slide 24
Performance overview op scherm	Op scherm: slide 24
Introduce perspectives op tafel. Jullie krijgen nu nog 3 pogingen om het	Op scherm: slide 25
Train button: klik	Op scherm: slide 24
Performance overview op scherm	Op scherm: slide 24
Train button: klik	Op scherm: slide 24
Performance overview op scherm	Op scherm: slide 24
Train button: klik	Op scherm: slide 24
Performance overview op scherm	Op scherm: slide 24
Pogingen zijn op. Zijn jullie tevreden nu? Wat zou jezelf anders willen zien? Wat voor meningsverschillen kunnen er zijn? Schrijf dit op	Op scherm: slide 26

Afronding

POST-KNOWLEDGE TEST

Nog weer 10 theoretische vragen achteraf, dus die zal ik zo opsturen. En graag deze invullen dan. Kan ook nu, wat uitkomt. Maar graag niet al te lang wachten.

APPENDIX C: EVALUATION PLAN

Link to Interface Prototype:

https://miro.com/app/board/uXjVPwVAcFw=?share_link_id=770227321580

APPENDIX D: PRE-QUESTIONNAIRE

Link to Interface Prototype:

<https://forms.gle/Fgm6zNDgcRB9dmz19>

APPENDIX E: POST KNOWLEDGE TEST

1. De rol van een scan auto is..
 - .. data verzamelen
 - .. controleren van mensen aan het doen zijn
 - .. boetes uitdelen
 - Al het bovenstaande
2. *Hoe maakt een machine vision systeem mensen in de stad onscherp (blur)?*
 - De mensen in de video worden nooit gedetecteerd, wat wazige plekken veroorzaakt.
 - Nadat het model de video heeft verwerkt, vervaagt iemand de mensen in de opgeslagen gegevens.
 - Het model detecteert eerst mensen in het beeld, om vervolgens voegt er dan een vervaging aan toe.
 - Nadat het model mensen heeft gedetecteerd, staan deze aangekaart (geflagged), waardoor iemand de mensen in de opgeslagen gegevens kan weghalen.
3. *We hebben een machine vision model de opdracht gegeven om katten te detecteren. Wat omschrijft het beste de term vals negatief in deze situatie?*
 - Detecties van katten met een lage detectie zekerheid
 - Detecties van katten die achteraf geen katten blijken te zijn
 - Detecties van katten die het model al eens eerder heeft gezien
 - De katten die het model niet heeft gedetecteerd
4. *De gevolgen van een detectie gemaakt door een scan auto zijn altijd hetzelfde.*
 - Waar, iedere detectie heeft dezelfde consequenties om het eerlijk te maken
 - Waar, iedere detectie leidt automatisch tot een boete
 - Niet waar, een detectie wordt eerst bekeken door een mens of deze correct is
 - Niet waar, alle detectie moeten meermaals gemaakt worden voordat er consequenties aan vast zitten
5. *Tijdens het ontwikkelen van een machine vision systeem, tunen de ontwikkelaars het model. Wat houdt dit in?*
 - Door het model te tunen, zijn ontwikkelaars op zoek naar de optimale prestatie van het model door verschillende hyperparameters aan te passen in een iteratief proces
 - Door het model te tunen, zijn ontwikkelaars opzoek naar patronen in de verzamelde data
 - Door het model te tunen, zijn ontwikkelaars aan het kijken naar verschillende mogelijke toepassingen voor het model
 - Door het model te tunen, zijn ontwikkelaars meer labels aan het toevoegen aan het model om te gaan detecteren
6. *Waarom evalueer je een model tijdens de ontwikkeling fase?*
 - Om erachter te komen hoe snel het model detecties kan maken
 - Om in te zien hoe het model nu presteert, zodat je het kan aanpassen
 - Om te zien hoe wat de consequenties van het model in de echte wereld zouden zijn
 - Dat doe je niet, dat doe je pas achteraf als het model echt in de buitenwereld heeft gewerkt

7. *Wat wordt bedoeld met de accuraatheid van een model?*
- De nauwkeurigheid van de voorspellingen die het model gaat doen
 - Het percentage van de voorspellingen dat het model correct had
 - Hoe kritisch het model naar zijn voorspellingen kijkt
 - De mate waarvoor het model klaar is om ingezet te worden
8. *Wat is een voordeel van een goed interpreteerbaar machine vision model?*
- Doordat het model beter te interpreteren is, zullen de resultaten accurater zijn
 - Je beter begrip krijgt van hoe een machine learning model een beslissing heeft genomen, waardoor je je eigen beslissingen beter kan onderbouwen
 - Doordat het model beter te interpreteren is, weet je precies hoe een bepaalde detectie tot stand is gekomen
9. *Je moet altijd een complex machine learning model gebruiken.*
- Niet waar, je hebt alleen een complex model nodig in een kritieke situatie
 - Niet waar, in sommige gevallen is er een afweging tussen complexiteit en transparantie/interpreteerbaarheid
 - Waar, complexe modellen presteren altijd beter
10. *In een hypothetisch medisch model, welk van de volgende elementen beschrijft de recall waarde het beste?*
- Onder de zieken, het aantal mensen dat correct als ziek werd gediagnosticeerd
 - Van de mensen die als ziek zijn gediagnosticeerd, is het aantal mensen dat werkelijk ziek is
 - Onder alle gezonde mensen, het aantal individuen gediagnosticeerd als gezond

APPENDIX F: SEMI-STRUCTURED INTERVIEW

SEMI-STRUCTURED FOLLOW-UP INTERVIEW/DISCUSSION

Super, top dank jullie wel. Ik heb nu een aantal vragen voor jullie.

1. Wat is jullie eerste reactie, ervaring?

De hoofd zoektocht van vandaag was het vinden van een acceptabel model. Zijn jullie allemaal tevreden over het laatste model wat jullie hebben samengesteld?

2. Zijn jullie allemaal tevreden over het laatste model wat jullie hebben samengesteld? Waarom wel of waarom niet?

- Waar verschillen jullie meningen? Wat had je zelf anders willen doen en waarom?
- Heeft hier discussie over plaatsgevonden?
- Hoe zijn jullie om gegaan met de verschillende perspectieven?

3. Wat speelt in jullie ogen allemaal of een model acceptabel is? Welke elementen spelen hier allemaal in mee? (systeem, model, gevolgen etc)

- Waren deze elementen duidelijk voor iedereen? Waarom wel of waarom niet? Wat maakte het duidelijk of onduidelijk?
- Waren er dingen die je miste, waar je nog meer over zou willen weten?

Het doel was om jullie over de trade-offs en wat belangrijker is te laten discussiëren tijdens de sessie.

4. Heeft iedereen het gevoel/idee gehad dat ze deel konden meepraten over wat acceptabel voor het model is? Waarom wel of waarom niet?

- Rol van kennis of informatie kunnen beschrijven?
- Hadden jullie genoeg informatie en kennis? (systeem, model, gevolgen etc)
- Hebben jullie deze informatie kunnen gebruiken in je argumentatie? Waarom wel of niet?
- Welke vragen had je nog willen stellen? Dingen die je zou willen weten?

5. Moeten burgers betrokken worden in dit proces, in deze beslissing? Hoe zou je dit omschrijven? Waarom wel of niet?

6. Wat is de rol van het ontwerp in deze discussie geweest?

- Hoe zou de discussie over trade-offs zijn gegaan als wat op tafel ligt was weggebleven?
- Stel we zouden dit nog een keer doen, wat zou jij dan anders doen? Wat zou je hetzelfde houden?
- Vinden jullie dat burgers betrokken moeten worden in dit proces? Denken jullie dat zo'n soort format zou helpen om samen te werken?
- Nog ander comments?