



Ontwerpgericht onderzoek: een smart- bin die zichzelf uitlegt

Hoe uitlegbare AI afvalscheiding begrijpelijk en betrouwbaar maakt op het moment van weggooien

Auteurs

2405109	-	Djody	Kort
2404043	-	Lars	Kleijnen
2403718	-	Boril	Vineshki
2404104	- Vince Lebourg		

Casus: Optimize - Duurzame Technologie in Beeld

Laatst Gewijzigd: 01-07-2026



Inhoudsopgave

Voorwoord	2
Samenvatting	2
1 Inleiding	3
1.1 Achtergrond	3
1.2 Aanleiding	3
1.3 Vooronderzoek	4
1.4 Doelstelling	4
1.5 Leeswijzer	4
2 Theoretisch kader	5
2.1 Ontwerpvrage	5
2.2 Theoretische uitgangspunten	5
2.3 Begripskloof rond AI (cluster A)	5
2.4 Afvalscheidingsgedrag (cluster B)	5
2.5 Technische realisatie (cluster C)	6
2.6 Deelconclusie onderzoek	6
3 Methode	6
3.1 Onderzoeksstrategie en samenhang	7
3.2 Relevance-cyclus: interviews en enquête	7
3.3 Rigor-cyclus: literatuuronderzoek en experts	8
3.4 Design-cyclus: bouw, A/B-test en kwaliteitsborging	8
4 Resultaten	9
4.1 Relevance-cyclus: de eisen vanuit de omgeving	9
4.2 Rigor-cyclus: de wetenschappelijke grounding	11
4.3 Design-cyclus: het artefact, gebouwd en geëvalueerd	12
5 Discussie	14
5.1 De rode draad en de gedocumenteerde koerswijziging	14
5.2 Reikwijdte van de enquête	14
5.3 Kracht en grens van de A/B-test	14
5.4 Niet uitgevoerde metingen	14
5.5 Aanbevelingen voor vervolgonderzoek	14
6 Conclusie	15
Literatuurlijst	16
Documentrelaties	19
Bijlagen	20
Uitgebreid drie-cyclusmodel	21

Voorwoord

Dit rapport is de afronding van een ontwerpgericht onderzoek dat het team Optimizers heeft uitgevoerd voor het Lectoraat Data Intelligence van Zuyd Hogeschool.

Het project raakte aan **twee onderwerpen die ons aanspraken**:

1. Kunstmatige intelligentie begrijpelijk maken voor mensen zonder voorkennis
2. Mensen helpen om afval beter te scheiden.

Wat begon als twee losse vraagstukken groeide gaandeweg uit tot één samenhangend ontwerp.

Wij danken onze opdrachtgever Koen Steeghs voor zijn betrokkenheid en de ruimte om te itereren, en onze begeleiders en de geraadpleegde experts voor hun kritische blik. Dit verslag laat zien hoe onderzoek en ontwerp elkaar stap voor stap hebben gestuurd.

Het team Optimizers, juli 2026.

Samenvatting

Het Lectoraat Data Intelligence van Zuyd Hogeschool wil duurzame technologie zichtbaar en begrijpelijk maken voor een breed publiek. Binnen het project Optimize zijn **twee knelpunten samengebracht**:

1. Een begripskloof rond kunstmatige intelligentie bij mensen zonder voorkennis
2. Aanhoudend onjuist afvalscheidingsgedrag in de publieke ruimte.

De centrale ontwerp vraag luidt hoe een interactief smart-bin-artefact met een uitlegbare AI-laag (explainable AI) tegelijk de werking van AI begrijpelijk kan maken en correct afvalscheidingsgedrag kan ondersteunen op het moment van handelen.

Het onderzoek volgt Design Science Research (Hevner, 2007), met interviews, een literatuuronderzoek, een enquête en een A/B-test als elkaar versterkende methoden. De enquête (N = 47) laat zien dat afvalscheiding een ingesleten gewoonte is, maar dat dertig procent onzeker is over de juiste bak: een zekerheidskloof, geen motivatieprobleem. De literatuur bevestigt dat een gelaagde uitleg zonder percentages, op het beslismoment en met een schuldvrije terugval, het best aansluit bij een leek.

Het gebouwde artefact, een smart-bin-kiosk met twee herkenningroutes (een streepjescodescanner en een Gemma-first-AI-route) en een route-stoplicht waarin de AI nooit als zekerheid wordt gepresenteerd, brengt deze inzichten samen. Een A/B-test op een toegevoegde richtingspijl laat zien dat bezoekers daarmee sneller kiezen, $t(9) = 3,80, p = 0,004$. Het advies is de richtingspijl op te nemen en het artefact in een veldopstelling verder te valideren.

1 Inleiding

1.1 Achtergrond

Dit project is uitgevoerd in opdracht van het Lectoraat Data Intelligence van Zuyd Hogeschool, binnen het thema Duurzame Technologie in Beeld. De opdrachtgever streeft naar een gelijkwaardige balans tussen duurzaamheid en de zichtbare werking van AI, in lijn met de Europese twin transition: het beleid om de digitale transitie (technologie) en de groene transitie (duurzaamheid) elkaar te laten versterken (Europese Commissie, 2019). Het project richt zich op een van de zeven duurzaamheidsdomeinen binnen dit kader, circulariteit en afval, omdat dit thema direct verweven is met het dagelijks leven.

De doelgroep is in de Stakeholderanalyse (Bijlage 1) **vastgelegd als Jan Modaal:**

Een brede groep volwassenen tussen ongeveer 17 en 65 jaar zonder specifieke voorkennis van AI, ICT of duurzaamheid, die in Zuid-Limburg woont, werkt of publieke evenementen bezoekt.

De Persona's (Bijlage 2) werken dit profiel uit naar concrete archetypen.

1.2 Aanleiding

De aanleiding kent **twee lagen die in elkaar grijpen:**

1. De eerste is een begripskloof rond AI en duurzaamheid, die leidt tot zowel overschatting als wantrouwen.
2. De tweede is verkeerd afvalscheidingsgedrag in de publieke ruimte.

Een analyse met de Five Times Why-methode en een Current Reality Tree in de Probleemanalyse (Bijlage 3) wees **op een gedeelde kernoorzaak:**

De afwezigheid van ervaringsgerichte feedbackmomenten waarin AI tastbaar wordt en het gedrag direct wordt teruggekoppeld. Beide ketens komen samen in de bekende kloof tussen houding en gedrag (Steg & Vlek, 2009).

De urgentie is concreet. In 2024 steeg het huishoudelijk afval naar 456 kilogram per inwoner, het einde van een jarenlange dalende trend (Centraal Bureau voor de Statistiek [CBS], 2025a, 2025b). Wanneer de vervuiling in het ingezamelde plastic te hoog is, wordt een hele lading afgekeurd, wat gemeenten direct geld kost (RWM N.V., 2024).

Tegelijk is de AI-geletterdheid van het brede publiek beperkt:

Vierentachtig procent van de Europeanen vraagt om zorgvuldige sturing op privacy en transparantie rond AI (Verian, 2025). Een eerdere proef met AI-afvaldetectie in de Zeeheldenbuurt (2023) werd technisch geslaagd bevonden, maar bewoners herkenden zich niet in de resultaten. Dat laat zien dat uitleg en herkenbaarheid even belangrijk zijn als de techniek.

1.3 Vooronderzoek

Een vooronderzoek, samengevat in de Quicksan, bakende de realistische ontwerpruimte af. Het zelf trainen van een volledig neuraal netwerk binnen de projecttijd bleek een te groot risico, waardoor een voorgetraind herkenningsmodel de realistische route werd. De besluiten van AI moeten voor leken vooral visueel begrijpelijk worden gemaakt, en de interactietijd bij publieke opstellingen is kort, met een richtwaarde van dertig tot honderdtwintig seconden (Wouters et al., 2016).

Privacy bleek een harde randvoorwaarde:

Door lokale beeldverwerking zonder opslag van beelden worden de zwaarste risico's rond de Algemene Verordening Gegevensbescherming (AVG) vermeden, al blijven de informatieplicht en de rechtsgrond gelden. De kernbegrippen die hieruit en uit het latere literatuuronderzoek als rode draad naar voren kwamen, zijn AI-geletterdheid, uitlegbare AI, de kloof tussen houding en gedrag en feedback op het beslismoment.

1.4 Doelstelling

De ontwerpuitdaging is een interactief artefact te ontwerpen dat op het moment van weggooien zowel het gedrag ondersteunt als de AI begrijpelijk maakt. De doelstelling is volgens de SMART-criteria uitgewerkt in het Plan van Aanpak (Grit, 2019): achterhalen welke uitlegvorm, gedragsondersteuning en technische realisatie passen bij een publiek zonder voorkennis, en dat vertalen naar een werkend, onderbouwd artefact, zodat de ontwerpkeuzes navolgbaar uit de bevindingen volgen. De fase-1-producten en de afbakening zijn door de opdrachtgever in het vierde interview bevestigd als geldig startpunt. De formele ontwerpvrage staat in H2.1.

1.5 Leeswijzer

- **H2 Theoretisch kader:**
Formuleert de centrale ontwerpvrage en vat de kennisbasis per cluster samen tot een deelconclusie.
- **H3 Methode:**
Plaatst het onderzoek in het Design Science Research-model en beschrijft per cyclus welke methoden zijn ingezet en hoe de kwaliteit is geborgd.
- **H4 Resultaten:**
Presenteert de bevindingen langs dezelfde drie cycli, met de enquête als statistisch zwaartepunt en het gebouwde artefact als afronding.
- **H5 Discussie:**
Weegt per thema de kwaliteit en de grenzen en geeft aanbevelingen voor vervolgonderzoek.
- **H6 Conclusie:**

Beantwoordt de centrale ontwerpvrage en geeft het advies aan de opdrachtgever. De samenhang met de overige projectdocumenten staat in de documentrelatietabel achterin, en de onderbouwende deliverables zijn als genummerde bijlagen opgenomen (zie het bijlagenregister).

2 Theoretisch kader

2.1 Ontwerpvraag

De centrale ontwerpvraag stuurt zowel het onderzoek als het ontwerp en luidt:

Hoe kan een interactief smart-bin-artefact met een uitlegbare AI-laag aan een breed publiek zonder AI-voorkennis tegelijk de werking van AI begrijpelijk maken en correct afvalscheidingsgedrag ondersteunen op het moment van handelen, in de publieke ruimte van Zuid-Limburg?

De vraag is uitgewerkt in **tien deelvragen** in **drie thematische clusters**:

1. Begripskloof rond AI (cluster A)
2. Afvalscheidingsgedrag (cluster B)
3. Technische realisatie (cluster C, verkennend)

De volledige deelvragen per cluster staan in het Literatuuronderzoek (Bijlage 4).

2.2 Theoretische uitgangspunten

De drie clusters zijn beantwoord op basis van **wetenschappelijke** en **grijze literatuur**, die samen de kennisbasis voor het ontwerp vormen. Naast deze literatuur rust de verantwoording op user-centered design, waarbij de gebruiker centraal staat in elke ontwerpstep (Norman & Draper, 1986; Internationale Organisatie voor Standaardisatie [ISO], 2019), en op een ethische toetsing met uitlegbaarheid als overkoepelende eis (Floridi et al., 2018).

De volgende paragrafen vatten de centrale bevindingen per cluster samen. De volledige onderbouwing staat in het Literatuuronderzoek (Bijlage 4).

2.3 Begripskloof rond AI (cluster A)

AI-geletterdheid bestaat uit meerdere onderdelen, en bij een korte publieksinteractie is voor de doelgroep alleen het **herkennings- en begripsniveau haalbaar**: begrijpen dat een AI-systeem kijkt, dat het niet onfeilbaar is en dat de uitkomst een inschatting is (Long & Magerko, 2020; Ng et al., 2021).

De beste uitleg voor een leek is gelaagd:

Een kader om het herkende voorwerp, een zekerheidsindicatie in kleur of woord (uitdrukkelijk geen percentage) en één zin uitleg in gewone taal. Ruwe waarschijnlijkheden en warmtekaarten die tonen waar het model naar kijkt, zijn voor leken ongeschikt (Adadi & Berrada, 2018). Een schuld vrije terugval bij twijfel beschermt het vertrouwen beter dan het verbergen van fouten.

2.4 Afvalscheidingsgedrag (cluster B)

De belangrijkste belemmering voor correct scheiden is een praktisch kennisgebrek en geen gebrek aan motivatie, waardoor informeren op het moment van handelen het meest oplevert (Barr, 2007; Kollmuss & Agyeman, 2002). Directe feedback op het beslismoment werkt het

best, en het artefact is daarmee vooral een hulpmiddel dat kennis biedt op het punt van actie, geen motivatie-instrument.

2.5 Technische realisatie (cluster C)

Dit cluster was verkennend. De literatuur wees eerst op een combinatie van twee herkenningmanieren: eerst een streepjescode opzoeken en, als dat niet lukt, het beeld herkennen met beeldherkenning (computer vision) als terugval, gedragen door een licht, voorgetraind model. Deze richting is tijdens de bouw onderbouwd bijgesteld (zie H4.3).

2.6 Deelconclusie onderzoek

De drie clusters komen samen in één ontwerpkader. De uitleg-laag wordt ontworpen voor nul-voorkennis en communiceert in lagen, waarbij falen zichtbaar wordt gemaakt in plaats van verborgen. Het artefact biedt vooral kennis op het punt van actie en is geen motivatie-instrument.

De uitleg-laag vervult daarbij **een dubbele functie**:

1. Het aanwijzen van de juiste bak (functionele feedback)
2. Het zichtbaar maken van de AI-werking (educatieve feedback)

Daarmee is de richting van het ontwerp onderbouwd: het artefact bundelt beide functies in één handeling, op het moment dat de bezoeker een voorwerp aanbiedt. De empirische bevestiging hiervan volgt uit de enquête (H4.1).

3 Methode

Het project is een ontwerpgericht onderzoek en volgt het drie-cyclusmodel van Design Science Research (Hevner, 2007; Hevner et al., 2004). Dit model past omdat het draait om het maken en beproeven van een artefact om een probleem in een echte omgeving op te lossen, en omdat de betrouwbaarheid ervan ontstaat uit het samenspel van drie cycli in plaats van uit één methode:

- De Relevance-cyclus haalt de eisen op uit de omgeving,
- De Rigor-cyclus verankert het ontwerp in wetenschappelijke kennis,
- De Design-cyclus ontwerpt, bouwt en evalueert stap voor stap (Peffers et al., 2007; Wieringa, 2014).

Figuur 1 (volgende pagina) vat de drie cycli samen. De volledige, uitgebreide versie staat in Figuur 6 in de bijlage. De resultaten in H4 volgen dezelfde drie cycli, zodat methode en resultaat één op één op elkaar aansluiten.



Figuur 1: Het drie-cyclusmodel van Design Science Research, toegepast op dit project.

3.1 Onderzoeksstrategie en samenhang

De onderzoeksstrategie vult dit raamwerk in: per cyclus zijn methoden gekozen die op elkaar voortbouwen. De interviews en de enquête leveren de eisen uit de omgeving in de Relevance-cyclus, het literatuuronderzoek en de expertinbreng vormen de kennisbasis in de Rigor-cyclus, en het bouwen en testen van het artefact vormt de ontwerp- en evaluatielus in de Design-cyclus.

De eisen zijn via een drie-bronnenmodel vertaald naar toetsbare requirements:

De user stories leveren de wens, de projectkaders de grens en het literatuuronderzoek de onderbouwing. De methodiek hierachter staat in de **Requirementsanalyse** en de methode-traceerbaarheid in de Methodologische Verantwoording.

Figuur 2 geeft de samenhang tussen de deelonderzoeken en het drie-bronnenmodel weer.



Figuur 2: Samenhang van de deelonderzoeken en het drie-bronnenmodel voor requirements.

3.2 Relevance-cyclus: interviews en enquête

Er zijn vier interviews gehouden, met een bewuste rolsplitsing om afhankelijkheid van één bron te beperken. De opdrachtgever, Koen Steeghs, is in de Relevance-cyclus benaderd als opdrachtgever (interviews 1 en 4) en in de Rigor-cyclus als domeinexpert (interview 3). De onafhankelijke expert Luuk Pernot bracht in het tweede interview een second opinion over de methodiek in. De interviews waren semi-gestructureerd (Brinkmann & Kvale, 2018) en per interview vastgelegd in transcript, notulen en verslag (Interviewverslagen 1-4, Bijlage 5).

De enquête is in het **Enquête-ontwerp** (Bijlage 6) opgezet in zes stappen op basis van de CBS-methodiek voor vragenlijstontwikkeling (CBS, z.d.) en getoetst op het responsproces (Tourangeau et al., 2000).

De vragenlijst telt dertien vragen in vier blokken:

Demografie, afvalscheiding, AI-kennis en houding.

De vragen zijn vooral kwantitatief, met een paar optionele open vragen als aanvulling. De doelgroep bestond uit bezoekers binnen het profiel Jan Modaal, met een streefaantal van twintig tot veertig respondenten (Baarda, 2014).

De analyse, uitgewerkt in de Enquête-analyse (Bijlage 7), volgt het meetniveau van de variabelen (Stevens, 1946): Likert-items zijn ordinaal behandeld met mediaan, modus en verdelingsvrije toetsen (de Winter & Dodou, 2010), terwijl de AI-herkenningsvraag als

telscore een t-toets toelaat. Voor proporties is een Wilson-betrouwbaarheidsinterval gebruikt (Brown et al., 2001) en voor samenhangen de Spearman-correlatie (de Winter et al., 2016), met nadruk op effectgrootte en betrouwbaarheidsinterval en niet op de p-waarde alleen (Cumming, 2014).

Van de zesenvijftig ruwe reacties vielen er acht buiten de doelgroep (leeftijd buiten 17 tot 65 jaar of onvolledig) en één wegens tegenstrijdige antwoorden, zodat **N = 47** binnen de doelgroep overblijft.

3.3 Rigor-cyclus: literatuuronderzoek en experts

De kennisbasis is opgebouwd met een vergelijkend bronnenonderzoek, waarin literatuurstudie en documentanalyse zijn gecombineerd om de clusters systematisch te beantwoorden. Per deelvraag is een vaste cyclus doorlopen: criteria afleiden, bronnen zoeken, bronnen wegen, bronnen vergelijken en een deelconclusie trekken.

Bronnen-triangulatie is bereikt door **vier categorieën te combineren**:

1. Academische (witte) literatuur
2. Grijze beleidsliteratuur
3. Statistische bronnen
4. Praktijkbronnen

Elke bron is beoordeeld met een zelfontwikkeld beoordelingsinstrument, de bronwegingsmatrix, om de geldigheid en de herhaalbaarheid te borgen. Aanvullend is expert-triangulatie ingezet, waarbij het opdrachtgever-perspectief is gecombineerd met een onafhankelijke expert om de bevindingen te versterken (Lincoln & Guba, 1985; Yin, 2018).

De volledige uitwerking staat in het Literatuuronderzoek (Bijlage 4).

3.4 Design-cyclus: bouw, A/B-test en kwaliteitsborging

De Design-cyclus omvat het stap voor stap bouwen en evalueren van het artefact.

Binnen deze cyclus is een A/B-test uitgevoerd naar één afgebakende aanpassing:

De aanwezigheid van een richtingspijl naar de juiste bak op het resultaatscherm. Variant A toont een neutrale wegwijzerstrook, variant B laat de doelbak oplichten met een pijl.

De test is within-subjects opgezet, wat betekent dat elke deelnemer beide varianten doorloopt, met counterbalancing (een wisselende volgorde) om volgorde-effecten te spreiden.

Gemeten zijn de **tijd tot weggooien** en de **succesratio**, aangevuld met een belevingsvraag. De opzet staat in het A/B-ontwerp en de uitvoering in de A/B-analyse (Bijlage 8). Het kader is evaluatie in Design Science Research (Venable et al., 2016) en de hypothese is onderbouwd vanuit aandachtssturing en wegwijzing (wayfinding) (Posner, 1980; Norman, 2013) en feedback op het beslismoment (Cakanlar et al., 2025).

De uitvoering week bewust af van de eerste, vooraf geplande opzet: door de within-subjects-keuze is een gepaarde t-toets op de tijd gebruikt en de McNemar-toets op de gepaarde succesuitkomst, een navolgbare bijstelling binnen de iteratieve DSR-aanpak.

Validiteit, betrouwbaarheid, reproduceerbaarheid en herleidbaarheid zijn projectbreed geborgd via triangulatie over methoden, bronnen en experts, en via een traceerbare keten van wens naar requirement naar test (Maxwell, 2013).

De ethische randvoorwaarden staan in de Ethische Afwegingen (Bijlage 9) langs vier lijnen: privacy, bias, paternalisme en toegankelijkheid, met uitlegbaarheid als overkoepelende eis (Floridi et al., 2018), juridisch verankerd in de AVG (Europees Parlement en Raad, 2016) en de AI-verordening (Europees Parlement en Raad, 2024). De vooraf geplande usability-meting met de System Usability Scale is binnen de projectperiode niet uitgevoerd en is daarmee een bewuste beperking.

4 Resultaten

Dit hoofdstuk presenteert de resultaten langs **dezelfde drie cycli als de methode**: **Eerst** de eisen uit de **Relevance-cyclus**, **dan** de grounding uit de **Rigor-cyclus**, en **ten slotte** het gebouwde en geëvalueerde artefact uit de **Design-cyclus**.

De onderbouwing per bevinding staat in het theoretisch kader en in het Literatuuronderzoek (Bijlage 4). Hier worden de kernpunten beknopt herhaald als argument voor de gemaakte ontwerpkeuzes.

4.1 Relevance-cyclus: de eisen vanuit de omgeving

4.1.1 Opdrachtgevervalidatie via interviews

De opdrachtgever bevestigde in het eerste interview het artefactconcept en de twee-laags probleemstelling en moedigde een iteratieve aanpak aan. In het vierde interview zijn de requirements op set-niveau en de wireframes formeel gevalideerd, en werd de uitlegvorm met kwalitatieve zekerheid zonder publiek percentage gewaardeerd. Het gesprek gaf een formele go plus drie nieuwe wensen: een aanmoedigingsprikkels, een terugkoppeling van de milieu-impact en een instelbaar kader. De gespreksverslagen staan in de Interviewverslagen 1-4 (Bijlage 5).

4.1.2 Kwantitatieve enquêteresultaten

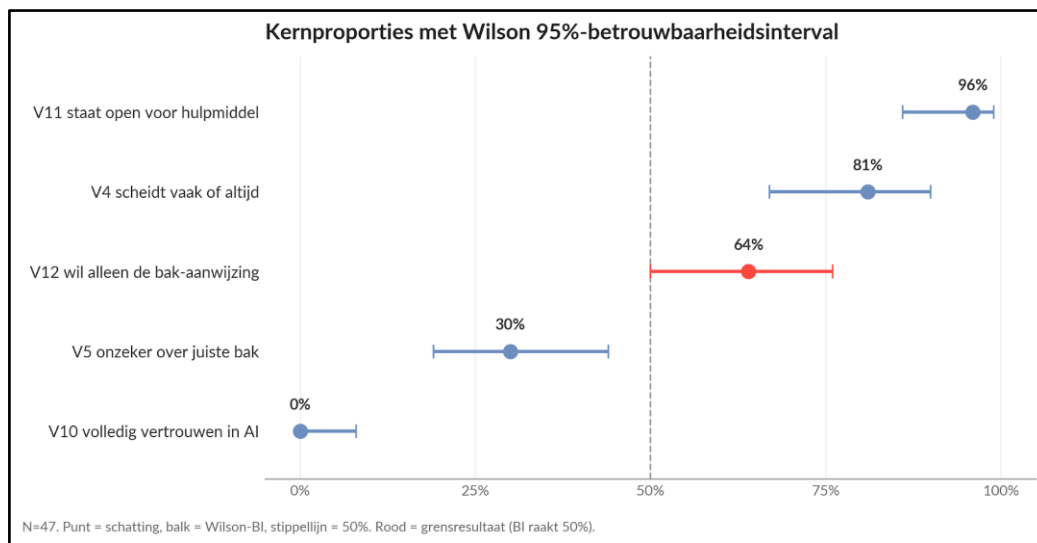
De enquête (N = 47) bevestigt het ontwerp kader. De details per deelvraag staan in Tabel 1 en Figuur 3. De doelgroep is bovengemiddeld AI-bewust (telscore 5,36 van 7, $t(46) = 10,43$, $p < 0,001$, $d = 1,52$), maar herkent verborgen, alledaagse AI vaak niet (aanbevelingssysteem 51%, gezichtsontgrendeling 38%).

Afvalscheiding is een ingesleten gewoonte (81% scheidt vaak of altijd), maar bijna een derde (30%) is onzeker over de juiste bak: een zekerheidskloof, geen motivatieprobleem (frequentie en zekerheid hangen samen, $\rho = 0,50$). De openheid voor een digitaal hulpmiddel is breed

(96%), terwijl niemand het AI-advies volledig vertrouwt (0%) en privacy (57%) een harde randvoorwaarde is. De bevindingen zijn indicatief vanwege de niet-aselecte steekproef.

Tabel 1: Kernresultaten van de enquête per deelvraag (N = 47).

Deelvraag	Kernresultaat	Toets / maat
DV1 AI-geletterdheid	telscore 5,36 > schaal midden 3,5; blinde vlek verborgen AI (51% / 38%)	$t(46) = 10,43, p < 0,001, d = 1,52$
DV2 Afvalscheiding en zekerheid	81% scheidt vaak of altijd, 30% onzeker over de bak	Wilson 67 tot 90%; rho 0,50
DV3 Vertrouwen en openheid	96% open, 0% volledig vertrouwen in AI-advies	Wilson 86 tot 99% / 0 tot 8%
DV4 Gewenste informatie	64% wil enkel de juiste bak; privacy 57% randvoorwaarde	Wilson 50 tot 76% (grensresultaat)



Figuur 3: Kernproporties uit de enquête met Wilson-betrouwbaarheidsintervallen (N = 47).

4.1.3 Van bevindingen naar requirements

De bevindingen zijn vertaald naar zes oplossingsdoelen (OD1 tot en met OD6). Deze doelen zijn via het drie-bronnenmodel uitgewerkt tot de **zeventig requirements en zevenentwintig user stories** in de User Stories & Requirements.

De must-have-requirements bepalen rechtstreeks de inhoud van het artefact:

- **Smart-bin-kiosk** met één scan-handeling: de bezoeker houdt een voorwerp voor de camera en krijgt direct advies over de juiste van zeven Nederlandse afvalstromen (uitgewerkt in het Functioneel Ontwerp en het Technisch Ontwerp).
- **Twee herkenningroutes:** een opzoekroute via een streepjescode en een AI-inschatting op basis van het camerabeeld.
- **Uitleg én zekerheid** als kern van de output, in plaats van enkel een classificatie.
- **Privacy by design** als harde randvoorwaarde (volgend uit de 57%-privacyeis): lokale beeldverwerking in een afgebakende scanzone, zonder opslag van beelden.
- **Toegankelijkheid:** jargonvrije uitleg op B1-niveau en tweetaligheid (Nederlands en Engels).

De zekerheidskloof uit de enquête koppelt direct aan het derde oplossingsdoel: ondersteuning op het beslismoment in plaats van meer aansporing.

4.2 Rigor-cyclus: de wetenschappelijke grounding

4.2.1 Literatuuronderzoek

De kernuitkomsten voor cluster A en B staan in H2.3 en H2.4: een gelaagde uitleg zonder percentages, kennis op het moment van handelen, en een schuldvrije terugval bij twijfel. De eerste technische richting, een combinatie van streepjescode en beeldherkenning, is tijdens de bouw bijgesteld (zie het artefact in H4.3).

4.2.2 Expertvalidatie

De expertinbreng versterkte de methodologische degelijkheid. De onafhankelijke expert adviseerde de doelgroep pragmatisch te begrenzen en het ontwerp en de enquête te richten op de minst AI-geletterde groep, omdat een te brede doelgroep ruis oplevert. De opdrachtgever in zijn expertrol (het derde interview) drong aan op één centrale ontwerpvrage in plaats van meerdere hoofdvragen, wat tot een herstructurering van het literatuuronderzoek leidde. Dezelfde expert nuanceerde de verhouding tussen kennis en motivatie, wees op vertekening bij zelfrapportage en op het risico van een niet-homogene dataset. De volledige gespreksverslagen staan in Bijlage 5.

4.2.3 Technische grounding en benchmark

De technische keuzes zijn met een eigen benchmark onderbouwd, met de prestatie afgezet tegen de vooraf gestelde eisen (Tabel 2). Op een testset van ruim duizend beelden behaalde een RF-DETR-detector een duidelijk hogere macro-F1 op bakniveau dan een YOLO11-model (0,705 tegenover 0,455). Een vision-taalmodel bleek een sterke maar trage classifier, met een hoge nauwkeurigheid maar een verwerkingstijd van ongeveer twaalf seconden per beeld, en is daarom alleen op het scanmoment ingezet.

De zekerheid van het model is gekalibreerd, waardoor de verwachte calibratiefout daalde van 0,228 naar 0,023. Dit zijn referentiewaarden uit de benchmark, geen geverifieerd eindstelsysteemresultaat. De eindstelsysteemverificatie staat in het Testplan en de architectuur en benchmark in het Technisch Ontwerp.

Tabel 2: Modelprestatie ten opzichte van de eisen (benchmark, referentiewaarden).

Maatstaf	Eis	Gemeten (referentie)
End-to-end macro-F1 (bak)	$\geq 0,45$	0,705
Detectie-recall (IoU $\geq 0,5$)	$\geq 0,55$	0,806
Bak-keuze-nauwkeurigheid	$\geq 0,85$	0,964

4.3 Design-cyclus: het artefact, gebouwd en geëvalueerd

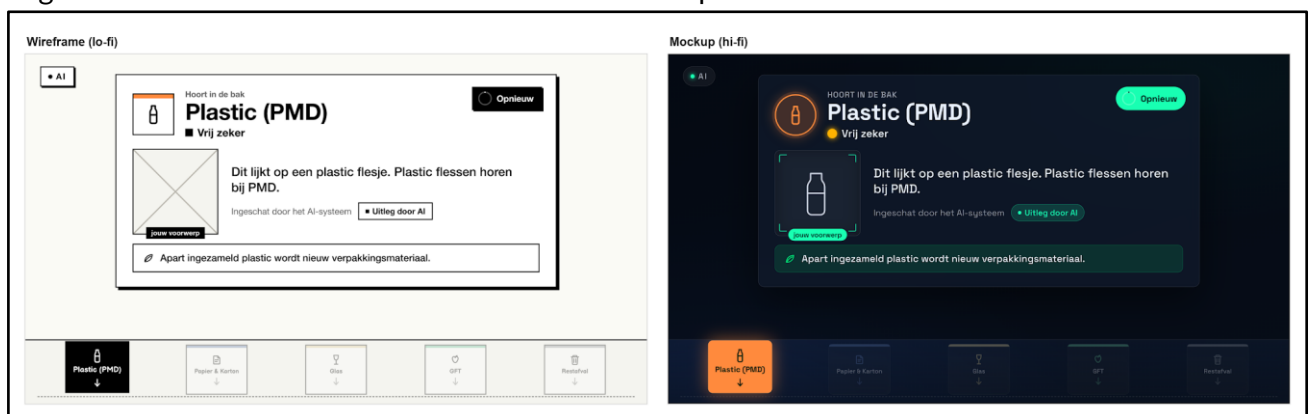
4.3.1 Het artefact zoals gebouwd

Het artefact is een smart-bin-kiosk die een voorwerp herkent en uitlegt in welke bak het hoort en waarom. De uitleg verschijnt in gewone taal met een route-stoplicht: een streepjescode-opzoeking levert een gezaghebbend groen resultaat, een AI-inschatting komt maximaal op oranje met de aanduiding dat het om uitleg door AI gaat, en onder de drempel dringt het systeem geen advies op maar laat het de bezoeker schuldvrij zelf kiezen. De AI wordt dus nooit als zekerheid gepresenteerd. De verwerking gebeurt lokaal, beeld verlaat de machine niet en er is geen beeldopslag, in lijn met de privacy-eisen.

Een publieksmodus toont kwalitatieve, jargonvrije zekerheid, en een conferentiemodus toont voor een vakpubliek ook technische informatie. De architectuur is Gemma-first: de eerste, op literatuur gebaseerde keuze voor een combinatie van streepjescode en beeldherkenning stuitte bij de bouw op onbekende voorwerpen en ontbrekende klassen, waarna een vision-taalmodel de classificatie overnam met de objectdetector als aandachtfunctie. Deze bijstelling is vastgelegd in het Beslisdocument Gemma-pivot. De architectuur en de twee herkenningroutes staan in het Technisch Ontwerp, de schermen en de scan-flow in het Functioneel Ontwerp, en de scope in de Demo Definitie.

Een eerlijke weergave hoort hierbij: de volledige eindversie staat op een aparte ontwikkelbranch, en de demonstratie-build toont vijf van de zeven bakken, met de tijdgestuurde scan-trigger en het sociaal-normatieve signaal nog niet geactiveerd. Deze keuzes vallen binnen de prototype-scope.

Figuur 4 toont het resultaatsscherm met het route-stoplicht.



Figuur 4: Resultaatscherm van het artefact met het route-stoplicht.

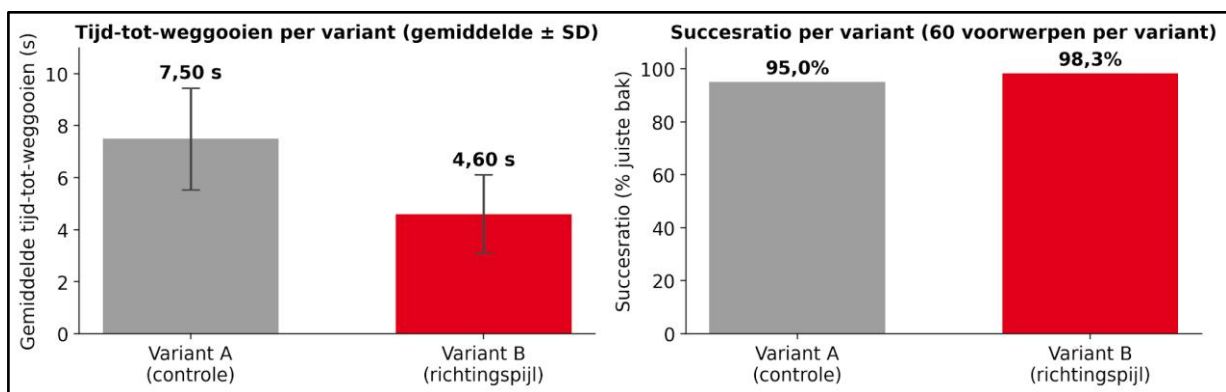
4.3.2 Resultaten van de A/B-test

De A/B-test (N = 10 deelnemers, zes voorwerpen per variant, samen 120 gepaarde waarnemingen) laat een duidelijk effect op de snelheid zien. Met de richtingspijl maakten deelnemers gemiddeld sneller een keuze: 4,60 seconden tegenover 7,50 seconden zonder pijl, een gepaard verschil van 2,90 seconden, $t(9) = 3,80$, $p = 0,004$, $dz = 1,20$, met een betrouwbaarheidsinterval van 1,17 tot 4,63 seconden. De succesratio was in beide varianten al hoog (57 van de 60 tegenover 59 van de 60) en het verschil is met de McNemar-toets niet

significant ($p = 0,500$), wat past bij een plafondeffect. Gezien de kleine deelnemersgroep ligt de nadruk op de effectgrootte en het interval, niet op de p-waarde alleen. Negen van de tien deelnemers gaven de voorkeur aan de variant met de pijl. Tabel 3 vat de uitkomsten samen en Figuur 5 toont de gemiddelde tijd en de succesratio per variant.

Tabel 3: Resultaten van de A/B-test op de richtingspijl (N = 10).

Maat	Variant A (neutraal)	Variant B (pijl)	Toets
Tijd tot weggooien (s)	7,50	4,60	$t(9) = 3,80$, $p = 0,004$, $dz = 1,20$
Succesratio	57/60	59/60	McNemar $p = 0,500$



Figuur 5: Gemiddelde tijd en succesratio per variant.

4.3.3 Triangulatie en de rode draad

Over de methoden heen ontstaat een consistent beeld. De kennislacune uit de literatuur, de zekerheidskloof uit de enquête en de validatie door de opdrachtgever wijzen samen naar dezelfde ontwerprichting: kennis op het moment van handelen, een uitleg zonder percentages en een schuld vrije terugval bij twijfel. Waar de methoden elkaar nuanceerden is dat zichtbaar gemaakt, en de enige afwijkende lijn, de technische richting, is bij de bouw bijgesteld. De richtingspijl voegt aan deze kern een gevalideerde versnelling op het beslismoment toe.

5 Discussie

Deze discussie weegt de waarde en de grenzen van het onderzoek per thema, en vertaalt die naar aanbevelingen voor vervolgonderzoek.

5.1 De rode draad en de gedocumenteerde koerswijziging

De kracht van het onderzoek zit in de wisselwerking tussen de cycli en in de open omgang met bijstelling. De koerswijziging naar een Gemma-first-architectuur (H4.3) is niet weggewerkt maar gedocumenteerd, en de conclusies over AI-geletterdheid, vertrouwen en afvalscheidingsgedrag bleven daarbij overeind. Alleen de technische realisatie en één detail van de uitlegvorm zijn bijgesteld.

5.2 Reikwijdte van de enquête

De enquête gebruikte een gelegenheidssteekproef en is niet aselekt, met een oververtegenwoordiging van jongere en hoogopgeleide respondenten, waardoor de resultaten indicatief zijn. Zelfrapportage kan het scheidingsgedrag hebben overschat, en het resultaat over de gewenste informatie (64%) is een grensresultaat. De robuuste bevindingen, zoals de sterke telscore ($d = 1,52$) en de zekerheidskloof, dragen de conclusies wel, maar generalisatie naar heel Jan Modaal vraagt een grotere en meer representatieve steekproef.

5.3 Kracht en grens van de A/B-test

Het snelheidseffect van de richtingspijl is robuust: het berust op 120 gepaarde waarnemingen, met een grote effectgrootte en een betrouwbaarheidsinterval dat de nul niet raakt. Tegelijk telde de test tien deelnemers, precies op de ondergrens van de vooraf afgesproken terugvalregel, waardoor de conclusie op de deelnemersgroep balanceert op die grens, en de tijd- en succesmeting berustte op handmatige observatie door de testleider. De test liet ook een grens van de interventie zelf zien: bij een meerdelig voorwerp kan één richtingspijl niet naar meerdere bakken tegelijk wijzen, waardoor in beide varianten enige twijfel bleef bestaan.

5.4 Niet uitgevoerde metingen

Twee vooraf overwogen onderdelen zijn binnen de projectperiode bewust niet uitgevoerd: de usability-meting met de System Usability Scale en het sociaal-normatieve signaal uit enkele requirements. Beide zijn geen blinde vlekken maar bekende, gedocumenteerde keuzes binnen de prototype-scope, die de reikwijdte van de conclusies begrenzen.

5.5 Aanbevelingen voor vervolgonderzoek

Vervolgonderzoek zou het artefact in een veldopstelling moeten toetsen met een grotere en meer aselechte steekproef, zodat de bevindingen uit het laboratorium naar de publieke ruimte kunnen worden vertaald. Daarnaast is het zinvol de usability-meting alsnog uit te voeren en de A/B-steekproef te vergroten, zodat ook de succesratio met voldoende onderscheidend vermogen kan worden getoetst. Meerdere voorwerpen verdienen een aparte test, omdat een enkele richtingspijl niet volstaat wanneer een voorwerp in onderdelen moet worden

gesorteerd. Tot slot zou een regionale dataset met de verwarpunten uit de praktijk de herkenning verder verbeteren.

6 Conclusie

Een interactief smart-bin-artefact kan de werking van AI begrijpelijk maken en correct afvalscheidingsgedrag ondersteunen wanneer het de uitleg op het moment van handelen levert, in gewone taal en zonder publieke percentages, en wanneer het bij twijfel eerlijk is over zijn onzekerheid in plaats van een advies op te dringen. Het onderzoek laat zien dat de doelgroep al gemotiveerd maar onzeker is, dat een leek baat heeft bij een gelaagde uitleg zonder cijfers, en dat een eenvoudige ruimtelijke aanwijzing zoals een richtingspijl het handelen meetbaar versnelt.

De belangrijkste ontwerpkeuzes volgen navolgbaar uit die bevindingen. De uitleg in lagen zonder percentages volgt uit het AI-geletterdheidsonderzoek en de voorkeur voor eenvoud in de enquête, de kennis op het beslismoment uit de zekerheidskloof, en de schuldvrije terugval waarin de AI nooit groen toont uit het vertrouwensonderzoek. De zichtbare AI-melding volgt uit de blinde vlek voor verborgen AI, de lokale en privacy-vriendelijke opstelling uit de privacy-randvoorwaarde, en de richtingspijl uit het wayfinding-onderzoek, bevestigd met de A/B-test.

Het advies aan de opdrachtgever is de richtingspijl op te nemen in het artefact, omdat de A/B-test een meetbare versnelling op het beslismoment liet zien ($t(9) = 3,80$, $p = 0,004$) en **negen van de tien** deelnemers er de voorkeur aan gaven.

Aanvullend lonen de drie wensen uit het vierde interview: een instelbare aanmoedigingsprikkels, een terugkoppeling van de milieu-impact en een instelbaar kader om het herkende voorwerp. Een geplande vijfde sessie, een walkthrough met de opdrachtgever, kan het prototype formeel laten accepteren. De DSR-aanpak zorgde ervoor dat het artefact niet op intuïtie maar op bevindingen rust, en dat een onderbouwde koerswijziging tijdens de bouw een gedocumenteerde stap werd in plaats van een breuk. Daarmee voldoet de oplossing aan de doelstelling: een werkend prototype, gedragen door onderbouwd ontwerpgericht onderzoek.

Literatuurlijst

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160.
<https://doi.org/10.1109/ACCESS.2018.2870052>
- Baarda, B. (2014). *Basisboek enquêteren* (3e druk). Noordhoff.
- Barr, S. (2007). Factors influencing environmental attitudes and behaviors: A UK case study of household waste management. *Environment and Behavior*, 39(4), 435-473.
- Brinkmann, S., & Kvale, S. (2018). *Doing interviews* (2nd ed.). SAGE.
- Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16(2), 101-133.
- Cakanlar, A., Hunter, M., & Nenkov, G. Y. (2025). Recycle right: How to decrease recycling contamination with informational point-of-disposal signage. *Journal of the Academy of Marketing Science*, 53(3). <https://doi.org/10.1007/s11747-024-01058-1>
- Centraal Bureau voor de Statistiek. (z.d.). *Vragenlijstontwikkeling*. Geraadpleegd op 3 juni 2026,
van <https://www.cbs.nl/nl-nl/onze-diensten/methoden/statistische-methoden/input/input/vragenlijstontwikkeling>
- Centraal Bureau voor de Statistiek. (2025a, 30 oktober). *4 kilo afval meer per inwoner in 2024*. CBS.
- Centraal Bureau voor de Statistiek. (2025b, 14 juli). *Einde aan dalende trend: meer afval ingezameld in 2024*. CBS.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29.
- Europees Parlement en Raad. (2016). *Verordening (EU) 2016/679 (Algemene Verordening Gegevensbescherming)*. Publicatieblad van de Europese Unie.
- Europees Parlement en Raad. (2024). *Verordening (EU) 2024/1689 (Artificial Intelligence Act)*. Publicatieblad van de Europese Unie.
- Europese Commissie. (2019). *The European Green Deal* (COM(2019) 640 final).
<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2019%3A640%3AFIN>

- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People: An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Grit, R. (2019). *Projectmanagement* (8e druk). Noordhoff.
- Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems*, 19(2), 87-92.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75-105.
- Internationale Organisatie voor Standaardisatie. (2019). *ISO 9241-210:2019. Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems*. ISO.
- Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, 8(3), 239-260.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). ACM. <https://doi.org/10.1145/3313831.3376727>
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). SAGE.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & 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>
- Norman, D. A. (2013). *The design of everyday things* (Rev. ed.). Basic Books.
- Norman, D. A., & Draper, S. W. (1986). *User centered system design*. Lawrence Erlbaum.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45-77.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3-25.

- RWM N.V. (2024). *Jaarverslag RWM N.V. 2023: Van uitvoeringsorganisatie naar partner*.
<https://rwm.nl/>
- Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology*, 29(3), 309-317.
<https://doi.org/10.1016/j.jenvp.2008.10.004>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677-680.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.
- Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, 25(1), 77-89.
- Verian. (2025). *Special Eurobarometer 554: Artificial intelligence and the future of work* [Survey conducted for the European Commission].
<https://europa.eu/eurobarometer/surveys/detail/3222>
- Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. Springer.
- de Winter, J. C. F., & Dodou, D. (2010). Five-point Likert items: t test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research, and Evaluation*, 15, 11.
- de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes. *Psychological Methods*, 21(3), 273-290.
- Wouters, N., Downs, J., Harrop, M., Cox, T., Oliveira, E., Webber, S., Vetere, F., & Vande Moere, A. (2016). Uncovering the honeypot effect: How audiences engage with public interactive systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (pp. 5-16). ACM. <https://doi.org/10.1145/2901790.2901796>
- Yin, R. K. (2018). *Case study research and applications* (6th ed.). SAGE.

Documentrelaties

Dit rapport is de afronding van een reeks zelfstandige projectdocumenten. Tabel 4 toont de documenten waarop het Eindrapport voortbouwt en hun relatie tot dit verslag.

Tabel 4: Documentrelaties van het Eindrapport.

Document	Relatie tot het Eindrapport
Quickscan	Eerste verkenning en vroege ontwerp-kiemen (input)
Plan van Aanpak	Projectkaders, doelstelling en planning (input)
Methodologische Verantwoording	DSR-methodologie en kwaliteitsborging (onderbouwing)
Requirementsanalyse	Methodiek van de requirements via het drie-bronnenmodel (onderbouwing)
User Stories & Requirements	De 27 user stories en 70 requirements (input)
Beslisdocument Gemma-pivot	Onderbouwing van de technische koerswijziging (onderbouwing)
Demo Definitie	Scope en demonstratieopzet van het artefact (complementair)
Technisch Ontwerp	Architectuur, herkenningroutes en benchmark (complementair)
Functioneel Ontwerp	Schermen, scan-flow en resultaatscherm (complementair)
A/B-ontwerp	Opzet en meetplan van de A/B-test (input)
Testplan	Teststrategie en eindsysteemverificatie van het artefact (complementair)

Bijlagen

De onderbouwende projectdeliverables zijn als genummerde bijlagen bij dit rapport meegeleverd. Tabel 5 vormt het bijlagenregister. Elke bijlage wordt in de hoofdtekst ten minste eenmaal aangehaald, in de volgorde van eerste vermelding.

Tabel 5: Bijlagenregister.

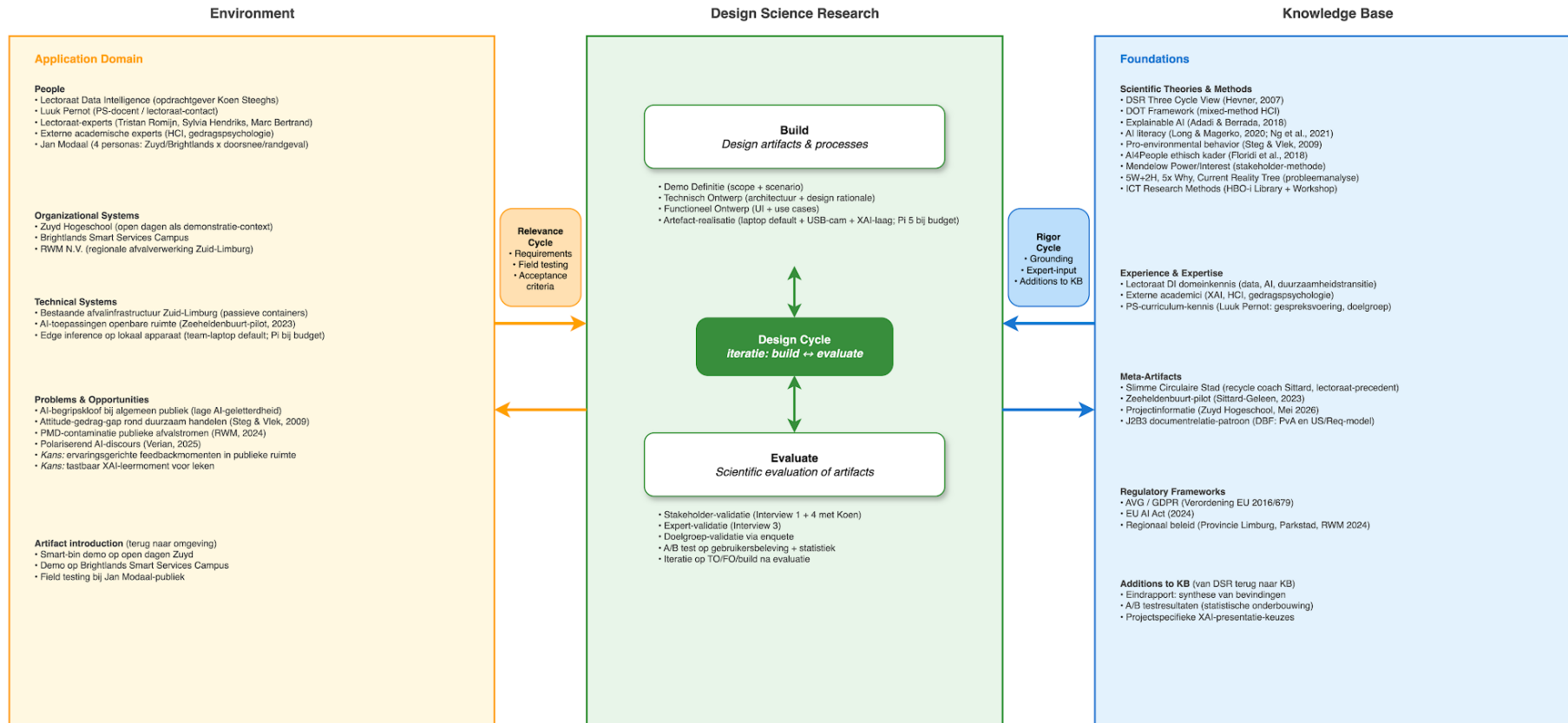
Nummer	Bijlage
Bijlage 1	Stakeholderanalyse
Bijlage 2	Persona's
Bijlage 3	Probleemanalyse
Bijlage 4	Literatuuronderzoek
Bijlage 5	Interviewverslagen 1-4
Bijlage 6	Enquête-ontwerp
Bijlage 7	Enquête-analyse
Bijlage 8	A/B-analyse
Bijlage 9	Ethische Afwegingen

Uitgebreid drie-cyclusmodel

Figuur 6: Het volledige drie-cyclusmodel van Design Science Research (naar Hevner, 2007)

Design Science Research: Hevner's Three Cycle View toegepast op dit project

Project Duurzame Technologie in Beeld, J2B4 MOPT-2026 Optimize, team Optimizers



Methodologische context. Hevner's Three Cycle View (Hevner, 2007) is het methodologisch raamwerk voor het ontwerpgericht onderzoek in dit project. De drie cycli (Relevance, Rigor, Design) draaien *continu en parallel*, niet sequentieel: gedurende het hele project worden requirements opgehaald uit de Environment (Relevance), wordt nieuwe kennis verworven uit de Knowledge Base (Rigor), en wordt het artefact iteratief gebouwd en geëvalueerd (Design).

Complementaire visualisatie. Voor de chronologische deliverable-flow (welke documenten in welke fase, met de planned-flow-pijlen en 3 specifieke DSR-iteratie-momenten), zie *documentrelatie – diagram – j2b4.drawio*. Beide diagrammen samen geven het volledige beeld: dit diagram = methodologie, het andere = project-planning.

Bron: Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19(2), 87-92.
Figuur aangepast aan projectcontext voor methode-hoofdstuk Eindrapport (opdracht 8) van J2B4 MOPT-2026 Optimize.