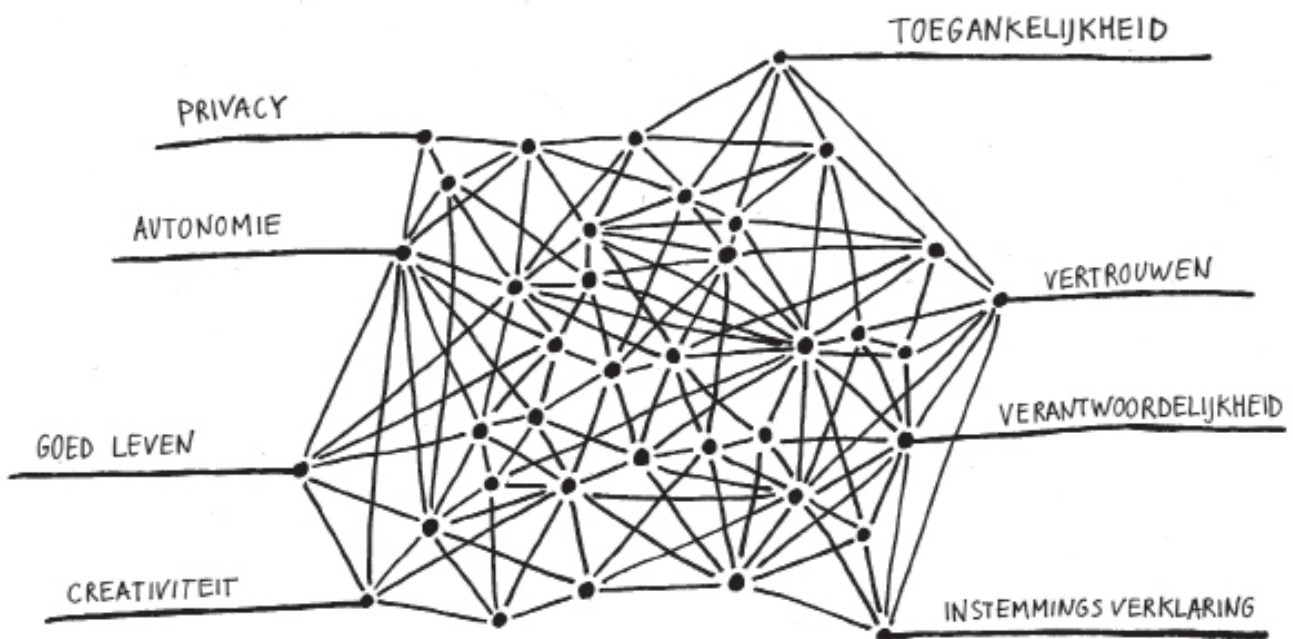


D e
E t h i s c h e
D a t a
A s s i s t e n t



Handleiding

Inventarisatie van **ethische kwesties** rond
dataproyecten bij overheden

Disclaimer

'De Ethische Data Assistent' is ontwikkeld door Data School, Universiteit Utrecht.

© Data School, Utrecht University
2026

DEDA - Versie 4.0
februari 2026
Voor updates kijk op:

deda.dataschool.nl

Noch de Universiteit Utrecht, noch Data School zijn aansprakelijk voor (mogelijke) schade in verband met, of voortkomend uit, het gebruik van materiaal of inhoud van de 'De Ethische Data Assistent'.

Utrecht, 2026.

Terwijl DEDA wordt gebruikt, leren we om het te verbeteren. Deze verbeteringen zullen in toekomstige versies worden geïmplementeerd. Wanneer u DEDA gebruikt en op- of aanmerkingen heeft, aarzel dan niet om deze met ons te delen. U kunt daarvoor altijd mailen naar:

dataschool@uu.nl

Concept :
Mirko Tobias Schäfer, Ph.D.
Aline Franzke, MA

Redactie:
Julia Straatman, MA
Iris Muis, MA
Hajo Bijleveld, BA
Ydo Bergstra, BSc
Lysa Ngouateu, BSc
Mirko Tobias Schäfer, Ph.D.
Bart Anthony Kamphorst, Ph.D.

Illustratie voorblad:
Sammy Hemerik

© Data School



**Universiteit
Utrecht**

DATA SCHOOL

Inhoudsopgave

Introductie

6 - 9

Vragen

10 - 31

TECHNISCHE OVERWEGINGEN

10 - 17



MAATSCHAPPELIJKE OVERWEGINGEN

18 - 25



UITKOMST

26 - 31



DATABRONNEN	12
ALGORITMEN	13
ANONIMISEREN	15
VISUALISEREN	16
TOEGANG EN HERGEBRUIK	17
PRIVACY	20
VOORINGENOMENHEID	21
TRANSPARANTE COMMUNICATIE	23
ORGANISATIE	24
TOEKOMSTSCENARIO'S	25
WAARDEN	28
CONCLUSIE	29
VERVOLGSTAPPEN	30
VERDERE INFORMATIE	31

Introductie

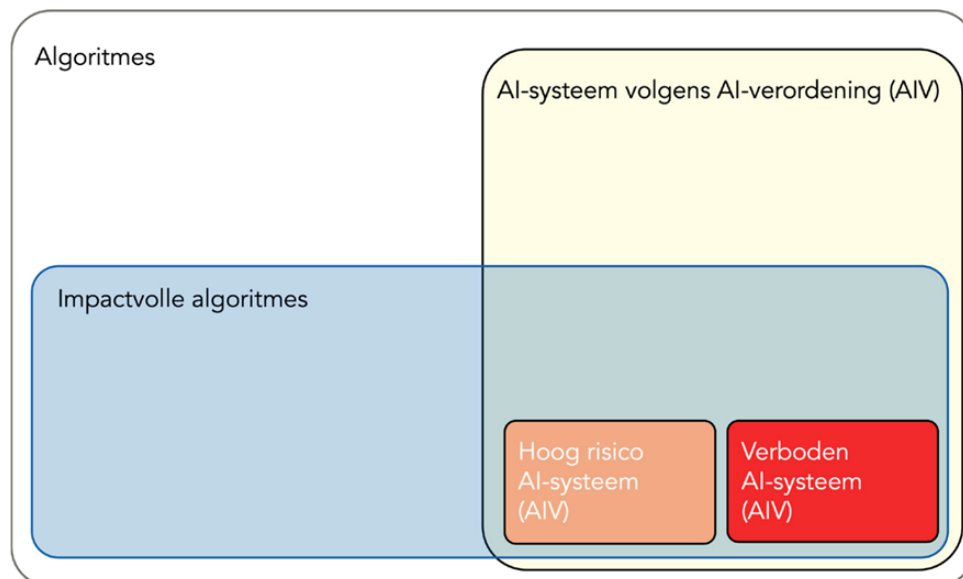
Waarom DEDA?

Data, algoritmen en AI beloven grote voordelen voor de publieke sector, van het faciliteren van ‘smart cities’, tot het vergroten van de economische welvaart. Echter zijn deze kansen niet zonder moeilijkheden. Deze moeilijkheden zijn makkelijk te negeren, maar kunnen er op de lange termijn voor zorgen dat goede bedoelingen tot slechte uitkomsten leiden. Zo kunnen er bijvoorbeeld problemen ontstaan rondom aansprakelijkheid wanneer er iets misgaat - wat kan leiden tot een rechtszaak - of kan een project in strijd zijn met de richtlijnen van goed bestuur.

Bedrijven en overheden liggen steeds vaker onder vuur vanwege de wijze waarop ze data en algoritmen gebruiken. Als reactie hierop zijn wetten en regulaties aangepast en wordt de inzet van AI door de Europese AI-verordening gereguleerd. Verzwaarde boetes voor het schenden van privacy zijn een voorbeeld van pogingen van de EU om een verantwoord gebruik van persoonlijke informatie af te dwingen. Daarnaast moeten de regulaties in de AI-verordening zorgen dat AI-systemen die in de EU worden gebruikt veilig,

transparant, traceerbaar, niet-discriminerend en milieuvriendelijk zijn. Ook algoritmen die niet door de AI-verordening als AI zijn geïdentificeerd kunnen op dezelfde manier impactvol zijn. Er zijn namelijk allerlei soorten problemen die voort kunnen komen uit dataprojecten. Daarom hanteert het Algoritmeregister van de Nederlandse Overheid de term “impactvolle algoritmen”. Hoe deze zich verhouden tot de in de AI-verordening opgenomen definitie van “high-risk AI” is te zien in afbeelding 1.

Zo kunnen datasets een twijfelachtige herkomst hebben, of uit hun context worden getrokken. Ook kan er sprake zijn van vooringenomenheid in de datasets, modellen en algoritmen. Daarnaast zijn er soms vragen omtrent (mogelijke) belangenverstrengeling van commerciële bedrijven en publieke instellingen. Het kan ook ontbreken aan een kritische evaluatie op de sociale impact van datagedreven beleid.



Afbeelding 1: 20250228 White paper Implementatie AI-verordening – definitie, Algorithm Audit (Feb 2025), geraadpleegd via https://algorithmaudit.eu/nl/knowledge-platform/knowledge-base/guidelines_ai_act_implementation/

De Ethische Data Assistent (DEDA) is een instrument dat helpt ethische knelpunten op te sporen en om een gevoel te ontwikkelen voor waardeconflicten binnen een dataproject of AI-project. Ook kan DEDA helpen inzicht te creëren in (publieke) waarden die worden geraakt, of om een ethisch besluitvormingsproces te documenteren. Het stelt organisaties in staat om te bepalen of een data- of AI-project bij de ambities en doelen van de organisatie en haar context past of deze juist ondermijnt. Ook levert het proces inzicht op de benodigde capaciteiten en randvoorwaarden om het project verantwoord te kunnen realiseren. DEDA helpt bij het verantwoordelijk om te gaan met data, modellen, algoritmen en meer.

Doel

De DEDA handleiding kan gebruikt worden samen met de DEDA poster. De handleiding heeft als doel meer uitleg te bieden bij bepaalde concepten en om wat dieper op sommige vragen in te gaan. Bij sommige vragen worden er ook voorbeelden geven.

Gebruik van de handleiding

1 Deze handleiding kunt u gebruiken als ondersteuning bij de DEDA poster. Op de poster zijn gekleurde vraagclusters te vinden. De vragen zijn procesmatig gestructureerd zodat ze achtereenvolgend technische overwegingen (blauw) en maatschappelijke overwegingen (groen) belichten. Voor elk vragencluster treft u in de handleiding uitleg aan. U kunt post-its gebruiken om het antwoord op de vragen op te schrijven, actiepunten op te schrijven en reacties te verzamelen op de poster.

2 In het onderdeel met technische vragen kunt u eventueel vragen overslaan die (nog) niet relevant zijn voor de fase waarin het project zich bevindt. Wel raden wij aan om dan tijdens het verloop van het project nog eens naar deze vragen te kijken.

De maatschappelijke overwegingen richten zich op vragen rondom privacy, communicatie, vooringenomenheid, transparante communicatie, organisatie en toekomsscenario's.

Opmerking: Wanneer u een vraag niet kunt beantwoorden omdat u hiervoor aanvullende details nodig heeft, is het aan te raden hier een actiepunt van te maken voor het beantwoorden van die vraag. Ook specifieke aandachtspunten bij die vraag kunt u noteren.

3 De laatste sectie op de DEDA poster vraagt extra aandacht voor waarden. Op basis van de waarden die belangrijk zijn binnen uw organisatie en voor de projectleden, kunt u beslissingen nemen over aandachtspunten van het project en mogelijke hindernissen die overkomen moeten worden. Voor dit doel biedt de conclusie enkele afsluitende vragen.

DEDA brengt alles rondom uw data- of AI-project globaal in kaart. Wanneer de conclusie uit het invullen van DEDA luidt dat er persoonsgegevens worden verwerkt, die tevens een hoog privacy risico voor de betrokken personen betreft, dan moet er een DPIA (Data Protection Impact Assessment) ingevuld worden. Het invullen van de DPIA is sinds 25 mei 2018 een verplichting op basis van de AVG (Algemene Verordening Gegevensbescherming). De AVG is van kracht in heel Europa en heeft directe werking in alle lidstaten. Meer informatie over de DPIA is te vinden bij de DEDA vragen omtrent Privacy (p. 20).

Voorbereiding DEDA sessie

Als organisator en/of moderator van een sessie met DEDA, zijn er een aantal zaken om van tevoren aan te denken. Allereerst is het van belang dat er een geschikte casus vanuit de eigen organisatie is. Dit kan van alles zijn, zolang het maar een dataproject betreft met mogelijke ethische valkuilen. Ten tweede is het cruciaal om een interdisciplinaire groep aan deelnemers uit te nodigen. Zo kan iedereen elkaars expertise aanvullen. Als houvast kan er rekening gehouden worden met de volgende factoren. Per DEDA poster kunnen vier tot acht mensen deelnemen. Trek drie tot vier uur uit voor de sessie en nodig deelnemers ruim van tevoren uit in een ruime vergaderzaal zodat er, indien nodig, groepjes gevormd kunnen worden. Het is aan te raden om de sessie te starten met een korte samenvatting van de casus zodat alle deelnemers goed op de hoogte zijn van de materie. Na afloop van de sessie kan de sessie plenair afgesloten worden door alle bevindingen te delen. De conclusie op de DEDA poster biedt als fundament voor verdere vervolgstappen van het project.

Vragen

TECHNISCHE OVERWEGINGEN



Uitleg

Het is belangrijk om na te gaan waar de data die u wilt gebruiken, vandaan komt. Wanneer datasets worden ingekocht, of het verzamelen van data wordt uitbesteed, is het soms moeilijk te begrijpen in welke context de data in eerste instantie zijn verzameld, wat de beperkingen van de dataset zijn en of deze data daadwerkelijk geschikt zijn voor het doel.

Soms is er sprake van een ‘hoe meer, hoe beter’ idee bij het verzamelen van data. Dit kan problematisch zijn wanneer bepaalde data niet relevant zijn voor de context van het project. Het is belangrijk om hier goed op te reflecteren. Een grotere dataset leidt niet altijd tot een betere kwaliteit van het onderzoek.

Daarnaast is de ‘houdbaarheidsdatum’ van data van belang. In de AVG staan voorschriften over de bewaartermijnen van data, maar het is ook noodzakelijk om bewust te zijn van de beperkte bruikbaarheid van sommige datasets. Hoelang blijven de data geschikt en relevant voor het beoogde project? Ethische reflecties over de oorsprong van de dataset moeten om deze reden beginnen met een reflectie van welke soort data er daadwerkelijk nodig zijn voor het project en hoelang deze data relevant blijven.

Tot slot is het van belang om te reflecteren op de vertegenwoordiging van verschillende groepen mensen in de dataset. Zo kan een oververtegenwoordiging of ondervertegenwoordiging van bepaalde sociale groepen leiden tot verschillen in uitkomsten voor deze groepen. Wanneer vooral één type data aanwezig is, leert het model niet goed voorspellen met andere soorten data, wat kan resulteren in vooringenomenheid. Een gebalanceerde dataset, met een representatieve verdeling in verschillende variabelen, zorgt ervoor dat het model zo nauwkeurig en eerlijk

mogelijke voorspellingen kan maken.

Daarnaast kan gereflecteerd worden op of de dataset bepaalde proxy-kenmerken bevat. Dat zijn indirecte kenmerken in de data die ogenschijnlijk neutraal lijken, maar toch indirect iets kunnen zeggen over bepaalde beschermde karakteristieken. Zo zou de ogenschijnlijk neutrale eigenschap ‘laaggeletterdheid’ als proxy kunnen dienen voor de beschermde eigenschap ‘afkomst’. Daarom is het noodzakelijk om goed op de hoogte te zijn van wat de dataset inhoudt, en wat dit ons vertelt.

Wanneer zowel de sociale groepen als de uitkomst labels (bijvoorbeeld laag of hoog risico) bekend zijn, kan vervolgens rekenkundig worden getoetst of er sprake is van vooringenomenheid. Dit kan met zogenoemde bias- en fairness-metingen, die laten zien of het model bepaalde groepen systematisch benadeelt.

Let op:

Onder het kopje Vooringenomenheid wordt later dieper ingegaan op mogelijke vooringenomenheid in het project. Als er een discussie ontstaat die te veel tijd inneemt en verder gaat dan vooringenomenheid in de data, stop de discussie dan hier en informeer de groep dat dit onderwerp later in meer detail zal worden behandeld.

Uitleg

Algoritmen kunnen data verwerken en inzichten op basis van die data ontwikkelen. Een voorbeeld hiervan is een algoritme dat bepaalt wanneer een stoplicht op groen moet springen of een algoritme dat inschat wie er wel of niet voor een uitkering in aanmerking komt. Algoritmen maken gebruik van wiskundige modellen. Deze modellen zijn echter maar zelden waarde vrij. Modellen worden numeriek uitgedrukt, en berekeningen en lijken daarom vaak neutraal. Echter zijn algoritmen vaak zo gemaakt dat er aan een numerieke waarde een normatief oordeel wordt gekoppeld, zoals 'risico op fraude' en 'ongewone gebeurtenis'. Het is van belang goed na te denken over waarom en hoe een type algoritme ingezet gaat worden, en wat voor invloed het heeft op deze normatieve oordelen.

Verdieping vraag 13 en 14: soorten algoritmen

Er zijn veel verschillende soorten algoritmen. Zo zijn er algoritmen die heel simpel en vaak goed uitlegbaar zijn, maar zijn er ook veel complexere en moeilijker uitlegbare algoritmes die wel nauwkeurigere resultaten bieden. Een simpel algoritme zou bijvoorbeeld een beslisboom kunnen zijn waar beleidsregels in zijn geprogrammeerd. Die regels zijn dus door mensen bedacht en vertaald naar een algoritmisch systeem. Een machine learning model is complexer en maakt keuzes op basis van data. Hoe deze modellen tot een bepaalde keuze komen, is lastiger na te gaan omdat het model zichzelf traint aan de hand van een grote hoeveelheid data. De regels om tot een uitkomst te komen, zijn hier niet vooraf door mensen geprogrammeerd. Het model doet dit zelf, door zijn interne gewichten in een leerproces aan te passen op basis van de trainingsdata, net zo lang totdat het goede voorspellingen kan maken. Het voordeel is dat een machine learning

model een bepaalde uitkomst kan creëren zonder dat mensen zelf beslisregels hebben moeten bedenken. Het nadeel is dat met deze modellen het vaak minder goed uit te leggen is hoe en waarom het tot een bepaalde uitkomst is gekomen.

Verdieping vraag 15

Algoritmen zullen steeds vaker besluitvorming informeren. Daarom is het belangrijk om te begrijpen hoe de output van een algoritme tot stand komt. Algoritmen moeten toegankelijk zijn voor toetsing door externe experts en voor verificatie van de resultaten. Dit is een van de pijlers voor transparantie. (Overheids) organisaties moeten in staat zijn om uit te leggen hoe hun modellen en algoritmen werken. Wat niet helpt, is dat het onduidelijk kan zijn wie precies de eigenaar is van een algoritme en wie er verantwoording draagt voor de uitkomsten ervan. Daarnaast worden modellen en algoritmen niet altijd openbaar gemaakt, waardoor transparante communicatie hierover moeilijk of onmogelijk kan zijn.

Verdieping vraag 16, 17 en 18

De uitkomst van een algoritme kan een numeriek getal zijn, zoals de waarde van een huis, maar kan ook een label of categorie zijn, zoals 'hoog risico'. Het is belangrijk om na te gaan in hoeverre het resultaat van een algoritme als leidend of aanvullend wordt gebruikt en in hoeverre een mens kan afwijken van de uitkomst. Zo schuilt er bijvoorbeeld een risico in 'automation bias', het fenomeen waarbij mensen het antwoord van een systeem te makkelijk overnemen omdat het systeem ogenschijnlijk het automatisch wel goed zou hebben. Om automation bias tegen te gaan, is er in sommige gevallen dan ook scholing nodig voor de

medewerkers die met (de uitkomsten van) het algoritme gaan werken.

Verdieping vraag 19 en 20: evaluatiecriteria en fouten

Om het functioneren van een algoritme te evalueren, wordt er vaak gebruik gemaakt van bepaalde criteria. Dit worden ook wel de evaluatiecriteria genoemd. Welke evaluatiecriteria relevant zijn voor uw algoritme, is afhankelijk van wat de uitkomst van het algoritme precies is. Is dat een binaire uitkomst (ja/nee, 1/0, positief/negatief)? Of is dat een non-binaire uitkomst (zoals het voorspellen van waarden tussen 0 en 1)?

Indien het algoritme een binaire uitkomst weergeeft, zoals de voorspellingen van binaire classificatiemodellen, wordt er vaak gewerkt met een zogenoemde 'confusion matrix'. Op basis van de voorspellingen zijn er een aantal belangrijke statistieken: 'precision', 'recall', en 'accuracy'. Precision gaat over het gedeelte aan positieve *voorspellingen* ($TP/(TP+FP)$), recall heeft betrekking op het daadwerkelijke aantal positieve *gevallen* ($TP/(TP+FN)$), en accuracy heeft betrekking op alle juiste *voorspellingen* ten opzichte van de gehele dataset ($(TN+TP)/N$). Vals positief is wat er gebeurt wanneer het model positief (1) voorspelt, en het daadwerkelijke label negatief (0) is. Vals negatief is het tegenovergestelde.

Om hiervan een concreet voorbeeld te geven: als een zwangerschapstest aangeeft dat iemand zwanger is, terwijl dit niet daadwerkelijk zo is,

spreeken we van een vals positief. De omgekeerde situatie, waarin de zwangerschapstest aangeeft dat iemand *niet* zwanger is, terwijl diegene dat *wél* is, wordt een vals negatief genoemd. Wanneer de zwangerschapstest de correcte staat van (al dan niet) zwanger zijn aangeeft, spreken we over echt positieven en echt negatieven.

Echter, niet ieder model classificeert in zulke categorieën of überhaupt op een binaire manier. Er zijn ook non-binaire uitkomsten mogelijk. Zo voorspelt een regressiemodel bijvoorbeeld continue waarden tussen 0 en 1, waardoor discrete categorisatie in klassen niet hanteerbaar is. Hierbij wordt vaak gebruik gemaakt van een 'fout-term', een functie waarmee achterhaald kan worden hoe goed of fout een output is. De meest gebruikte term is 'Mean Squared Error (MSE)' of 'R²(R-squared)'. De eerste meet de afstand tot de juiste uitkomst en straft grotere fouten zwaarder. De tweede verklaart vooral de variantie binnen de dataset.

		Real Label	
		Positive	Negative
Predicted Label	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

ANONIMISEREN

Uitleg

Anonimiseren is het onherleidbaar maken van persoonsgegevens in data. Dit is onomkeerbaar. Een mogelijke techniek die hierbij gebruikt kan worden, is generaliseren, bijvoorbeeld door alle persoonlijke identificatiegegevens te reduceren tot leeftijdsgroepen. Een voorbeeld van anonimiseren, is het veranderen van burgerservicenummers binnen een dataset naar een willekeurig ander nummer, waarna de originele burgerservicenummers worden verwijderd. Hierbij is het natuurlijk wel van belang dat het nieuwe willekeurige nummer niet gekoppeld is aan herleidbare persoonsgegevens zoals voornaam en achternaam. Let op: door het combineren van overlappende datapunten (zoals postcode, leeftijd, geslacht, etc.) met andere datasets kan het nog wél mogelijk zijn om de identiteit van personen in de geanonimiseerde dataset te achterhalen.

Bij pseudonimiseren wordt een dataset op vergelijkbare manier onherleidbaar gemaakt, maar kan het proces nog omgekeerd worden. Wanneer we terugkeren naar bovengenoemd voorbeeld, dan wordt in het geval van pseudonimiseren het burgerservicenummer niet volledig verwijderd. In dat geval zal er minimaal één persoon nodig zijn die eventueel de pseudonimisering terug kan draaien. Die persoon moet de sleutel hebben om het proces terug te draaien. Wanneer dit gebeurt, wordt bijvoorbeeld het willekeurige nummer weer vervangen door het burgerservicenummer, mogelijk zelfs inclusief persoonsgegevens zoals voornaam en achternaam.

VISUALISEREN

Uitleg

Voor sommige projecten is het nodig dat data, of de uitkomsten van dataprojecten, gevisualiseerd worden. Dit kan op veel uiteenlopende manieren gebeuren, zoals middels infographics, data-visualisaties of dashboards. Houd in gedachten dat een visualisatie ook altijd de interpretatie van de data stuurt en daarmee niet neutraal is. Denk bijvoorbeeld aan het gebruik van kleur (rood versus groen) of hoe het gebruik van de schaal van een grafiek invloed kan hebben op de overgedragen boodschap.

Bij deze vraag is het vooral de bedoeling om te reflecteren op de manier waarop de data of de resultaten in beeld worden gebracht. Is dit de beste manier? Wat zijn redenen om voor een (minder) voor de hand liggende manier van visualiseren te kiezen? Kan een visualisatie verkeerd worden opgevat? Welke visualisatiestijl en technieken worden er gebruikt en drukken deze vooroordelen of vooringenomenheid uit? Houd in gedachten dat dezelfde dataset die op verschillende manieren gevisualiseerd wordt ook kan leiden tot verschillende 'lezingen' van de data. Daarnaast is het belangrijk om erbij stil te staan dat niet alle resultaten op basis van bepaalde datasets geschikt zijn om te visualiseren.

TOEGANG EN HERGEBRUIK

Uitleg

‘Toegang’ kan veel verschillende dingen betekenen, afhankelijk van de context. Hier doelen we op de kwestie van toegang tot verzamelde en gearhiveerde datasets binnen uw organisatie. Het toegankelijk maken van data binnen een bedrijf of gemeente, of het toegankelijk maken van data voor iedereen (open toegang), gaat gepaard met verschillende overwegingen. Enerzijds kan open toegang de transparantie en het vertrouwen vergroten, anderzijds kan het vertrouwen juist geschaad worden als de data niet geschikt blijken om met iedereen te delen. In sommige datasets zitten namelijk vertrouwelijke informatie en/of persoonsgegevens. De vertrouwelijkheid van deze informatie dient gewaarborgd te worden.

Daarnaast moet er goed nagedacht worden over of de resultaten op enige manier hergebruikt worden en wat voor valkuilen dat kan presenteren. Bij het beantwoorden van de vragen bij dit onderwerp kan het helpen om de voor- en nadelen van het hergebruik van data tegen elkaar af te wegen.

Soms kunnen datasets niet alleen gunstig zijn voor één project, maar kunnen ze ook hergebruikt worden voor andere doeleinden en projecten. Echter, het is goed mogelijk dat data die voor een specifiek project zijn verzameld, de validiteit verliezen wanneer deze hergebruikt worden voor een ander project. Ook zijn er beperkingen in de AVG voor het gebruik van data voor andere doeleinden dan de aangegeven doeleinden ten tijde van de dataverzameling. Dat betreft ook het hergebruiken van de dataset door derde partijen. Bovendien kan een externe partij met toegang tot de data de kennispropositie en/of betrouwbaarheid van de organisatie ondermijnen.

MAATSCHAPPELIJKE OVERWEGINGEN



Uitleg

De privacy van ieder levend persoon wordt beschermd door de AVG. Door deze wet wordt de inbreuk op privacy of het lekken van persoonlijke informatie streng gereguleerd en kan het schenden van deze wet met hoge boetes worden bestraft. Zelfs als er een algemeen idee heerst dat mensen achteloos omgaan met hun recht op privacy door zich aan te melden op sociale media, of door openlijk in te gaan op intieme details van hun persoonlijke leven, betekent dit niet dat het recht op privacy zijn urgentie verliest. Privacy blijft essentieel voor de democratie. Het is immers aan mensen zelf om te besluiten welke informatie ze delen.

Over de Data Protection Impact Assessment (DPIA)

- *De DPIA is een instrument om van tevoren risico's op het gebied van privacy in kaart te brengen. Daaropvolgend kunnen er maatregelen genomen worden om deze risico's te verkleinen.*
- *Het invullen van de DPIA is niet altijd verplicht. De DPIA is verplicht op grond van de AVG als er een hoog privacy risico is voor de betrokken personen wiens gegevens worden verwerkt.*
- *Niet elk dataproject is een verwerking van persoonsgegevens.*
- *Gegevens uit DEDA kunnen behulpzaam zijn bij het invullen van een DPIA.*

VOORINGENOMENHEID

Uitleg

‘Bias’, ofwel vooringenomenheid, is een groot probleem in dataprojecten. Een dataset, model, algoritme of organisatie met vooringenomenheid produceert ongelijke resultaten voor verschillende (sociale) groepen. Als deze vooroordelen niet herkend en vervolgens voorkomen of opgelost worden, kan er discriminatie ontstaan. Dit kan raken aan de waarde van gelijkheid en non-discriminatie. Dit gebeurt bijvoorbeeld als bepaalde vooroordelen (onbedoeld) overgenomen zijn in de data en daarom niet een eerlijke verwachting reflecteren, als het algoritme vooringenomen beslisregels of indicatoren bevat of bijvoorbeeld als de uitkomsten vaak onterecht worden aangenomen voor juist.

Verdieping vraag 32

Naast concrete tekenen van discriminatie kunt u ook een onderbuikgevoel over uw project hebben. Onderbuikgevoelens kunnen erg nuttig zijn in het opsporen van mogelijke bias, door vroeg al te herkennen dat er ‘iets’ in het project niet helemaal in de haak steekt. Het is belangrijk om naar dit soort gevoelens te luisteren en deze te onderzoeken. Als deze gevoelens namelijk wel degelijk berusten op een fout in de data of in het project, kunnen waardevolle oplossingen zich al vroeg aandienen en kan potentieel leed voorkomen worden. Bovendien, ook als deze gevoelens ongegrond blijken te zijn, is het toch vaak een goede indicatie voor hoe het algoritme door derden wordt ervaren.

Verdieping vraag 33

Er zijn verschillende manieren waarop vooringenomenheid kan terugkomen in uw dataproject. Al deze manifestaties van bias kunnen op hun eigen manier bijdragen aan een ongelijke behandeling van sociale groepen. Op deze pagina zijn een aantal beschrijvingen van vooringenomenheid te vinden voor verschillende fases in het project. Houd in gedachten dat dit slechts enkele voorbeelden van vele zijn. De lijst voorbeelden is dus niet uitputtend.

Bias in de data

Onder het kopje Databronnen is al ingegaan op de onder of over representatie van bepaalde groepen mensen in de data en eventuele proxy variabelen in de datasets die zouden kunnen zorgen voor vooringenomenheid in het model. Er kan daarnaast ook nagedacht worden over vooringenomenheid in de selectie van de datasets. Zo kan namelijk ‘selectie bias’ optreden wanneer de keuze van uw databron een vooringenomen keuze is op zich. Zo kan selectie bias bijvoorbeeld optreden als er een algoritme ingezet wordt om verkeersplanning in te regelen, maar er qua databron alleen gebruik wordt gemaakt van data afkomstig van smartphonegebruikers. In dit geval worden groepen die minder of geen smartphones gebruiken (ouderen, digitaal laaggeletterden) niet meegenomen in het algoritme. Beleidsbeslissingen die daarop volgen, zijn dan gebaseerd op een onvolledig beeld van mobiliteit. Wat zich als een objectief gegeven kan voordoen (‘mobiliteit van de stad’), is wellicht beïnvloed door het soort data waarmee gewerkt wordt (louter smartphonegebruikers). ‘Random sampling’, controlegroepen (wanneer mogelijk) en gesprekken met uw team kunnen het risico op een vooringenomen selectie minimaliseren.

Bias in het algoritme

Vooringenomenheid kan zich ook verschuilen in de manier waarop het algoritme werkt. Wellicht is een bepaalde groep niet in staat van het algoritme gebruik te maken, of zijn er beslisregels of meetmethodes gebruikt die voor vooringenomenheid kunnen zorgen.

Bias tijdens inzet van het algoritme

Vooringenomenheid kan ook ontstaan bij het interpreteren van de uitkomsten. Bijvoorbeeld in de vorm van 'confirmation bias' (bevestigingsbias) of in de vorm van automation bias. Automation bias is al eerder besproken onder het kopje Algoritmen. Confirmation bias slaat op onze menselijke geneigdheid om op zoek te gaan naar bevestiging van al eerder bestaande ideeën, en daarin ook voorkeur te geven aan stellingen die ons gedachtegoed bevestigingen (in tegenstelling tot stellingen die ons gedachtegoed juist ter discussie stellen). Confirmation bias wordt vaak versterkt door een tunnelvisie. Hierbij is men geneigd meningen te negeren die niet overeenkomen met het eigen standpunt, ook al kunnen deze afwijkende meningen belangrijk en valide zijn. Deze neiging kan problemen veroorzaken in het gebruik van data, omdat belangrijke externe standpunten, afwijkende interpretaties en/of andere zorgen, missen of niet gehoord worden.

Verdieping vraag 34

Een 'feedback loop' is ook een manier waarop vooringenomenheid kan ontstaan. Een feedback loop bij dataprojecten houdt in dat de resultaten van het project op de een of andere manier weer in het project gebruikt worden en ingezet worden als nieuwe data. Dit kan een bewuste keuze zijn, maar kan ook met minder bedachte rade plaatsvinden. In ieder geval moet men waakzaam zijn voor een zichzelf versterkende feedback loop, waarbij de uitkomsten van een eerdere iteratie impact hebben om toekomstige beslissingen.

Een klassiek voorbeeld van een feedback loop is een algoritmisch systeem dat helpt om de route van politiepatrouilles te bepalen. Als er meer misdaad wordt gemeld of gevonden in Wijk A, zal het systeem aangeven om vaker in Wijk A te patrouilleren. Hierdoor zal er meer politieaanwezigheid zijn in Wijk A, waardoor er vervolgens ook meer misdaad wordt gevonden in die wijk. Op basis van die stijgende meldingen, wordt er vervolgens nog meer gepatrouilleerd in de wijk. Ook kan dit tot reacties van inwoners leiden die op hun beurt weer als reden voor meer politieaanwezigheid gebruikt worden. Dit kan voor een zichzelf versterkend effect zorgen, waarbij Wijk A overmatig gepatrouilleerd wordt en andere wijken over het hoofd gezien worden.

TRANSPARANTE COMMUNICATIE

Uitleg

Het komt vaak voor dat er pas over communicatie en het betrekken van belanghebbenden wordt nagedacht wanneer er dingen misgaan en er kritische vragen worden gesteld. Het is noodzakelijk om niet alleen binnen de organisatie eenduidig te communiceren, maar ook naar buiten toe. Dit helpt bij het uitdragen van verantwoordelijkheid over het project en kan bijdragen aan het wekken van vertrouwen bij partijen die niet direct betrokken zijn bij het project, maar er wel mee te maken krijgen. Onthoud dat politieke partijen, burgers, advocaten of activisten hun rechten kunnen gebruiken om navraag te doen naar uw dataprojecten. Het is daarom belangrijk om na te denken over de mate waarin belanghebbenden betrokken worden en over transparantie naar buiten toe.

Verdieping vraag 36

Het is niet altijd makkelijk om transparant te zijn. Zo kunnen modellen en algoritmen zeer complex zijn. Vaak is een grote mate van kennis over statistiek en datawetenschap nodig om deze te kunnen begrijpen. Transparantie betekent in dat soort situaties niet altijd dat modellen en algoritmen vertaald moeten worden in begrijpelijke taal, maar dat ze wel toegankelijk moeten zijn voor kritische bevraging. Ten slotte is het binnen dataprojecten ook mogelijk om té transparant te zijn, bijvoorbeeld met betrekking tot de dataset. Als dit gebeurt, kan er onbedoeld een datalek ontstaan waarmee er te veel informatie verstrekt kan worden aan mensen met slechte bedoelingen.

ORGANISATIE

Uitleg

Binnen een organisatie zijn modellen en algoritmen ook onderwerp van verantwoording en behoorlijk bestuur. Steeds vaker worden overheidsinstanties gevraagd hun modellen en algoritmen te verantwoorden. Hierbij is het noodzakelijk dat duidelijk is hoe verantwoordelijkheden belegd zijn, ook omdat het soms onduidelijk kan zijn wie de eigenaar van een algoritme is. Dataprojecten hebben daarnaast vaak invloed op de dagelijkse praktijk binnen een organisatie. Ze kunnen werkprocessen veranderen, taken verschuiven of de manier waarop medewerkers hun werk ervaren, beïnvloeden.

AI. Het is dus van belang dat medewerkers voldoende 'AI-geletterd' zijn. Deze term, ook genoemd in de Europese AI-verordening, verwijst naar het begrip van medewerkers van hoe systemen functioneren, welke risico's eraan verbonden zijn en welke verantwoordelijkheden zij dragen bij het gebruik ervan. AI-geletterdheid is echter breder toepasbaar dan alleen op AI-systemen en relevant voor elke technologie die binnen een organisatie wordt geïntroduceerd.

Verdieping vraag 42

Dataprojecten kunnen directe invloed hebben op medewerkers binnen de organisatie. Een belangrijk aandachtspunt daarbij is 'deskilling'. Dit is het verschijnsel waarbij gespecialiseerd werk geleidelijk wordt overgenomen door technologie, waardoor de vaardigheden van medewerkers kunnen teruglopen, met nadelige gevolgen voor de kwaliteit. Dit kan ook gevolgen hebben voor onder andere werkplezier of werkgelegenheid. Het is daarom van belang om ook na te denken over de impact van uw project op uw medewerkers. Houd er ook rekening mee dat AI-projecten vaak vastlopen omdat er te weinig aandacht is besteed aan de implementatie in de organisatie en het meenemen van de medewerkers.

Verdieping vraag 43

Verantwoord en behoorlijk bestuur vraagt om zorgvuldig en deskundig gebruik van data en

TOEKOMSTSCENARIO'S

Uitleg

Hoewel sommige dataprojecten maar korte tijd relevant of in gebruik zijn, komt het vaak voor dat projecten op de midden tot lange termijn effecten dragen. Het is daardoor van belang om nu al over deze langetermijneffecten na te denken, en of deze de huidige stand van zaken zouden veranderen. Daarbij spelen thema's als 'function creep', datasoevereiniteit en klimaatimpact een belangrijke rol. Het is dan ook nodig om tussentijdse evaluatiemomenten te houden, om ervoor te zorgen dat het dataproject nog steeds doet waarvoor het bedoeld is en er geen (onbedoelde) negatieve effecten aan kleven.

Verdieping vraag 44

'Function creep' houdt in dat informatie wordt gebruikt voor een doel dat niet het originele toegewezen doel is. Een voorbeeld van function creep is als volgt: op kantoor wordt er een beveiligingssysteem geïnstalleerd, waarvoor werknemers moeten in- en uitchecken met als doel een beveiligde toegang te faciliteren. Echter, als deze informatie vervolgens gebruikt wordt om individuele werknemers te traceren of te monitoren, spreken we van function creep. Dit is namelijk een inbreuk op privacy en waardigheid die voorbijgaat aan het vastgestelde doel van het creëren van een beveiligde toegang. Door vroegtijdig te verkennen welke alternatieve toepassingen denkbaar zijn, kunnen duidelijke grenzen worden gesteld aan het gebruik van gegevens. Dit helpt misbruik te voorkomen, bevordert transparantie en beschermt betrokkenen tegen onverwachte negatieve gevolgen.

Verdieping vraag 45

Datasoevereiniteit gaat over het behoud van controle over data. Waar wordt de data opgeslagen? Welke partijen krijgen toegang tot de data? Welke wetten zijn op het databehoud van toepassing? Krijgen derde partijen de data in hun bezit en kunnen zij deze verder bewerken? Is er kans op afhankelijkheid van een bepaalde leverancier? Omdat digitale omgevingen steeds meer afhankelijk worden van externe (internationale) leveranciers en infrastructuren is het van belang dat datasoevereiniteit nagestreefd wordt. Dit draagt bij aan het waarborgen van onder andere de waarden van autonomie en veiligheid.

Verdieping vraag 46

Digitale projecten hebben een impact op het klimaat door hun energiegebruik, datacenterbelasting en de ecologische voetafdruk van hardware. Tegelijkertijd kunnen dataprojecten ook juist helpen om duurzaamheid te bevorderen, bijvoorbeeld doordat het processen efficiënter maakt. Het is daarom van belang te kijken naar de milieu-impact van een dataproject om klimaatverantwoordelijke keuzes te maken.

UITKOMST



UITKOMST

WAARDEN

Uitleg

Aan het begin van de workshop heeft u een aantal waarden opgeschreven. Pak deze er weer bij en ga voor uzelf na of deze waarden in het project gerepresenteerd zijn, en zo ja, waar.

Ga alle leden van de projectgroep af en laat ze vertellen of, en waar, de waarden in het project gerepresenteerd zijn. Degene die de beurt heeft, mag diens waarden op de poster plakken op een plek waar diegene denkt dat deze het meeste positief terugkomen. Zo kan de waarde “eigenaarschap” bijvoorbeeld bij de vragen over Organisatie of bij de vragen over Databronnen worden geplakt. Als u aan de beurt bent om uw waarden op te plakken, leg dan ook aan de rest van uw groep uit waarom u vindt dat de waarden op die plek in het project horen. Hierover in gesprek gaan als groep kan leiden tot nieuwe inzichten.

Het kan ook zijn dat u vindt dat (één van) de waarden die u hebt genoteerd overal in het project passen. Dan kunt u die waarde in het midden plakken.

Ten slotte is het ook nog mogelijk dat (één van) de waarden die u heeft opgeschreven helemaal niet in het project gerepresenteerd zijn.

Bespreek dan met elkaar of deze waarde van belang zou kunnen zijn voor het project en of het nodig is bepaalde aspecten van het project te veranderen zodat de waarde wel kan worden gerepresenteerd. Het kan voorkomen dat een waarde minder relevant is voor een project. Zo is de waarde “loyaliteit” bijvoorbeeld minder relevant bij een dataproject dat passanten binnen een bepaald gebied anoniem telt. Het kan echter ook voorkomen dat een waarde wél

relevant is voor het project, maar niet positief terugkomt. Dit kan bijvoorbeeld het geval zijn wanneer “transparantie” als belangrijke waarde is opgeschreven, maar het bij het beantwoorden van de vragen blijkt dat het project helemaal niet transparant is. Bespreek in zo’n geval als groep wat er mogelijk veranderd moet worden aan het project om gehoor te geven aan die waarde.

UITKOMST

CONCLUSIE

Uitleg

Nadat alle waarden zijn behandeld, kan de groep door naar het laatste, afsluitende onderdeel. De Conclusie biedt vier afsluitende vragen die de gehele sessie samenvatten en de basis vormen voor de verdere vervolgstappen.

UITKOMST

VERVOLGSTAPPEN

Uitleg

U kunt de DEDA poster, inclusief alle antwoorden en actiepunten, gebruiken als basis voor of ter ondersteuning van een rapport over het dataproject.

De antwoorden en actiepunten kunnen dienen als handvatten die de verantwoording van het project makkelijker maken. Zo kunnen de antwoorden helpen om uit te leggen welke keuzes er zijn gemaakt, en minstens even belangrijk, waarom ervoor is gekozen om bepaalde dingen niet te doen.

UITKOMST

VERDERE INFORMATIE

Uitleg

Voor verdere informatie over DEDA, verwijzen we naar de volgende publicaties:

Franzke, A., Muis, I. M., & Schaefer, M. T. (2021). Data Ethics Decision Aid (DEDA): a dialogical framework for ethical inquiry of AI and data projects in the Netherlands. *Ethics and Information Technology*, 23(3), 551-567. Article 3. <https://link.springer.com/article/10.1007/s10676-020-09577-5>

Siffels, L., van den Berg, D., Schaefer, M., & Muis, I. (2022). Public Values and Technological Change: Mapping how Municipalities Grapple with Data Ethics. In *New Perspectives in Critical Data Studies: The Ambivalences of Data Power* (pp. 243-266). (Transforming Communications – Studies in Cross-Media Research). Palgrave Macmillan. https://link.springer.com/chapter/10.1007/978-3-030-96180-0_11

Colofon

Concept:

Mirko Tobias Schäfer, Ph.D.
Aline Franzke, MA

Redactie:

Julia Straatman, MA
Iris Muis, MA
Hajo Bijleveld, BA
Ydo Bergstra, BSc
Lysa Ngouateu, BSc
Mirko Tobias Schäfer, Ph.D.
Bart Anthony Kamphorst, Ph.D.

Illustratie voorblad:

Sammy Hemerik

Universiteit Utrecht
Data School
Drift 13, kamer 0.01
3512 BR Utrecht
Netherlands

'De Ethische Data Assistent' is ontwikkeld door de Utrecht Data School en de Universiteit Utrecht in opdracht van de Gemeente Utrecht. © Data School, Utrecht University 2026

DEDA - Versie 4.0
februari 2026
Voor updates kijk op:

deda.dataschool.nl

Noch de Universiteit Utrecht, noch de Data School, noch de Gemeente Utrecht zijn aansprakelijk voor (mogelijke) schade in verband met, of voortkomend uit, het gebruik van materiaal of inhoud van de 'De Ethische Data Assistent'.

Utrecht, 2026.

Terwijl DEDA wordt gebruikt, leren we om het te verbeteren. Deze verbeteringen zullen in toekomstige versies worden geïmplementeerd. Wanneer u DEDA gebruikt en op- of aanmerkingen heeft, aarzel dan niet om deze met ons te delen. U kunt daarvoor altijd mailen naar:

dataschool@uu.nl





**Universiteit
Utrecht**

DATA SCHOOL