

Reflectierapport: Pilot AI ethiek volwassenheidsmodel

Inzichten in de inzet van een AI ethiek volwassenheidsmodel binnen onderwijsinstellingen



Auteur(s): SURF
Versie: 1.0
Datum: 14 november 2024
Kenmerk: 10.5281/zenodo.14040030

Deze publicatie is gelicenseerd onder een Creative Commons Naamsvermelding 4.0 Internationaal.

Inhoudsopgave

1 Management samenvatting	3
2 Inleiding	4
2.1 Achtergrond	4
2.2 AI ethiek volwassenheidsmodel	5
3 Aanpak van de pilot	7
3.1 Uitvoering van de pilot	7
4 Resultaten van de pilot	8
4.1 De AI ethiek volwassenheid van de deelnemende instellingen	8
4.1.1 <i>Dimensie 1: Cultuur en Bewustwording</i>	8
4.1.2 <i>Dimensie 2: Beleid</i>	9
4.1.3 <i>Dimensie 3: Governance</i>	9
4.1.4 <i>Dimensie 4: Communicatie en Training</i>	10
4.1.5 <i>Dimensie 5: Ontwikkelingsprocessen</i>	10
4.1.6 <i>Dimensie 6: Tooling</i>	11
4.2 Conclusie	11
5 Reflectie	12
5.1 Ervaringen met het AI ethiek volwassenheidsmodel	12
5.2 Ervaringen met het uitvoeren van de Pilot	13
6 Conclusie	14
6.1 Dit hebben we geleerd uit de pilot	14
6.2 Aanbevelingen voor vervolg	14
7 Relevante links & activiteiten	16
7.1 Leeslijst	16
7.2 Relevante sector activiteiten binnen SURF en Npuls	16
Bijlage 1 AI ethiekvolwassenheidsmodel	17
Colofon	18

1 Management samenvatting

Onderwijsinstellingen zien met de snelle opkomst van AI heel veel op zich af komen. Naast de enorme impact die AI heeft op alle aspecten van het onderwijs komt er ook nog eens aanvullende wet- en regelgeving aan en dient het onderwijs op een ethische en verantwoorde manier om te gaan met AI. Dat kan het thema AI en de vraag over verantwoord gebruik tot een ongrijpbaar thema maken waar eigenaarschap, strategie en structuur lastig voor de definiëren is. Instellingen hebben een verantwoordelijkheid om te zorgen dat AI-systemen volgens de wet én ethisch worden toegepast.

Het AI ethiek volwassenheidsmodel biedt een raamwerk om je organisatie in te richten op het structureel borgen van ethische afwegingen bij de inzet van AI en algoritmen. Het is ontwikkeld om op een holistische manier alle aspecten van ethische AI te vertalen naar de realiteit van organisaties. Door de koppeling tussen kaders en organisatiepraktijk kunnen instellingen met het model ethische AI systematisch benaderen en integreren in de bredere strategische doelstellingen van onderwijsinstellingen. Om dit concreet te maken onderscheidt het model zes dimensies: bewustzijn en cultuur, beleid, governance, communicatie en training, ontwikkelprocessen, en tooling. Dit zijn de organisatorische randvoorwaarden om op een betekenisvolle manier ethiek een plek te geven in de inkoop, ontwikkeling en gebruik van AI systemen.

Binnen de pilot met het AI ethiek volwassenheidsmodel (georganiseerd door SURF, EDSA en de EUR) hebben 11 instellingen dit model toegepast om ervaring op te doen. De pilot bood een startpunt voor gesprekken over ethische vraagstukken en heeft bij sommige deelnemers bijgedragen aan het verhogen van het bewustzijn en de betrokkenheid binnen de instelling. Het model bood een gestructureerde leidraad, waardoor het voor deelnemende instellingen ingezet kon worden als 'good practice' voor het vormgeven van hun eigen AI ethiek aanpak. Eigenlijk kunnen we concluderen dat de pilot ervoor zorgde dat het gesprek plaats vond en het model ervoor zorgde dat dat een nuttig gesprek was.

Bovenal werd duidelijk dat het tijd is om serieus te kijken naar hoe men regie kan voeren over AI, ethiek, en de AI verordening binnen de organisatie.

2 Inleiding

In 2024 heeft SURF, in samenwerking met de Ethische Data Science Associatie (EDSA) en de Erasmus Universiteit (EUR) een pilot georganiseerd om de toepassing van het AI ethiek volwassenheidsmodel (Krijger et al., 2022) in de onderwijssector te onderzoeken. Het doel van de pilot was onderzoeken hoe dit model kan bijdragen aan de verantwoorde inzet van AI binnen onderwijsinstellingen. De intentie was om inzicht te krijgen in de meerwaarde van het model, te leren van de praktijkervaringen, en te inspireren tot bredere toepassing. Dit rapport biedt een reflectie op de leringen en ervaringen uit deze pilot en is bedoeld om beslissers binnen mbo-, hbo- en wo-instellingen, zoals directeuren ICT, CIO's en informatiemanagers, te informeren en aan te zetten tot actie.

Sector	Deelnemende instellingen ¹
Universiteit	Wageningen Universiteit & Research
	Universiteit Utrecht
	Radboud Universiteit
	Radboud UMC
	Erasmus Universiteit Rotterdam ²
HBO	Hogeschool van Amsterdam
	Hanze
	Hogeschool Utrecht
MBO	ROC Amsterdam-Flevoland
	ROC Mondriaan
Overig	SURF

Tabel 1 Deelnemende instellingen

2.1 Achtergrond

Iedere dag brengt ons nieuwe ontwikkelingen van AI die de onderwijssector raken. Die opkomst biedt zowel kansen als uitdagingen voor onderwijsinstellingen. Onderwijsinstellingen staan voor de taak om AI op een manier te gebruiken die hun (publieke) waarden respecteert en de gebruikers, zoals studenten en docenten, ten goede komt. Daarbij hoort hun maatschappelijke taak van het verzorgen van goed onderwijs centraal te staan.

De digitale transformatie in het onderwijs biedt tal van voordelen, zoals verbeterde toegankelijkheid en efficiëntere leermethoden, maar legt tegelijkertijd druk op belangrijke publieke waarden, zoals eerlijkheid, privacy, inclusiviteit en gelijkheid. Ook zijn er zorgen over de groeiende invloed van dominante marktpartijen op de manier waarop wij ons onderwijs vormgeven.³

De opkomst van AI roept tal van ethische vragen op over besluitvorming en rechtvaardigheid. AI-systemen kunnen bestaande vooroordelen versterken en nieuwe creëren, wat kan leiden tot discriminatie en een toename van ongelijkheid in de samenleving. Deze zorgen worden onderstreept door incidenten zoals de toeslagen-affaire of

¹ Mocht je willen weten wie er vanuit jouw instelling contactpersoon was voor de pilot, stuur dan een mail met die vraag naar responsibletech@surf.nl

² De EUR heeft de pilot zelfstandig uitgevoerd voordat de SURF pilot plaatsvond, hun resultaten zijn meegenomen in dit rapport.

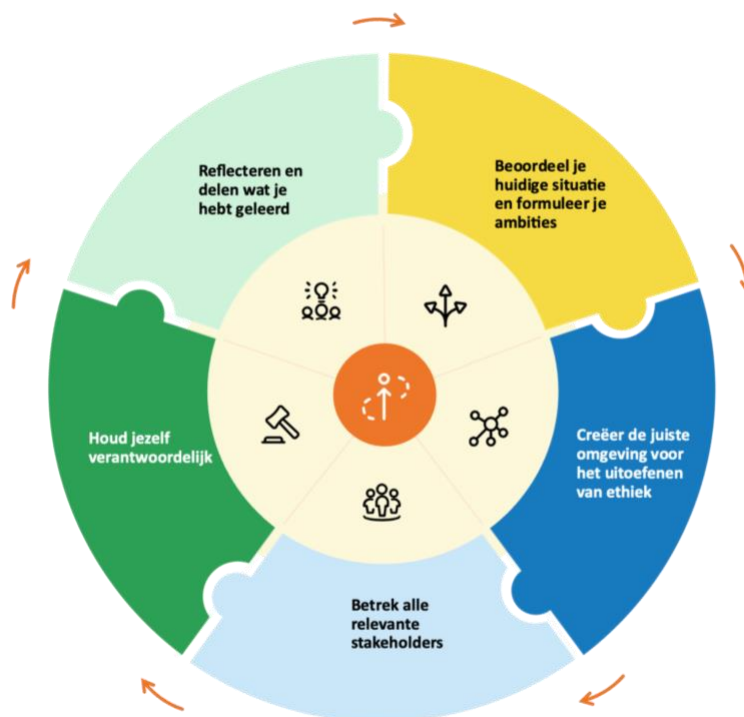
³ "Digitalisering bedreigt onze universiteit. Het is tijd om een grens te trekken," de Volkskrant, December 22, 2019, <https://www.volkskrant.nl/columns-opinie/digitalisering-bedreigt-onze-universiteit-het-is-tijd-om-een-grens-te-trekken~bff87dc9/>.

discussies rondom proctoring waarbij een student haar onderwijsinstelling beschuldigde van discriminatie door de inzet van proctoringsoftware. Hoewel discriminatie in dit specifieke geval niet werd bewezen, benadrukte het College voor de Rechten van de Mens de zorgplicht van instellingen om te waarborgen dat de technologieën die zij gebruiken, niet-discriminerend zijn.⁴

De noodzaak van een ethische benadering van AI wordt erkend door de EU. In 2019 publiceerde de 'High-Level Expert Group' ethische richtlijnen voor betrouwbare AI, welke voorschreven dat AI-toepassingen wettig, ethisch, en robuust moeten zijn.⁵ In 2022 publiceerde de EU een richtlijn voor het ethisch gebruik van AI en data in het onderwijs.⁶ Deze richtlijn benadrukt de nauwe relatie tussen AI-ethiek en onderwijsethiek, en behandelt ethische dimensies zoals menselijke agency⁷, rechtvaardigheid, menselijkheid en gerechtvaardigde keuze.

Het belang van ethische omgang met AI wordt verder versterkt door de invoering van wetgeving zoals de AI verordening, welke in juli 2024 door de Europese Unie is aangenomen.⁸ In deze wet worden de verschillende ethische principes wettelijk verankerd. De nieuwe AI verordening en bestaande GDPR wetgeving zorgen er voor dat onderwijsinstellingen worden geconfronteerd met de noodzaak om uitgebreide AI-governance-strategieën te ontwikkelen, waarbij ethische overwegingen centraal staan.

Kortom, onderwijsinstellingen hebben een uitdagende verantwoordelijkheid voor de manier waarop zij AI systemen inzetten binnen de organisatie. Binnen SURF hebben we een model ontwikkeld om organisaties op pragmatische wijze te helpen bij deze verantwoordelijkheid.



Figuur 1 Een 5 stappen model voor Responsible Tech

⁴ <https://www.mensenrechten.nl/actueel/nieuws/2023/10/17/student-niet-gediscrimineerd-door-tentamenssoftware-proctorio-maar-vu-had-de-klacht-zorgvuldiger-moeten-behandelen>

⁵ <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

⁶ <https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1/language-en>

⁷ De capaciteit om vorm te geven aan het eigen leven, de handelingsmogelijkheid van een individu. [https://nl.wikipedia.org/wiki/Agency_\(sociologie\)](https://nl.wikipedia.org/wiki/Agency_(sociologie))

⁸ Wil je meer weten over de AI verordening? <https://edu.nl/x8pd4>

2.2 AI ethiek volwassenheidsmodel

Het AI ethiek volwassenheidsmodel is ontwikkeld door EDSA (Ethical Data Science Association) en biedt een raamwerk dat organisaties helpt bij het operationaliseren van ethische principes rond AI.⁹ In tegenstelling tot de traditionele benadering van AI-ethiek, die vaak gericht is op abstracte richtlijnen en principes, richt dit model zich op de praktische implementatie binnen organisaties. Het model benadrukt zes belangrijke dimensies: bewustzijn en cultuur, beleid, governance, communicatie en training, ontwikkelprocessen, en tooling. Deze dimensies bieden een holistisch perspectief dat organisaties helpt om ethiek integraal onderdeel te maken van hun dagelijkse praktijk en strategische doelen. Op deze zes dimensies zijn vijf niveaus van volwassenheid mogelijk variërend van level 1, waarin enkel gefragmenteerd en zonder formele status activiteiten op deze dimensie worden ontplooid, tot level 5 waarin activiteiten rond de dimensie volledig geïntegreerd zijn in de structuur en werkwijze van de organisatie en doorlopend geëvalueerd en geoptimaliseerd worden.



Figuur 2 De zes dimensies van AI ethiek volwassenheid

Het model is ontwikkeld op basis van literatuuronderzoek en inzichten uit leerbijeenkomsten met verschillende publieke, private en onderzoeksinstituten in Nederland. Door de centrale thema's te halen uit tientallen gesprekken over ethische dilemma's bij AI en hoe organisaties boards, teams en beleid gebruiken om deze dilemma's goed te managen kon een organisatiemodel ontwikkeld worden dat inzicht geeft in welke terreinen essentieel zijn om ethiek te borgen. Het model is ontworpen om organisaties te ondersteunen bij het begrijpen en implementeren van AI-ethiek, door een duidelijke structuur te bieden voor het evalueren en verbeteren van ethische praktijken. Het benadrukt het belang van een gecoördineerde aanpak waarbij verschillende aspecten van de organisatie betrokken zijn, van beleidsontwikkeling en governance tot training en communicatie. De centrale gedachte is dat vooruitgang op het gebied van ethische AI pas tot stand kan komen als op al deze zes dimensies actie wordt ondernomen en dat, afzonderlijk, focus op één enkele dimensie maar beperkt effect zal hebben. Hierdoor wordt ethische AI niet alleen iets wat zich op het technische of compliance domein afspeelt, maar iets waar de hele organisatie bij betrokken is.

Het model is bedoeld om als organisatie "grip" te krijgen op AI en ethiek door het concreet in kaart brengen van aandachtsgebieden en activiteiten. Een hogere volwassenheid stelt de organisatie in staat op incidenten en ontwikkelingen te reageren.¹⁰ Het model is gebaseerd op een holistische gedachte dat het hebben van principes niet voldoende is, en dat er procedures en structuren vereist zijn om deze principes daadwerkelijk in de praktijk toe te passen.

⁹ Krijger, J., Thuis, T., de Ruiter, M. et al. The AI ethics maturity model: a holistic approach to advancing ethical data science in organizations. *AI Ethics* 3, 355–367 (2023). <https://doi.org/10.1007/s43681-022-00228-7>

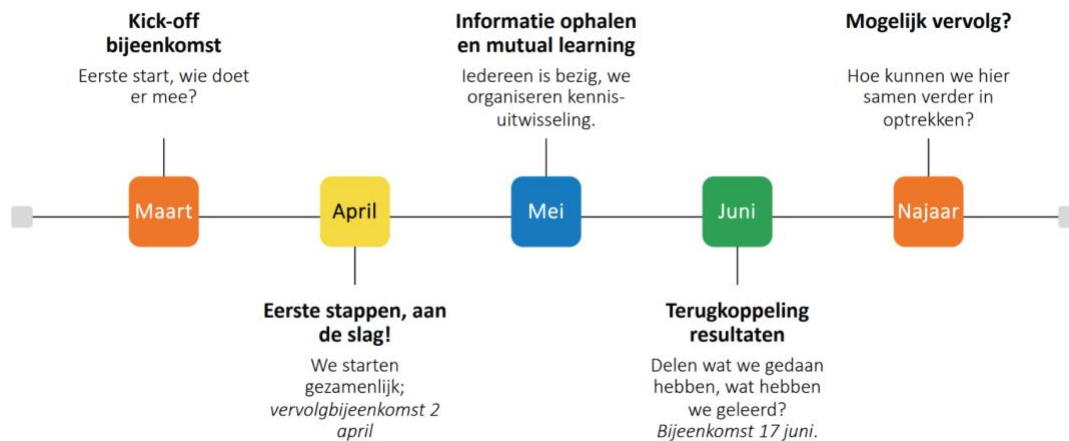
¹⁰ De verschillende volwassenheidsniveaus zijn beschreven in appendix 1.

Deze benadering sluit aan bij de AI verordening. Diverse AI toepassingen in het onderwijs worden in de wet als hoog risico worden getypeerd¹¹ wat betekent dat organisaties in het onderwijs aan strenge eisen zullen moeten voldoen op o.a. het gebied van risicomanagement. Het volwassenheidsmodel kan (mogelijk) daarin helpen om organisaties voor te bereiden door de relevante stakeholders intern bij elkaar te brengen en bijvoorbeeld impact assessments of conformiteitsassessments logisch te verbinden aan andere activiteiten in de AI governance. Organisaties die bijvoorbeeld al hoog scoren op de dimensie governance en beleid zijn daarmee beter gepositioneerd om te reageren op de veranderingen en eisen die de AI verordening met zich mee brengt.

¹¹ De AI verordening beschrijft AI systemen die *toegang tot het onderwijs regelen, die leeruitkomsten evalueren, onderwijsniveau beoordelen, en het monitoren van ongeoorloofd gedrag van studenten tijdens toetsing* allemaal als hoog-risico.

3 Aanpak van de pilot

De pilot is georganiseerd door SURF in samenwerking met EDSA en de Erasmus Universiteit Rotterdam. Alle bij SURF aangesloten instellingen werden uitgenodigd om deel te nemen. De 11 deelnemende instellingen (zie tabel 1) voerden een informele self-assessment uit aan de hand van het AI ethiek volwassenheidsmodel. Zij onderzochten hun huidige staat van AI-ethiek implementatie, kregen inzicht in waar verbeteringen mogelijk zijn en deelden hun bevindingen en ervaringen met elkaar (onder de Chatham House Rules om onderling delen te stimuleren). Het doel was niet alleen om te leren hoe het model in de praktijk werkt, maar ook om te begrijpen hoe het model kan helpen bij het vormgeven van een ethisch verantwoorde inzet van AI binnen onderwijs- en onderzoekinstellingen. Dit alles in een korte periode van 11 weken.



Figuur 3 De tijdslijn van de pilot

De deelnemers kregen een informatiepakket met een Excel-format om het volwassenheidsmodel in te kunnen vullen. Gedurende de pilot werd een ondersteuning geboden door middel van een discussieplatform in een Teams-omgeving en georganiseerde 'mutual learning'-sessies in de vorm van vier vragen-momenten. Daarnaast was de organisatie gedurende de pilot beschikbaar als vraagbaak voor de deelnemers.

3.1 Uitvoering van de pilot

Deelnemende instellingen zijn vervolgens voortvarend aan de slag gegaan met het uitvoeren van de pilot. De aanpak varieerde per instelling afhankelijk van beschikbare resources, prioriteit, en plaats binnen de organisatie. Dit liep van focus-sessies binnen een team tot grootschalige uitvraag in de hele organisatie.

Verschillende methodes die gebruikt werden door deelnemers zijn:

- Binnen het (diverse) pilot-team zelfstandig het model invullen en vervolgens doorspreken.
- Uitzet van enquêtes binnen de organisatie (bijv. onder digicoaches).
- Gespreksvoering met organisatie-achterban, waarbij met model een structuur was voor het gesprek.
- Verschillende stakeholders verzoeken het model zelfstandig in te vullen, om vervolgens de modellen te vergelijken.
- Geïntegreerde aanpak samen met business information management (BIM).
- Georganiseerde sessies met key-stakeholders uit de organisatie met een workshop gebaseerd op het model.

4 Resultaten van de pilot

Hoewel het gebruik van het model en het onderzoeken van de meerwaarde ervan centraal stond in de pilot willen we niet voorbijgaan aan de resultaten en eerste inzichten die het gebruik van de pilot heeft opgeleverd voor de deelnemende instellingen. In dit hoofdstuk zullen we kort de scores van de deelnemende instellingen bespreken, gekoppeld aan wat algemene inzichten die naar voren zijn gekomen.

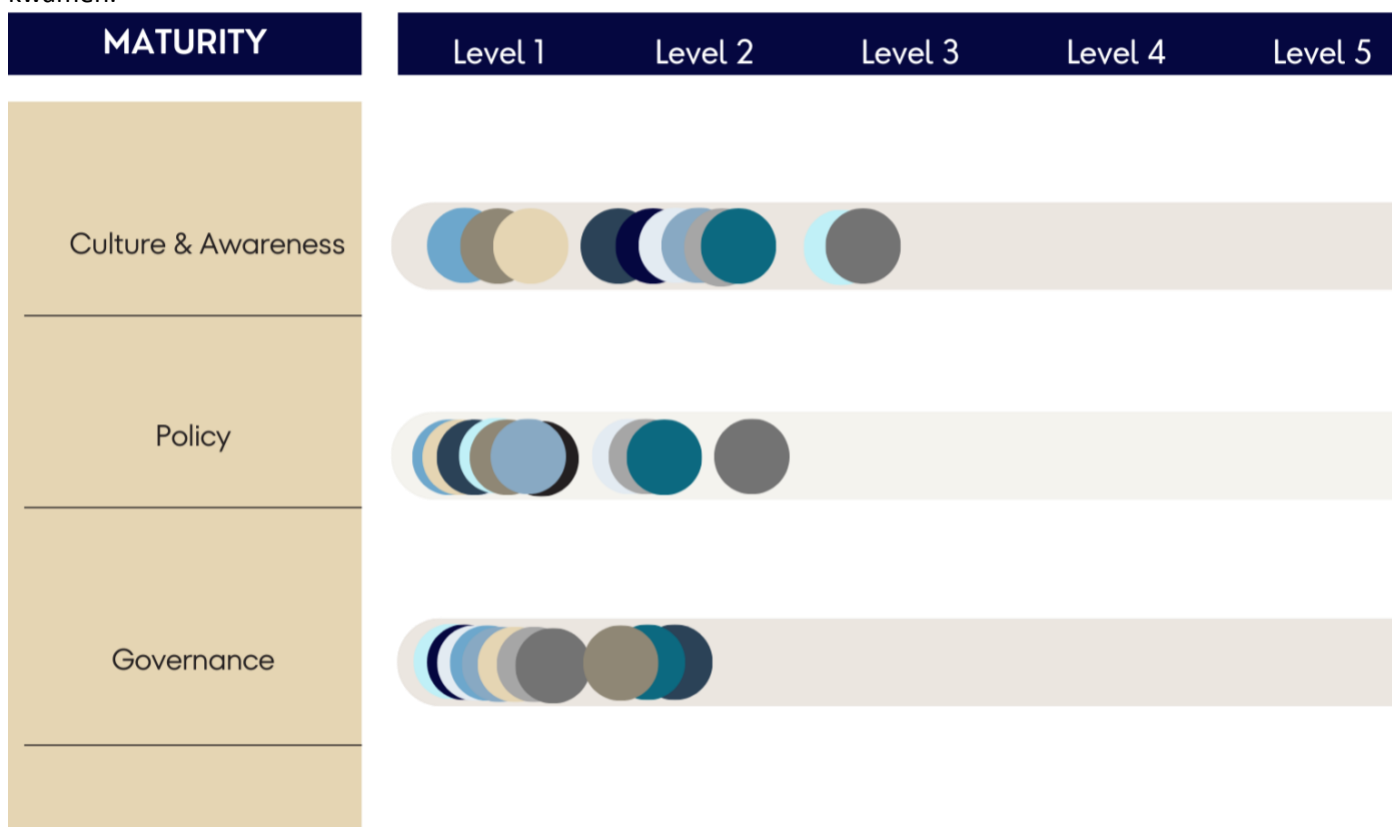
4.1 De AI ethiek volwassenheid van de deelnemende instellingen

De resultaten van de pilot zijn geaggregeerd en geanalyseerd door EDSA en geven een algemene indruk van de ingevulde modellen. Let wel, dit kan niet gezien worden als representatief voor de sector en er kunnen geen resultaten voor individuele onderwijsinstellingen uit gehaald worden. De resultaten zijn gebaseerd op de 11 deelnemende onderwijsinstellingen.

De gemiddelde scores:¹²

1. Cultuur en bewustwording: 1,9
2. Beleid: 1,3
3. Governance: 1,4
4. Communicatie en training: 1,6
5. Ontwikkelprocessen: 1,4
6. Tooling: 1,4

Hieronder wordt per dimensie kort toegelicht wat deze scores betekenen en welke inzichten hier uit naar voren kwamen.



Figuur 4 Overzicht van de verdeling van scores door deelnemende instellingen voor de eerste 3 dimensies.

¹² Het doel van deze pilot was niet

4.1.1 Dimensie 1: Cultuur en Bewustwording

De eerste dimensie, Cultuur en Bewustwording, richt zich op het bewustzijn van ethische aspecten van AI binnen organisaties, die ruimte die vanuit management geboden wordt voor ethiek en de cultuur waarin deze ethische praktijken zijn ingebed. Uit de pilot bleek dat deze dimensie relatief goed ontwikkeld is. Van alle dimensies scoren onderwijsinstellingen hier het hoogst op. Bij de meeste, maar nog niet alle, organisaties is de stap genomen van het eerste niveau, waar gefragmenteerd bewustzijn is zonder concrete organisatie, naar het tweede niveau waarop meer intern de afstemming wordt gezocht over ethische AI. Dat neemt niet weg dat er nog werk aan de winkel is wat betreft het formaliseren en verder uitrollen van activiteiten op deze dimensie, die nu nog vooral *ad hoc* zijn ingestoken.

In de pilot kwam naar voren dat er op deze dimensie behoefte is aan een algemeen bewustwordingsplan met gerichte kennissessies per team of rol. Het introduceren van vaste 'demo' of 'stand-up' sessies kan helpen het bewustzijn te vergroten. Ook werd gesuggereerd dat uitkomsten van het volwassenheidsmodel en de resultaten van de pilot zouden moeten worden gedeeld met het leiderschap binnen in de instelling (e.g. CvB) om de diverse facetten van verantwoorde inzet van AI (ethiek) bij hen inzichtelijk te maken.

4.1.2 Dimensie 2: Beleid

De tweede dimensie, Beleid, richt zich op de aanwezigheid en kwaliteit van beleidskaders rond ethische richtlijnen voor AI binnen organisaties. Ten opzichte van de eerste dimensie waren de scores hier aanzienlijk lager en op deze dimensie verkeren nagenoeg alle deelnemende organisaties nog op het eerste *ad hoc* niveau.

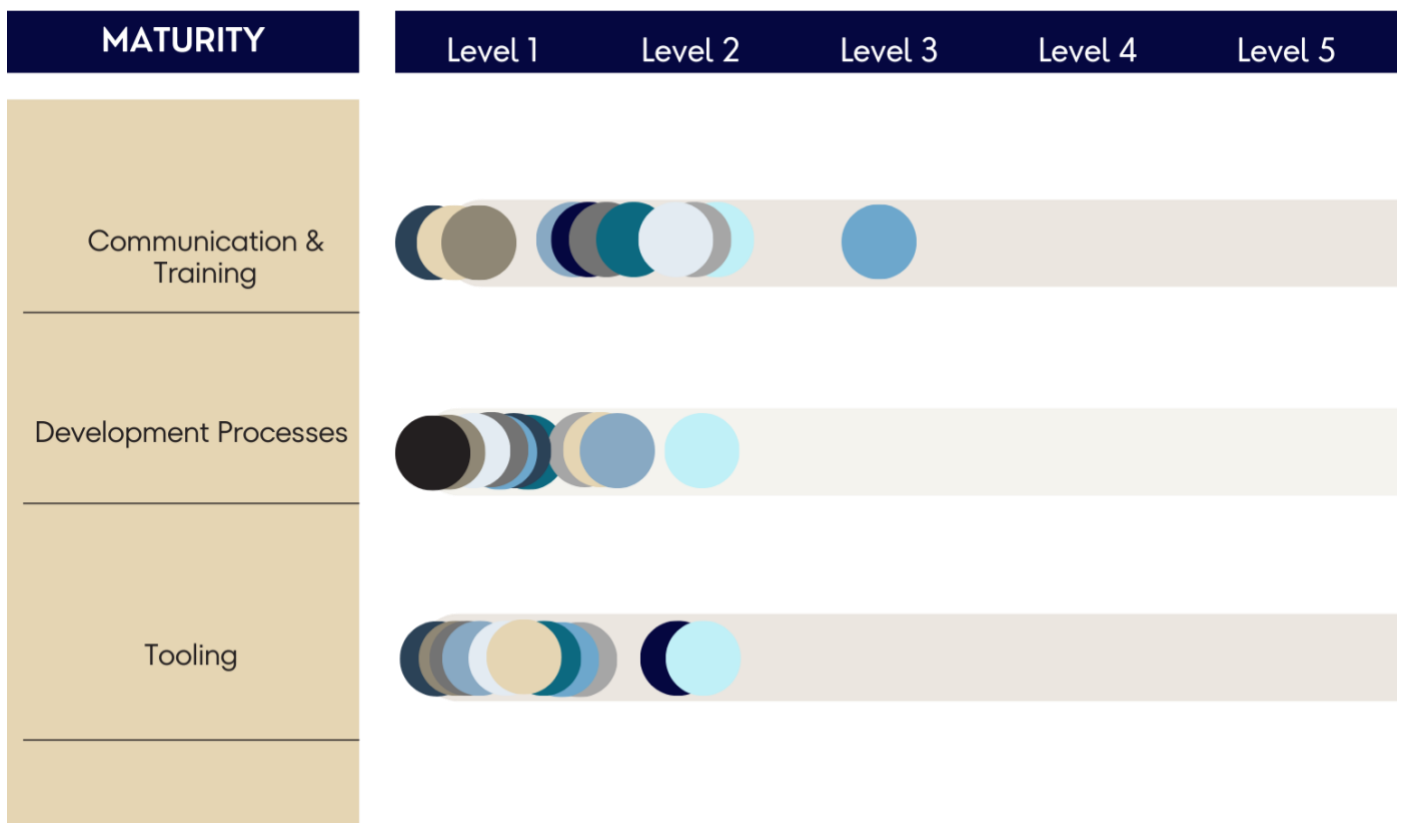
Aangezien in het assessment indicaties naar voren kwamen dat men snel vooruitgang wil boeken op deze dimensie liggen hier kansen voor verdere (sectorale) samenwerking.¹³ Zo kan gezamenlijk gewerkt worden aan het opstellen van specifieke kaders voor verantwoord AI-gebruik of kan men werken aan het ontwikkelen van een sector brede template of blauwdruk voor AI beleid die kan bijdragen aan uniform beleid. Ook is er behoefte aan onderzoek naar best practices en een inventarisatie van bestaande beleidskaders die als basis kunnen dienen.

4.1.3 Dimensie 3: Governance

Deze dimensie onderzoekt de aanwezigheid van organisatorische structuren en procedures zoals formele checks in het proces of specifieke boards om ethisch gebruik van AI te waarborgen. De resultaten laten een vergelijkbaar beeld zien als bij de beleidsdimensie met iets meer onderwijsinstellingen die stappen zetten in de richting van governance of zich oriënteren op hoe ethische governance rond AI kan aansluiten bij andere vormen van governance zoals data governance.

Het integreren van AI-governance binnen bestaande werkwijzen is essentieel en organisaties zouden moeten inventariseren welke kritieke afwegingen en checks nodig zijn en wie daarbij betrokken moet worden. Daarnaast kan kennis gedeeld worden over hoe ethische checks op AI gebruik gekoppeld kunnen worden aan bestaande processen of boards die onderwijsinstellingen al in gebruik hebben. Het formaliseren van werkgroepen en delen van inzichten om tot best practices te komen is een aanbevolen volgende stap.

¹³ De leergang AI beleid (najaar 2024) kan gezien worden als een eerste activiteit op dit vlak. Zie ook: <https://communities.surf.nl/ai-in-education/artikel/surf-organiseert-de-leergang-ai-beleid-voor-onderwijs>



Figuur 5 Overzicht van de verdeling van scores door deelnemende instellingen voor de laatste 3 dimensies.

4.1.4 Dimensie 4: Communicatie en Training

De vierde dimensie richt zich op de communicatie en training rondom de ethische aspecten van AI in alle lagen van de organisatie. Dat wil zeggen dat er intern en extern gecommuniceerd wordt over verantwoord gebruik van AI maar ook dat medewerkers in allerlei rollen, van onderzoeker, docent tot privacy office en bestuur, kennis hebben van de ethische aspecten van AI. In lijn met de relatief hogere score op Cultuur en Bewustzijn zien we ook op deze dimensie een wat hogere score ten opzichte van de andere dimensies. Instellingen zijn bezig om het bewustzijn om te zetten in concrete trainingen zoeken naar manieren om te communiceren hierover. Wederom blijft dit vaak *ad hoc* en zonder beleid of leerpad voor ethische AI per rol.

Ook hier leert de pilot dat er kansen liggen voor samenwerking. Het is belangrijk om bestaande training te integreren en te verkennen wat sectoroverstijgende (mbo, hbo, en wo) training zou kunnen omvatten. Ook is het zinvol om de behoefte aan communicatie binnen organisaties te inventariseren, evenals de vraag naar informatie over AI ethiek. Delen en co-creatie van aanpakken en vraagstukken binnen de sector kan leiden tot externe validatie en verdere ontwikkeling.

4.1.5 Dimensie 5: Ontwikkelingsprocessen

Deze dimensie richt zich op de ontwikkelingspraktijken rondom AI en inbedding van ethiek in ontwikkel- en inkoopprocedures om ethisch gebruik van AI te bevorderen. Omdat onderwijsinstellingen zelf beperkt AI toepassingen ontwikkelen voor hun bedrijfsvoering is dit een wat lastigere dimensie. Binnen onderzoek en onderwijs wordt wel AI ingezet maar dit zijn meestal ingekochte pakketten. Over het algemeen wordt aan de inkoop- en ontwikkelkant geen plek ingericht voor de ethische principes of afwegingen.

Er is echter wel behoefte om ethische aspecten in elke stap van het ontwikkelingsproces mee te nemen. Ook bij inkoop en implementatie van AI applicaties van derden. Daarom is het belangrijk om aandacht te besteden aan proces en transparantie voor weloverwogen beslissingen voor inkoop en implementatie. Dit draagt bij aan duidelijkheid over waar AI wordt ontwikkeld en onderhouden binnen de organisatie.

4.1.6 Dimensie 6: Tooling

De laatste dimensie richt zich op het gebruik van tools en methoden om ethische aspecten van AI te identificeren en te beheren. Denk hierbij aan technische hulpmiddelen om vooringenomenheid te signaleren maar ook aan bijvoorbeeld een impact assessment als het IAMA¹⁴ om toepassingen mee te beoordelen. Op deze dimensie wordt over het algemeen erg laag gescoord. Het overzicht ontbreekt en beleid rond hoe deze tools structureel ingezet kunnen worden moet nog vorm krijgen.

Er is behoefte aan een inventarisatie en ontwikkeling van een overzicht van beschikbare tools en hoe deze te gebruiken binnen de sector.¹⁵ Verder zou een assessment kunnen worden uitgevoerd om te bepalen in hoeverre beschikbare tools kunnen bijdragen aan het verhogen van de volwassenheidsniveaus op andere dimensies.

4.2 Conclusie

De resultaten van de pilot tonen aan dat de volwassenheid van AI-ethiek binnen de deelnemende organisaties varieert. De gemiddelde scores laten zien dat de deelnemers nog aan het begin staan. Men scoort relatief hoog op cultuur en bewustzijn van ethische kwesties in vergelijking met andere dimensies. Dit bewustzijn moet echter worden omgezet in concrete acties, met name op het gebied van beleid, governance en training. Dit zou bijvoorbeeld kunnen door specifieke kaders voor AI-gebruik, inclusief generatieve AI, te ontwikkelen en te implementeren. Sectorale samenwerking kan een sleutelrol spelen in het ontwikkelen van standaardrichtlijnen en best practices zodat niet iedere organisatie zelf het wiel hoeft uit te vinden.

AANBEVELINGEN VOOR DE SECTOR NAAR AANLEIDING VAN DE ANALYSE

1. Ontwikkel een algemeen bewustwordingsplan met gerichte kennissessies per team of rol.
2. Introduceer vaste 'demo' of 'stand-up' sessies om het bewustzijn te vergroten.
3. Deel de uitkomsten van het volwassenheidsmodel en de pilotresultaten met het leiderschap (CvB).
4. Werk gezamenlijk aan het opstellen van specifieke kaders voor verantwoord AI-gebruik.
5. Ontwikkel een sectorbrede template of blauwdruk voor AI-beleid.
6. Onderzoek best practices en inventariseer bestaande beleidskaders.
7. Integreer AI-governance binnen bestaande werkwijzen.
8. Inventariseer welke kritieke afwegingen en checks nodig zijn en wie betrokken moet worden.
9. Deel kennis over hoe ethische checks op AI gebruik kunnen worden gekoppeld aan bestaande processen.
10. Formaliseer werkgroepen en deel inzichten om tot best practices te komen.
11. Integreer bestaande trainingen en verken sectoroverstijgende training.
12. Inventariseer de communicatiebehoefte binnen organisaties over AI-ethiek.
13. Deel en co-creëer aanpakken en vraagstukken binnen de sector voor externe validatie en verdere ontwikkeling.
14. Neem ethische aspecten mee in elke stap van het ontwikkelingsproces, ook bij inkoop en implementatie van AI-applicaties.
15. Besteed aandacht aan proces en transparantie voor weloverwogen beslissingen bij AI-inkoop en implementatie.
16. Inventariseer beschikbare tools om ethische aspecten van AI te identificeren en te beheren.
17. Ontwikkel een overzicht van beschikbare tools en hoe deze te gebruiken binnen de sector.
18. Voer een assessment uit om te bepalen hoe tools kunnen bijdragen aan het verhogen van de volwassenheidsniveaus.
19. Zet bewustwording om in concrete acties op het gebied van beleid, governance en training.
20. Werk samen aan het ontwikkelen van standaardrichtlijnen en best practices voor AI-gebruik.

¹⁴ Impact Assessment Mensenrechten en Algoritmen (IAMA) van de Rijksoverheid en Data School

<https://www.rijksoverheid.nl/documenten/rapporten/2021/02/25/impact-assessment-mensenrechten-en-algoritmes>

¹⁵ Een eerste overzicht van bestaande tools en methodes voor verantwoorde toepassing van AI, vind je hier: <https://edu.nl/cu9j4>

5 Reflectie

Het doel van de pilot was om inzicht te krijgen in de meerwaarde van het model, te leren van de praktijkervaringen, en te inspireren tot bredere toepassing. In welke mate is dit gelukt? Wat nemen we mee uit deze ervaring? In deze sectie reflecteren we kort op de gedeelde ervaringen van de deelnemers over de inzet van het AI ethiek volwassenheidsmodel én op welke manier een pilot helpt om hier mee aan de slag te gaan.

5.1 Ervaringen met het AI ethiek volwassenheidsmodel

Het AI ethiek volwassenheidsmodel werd door deelnemers als een nuttig instrument ervaren. Door met het model aan de slag te gaan werden instellingen geholpen met een goed begin, een gestructureerde aanpak, en een mogelijkheid om inzicht te krijgen in hun organisatie.

1. Het AI-ethiek volwassenheidsmodel werd door deelnemers als een nuttige eerste stap ervaren. Het hielp organisaties om in beweging te komen en hun huidige staat van ethische volwassenheid in kaart te brengen, waardoor ze meteen van start konden gaan.
2. Het model biedt houvast en een gestructureerde aanpak voor het benaderen van ethische vraagstukken. Het maakt het concreet en eenvoudig om verschillende mensen binnen de organisatie te betrekken, door hen te vragen naar hun perspectief op de ethische volwassenheid van AI. Daarnaast stimuleert het de toewijzing van eigenaarschap, wat zorgt voor meer betrokkenheid bij zowel bestuurders als medewerkers op verschillende niveaus. Het model kan ook dienen als een basis voor het voldoen aan de AI Act, omdat het helpt bij het vormgeven van compliance.
3. Het gebruik van het model biedt instellingen een waardevol inzicht door een snapshot te geven van de huidige situatie en verschillende perspectieven binnen de organisatie vast te leggen. Het maakt duidelijk waar de organisatie al bewust bezig is met AI-ethiek en waar nog verbeteringen nodig zijn. Dit helpt ook bij het formuleren van de ambitie.

Het doen van een pilot heeft ook als doel om te achterhalen waar het model minder goed verbindt met de uitdagingen bij de instellingen. Hier noemde men vooral de volgende zaken: hoe werkt een centrale aanpak in een gefragmenteerde organisatie, welke rol heeft inkoop binnen het model, de holistische aanpak beantwoordt geen ethische vragen, en de 3-5 jaren ambitie is lastig in te vullen.

1. De vaak gefragmenteerde organisatiestructuur van onderwijsinstellingen met een breed portfolio aan activiteiten maakt het lastig om een gecentraliseerd governance model toe te passen. Het bleek vaak niet gemakkelijk om naast de bedrijfsvoeringafdelingen ook te zorgen voor volledig overzicht van de primaire processen (onderwijs en onderzoek). Daarnaast werd duidelijk dat sommige beleidsfuncties, zoals privacy & security, wel centraal binnen instellingen zijn opgezet, terwijl ethische en waarde-gedreven initiatieven vaak lokaal georganiseerd zijn. Dit was bijvoorbeeld zichtbaar in de verschillen tussen afdelingen zoals onderzoek en bedrijfsvoering, waarbij ethische kwesties binnen onderzoek (niet alleen voor wat betreft gebruik van AI) vaak beter georganiseerd zijn dan binnen de bredere bedrijfsvoering.
2. De dimensie Ontwikkelprocessen bleek voor instellingen lastig om invulling aan te geven. De huidige beschrijving van het model geeft maar beperkt invulling aan inkoopprocessen, welke regelmatig voorkomen bij instellingen dan interne ontwikkeling van AI systemen (voor intern gebruik).
3. De holistische organisatie aanpak geeft geen antwoord op ethische vragen. Dat is een bewuste ontwerp keuze; maar dit betekent wel dat organisaties, naast de invulling van dit model binnen de instelling, zelf nog aan de slag moeten met de ethische vraagstukken. Daaronder hoort bijvoorbeeld het uitwerken van beleid en strategie op dit vlak.
4. Men vond het erg lastig om de 3-5 jaren ambitie binnen het model in te vullen en heeft dat dus ook beperkt gedaan. Naarmate het thema AI ethiek meer binnen de instelling geborgd is, zal dit waarschijnlijk eenvoudiger worden.

5.2 Ervaringen met het uitvoeren van de Pilot

De pilot heeft aangetoond dat het AI ethiek volwassenheidsmodel een effectieve katalysator is voor onderwijsinstellingen om prioriteit te geven aan AI-ethiek. Deelname aan de pilot forceerde deelnemers binnen de instelling het gesprek aan te gaan. Daarnaast benadrukte de pilot het belang van samenwerking en uitwisseling van inzichten tussen instellingen. Door ervaringen te delen, werden deelnemers niet alleen geïnspireerd door elkaars aanpak, maar kregen ze ook een beter inzicht in gemeenschappelijke uitdagingen.

Ondanks de positieve resultaten waren er ook enkele uitdagingen. De korte duur van de pilot, slechts elf weken, maakte het moeilijk om diepgaande analyses uit te voeren binnen de organisaties. Deze beperkte tijdspanne had echter ook een voordeel: deelnemers werden gedwongen zich te focussen en niet afgeleid te worden door een langdurig proces.

Een andere uitdaging was de variatie in aanpak onder de deelnemers. Sommige instellingen werkten met een centrale aanpak voor het invullen van het model, vaak door middel van workshops of plenaire sessies, terwijl anderen de betrokken medewerkers het model individueel lieten invullen. Deze variatie leidde tot enige inconsistentie in de diepgang en de volledigheid van de resultaten. Daarnaast bleek de kennis die nodig was om alle vragen in het model te beantwoorden vaak gefragmenteerd te zijn binnen de instellingen, wat de coördinatie bemoeilijkte (maar ook waardevolle inzichten opleverde bij het ophalen).

Eigenlijk kunnen we concluderen dat de pilot ervoor zorgde dat het gesprek plaats vond en het model ervoor zorgde dat dat een nuttig gesprek was.

6 Conclusie

Door de huidige technologische ontwikkelingen is het cruciaal om binnen de organisatie de juiste structuur te ontwikkelen voor het beantwoorden en borgen van ethische vraagstukken. Het is van belang om serieus aan de slag te gaan met het vormgeven van interne ethische governance bij onderwijs- en onderzoek instellingen. Dit is noodzakelijk omdat ethische vraagstukken steeds breder binnen de organisatie gevoerd moeten worden. Dit gaat verder dan de bestaande ethische commissies voor bijv. onderzoek.

Doormiddel van de pilot hebben we laten zien dat het AI ethiek volwassenheidsmodel hier aan bijdraagt. Maar ongeacht de methode is het belangrijk dat dit gestructureerd gebeurt – anders is AI ethiek niet te waarborgen.

6.1 Dit hebben we geleerd uit de pilot

De pilot heeft ons geholpen inzicht te krijgen in hoe onderwijsinstellingen omgaan met AI-ethiek. De pilot bood een startpunt voor gesprekken over ethische vraagstukken en heeft bij sommige deelnemers bijgedragen aan het verhogen van het bewustzijn en de betrokkenheid binnen de instelling. Het model bood een gestructureerde leidraad, waardoor het voor deelnemende instellingen ingezet kon worden als ‘good practice’ voor het vormgeven van hun eigen AI ethiek aanpak.

De ervaringen uit de pilot benadrukken de waarde van samenwerking tussen instellingen. De deelnemers vonden het prettig om gezamenlijk het proces te doorlopen, wat heeft geleid tot een beter begrip van gemeenschappelijke uitdagingen. Dit zien we ook terug in de behoefte om samen te blijven werken.

6.2 Aanbevelingen voor vervolg

Uit de ervaringen van de pilot komen nog de volgende aanbevelingen voor vervolgstappen:

Aanbevelingen voor het AI ethiek volwassenheidsmodel:

1. Aanvullen van het model met sector behoeftes: Het model kan worden aangepast om ook aan te sluiten bij de specifiekere vraagstukken van onderwijsinstellingen, bijvoorbeeld de ethische implicaties bij de inkoop van AI-systemen en de inzet van ingekochte AI-systemen.
2. Onderzoeken of het model een rol kan spelen in AI-act compliance: Onderwijsinstellingen moeten op korte termijn voldoen aan de AI-act, dit brengt vraagstukken met zich mee bij ‘hoog-risico’ toepassingen van AI in het onderwijs. Onderzoek en maak duidelijk of het AI ethiek volwassenheidsmodel hier ook invulling aan kan geven.

Aanbevelingen voor onderwijsinstellingen:

1. Het organisatie-perspectief is waardevol: Het is belangrijk om de holistische aard van het model te behouden, waarbij zowel onderzoek, onderwijs als bedrijfsvoering worden meegenomen. Dit zal instellingen helpen om een volledig beeld te krijgen van hun AI ethische volwassenheid.
2. Langetermijnvisie ontwikkelen: Instellingen moeten worden aangemoedigd om een realistische langetermijnvisie te ontwikkelen voor AI-ethiek, met duidelijke mijlpalen en doelen voor de komende jaren.
3. Zorg voor een ethische governance-structuur die bijdraagt aan het sturen en controleren van de organisatie op ethische vraagstukken: dit zorgt ervoor dat instellingen structureel in staat zijn om verantwoordelijkheid te nemen voor hun huidige en toekomstige applicatielandschap. De aanbevelingen geformuleerd op pagina 11 kunnen hier een eerste bijdrage aan doen.

Aanbevelingen voor organisatie overstijgende samenwerking:

1. Stimuleren van samenwerking en mogelijk vormen van een community: Alle deelnemers staan nog aan het begin. Daarom is het waardevol om samenwerking en uitwisseling van best practices tussen instellingen te bevorderen. Dit kan bijdragen aan de verdere ontwikkeling van sectorale standaarden en het versterken van de ethische AI-praktijk binnen het hoger onderwijs.
2. Proactief reageren op ontwikkelingen: Gezien de snelle ontwikkelingen is het belangrijk om vooruit te kijken en flexibel te kunnen reageren op nieuwe ontwikkelingen.
3. Blijvend op de agenda van thema op sectorniveau: we zien dat dit thema niet vanzelf blijvend op de agenda komt en dat er gebaad is bij doorlopende samenwerking. Daarom is het belangrijk dat sectorale organisaties

(bijv. koepelorganisaties, SURF, Npuls) dit thema blijvend op de agenda zetten – een voorbeeld hiervan is de werkagenda *Samen waardengericht onderweg* van UNL en SURF.¹⁶

¹⁶ https://www.surf.nl/files/2024-06/a5_boek_surf_samen-waardengericht-onderweg_def.pdf

7 Relevante links & activiteiten

7.1 Leeslijst

- Waardenwijzer (SURF & Kennisnet) <https://www.surf.nl/themas/publieke-waarden/waardenwijzer>
- Responsible Tech: On public values and emerging technologies (SURF, 2023) [responsible-tech_on-public-values-and-emerging-technologies_en_def.pdf](https://www.surf.nl/files/publications/responsible-tech_on-public-values-and-emerging-technologies_en_def.pdf) (surf.nl)
- The state of AI and the modern educational institution: AI explained in the context of the educational sector (Npuls, 2023) <https://npuls.nl/wp-content/uploads/2024/04/Npuls-Startnota-State-of-AI-B5-EN.pdf>
- SURF Sounds Responsible AI met Peter-Paul Verbeek en Duuk Baten <https://www.surf.nl/podcast/podcast-responsible-ai-in-onderwijs-en-onderzoek>
- Advies publieke waarden voor het onderwijs (UNL, 2021) <https://www.universiteitenvannederland.nl/files/publications/Advies%20werkgroep%20publieke%20waarden%20onderwijs.pdf>
- Werkagenda Samen waardengericht onderweg (SURF & UNL, 2024) <https://www.surf.nl/themas/publieke-waarden/werkagenda-samen-waardengericht-onderweg>

7.2 Relevante sector activiteiten binnen SURF en Npuls

Lopende ontwikkelingen

- Project Onderzoek en Ontwerp Algoritmeregister (Npuls): <https://community-data-ai.npuls.nl/groups/view/fe114dfb-4eba-4164-950e-5e33f2d7b467/project-onderzoek-en-ontwerp-algoritmeregister>
- Project Referentiekader Privacy en Ethiek 2.0 (Npuls): <https://community-data-ai.npuls.nl/groups/view/337ee9fb-b9dc-4ee1-b4d5-33c1b5a881d6/project-referentiekader-privacy-en-ethiek-20>
- Innovatiezone gemeenschappelijke digitale soevereiniteit met daarin de roadmap AI (SURF): <https://www.surf.nl/nieuws/gemeenschappelijke-digitale-soevereiniteit-speerpunt-cooperatie-surf>

Communities

- Community platform van de pilothub studiedata en AI: <https://community-data-ai.npuls.nl/>
- SURF community AI in Education en SIG AI in Education: <https://communities.surf.nl/ai-in-education>
- SURF community Publieke Waarden: <https://communities.surf.nl/publieke-waarden>
- Responsible Tech discussie mailing list: <https://list.surfnet.nl/mailman/listinfo/ai-act-discussie> (schrijf je in met een instellings e-mail-adres)
- AI act discussie mailing list: <https://list.surfnet.nl/mailman/listinfo/tech-policy> (schrijf je in met een instellings e-mail-adres)
- Voor de security en privacy community kan je terecht bij SCIPR: <https://www.surf.nl/scipr-samenwerken-aan-betere-informatiebeveiliging-en-privacy>

Bijlage 1 AI ethiek volwassenheidsmodel

AI Ethical Maturity - Level Overview

Dimension	Level 1	Level 2	Level 3	Level 4	Level 5
Awareness & Culture	Awareness of data on an individual level out of personal interest	Fragmented attention throughout the organization	Focused and synthesised awareness through the formation of specific working groups or task forces	Organization wide support and representative multidisciplinary working groups	Buy-in from senior, middle and junior management, broad support and active involvement of developers, business and management
Policy	Minimal to no policy available for warranting ethics in data science	There is a demand for policy. Conversations have started and there is a first concept on the policy	Policy for ethical data science is available. A person assigned for the implementation and monitoring of the policy aspects	Policy is implemented in most parts of the organization. A central point is initiated for questions, monitoring, and feedback	Policy on data science ethics is widely implemented and monitored throughout the organization
Governance	Only legally mandatory checks	Additional robustness and model validation checks, not formally required	Specific ethical checks in the design phase or post hoc, not formally required	Formally required ethical checks throughout data science lifecycle, governance committees are appointed	Fully integrated and supported AI ethics governance structure with formally required checks, procedures, and operating governance committees
Communication & Training	Minimal to no communication; employees improve their understanding based on own initiatives	Initiative for training and communication only in small teams involved in data science processes	Incorporation of training and communication not only inside data science teams but also key stakeholders (e.g. C-suite) in line with established ethical framework	Company-wide sessions as well as the regular training of core team members. Communication about the ethical aspects is becoming a part of the daily tasks and activities	Communication happens outside of the company to customers and citizens. There is a fully developed training module that includes a schedule for regular training for different types of users in the organization
Development Processes	No structural approach to data science, or ethics in the lifecycle phases	Initiative for a structured data science approach mainly focusing on technical design choices in the development process	Relatively structured data science approach with ethical design choices were requested (on demand)	Structured approach, with alignment of ethical data science aspect to different phases in the data science lifecycle	Integration in the entire data science workflow where specific activities are implemented in and aligned with distinct lifecycle phases
Tooling	No or minimal tooling is used	There is demand for insights into the ethical aspects of data science. First ideas are gathered and translated into possible analysis/tooling	First methods and tools for generating insights into the ethical aspects are implemented and adopted	Tooling is available for and adopted by multiple stakeholders in the organisation to monitor, discuss, and improve ethical data science aspects	Wide adoption of tooling where both internal and external stakeholder are using the available tooling to proactively monitor, discuss, and improve ethical data science aspects

Figuur 6 Het AI ethiek volwassenheidsmodel

Colofon

Neem contact met ons op via responsible-tech@surf.nl

Auteurs

Duuk Baten, SURF

Annalies Kootstra, Hanze

Joris Krijger, Ethische Data Science Associatie (EDSA)

Heleen van der Laan, ROC Amsterdam-Flevoland

Tina van der Linden, Hogeschool Utrecht

Met dank voor de input van

Tamara Thuis, Ethische Data Science Associatie (EDSA)

Bo Schijven, Erasmus Universiteit Rotterdam

Sanderijn Kuijvenhoven, SURF

Alle deelnemers aan de verschillende instellingen

Visualisaties

Voorpagina: SURF

Figuur 1: SURF

Figuur 2: EDSA

Figuur 3: SURF

Figuur 4: EDSA & EUR

Figuur 5: EDSA

Figuur 6: EDSA & EUR

Gewenste aanhaling

SURF (2024). Reflectierapport: Pilot AI ethiek volwassenheidsmodel. Utrecht. (Auteurs: Baten, D., Kootstra, A., Krijger, J., van der Laan, H., van der Linden, T.) <https://doi.org/10.5281/zenodo.14040030>

Deze discussienota is een publicatie van SURF

14 november 2024



Deze publicatie is gelicentieerd onder
Creative Commons Attribution 4.0 International.