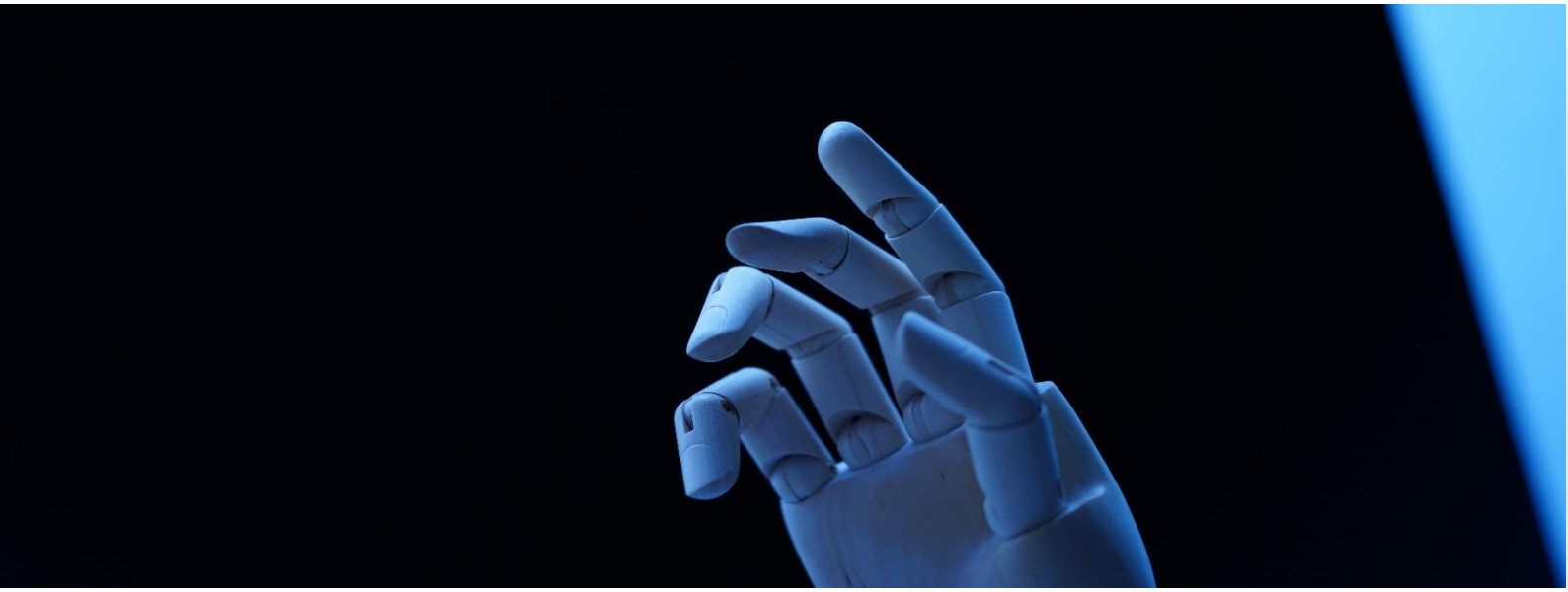
Rathenau Instituut

Zo brengen we AI in de praktijk vanuit Europese waarden



Hoeveel ruimte geven we de techniek en de bedrijven achter kunstmatige intelligentie (AI)? Voor die vraag is steeds meer aandacht. De afgelopen maanden publiceerden allerlei organisaties hun eigen ethische codes voor AI. Nu is het tijd voor gezamenlijke regels en wetgeving. Dit vraagt om technologische, sociale en juridische innovatie, met respect voor Europese waarden.

Door Roos de Jong, Linda Kool en Rinie van Est

Inhoud

Introductie

Dit artikel is geschreven in aanloop naar de Nederland Digitaal Dag op 21 maart 2019. Tijdens die digitale top spreken diverse partijen uit het bedrijfsleven, wetenschap, overheid en maatschappelijke organisaties over de verdere ontwikkeling van de Nationale Digitaliseringsstrategie en dus over de vraag hoe de digitale transitie kan worden vormgegeven.

Begin 2017 vroeg het Rathenau Instituut met het rapport Opwaarderen aandacht voor een groot scala maatschappelijke en ethische kwesties die door digitalisering op ons afkomen. Inmiddels groeit het bewustzijn voor deze kwesties mede door de huidige belangstelling voor AI en ethiek. Ook voor de digitale top is AI het overkoepelende thema.

In dit artikel plaatsen we AI in de bredere digitale transitie. We spreken liever van AI-systemen. AI-systemen zijn geautomatiseerde beslissystemen en brengen zodoende tal van politieke vragen met zich mee, zoals: wie bezit kennis om besluiten te nemen en wie bepaalt wie de kennis bezit?

Inmiddels wordt breed erkend dat AI diep kan ingrijpen op ons mens-zijn. De vele ethische codes die afgelopen twee jaar door diverse partijen zijn ontwikkeld, hebben hieraan bijgedragen. Na dit cruciale proces van ‘waarderen,’ waarbij doelen, grenzen en randvoorwaarden zijn verhelderd, is het nu tijd om richting *gezamenlijke* regels en wetgeving te bewegen voor de ontwikkeling en het gebruik van AI.

We geven daarvoor drie handvatten:
- neem standaarden en wetgeving als uitgangspunt;
- zoek nadrukkelijker de verbinding op tussen innovatiebeleid en waarden, ook in de nationale digitaliseringsstrategie; en
- zet het waardenperspectief centraal bij experimenten in de praktijk.

1. Wat is AI?

Elke dag tonen de media nieuwe mogelijkheden van AI. Denk aan het nauwkeurig vaststellen van kanker, grootschalige testen met zelfrijdende auto’s, en drones die op een dag wellicht onze postpakketten bezorgen. We worden ook geconfronteerd met de negatieve kanten: ongrijpbare algoritmen spelen een rol bij de verspreiding van desinformatie, slimme softwaresystemen blijken te kunnen discrimineren en slimme apparaten in ons huis zijn eenvoudig te hacken.

We associëren AI dikwijls met futuristische toepassingen of met science fiction. Maar veel AI-toepassingen gebruiken we al lange tijd. Denk aan spamfilters in je mailbox, online zoekmachines en aanbevelingen die we krijgen bij Bol.com of Netflix. Die toepassingen zijn volledig geïntegreerd in ons dagelijks leven en mede mogelijk gemaakt door AI.

1.1. Een korte geschiedenis van AI

AI is niet nieuw. In de jaren ’50 verkende een generatie wetenschappers, wiskundigen en filosofen het concept van kunstmatige intelligentie. Ook zijn de basale AI-technieken die gebruikt worden binnen **machine learning** en **deep learning** (zie kader ‘De ene AI is de andere niet’) al lange tijd bekend. Door de toegenomen rekenkracht van computers en de grote hoeveelheden data zijn machine learning en deep learning de laatste twee decennia echt in een stroomversnelling geraakt. Inmiddels kunnen AI-

systemen in zeer specifieke vaardigheden beter zijn dan mensen. Zo versloeg in 1997 IBM's schaakcomputer DeepBlue een van 's werelds beste schakers, Gary Kasparov, en in 2017 won Google's AlphaGo van kampioen Go-speler Ke Jie.

AI is deel van een systeem

AI is niet één technologie, maar is beter te begrijpen als een 'cybernetisch systeem' dat waarneemt, analyseert (denkt) en handelt, en daarvan kan leren. Diverse technologieën (sensoren, big data, robotlichamen, of een app) zorgen er samen voor dat een systeem of machine in een juiste omgeving een zekere mate van intelligent gedrag vertoont. Het is de combinatie van alomtegenwoordige, met het internet verbonden apparaten en sensoren. Die combinatie is in meer of mindere mate in staat zelfstandig acties uit te voeren.

AI valt daarom niet los te zien van andere digitale technologieën, zoals robotica, Internet of Things, digitale platformen, biometrie, virtual en augmented reality, persuasieve technologie en big data. AI is te zien als het 'brein' achter diverse 'slimme' toepassingen:

- Een zelfsturende auto is bijvoorbeeld een combinatie van Internet of Things en AI.
- Netflix-aanbevelingen zijn een combinatie van big data, AI en een platform.

In de rapporten van het Rathenau Instituut over de digitale samenleving spreken we daarom over intelligente machines en digitalisering, om te benadrukken dat de vraagstukken met dit brede cluster van technologieën te maken hebben. In diverse rapporten bespreken we de betekenis van toepassingen als zelfsturende auto's, drones, slimme zorgapplicaties en platformen.

De ene AI is de andere niet

In de media begint een discussie over AI vaak met toekomstvisies over situaties waarin AI de menselijke intelligentie overstijgt. Dat wordt ook wel **‘technologische singulariteit’** genoemd. Maar zo ver is het nog lang niet. AlphaGo van Google heeft bewezen beter te zijn in het spel Go dan mensen, maar kan niet opeens een ander spelletje met je spelen. Ook kan de AI die huidkanker beter kan herkennen dan ervaren dermatologen, niet ingezet worden voor een andere taak. Dit laat zien dat de meeste AI is gespecialiseerd in één specifieke vaardigheid.

In 2018 kwam AlphaZero. Dat systeem kan meerdere spellen heel goed spelen. In tegenstelling tot AlphaGo, kan AlphaZero zowel met schaken, Go als Shogi winnen van professionele menselijke tegenstanders. Een essentiëler verschil tussen AlphaGo en AlphaZero gaat over de manieren waarop ze leren. AlphaGo is getraind op basis van de spellen gespeeld door mensen, terwijl AlphaZero leerde door tegen zichzelf te spelen.

Tegenwoordig gaat het vaak over het verschil tussen **‘rule-based AI’** en **‘machine learning’**. Rule-based AI is gebaseerd op geprogrammeerde ‘als dit, dan dat’-instructies. Een dergelijk systeem leert niet van zichzelf, maar vertoont wel intelligent gedrag door de omgeving te analyseren en acties te ondernemen – met enige mate van autonomie – om specifieke doelen te bereiken zie definitie van AI van de Europese Commissie).

Machine learning is gebaseerd op het detecteren en leren van patronen in data. Het gaat om de ontwikkeling van software die de eigen prestaties verbetert, en leunt sterk op statistische wetenschap**. Deep learning** is een vorm van machine learning, gebaseerd op neurale netwerken (specifieker op de biologie van ons brein). Hierbij draait het om het combineren van gewichten met input om de input te classificeren en te clusteren.

Hoewel de publieke en politieke discussie zich momenteel voornamelijk concentreert op vormen van machine learning, is het van belang om te begrijpen dat machine learning slechts één type AI is. Machine learning en deep learning vervangen niet andere typen AI; vaak gaat het binnen een AI-systeem om een combinatie van meerdere AI-technieken, zoals beslisbomen of logisch redeneren.

1.2. Slimme of domme AI: geautomatiseerde besluitvorming

AI-systemen zijn vaak goed in een specifieke vaardigheid. Maar in bepaalde situaties levert de software geen betrouwbare resultaten en is ze gemakkelijk te foppen. Door slechts een paar pixels te veranderen of een zichtbaar verstorend patroon over een afbeelding te ‘plakken’, kan een slim systeem tot een dom antwoord komen. Zo kan automatische beeldherkenning door wat ruis een panda ineens betitelen als een mensaap (zie voor een mooi overzicht overzicht dit artikel in de Volkskrant).

Wat voor mensen simpel lijkt, kan voor een AI-systeem heel lastig zijn. Dat een zes-jarig kind een panda niet zal aanzien voor een mensaap en diepte en schaduwen begrijpt op tekeningen in 2D, is het resultaat van een complex brein dat zich miljoenen jaren heeft geëvolueerd.

Diverse slimme systemen zijn afhankelijk van de data die zij als input krijgen. Als die fouten bevatten, gaat het systeem ook de mist in. Zo moest Microsoft zijn slimme chatbot Taj binnen 24 uur stopzetten toen mensen haar racistische uitingen voerden en Taj zelf racistische en opruiende tweets ging plaatsen. Vooringenomenheid van data blijkt een lastig te tackelen probleem. Zo moest Amazon een systeem voor het automatisch selecteren van sollicitanten stoppen, omdat het vrouwen onderaan de stapel legde. Het systeem baseerde zich op historische data (in de ICT-industrie werken meer mannen dan vrouwen), maar het was niet de bedoeling dat het systeem dat als selectiecriterium zou nemen.

1.3. Wie is verantwoordelijk voor beslissingen door AI?

Hoe slim of dom een AI-systeem ook precies is, ondertussen zetten overheden, bedrijven en andere organisaties AI-systemen in voor allerlei doeleinden: autorijden, ondersteuning van de arts, kredietbepaling, opsporing van fraude en voor oorlogsdoeleinden. Een belangrijk deel van de discussie spitst zich daarom toe op de beslissingen die deze systemen nemen: AI-systemen zijn beslissystemen. Dat roept de volgende vragen op:

Hoe komen de systemen tot een advies of beslissing? Is dat te controleren, bijvoorbeeld door toezichthouders, of voor degene op wie het besluit betrekking heeft? In hoeverre is er een mens bij het besluit betrokken? En wie is verantwoordelijk voor de genomen beslissing? Heb je straks als arts wat uit te leggen als je de aanbevelingen van het systeem niet heb opgevolgd? Of als je als automobilist de aanwijzingen van je slimme auto niet hebt opgevolgd?

We maken graag dankbaar gebruik van het gemak dat digitale diensten (met behulp van AI) bieden. Maar achter die diensten gaat meer schuil . De verzameling en exploitatie van data over ons gedrag zijn alomtegenwoordig, en worden ingezet ter beïnvloeding van ons gedrag. Harvard-professor Shoshana Zuboff bestudeert de

bedrijfsmodellen en strategieën van grote technologiebedrijven. In haar boek *The Age of Surveillance Capitalism* (2019) stelt zij drie cruciale vragen:

1. Wie heeft de kennis?
2. Wie beslist wie de kennis heeft?
3. Wie beslist over wie er mag beslissen wie de kennis heeft?

In de volgende paragraaf gaan we nader in op de maatschappelijke en ethische kwesties rondom AI en automatische besluitvorming. In paragraaf 3 gaan we dieper in op de hierbovengenoemde machtsvragen.

2. Van individuele ethische codes naar Europese kaders

In het rapport Opwaarderen laten we zien dat digitalisering gepaard gaat met een groot scala aan maatschappelijke en ethische kwesties (zie tabel 1). Naast privacy en veiligheid, spelen ook kwesties zoals controle over technologie, rechtvaardigheid, menselijke waardigheid en ongelijke machtsverhoudingen. In het rapport Human rights in the robot age beschreven we hoe mensenrechten als non-discriminatie en recht op eerlijk proces door digitalisering (onder andere door AI) onder druk kunnen komen te staan.

Thema	Maatschappelijke en ethische vraagstukken
Privacy	Gegevensbescherming, digitaal huisrecht, mentale privacy, surveillance, doelverschuiving
Autonomie	Keuzevrijheid, vrijheid van meningsuiting, manipulatie (verspreiding van desinformatie, *microtargeting*), bescherming democratie, paternalisme, vaardigheden, grenzen zelfredzaamheid
Veiligheid	Informatieveiligheid, identiteitsfraude, fysieke veiligheid
Controle over technologie	Controle over en inzicht in algoritmen, verantwoordelijkheid, voorspelbaarheid
Menselijke waardigheid	Dehumanisatie, instrumentalisering, de-skilling, de-socialisatie, werkloosheid
Rechtvaardigheid	Discriminatie, uitsluiting, gelijke behandeling, stigmatisering
Machtsverhoudingen	Oneerlijke concurrentie, uitbuiting, relatie consument-bedrijf, relatie bedrijf-platform

De afgelopen twee jaar is het maatschappelijke bewustzijn voor deze kwesties enorm toegenomen. In het rapport Doelgericht digitaliseren geven we een gedetailleerd overzicht van initiatieven van diverse partijen. De discussie rondom AI speelt daarbij een aanjagende rol.

Binnen de wetenschap, het bedrijfsleven en de overheid zijn er veel initiatieven geweest om vanuit ethisch perspectief naar AI te kijken. Dit heeft geleidt tot een breed scala aan ethische codes. Hieronder zullen we enkele daarvan bespreken.

- In januari 2017 organiseerde het Future of Life instituut (bestaande uit wetenschappers en technologiebedrijven en gericht op het voorkomen of verminderen van risico's van AI) de Asilomar Conference on Beneficial AI. Deze conferentie leverde een lijst principes op waaraan onderzoekers uit de academische wereld en het bedrijfsleven zich willen committeren. De 23 principes hebben betrekking op onder andere onderzoeksstrategieën: bijvoorbeeld dat het doel van AI-onderzoek is om nuttige en 'niet ongerichte' intelligentie te ontwikkelen. Met betrekking tot de doelen en gedragingen van AI-systemen stelt men dat deze verenigbaar moeten zijn met idealen van menselijke waardigheid, gegevensrechten, vrijheden en culturele diversiteit. Daarnaast moet een wedloop in dodelijke autonome wapens vermeden worden.

- Vanuit beroepsverenigingen, bijvoorbeeld IEEE, en technologiebedrijven als Google en Telefónica, verschenen in 2018 verklaringen en ethische richtlijnen voor AI. Google noemt als belangrijke kwesties bijvoorbeeld uitlegbaarheid, eerlijkheid, veiligheid, mens-machine-samenwerking en aansprakelijkheid. Volgens de codes moet AI menselijk welzijn bevorderen. Ook benoemen de codes ambities met betrekking tot 'inclusieve representatie', eerlijke behandeling, betrouwbaarheid, uitlegbaarheid, veiligheid, verantwoordelijkheid en aansprakelijkheid.

- In Nederland bracht het platform voor de informatiesamenleving van bedrijfsleven, overheid en maatschappelijke organisaties, ECP, eind 2018 een AI Impact Assessment tool uit. Het neemt negen waarden als uitgangspunt, gebaseerd op de Europese adviesgroep in hun advies over AI, robotica en autonome systemen (zie onder).

- Ook binnen het overheidsdomein was er aandacht voor ethiek en AI. In Europa bracht de ethische adviesgroep (European Group on Ethics in Science and New Technologies, EGE) in rapport 'Statement on Artificial Intelligence, Robotics and Autonomous Systems' uit. Daarin formuleert de adviesgroep een aantal fundamentele ethische principes en democratische vereisten. Deze zijn gebaseerd op de waarden die zijn vastgelegd in Europese verdragen en het Handvest van de grondrechten van de Europese Unie: menselijke waardigheid, autonomie, verantwoordelijkheid, rechtvaardigheid, gelijke toegang en solidariteit, democratie, rechtsstaat, verantwoording en aansprakelijkheid, veiligheid, lichamelijke en mentale integriteit, bescherming van data en privacy, en duurzaamheid.

- In juni 2018 werd de Europese High-Level Expert Group on AI (de AI HLEG) opgericht – bestaande uit academici, en actoren uit het bedrijfsleven en maatschappelijk middenveld – om de Europese Commissie te ondersteunen bij het opstellen van een AI-strategie. De EU wil wereldleider worden in het ontwikkelen

en toepassen van verantwoorde, ‘human-centric’ AI. In december 2018 publiceerde de AI HLEG een eerste concept met ‘Ethische richtsnoeren voor betrouwbare kunstmatige intelligentie’; in april volgt naar verwachting een definitieve versie.

Voor betrouwbare AI moeten volgens de AI HLEG twee voorwaarden vervuld worden:

Ten eerste moeten de grondrechten, de toepasselijke regelgeving en fundamentele beginselen en waarden die een ‘ethisch doel’ waarborgen, worden gerespecteerd. Om ervoor te zorgen dat de mens centraal staat, moeten de mogelijke effecten van AI op het individu en de gemeenschap getoetst worden aan maatschappelijke normen en waarden en vijf ethische beginselen: beneficence (weldoen), non maleficence (niet schaden), de autonome wil van de mens, rechtvaardigheid en explicability (verantwoording).

Ten tweede moet de technologie robuust en betrouwbaar zijn. Daarbij gaat het om zaken als verantwoordingsplicht, data governance, design for all, beheer van AI-autonomie (menselijk toezicht), non-discriminatie, eerbiediging van de autonome wil van de mens, respect voor de privacy, robuustheid, veiligheid en transparantie.

2.1. Consensus over AI en Europese waarden

De ethische codes die we hebben onderzocht laten een overlap zien in kwesties en waarden die kunnen worden geraakt door AI. Daarbij vormen centrale Europese waarden, zoals vastgesteld in mensenrechtenverdragen en het Europese Handvest van de Europese Commissie, de kern.

We zien ook dat het brede spectrum aan kwesties in de AI-discussie sterk overeenkomt met de kwesties die in het rapport Opwaarderen zijn gesignaleerd. Dit laat zien dat de discussie over AI inderdaad heeft bijgedragen aan een verbreding van het maatschappelijke debat over digitalisering.

De hoogleraren ethiek Floridi, Dignum e.a. stellen dat veel geformuleerde principes rondom het gebruik van AI sterk overeenstemmen met belangrijke bio-ethische principes, zoals weldoen, niet-schaden, autonomie en rechtvaardigheid. Bio-ethiek gaat over technologieën die ingrijpen in het lichaam en brein van de mens. In de rapporten Intieme technologie en Human rights in the robot age liet het Rathenau Instituut zien dat digitale technologie evengoed sterk ingrijpt in ons mens-zijn: ons denken, sociale leven, en handelen. De discussie over AI en ethiek toont dat vele partijen nu erkennen dat AI op een fundamentele manier in ons leven kan ingrijpen.

De erkenning dat de medische technologie allerlei bio-ethische kwesties met zich meebrengt, heeft grote invloed (gehad) op de manier waarop de samenleving daarmee omgaat, en op de rol van regulering. Op eenzelfde manier zal het brede besef dat AI,

en digitalisering in het algemeen, tal van bio-ethische kwesties oproepen grote gevolgen hebben voor hoe we vorm geven aan het beleid en aan het toezicht op de ontwikkeling en het gebruik van ICT-systemen. Ethische codes alleen zullen daarbij niet genoeg zijn; er zal ook wetgeving nodig zijn om AI daadwerkelijk ‘human-centric’ en betrouwbaar te maken.

3. Opkomst geopolitieke discussie

Ons onderzoek naar de diverse ethische codes laat zien dat een groot maatschappelijk vraagstuk veelal niet benoemd wordt: vragen over ongelijke machtsverhoudingen. Zoals we hierboven al aangaven, gaat AI over het nemen van beslissingen: wie neemt de beslissing en wie besluit wie de beslissing mag nemen? In de laatste anderhalf jaar zijn diverse machtsvragen als nieuwe dimensie van het AI-debat opgekomen.

3.1. AI-wedloop: techno-economisch, socio-politiek en militair

Velen zien AI als een algemene technologie die vele nieuwe producten en diensten mogelijk maakt. Als een technologie die invloed zal hebben op domeinen als mobiliteit, zorg, opsporing, dienstverlening, energie, onderwijs, arbeid, landbouw en rechtspraak. AI wordt daarom ook wel gezien als de motor van de economische groei, technologische innovatie en de sleutel voor militaire technologie.

Op economisch vlak spelen dilemma’s van economische en technologische (on)afhankelijkheid. Door winner-takes-all-dynamieken hebben bedrijven in China en de VS toegang tot grote hoeveelheden gegevens. Ze gebruiken deze om hun algoritmen en producten en diensten beter te maken. Zo krijgen die bedrijven een steeds groter marktaandeel. Enkele grote technologiebedrijven herpositioneren zich als AI-bedrijven. Denk aan Facebook, Apple, Amazon, Netflix en Google in de VS en aan Baidu, Alibaba en Tencent in China. Zij domineren de AI-markt en hebben dikwijls een groter R&D-budget dan landen als geheel (totaal van publieke en private investeringen in R&D). Amazon spendeerde in 2017 bijvoorbeeld 22,6 miljard dollar aan R&D, en Alphabet 16,6 miljard dollar, aldus Bloomberg.

Daarnaast spelen socio-politieke vragen en implicaties voor de internationale orde, aldus de Verenigde Naties. In China draagt AI bij aan het beoordelen en bijsturen van gedrag van burgers. In liberale democratieën is discussie ontstaan over de rol van AI en op de invloed van het publieke en politieke debat, bijvoorbeeld via politieke micro-targeting en de verspreiding van desinformatie. De internationale gemeenschap spreekt daarom over de bescherming van fundamentele mensenrechten en de soevereiniteit van het internet: welke rol heeft de staat bij het beschermen van het internet tegen buitenlandse invloeden?

Tot slot is er aandacht voor de implicaties voor defensie en militaire machtsverhoudingen. Sommige deskundigen waarschuwen dat de internationale consequenties van AI vergelijkbaar zijn met de impact van de ontwikkeling van nucleaire wapens in de vorige eeuw. In de Asilomar Conference on Beneficial AI hebben diverse wetenschappers en bedrijven afgesproken om een wapenwedloop in autonome wapensystemen te voorkomen, maar er is nog geen internationaal verbod. Mocht één staat besluiten dodelijke autonome wapens op te nemen in hun strijdkrachten, dan zullen andere staten zich mogelijk gedwongen voelen om te volgen. Ook met AI vervalste video- of audiofragmenten (zogenoemde deep fakes, zie dit voorbeeld) kunnen internationale verhoudingen op scherp zetten.

3.2. Nationale AI-strategieën

Het belang van deze machtsvragen wordt nu breed ingezien. Diverse landen publiceren daarom nationale AI-strategieën. Vaak uiten zij daarin de ambitie om wereldleider in AI te blijven of te worden, zoals VS en China (zie de kaders hieronder). Ook diverse Europese landen kwamen met een dergelijke strategie, waaronder Frankrijk in maart 2018. De Franse president Emmanuel Macron benoemt in zijn strategie Europese waarden als een uitgangspunt.

De Europese Commissie (EC) volgde in april 2018. De EU stelt Europese waarden voorop bij het innoveren en gebruik van AI, en kiest zo haar eigen benadering. De EC moedigt elke lidstaat aan zelf een AI-strategie te ontwikkelen. In Nederland is het ‘strategisch actieplan AI’ in voorbereiding. Dit wordt voor de zomer verwacht.

China – AI ingezet voor bestuurlijke doelen

In juli 2017 maakte China met het ‘New Generation Artificial Intelligence Development Plan’ de ambitie bekend om in 2030 de leidende AI-macht te zijn. In deze nationale AI-strategie maakt China een nationale prioriteit van de ontwikkeling van de AI-sector. Volgens die strategie moeten AI-technologieën en -toepassingen ‘made in China’ tegen 2020 vergelijkbaar zijn met AI uit de rest van de wereld. Hetzelfde geldt voor Chinese bedrijven en onderzoeksfaciliteiten. Vijf jaar later moeten doorbraken in specifieke AI-disciplines leiden tot een koploperspositie. In de laatste fase rekent China erop ’s werelds belangrijkste AI-innovatiecentrum te zijn.

Verenigde Staten – AI gestuurd door de private sector

Ook het Witte Huis heeft Amerikaans leiderschap in AI tot prioriteit gemaakt. Zo bracht de regering-Obama verschillende invloedrijke AI-rapporten uit, zoals 'Preparing for the Future of Artificial Intelligence, National Artificial Intelligence Research and Development Strategic Plan' en 'Artificial Intelligence, Automation, and the Economy'. De regering-Trump maakte AI een R&D-prioriteit, met veel ruimte voor Amerikaanse bedrijven. In mei 2018 kwamen vertegenwoordigers uit de industrie, de academische wereld en de regering bijeen voor een top over AI. Hieruit komen vier centrale doelen: (1) het nationale R&D-ecosysteem versterken, (2) Amerikaanse werknemers ondersteunen om ten volle te profiteren van de voordelen van AI, (3) barrières voor AI-innovaties wegnemen, en (4) hoogwaardige, sectorspecifieke AI-toepassingen mogelijk maken. In februari 2019 kwam Trump met een 'executive order' om om Amerika's positie als wereldleider te behouden.

Europa – betrouwbare AI waarbij de mens centraal staat

In Europa zijn er diverse landen met een nationale AI-strategie. In april 2018 ondertekenden 25 EU-lidstaten een verklaring om samen te gaan werken aan de ontwikkeling van AI. Hierop volgde de mededeling 'Kunstmatige Intelligentie voor Europa' van de Europese Commissie waarin een Europese AI-aanpak uiteen wordt gezet. De voorgestelde aanpak is drieledig: (1) investeren in onderzoek en innovatie om de technologische en industriële capaciteit van de EU te vergroten en het gebruik van AI op te voeren in de hele economie, (2) onderwijs moderniseren en anticiperen op verschuivingen op de arbeidsmarkt en sociaal-economische veranderingen, (3) zorgen voor een passend ethisch en juridisch kader.

Met dit derde punt onderscheidt Europa zich van de VS en China. In december 2018 verscheen het gecoördineerde AI-plan met als doel wereldleider te worden op het gebied van verantwoorde ontwikkeling en toepassing van AI. De Europese Commissie zet in op AI waarbij de mens centraal staat, gebaseerd op Europese normen en waarden. Behalve een definitieve versie van de eerder genoemde 'ethische richtsnoeren voor AI' wordt dit jaar ook nog een richtlijn over productaansprakelijkheid uitgebracht. Verschillende websites bieden een mooi overzicht van nationale AI-strategieën; zie bijvoorbeeld de webpagina's van de OESO en het Future of Life Insititute, en dit artikel op Medium.

3.3. Rechtsstaat versus machtige bedrijven

De AI-wedloop lijkt vooral betrekking te hebben op landen als de Verenigde Staten en China. De EU plaatst zichzelf op het wereldtoneel door zich te richten op een waardengedreven benadering. Tegelijkertijd zijn er zorgen over de macht van landen, gezien de groeiende macht van grote technologiebedrijven. Er is een unieke concentratie van macht ontstaan bij tech-giganten. Dat komt, schrijft strategisch adviseur Paul Nemitz van de Europese Commissie, doordat die grote investeringen kunnen doen en steeds meer controle hebben over de infrastructuren van het publieke discours. Ook kunnen ze persoonlijke gegevens verzamelen; profileren; en de ontwikkeling van AI-diensten domineren.

Technologiebedrijven spelen met het formuleren van ethische codes strategisch in op de talrijke ethische en maatschappelijke kwesties die in de maatschappij leven. Dat is belangrijk, maar lost niet alle problemen op. Partijen werken ten eerste samen in een keten: wiens code is dan eigenlijk van toepassing? Wie overziet de diverse codes en stemt af bij verschillen? Er is dus ook gezamenlijke actie nodig. Partijen als de Verenigde Naties en ISO werken daarom aan collectieve standaarden.

Ten tweede zijn ethische codes *aanvullend* op bestaande wetgeving. Maatschappelijke en ethische kwesties kunnen niet alleen beslecht worden met het opstellen van codes; bedrijven moeten zich ook verhouden tot bestaande wetgeving.

In de volgende paragraaf gaan we nader in op hoe partijen vanuit Europese waarden met AI kunnen innoveren.

4. Naar waardengedreven innovatie

Aangezien AI raakt aan vele publieke waarden, is de uitdaging om vanuit gedeelde publieke waarden gericht vorm te geven aan innovatie. Het Rathenau Instituut heeft de afgelopen jaren veel onderzoek gedaan naar diverse aspecten van waardengedreven innovatie (zie de rapporten Waardevol Digitaliseren, Bedrijf zoekt universiteit, Living labs in Nederland en Gezondheid Centraal). Daarin maken we duidelijk dat het bij innovatiebeleid niet uitsluitend gaat om het ontwikkelen van nieuwe technologische toepassingen, maar ook om het doel dat die technologische toepassingen dienen: de maatschappelijke opgave.

Daarom is er bij waardengedreven innovatie aandacht voor de ontwikkeling van passende verdienmodellen, passende wet- en regelgeving en maatschappelijke inbedding van nieuwe toepassingen. Deze manier van innoveren erkent in een vroeg stadium de complexiteit van innovatie. We bespreken drie manieren waarop innovatie vanuit waarden kan worden vormgegeven.
Neem wetgeving als uitgangspunt

De lawine aan ethische codes wekt mogelijk de indruk dat de ontwikkeling van AI in een wettelijk vacuüm plaatsvindt. Dat is uiteraard niet zo. Er is veel bestaande wetgeving waar ook de ontwikkeling en het gebruik van AI zich toe moet verhouden. Het gaat daarbij om fundamentele rechten zoals grondrechten, maar ook om specifieke wetgeving, zoals de Algemene Verordening Persoonsgegevens, of sectorspecifieke wetgeving in domeinen als de zorg of de vervoersmarkt.

De recente geschiedenis van digitalisering laat zien dat diverse IT-platformen weinig respect tonen voor de wet. Door bestaande wetgeving als verouderd weg te zetten, proberen technologiebedrijven diverse wettelijke verantwoordelijkheden te ontlopen. Dat leidt tot onduidelijkheid over rechten, plichten en verantwoordelijkheden (zie ook ons rapport Eerlijk delen).

In diverse rechtszaken scheppen rechters nu duidelijkheid, bijvoorbeeld het arrest van het Europese Hof van Justitie over de verantwoordelijkheden van Uber. Het Hof oordeelde dat Uber een vervoersdienst aanbiedt in de zin van het EU-recht. Dat betekent dat de EU-lidstaten nationaal mogen bepalen onder welke voorwaarden deze dienst mag worden verleend. Ook toezichthouders spelen een belangrijke rol in het ophelderen van juridische onduidelijkheden.

Daarnaast speelt dat de platformen vaak niet precies passen in de bestaande wettelijke categorieën. Dat leidt tot conceptuele onzekerheid, en beleidsonzekerheid: is Facebook een socialmediaplatform, of een nieuwsbedrijf met bijbehorende verantwoordelijkheden? Daarom is vaak ook een update van bestaande wetgeving nodig. De Europese Commissie bereidt op dit moment de vernieuwing van een groot aantal juridische kaders voor, onder meer voor het consumentenrecht, auteursrecht, audiovisuele media, privacy, digitale veiligheid en het mededingingsrecht.
Innovatiebeleid: zet publieke waarden centraal

De afgelopen jaren is meer aandacht gekomen voor ethiek in het innovatiebeleid. In juni 2018 verscheen bijvoorbeeld de nationale digitaliseringsstrategie van het kabinet. Daarin komen veel kwesties aan bod, waaronder privacy, cybersecurity en een eerlijke data-economie. De laatste paragraaf daarin gaat over grondrechten en ethiek. Op dit moment werkt het kabinet twee visies uit over AI, een strategisch actieplan AI, en een visie over grondrechten en AI. Deze laatste is een bouwsteen voor het actieplan.

Het is belangrijk om ethiek niet los of als sluitstuk te zien van innovatieprogramma's, maar als integraal onderdeel hiervan. De uitdaging is om gedeelde publieke waarden als uitgangspunt te nemen. In andere Europese landen zien we daar voorbeelden van. Op het gebied van mobiliteit heeft het Verenigd Koninkrijk in vroeg stadium cybersecuritystandaarden opgesteld voor de zelfrijdende auto, om leidend te worden in de ontwikkeling hiervan. Duitsland heeft ethische richtlijnen opgesteld voor de ontwikkeling van de zelfrijdende auto. Ook Nederland kan via het stellen van randvoorwaarden op het gebied van privacy, cybersecurity, transparantie en andere

uitgangspunten, vorm en richting geven aan innovatie, bijvoorbeeld bij experimenteerruimtes of bij het verlenen van (tijdelijke) vergunningen door toezichthouders (‘regulatory sandboxes’).

4.1. De praktijk: zet tegelijkertijd in op technologische en juridische innovatie

In de praktijk blijkt dat technologische innovatie niet los kan worden gezien van verdienmodellen en regelgeving; deze ontwikkelen zich in samenhang. Zo kregen innovatieve steden als Eindhoven en Amsterdam te maken met vragen over het inwinnen en gebruik van sensordata in de openbare ruimte. Wie heeft zeggenschap over die gegevens? Voor welke doelen mogen ze gebruikt worden? Hoe kan een datamonopolie voorkomen worden? Amsterdam en Eindhoven vroegen daarom om de ontwikkeling van nationale spelregels. Inmiddels is hier een handreiking voor..

Ook in de zorg is een ontwikkeling te zien waarin innovatie is ingebed in een lokale zorgcontext tussen artsen, patiënten, wetenschappers en ontwikkelaars. Dit komt de kwaliteit van de nieuwe toepassingen ten goede. Het draait niet langer om de kwantiteit van de data, maar om de kwaliteit, en het hogere doel: een betere gezondheid (zie ook ons rapport Gezondheid Centraal).

5. Conclusie

De toegenomen aandacht voor AI en ethiek heeft het publieke debat over digitalisering verbreed. Diverse actoren hebben zich gecommitteerd aan het formuleren en onderschrijven van ethische codes en principes voor de ontwikkeling en inzet van AI. Het is goed om te zien dat de discussie over waarden internationaal wordt opgepakt en dat er veel aandacht is voor ethiek. Tegelijk dreigt die massale aandacht voor ethiek de vervolgstappen in de weg te staan. Nu is het tijd voor de praktijk.

Individuele ethische codes vormen slechts een deel van het verhaal. Afstemming en collectieve standaarden zijn ook van belang. Deze beweging komt nu op gang. Daarnaast vormen ethische codes geen afdwingbare en door een democratisch proces gelegitimeerde regels. In een democratische rechtsstaat moet wetgeving een centrale rol spelen bij de ontwikkeling van AI. Zeker gezien de veranderende geopolitieke situaties en de verregaande macht van grote techbedrijven is de rol van wetgeving en het democratische proces om daartoe te komen essentieel.

Dit betekent dat de verbinding tussen innovatie en publieke waarden veel nadrukkelijker opgezocht moet worden. In de praktijk zien we dit steeds meer gebeuren. Innoveren gaat niet alleen om technologische innovatie, maar ook om sociale, economische en juridische innovatie. In dit artikel zijn diverse voorbeelden naar voren gekomen die laten zien dat samen innoveren moet én kan.

Meer lezen

Lees ook onze publicaties over dit onderwerp:

- Eerlijk delen (2017)
- Opwaarderen (2017)
- Living labs in Nederland (2017)
- Doelgericht digitaliseren (2018)
- Wethouders en raadsleden, durf te vragen (2018)
- Beschaafde Bits (2018)
- Nieuwe regels voor kunstmatige intelligentie? (2018)
- Bedrijf zoekt universiteit (2018)
- Gezondheid Centraal (2019)
- Overzicht van ethische codes en principes voor AI (2019)

Auteurs
Roos de Jong, Linda Kool en Rinie van Est

Bij voorkeur citeren als:
Jong, R. de, L. Kool en R. van Est (2019). *Zo brengen we AI in de praktijk vanuit Europese waarden.* Rathenau Instituut: Den Haag

Contactgegevens
Anna van Saksenlaan 51
Postbus 95366
2509 CJ Den Haag
070-342 15 42
info@rathenau.nl
www.rathenau.nl

Het Rathenau Instituut stimuleert de publieke en politieke meningsvorming over de maatschappelijke aspecten van wetenschap en technologie. We doen onderzoek en organiseren het debat over wetenschap, innovatie en nieuwe technologieën.

Rathenau Instituut