ETHISCHE AI ONTWIKKELEN? ZEVEN ASPECTEN DIE OOK VOOR MKB VAN BELANG ZIJN

Ethiek is als ‘het milieu’. Je kunt er moeilijk tegen zijn. Tegelijk merken we bij bedrijven en collega’s nog wel eens weerstand als het onderwerp ter sprake komt. Ethiek vergt tijd, aandacht en nadenken. Ethische vragen zijn niet altijd eenvoudig en snel te beantwoorden. En dat is lastig wanneer iemand snel aan de slag wil met een mooi idee voor een AI-systeem.

Toch is het ook in het voordeel van Mkb’ers om op tijd te kijken naar ethische aspecten van de systemen die je ontwikkelt. Door op tijd (vooraf!) en met meerdere disciplines de juiste vragen te bespreken, wordt de opzet en het ontwerp van het AI-systeem sterker. Bovendien levert het bij voorbaat antwoorden op indringende vragen van toekomstige klanten.

Zeven aspecten voor ethische AI In de kern gaat het bij het ontwikkelen van AI om de volgende zeven aspecten:

1. Privacy. Verwerkt het AI-systeem persoonsgegevens, of bevat de trainingsset data die herleidbaar kunnen zijn naar personen? Dan is de AVG van toepassing. Sinds de invoering van de AVG hebben de meeste bedrijven dit onderwerp wel op de radar staan en een privacyfunctionaris aangesteld. Betrek die dan ook bij het project. Is er geen sprake van persoonsgegevens dan wordt het leven een stuk makkelijker. Maar let op. Data die op het eerste gezicht anoniem is, kunnen soms toch herleidbaar zijn. Bijvoorbeeld data van pompen van rioolgemalen lijken anoniem, maar wanneer ze slechts 1 of 2 huizen bedienen dan geven ze toch een indicatie of iemand met vakantie is.
2. Beveiliging. Is het AI-systeem voldoende beveiligd? Hoe kwetsbaar is het systeem voor aanvallen? Specifieke aanvallen voor AI-systemen zijn bijvoorbeeld het expres aanbieden van gemanipuleerde data in training of productie. Bij een recent geval wisten ‘hackers’ de autopilot van Tesla te verstoren met een stukje tape op een verkeersbord.
3. Eerlijkheid. Hieronder vallen zaken als diversiteit en het vermijden van onterechte vooroordelen (‘bias’). Als een AI-systeem wordt ingezet voor een bepaalde groep personen, dan moet de trainingsdataset ook representatief zijn voor die groep personen. Dat lijkt evident, maar in de praktijk kunnen er toch subtiele en onbewust afwijkingen in de verzamelde data zitten. Bijvoorbeeld omdat de oorspronkelijke bron van de data al onbewuste bias bevat. Een AI getraind op internet afbeeldingen die werd gevraagd om een gezicht van een Amerikaanse politica aan te vullen tot een volledige foto, kwam met een bikinifoto aanzetten...
4. Autonomie en toezicht. Dit aspect gaat over de menselijke maat. Blijven mensen zelf verantwoordelijk voor hun beslissingen? Weten mensen dat ze met AI te maken hebben? Is er menselijke toezicht op het systeem? *Computer says no* is een scenario dat we moeten vermijden.
5. Transparantie. Dit is een lastige. In hoeverre kan de gebruiker inzien waarom een AI-systeem tot een bepaalde beslissing of voorspelling is gekomen? Voor de huidige AI-systemen op basis van deep learning is dat vaak erg moeilijk. Dit is een actief academisch onderzoeksgebied waar steeds meer tooling wordt ontwikkeld, maar de vertaalslag naar de praktijk moet voor een groot deel nog gemaakt worden. We zullen hier in het blog nog meer over schrijven.
6. Maatschappelijke impact. Wat is de impact van het systeem op milieu, maatschappij, democratie

relevant zijn. Toch kan het goed zijn om de vraag te stellen ‘Wat als we plotseling een miljoen gebruikers hebben?’ Denk bijvoorbeeld nog eens terug aan de verrassende effecten toen PokémonGo opeens een rage werd.

7. Verantwoordelijkheid. Zijn er mechanismes in werking om te zorgen voor verantwoordelijke ontwikkeling en gebruik van het systeem? Zijn risico’s in kaart gebracht en zijn er maatregelen getroffen? Is het systeem te auditen? Als het systeem ongewenste impact heeft, is er dan een manier om dat te ontdekken en een procedure om verbeteringen door te voeren. Bij de Toeslagenaffaire heeft dit in elk geval niet gewerkt.

Aanpak en hulpmiddel Hoe kun je als Mkb’er met deze zeven aspecten concreet aan de slag? Uit de beschrijvingen is in elk geval wel duidelijk dat het goed is om ethische aspecten vanuit verschillende invalshoeken te bekijken. Betrek daarom meerdere rollen zoals toekomstige gebruikers, ontwikkelaars, ontwerpers en ethical hackers en bespreek hoe zij tegen de verschillende aspecten aankijken.

Het is dan wel prettig om een hulpmiddel te hebben dat als leidraad kan dienen. Een mooi hulpmiddel is de *Assessment List for Trustworthy AI (ALTAI),* een checklist ontwikkeld door experts van de EU. Deze doorwrochte checklist is echter 15 bladzijden lang en bevat meer dan 100 vragen. In het KI Agile project maken we daarom een ‘MKB-vriendelijke’ variant van deze vragenlijst. De variant testen we bij onze MKB-partners in het project. Binnenkort delen we de eerste statische versie via deze site en daarna maken we er een interactieve tool van. Dus watch this space!

Auteur: Jan Balje