

"Kunstmatige intelligentie moet ten dienste staan van de mensheid, niet andersom": een certificering voor ethische en veilige AI



Getty Images

Jennifer Mertens

Gezien de opkomst van kunstmatige intelligentie in onze samenleving worden er steeds meer initiatieven genomen om deze technologie, waarvan de werking ongewenste effecten kan hebben, te reguleren.

AI is tegenwoordig alomtegenwoordig en gezien de groeiende belangstelling voor deze technologie zal het een steeds grotere rol gaan spelen in ons leven, zowel in ons dagelijks bestaan als op meer gevoelige gebieden, zoals financiën, defensie, media en gezondheidszorg. Dit gebruik roept vragen op, met name op het gebied van risico's, verantwoordelijkheid, maar ook toegang tot en beheer van de gegevens die nodig zijn om het te trainen.

Al maandenlang worden er stemmen gehoord die waarschuwen voor de risico's van het gebruik ervan en vragen om kaders om misbruik te voorkomen. Dit heeft onder meer geleid tot de totstandkoming van de AI-wet in de Europese Unie. Maar er worden ook andere, minder dwingende initiatieven ontwikkeld, zoals het GAIGI ([Geneva Artificial Intelligence Governance Institute](#)).

Regulering van gevoelige AI

Dit initiatief, dat in Genève werd gelanceerd door dokter Alex Mazolo, heeft tot doel gevoelige kunstmatige intelligentie te reguleren door middel van een certificering op verschillende niveaus. "Het idee is dat technologie ten dienste moet staan van de mensheid, en niet andersom", legt de initiatiefnemer uit. En de tijd dringt, want "gevoelige AI is geen futuristisch concept: het sorteert, verzorgt, bewaakt en beslist nu al. De toepassingen ervan moeten worden omkaderd door streng toezicht, waarbij menselijke waardigheid, traceerbaarheid en verantwoordelijkheid centraal staan."

Met dit doel voor ogen is de GAIGI opgericht, om "de betrouwbaarheid en veiligheid van zogenaamde gevoelige AI-systemen te waarborgen". De organisatie stelt een flexibel handvest voor, een moreel kompas om bedrijven en sectoren te helpen "te navigeren naar betrouwbare, rechtvaardige en duurzame AI". Het is een internationaal initiatief dat een antwoord wil bieden op de fragmentatie van de regelgeving die door de verschillende landen en organisaties al dan niet is ingevoerd.

"We willen iedereen en alle landen bereiken, omdat we willen dat GAIGI internationaal wordt, simpelweg omdat we geloven dat er geen Amerikaanse, Europese of Chinese AI bestaat, maar **een AI ten dienste van de mensheid**", legt dr. Alex Mazolo uit. "Daarom zijn we in gesprek met alle partijen over de hele wereld, en niet alleen met regelgevende instanties. We willen met alle spelers, groot en klein, in gesprek gaan, zodat zoveel mogelijk partijen onze certificering willen verkrijgen."

GAIGI is pas in juli van dit jaar opgericht en staat nog in de kinderschoenen, maar merkt dat er vraag naar is in de sector: "Na een conferentie in Marokko vorige week, een hier in Brussel en een andere in Parijs in de komende dagen, zullen we zeer binnenkort voor de WTO spreken, wat bewijst dat er echt belangstelling is voor een certificering van AI."

Hoe vertaalt zich dat concreet?

Het behalen van de GAIGI-certificering toont aan dat een bedrijf of, eenvoudigweg, een AI betrouwbaar is, omdat het voldoet aan "technische, ethische en bestuursnormen die in Genève zijn ontwikkeld". Een certificering die wordt verkregen na een grondige controle en wordt geverifieerd door "regelmatige audits, aanzienlijk menselijk toezicht en transparante processen".

In tegenstelling tot een regelgevende instantie die actoren kan verplichten zich aan haar wetgeving te houden, wordt de GAIGI-certificering echter op vrijwillige basis verleend en "is zij een bewijs van de betrouwbaarheid van de actor en zijn ethisch gebruik van AI". Ontwikkelaars of bedrijven kunnen dus vrijwillig besluiten zich te onderwerpen aan het handvest van het initiatief: het eerste niveau.

Daarna kunnen zij de 'tweede niveau'-certificering verkrijgen, die aangeeft dat zij voldoen aan de richtlijnen die door deskundigen voor elke sector (gezondheidszorg, defensie, financiën, media) zijn opgesteld.

Ten slotte garandeert het hoogste certificeringsniveau, het zogenaamde algoritmische niveau, dat de AI door experts is getest en dat de antwoorden die het systeem geeft geen problemen zullen opleveren. "Tegenwoordig bestaan er middelen en instrumenten om AI te testen en te voorspellen wat het zal zeggen, met name in Zwitserland met de ISO", die juist internationale technische normen ontwikkelt voor het testen, controleren en certificeren van gevoelige algoritmen.



Economisch belang

Echter, hoe kunnen we giganten zoals Meta, Google of OpenAI ertoe aanzetten om een van deze certificeringen te verkrijgen, aangezien er, in tegenstelling tot bij regelgeving, geen sancties zijn voorzien? Dat is niet het doel van het initiatief. "We moeten eerst beleidsmakers, actoren uit verschillende sectoren en bedrijven overtuigen om ons charter te onderschrijven, maar ook het publiek informeren en voorlichten, zodat het de voorkeur geeft aan tools die als ethisch en veilig zijn gecertificeerd. Op die manier zullen de AI-giganten indirect worden gedwongen om aan de eisen van onze certificering te voldoen."

Bovendien "zien sommigen regelgeving als een belemmering voor innovatie, wat wij juist niet willen. Wij willen de technologische ontwikkeling van AI ondersteunen door vertrouwen in deze tool op te bouwen. Pas als mensen er zeker van zijn dat ze erop kunnen vertrouwen, zullen ze het in hun dagelijks leven of werk gaan gebruiken", concludeert Alex Mazolo.

De GAIGI zal haar certificeringssysteem officieel lanceren aan het einde van september, wanneer deskundigen uit verschillende gevoelige sectoren zich bij het project hebben aangesloten – en er zijn veel kandidaten, zo deelt de initiatiefnemer van het project ons mee –, met een gedetailleerd tweejarig tijdschema om "de mensheid te beschermen tegen algoritmische misstanden" en de lancering, vanaf oktober, van een eerste proefproject waarin de richtlijnen voor de mediasector worden beschreven.

Bron: [Raadpleging van 12/05/2025 [Een certificering voor ethische en veilige AI](#)]