

4

Vragenuur

Vragenuur

Vragen Zwinkels

Vragen van het lid Zwinkels aan de staatssecretaris Digitale Economie en Soevereiniteit over **het bericht "AI-assistenten liegen en bedriegen, ondanks wetten en regels"**.

De voorzitter:

Het woord is aan mevrouw Zwinkels voor haar mondelinge vragen namens de fractie van het CDA aan de staatssecretaris van Economische Zaken en Klimaat. Er werd al even naar verwezen. Van harte welkom in het parlement. Mevrouw Zwinkels, u heeft het woord.

Mevrouw **Zwinkels** (CDA):

Dank u wel, voorzitter. Voor het CDA is ten aanzien van AI één ding fundamenteel, namelijk dat technologie de mens moet dienen en niet andersom. AI is steeds meer verbonden met ons dagelijks leven, niet alleen als handige teksthulp, maar ook als klantenservicemedewerker, als verkoophulp, als persoonlijk assistent en zelfs als adviseur bij ingewikkelde keuzes. AI-technologie biedt kansen; dat ziet het CDA ook. Wij willen dat Nederland vooroploopt in nieuwe technologie, innovatie en digitale weerbaarheid. Maar dat AI-assistenten liegen en bedriegen, ondanks wetten en regels, laat zien dat het niet vanzelf gaat. Zelfs het best scorende model overtrad volgens het onderzoek 46% van de scenario's uit de wet. Een AI-assistent die zich voordoeft als mens, terwijl het altijd helder moet zijn of je met AI praat of niet, raakt op een gegeven moment aan het vertrouwen. Zoals paus Leo onlangs al waarschuwde, mag technologische vooruitgang nooit losraken van menselijke waardigheid.

Voor het CDA zit hier dan ook de diepere zorg, want AI kan een nieuwe ongelijkheid veroorzaken. Aan de ene kant zijn er mensen die weten hoe AI werkt en hoe je het kunt controleren, maar aan de andere kant zijn er ook mensen die niet doorhebben dat zij worden gestuurd, beïnvloed of bedrogen. Juist kwetsbare mensen kunnen dan in de val lopen.

Mijn eerste vraag aan de staatssecretaris is daarom: deelt zij de zorg van het CDA dat AI-toepassingen alleen verantwoord kunnen groeien als menswaardigheid, veiligheid en bescherming van kwetsbare mensen centraal staan? Graag een reflectie.

De voorzitter:

Het woord is aan de staatssecretaris.

Staatssecretaris **Aerdt**:

Voorzitter. Ja, de inzet van het kabinet is natuurlijk dat AI verantwoord ingezet wordt. Ik vind het dan ook zorgelijk dat het erop lijkt dat je die AI-modellen opdrachten kan geven waarmee bijvoorbeeld de AI-verordening of andere wetgeving wordt overtreden. Dat maakt de uitkomsten van de opdrachten weinig waardevol, maar het is ook zorgelijk, ook voor wat betreft het vertrouwen dat men in AI heeft. Ik

hecht ook groot belang aan die waarden, die ook met die Europese wetgeving worden beschermd. Zeker ook het vertrouwen in de diffusion, het gebruik van AI, is daarin een hele belangrijke.

Mevrouw **Zwinkels** (CDA):

Ik ben heel erg blij om dat te horen. Mijn vervolgvraag is dan: op welke wijze borgt het kabinet dat die AI-agenten zich daadwerkelijk houden aan de bestaande wet- en regelgeving? En wie is er juridisch verantwoordelijk wanneer een autonome AI-agent de wet overtreedt? Kan de staatssecretaris daar ook op reflecteren?

Staatssecretaris **Aerdt**:

Een aantal zaken. Misschien is het eerst goed om te vermelden dat de AI-verordening deels al in werking is getreden en deels dus nog niet. Bepaalde praktijken, zoals misleiding of manipulatie — daar had u het net over — zijn op dit moment al verboden. In appendix 3, zeg ik uit m'n hoofd, is een lijst opgenomen met AI die echt niet gebruikt mag worden. Het lastige is natuurlijk, zeker bij dit soort zaken ... Wetgeving loopt natuurlijk eigenlijk per definitie een beetje achter de feiten aan, maar zeker bij AI is dat er ook echt nog wel eentje om goed in het achterhoofd te houden. De toezichthouders in het kader van die AI-verordening moeten nog worden aangewezen. Uw Kamer heeft hier volgende week, zeg ik uit m'n hoofd, een schriftelijk overleg over, om hier ook heel specifiek op in te gaan.

Precies dat stukje over die ... In principe zijn natuurlijk de ontwikkelaars van de taalmodellen verantwoordelijk voor hoe zij het model opstellen en hebben mensen zelf natuurlijk ook een verantwoordelijkheid als het gaat om de vragen die zij stellen, maar die verantwoordelijkheid kan natuurlijk niet volledig bij mensen gelegd worden. Daarom is in die AI-verordening ook opgenomen dat bedrijven een gebruikersinstructie moeten meegeven. Maar heel specifiek dat juridische stukje ... Dat is ook echt nog de vraag die wij hebben en ook in Europees verband oppakken.

Mevrouw **Zwinkels** (CDA):

Het is goed om te horen dat er al een aantal zaken lopen en dat er ook nog een aantal openstaande punten zijn. Daarbij werden de toezichthouders al genoemd. Ik zou ook willen weten: hoe wordt er op dit moment op toegezien dat AI-systemen, voordat zij worden ingezet richting burgers of consumenten, daadwerkelijk zijn getest op risico's zoals misleiding, manipulatie en andere onverantwoorde effecten?

Staatssecretaris **Aerdt**:

Misschien is het goed om te vermelden dat niet alle AI-systemen vooraf getest hoeven te worden. Dit gaat eigenlijk risicogestuurd, zoals we dat bijvoorbeeld ook zien bij de Nederlandse Voedsel- en Warenautoriteit. Daarom gaf ik ook aan: sommige dingen mogen volgens de AI-verordening niet. Sommige, met een heel hoog risico, moeten inderdaad getest worden. Maar aan de andere kant heb je natuurlijk ook de innovatiekant. Je kan niet alle systemen testen voordat er wat mee gebeurt. Als je het bij wijze van spreken over een simpele Excelsheet hebt, dan hoeft dat niet van tevoren getest te worden. Het is dus, denk ik, wel goed om de gelaagdheid daarin aan te geven. We zitten op

dit moment ook een beetje in een vacuüm, want een deel van die AI-verordening is inderdaad al in werking getreden, maar de toezichthouders zijn nog niet aangewezen.

Mevrouw Zwinkels (CDA):

Je kunt dan wel met elkaar betwisten over in hoeverre dit "hoog risico" betreft, want wat ons betreft maken we ons hier zeker zorgen over en grijpen we hierop in.

Maar ik zou, gelet op de tijd, graag nog een laatste vraag willen stellen. Dat is eigenlijk een vraag om een toezegging: is de staatssecretaris bereid om toe te zeggen om met een plan van aanpak "menswaardige en veilige AI" te komen, waarin staat hoe burgers worden beschermd tegen die misleiding, hoe bewust gebruik van AI wordt versterkt en hoe kwetsbare groepen niet de dupe worden?

Staatssecretaris Aerdts:

Ik denk eigenlijk dat we dat al hebben. In de Strategie Digitale Economie wordt specifiek over verantwoord gebruik van AI gesproken. Ook de Nederlandse Digitaliseringsstrategie neemt dit mee, net als diverse Europese wetgeving. Ik spreek heel uitgebreid met mijn collega van OCW over de digitale vaardigheden en AI-vaardigheden die we willen meegeven. Ik denk dus dat we dit al doen in de verschillende plannen en strategieën die we hebben. Tijdens ons eerste wetgevingsoverleg heb ik er ook met mevrouw Zwinkels over gesproken dat ik drie A4'tjes heb volgeschreven met alle plannen en strategieën die we hebben. Ik zou u om het vertrouwen willen vragen dat we met deze strategieën aan de slag kunnen gaan, juist ook om het tempo erin te houden.

Mevrouw Zwinkels (CDA):

Dat laatste vind ik toch wel jammer om te horen. Ik weet dat de staatssecretaris hard werkt als het gaat om digitale autonomie en cybersecurity, maar het gaat mij om de ethische vraag, om de menswaardigheid. Wij vinden dat dat op dit moment echt nog onvoldoende in de plannen is verwerkt. Wij zien hier echt graag meer actie op tegemoet en ook meer visie van het kabinet. Ik zou dan toch weer in herhaling willen vallen en de staatssecretaris het volgende willen vragen. Als er geen apart plan van aanpak kan worden verwacht, kunnen wij zelf voorstellen doen, maar is zij dan in ieder geval bereid om in een eerstvolgende brief over de digitale strategie expliciet hierop te reflecteren?

Staatssecretaris Aerdts:

Ja, dat kan ik toezeggen.

Mevrouw Zwinkels (CDA):

Fijn, dank u wel.

Mevrouw Kathmann (GroenLinks-PvdA):

Chatbots die liegen en bedriegen kunnen bizarre schade toebrengen. Er zijn chatbots die tegen jongeren met suïcidale neigingen ... Eerst doen ze zich voor als een mens. Dan worden de jongeren niet doorverwezen naar goede hulp en wordt er door die chatbots alleen maar geroepen: praat erover. Dat is echt gigantisch zorgelijk. Je moet er gewoon op kunnen vertrouwen dat je op belangrijke en

moeilijke momenten in je leven goede hulp krijgt. Daarom vind ik ook echt dat het recht op menselijk contact bij wet moet worden afgedwongen. Op 28 juni hebben we hier een commissiedebat over. Kan de staatssecretaris toezeggen om vóór 28 juni in gesprek te gaan met zorg- en veiligheidspartners en misschien met nog meer partners die ze kent die hierbij moeten aansluiten, om met een overzicht te komen van alle mogelijkheden die we hebben om ligende chatbots in de ban te doen?

Staatssecretaris Aerdts:

Ik wil dat zeker oppakken in de voorbereiding van dit debat. Ik denk dat ik dit deels samen met mijn collegae moet doen. Ik weet niet of het een uitputtend overzicht van alle mogelijkheden zal zijn, maar we zullen dat voor het debat oppakken.

De heer Verkuijlen (VVD):

Regels die niet worden gehandhaafd zijn krachteloos. Ik denk dat de toezichthoudende instellingen daarom hard nodig zijn. Maar er komen nog krachtigere AI-modellen aan. Ik zag vandaag in antwoorden op vragen van de Kamer dat het kabinet ervan overtuigd is dat de regels ook gaan werken voor die krachtige modellen. Naar aanleiding van dit bericht ben ik wel benieuwd of de staatssecretaris nog steeds die mening is toegedaan.

Staatssecretaris Aerdts:

De AI-verordening moet nog in werking treden, ook als het om het toezicht gaat. We spreken hier volgende week in de Kamer over, maar de toezichthouders moeten nog worden aangewezen. Dit is natuurlijk nadrukkelijk een zorg die het kabinet deelt en ook zal opvolgen op het moment dat de verordening in werking is getreden en de uitvoeringswet in Nederland van kracht is geworden. Hier houden we nadrukkelijk een vinger aan de pols.

Mevrouw El Boujdaini (D66):

Het is heel onwenselijk als AI zich voordoet als mens. We zien ook dat het impact heeft op het dagelijks gebruik, omdat mensen het ook gebruiken als psycholoog of voor psychologisch advies. Dat is enorm gevaarlijk. Daarom is het des te belangrijker dat we inzicht krijgen in hoe zo'n AI eruitziet. Zouden we daarom meer kunnen doen op het gebied van de transparantienorm om meer inzicht te krijgen in de algoritmes van dit soort AI's?

Staatssecretaris Aerdts:

Wat dat betreft ben ik heel blij met de AI-verordening, want die kijkt naar de hele AI-keten. Je moet weten dat content via AI gegenereerd is. Ik sprak net al over gebruikersinstructies. Als besluiten met een hoog risico door AI-systemen worden genomen, moet dat van tevoren bekend zijn. Er wordt ook naar de systeemrisico's gekeken. Ik denk dat dit heel erg gaat helpen in de transparantie van de normen. Ik deel de zorg die u hier aangeeft.

De voorzitter:

Ik dank de staatssecretaris voor de beantwoording van de gestelde vragen. Ik schors een enkel ogenblik, waarna we

overgaan tot de stemmingen. De vergadering is kort geschorst.

De vergadering wordt enkele ogenblikken geschorst.