Tweede Kamer der Staten-Generaal

2

Vergaderjaar 2025–2026

Aanhangsel van de Handelingen

Vragen gesteld door de leden der Kamer, met de daarop door de regering gegeven antwoorden

1796

Vragen van het lid **Zwinkels** (CDA) aan de Staatssecretaris van Binnenlandse Zaken en Koninkrijksrelaties over *het artikel «Snelgroeiende autonome AI-assistent is een «disaster waiting to happen»»* (ingezonden 6 februari 2026).

Antwoord van Staatssecretaris **Aerdts** (Economische Zaken en Klimaat) (ontvangen 29 april 2026).

Vraag 1
Bent u bekend met het artikel «Snelgroeiende autonome AI-assistent is een «disaster waiting to happen»»?[1]

Antwoord 1
Ja, ik heb kennisgenomen van dit artikel.

Vraag 2
Deelt u de zorgen van experts dat steeds autonomer opererende AI-assistenten risico's vormen voor veiligheid, privacy, menselijke controle en mentale gezondheid? En kunt u daarbij aangeven welke risico's u het meest urgent acht?

Antwoord 2
Ja, die zorgen deel ik. In de Verzamelbrief van 18 december 2025[2] heeft de voormalig Staatssecretaris Digitalisering en Koninkrijksrelaties uw Kamer geïnformeerd over de resultaten van de forecast «Zicht op de Digitale Toekomst», welke TNO heeft ontwikkeld op zijn verzoek. TNO heeft daarin ook gekeken naar de impact van autonomer opererende AI-assistenten («Agentic AI») op publieke waarden zoals privacy, autonomie en democratie. Een risico is bijvoorbeeld dat gebruikers de controle verliezen op wat AI-assistenten doen. Daarnaast verzamelt en verwerkt het veel persoonlijke gegevens van gebruikers. Omdat deze AI-assistenten zelfstandig werken is het voor gebruikers moeilijk te overzien welke data waar terechtkomt en wat er met die data gebeurt. Een ander risico is dat het niet meer scherp wordt wie verantwoordelijk is voor de handelingen van de AI-assistent. Ook op het gebied van privacy en (cyber)veiligheid zijn er risico's. Dat geldt met name

[1] NRC, 4 februari 2026, Snelgroeiende autonome AI-assistent is een «disaster waiting to happen», https://www.nrc.nl/nieuws/2026/02/03/snelgroeiende-autonome-ai-assistent-die-taken-van-gebruikers-overneemt-is-een-disaster-waiting-to-happen-a4919418

[2] Kamerstukken II 2025/26, 26 643, nr. 1450

ah-tk-20252026-1796
ISSN 0921 - 7398
's-Gravenhage 2026

voor de meer experimentele vormen van AI-assistenten waar in het artikel naar verwezen wordt, en waar ook de Autoriteit Persoonsgegevens recent voor heeft gewaarschuwd.[3] Op het moment dat autonome AI-assistenten breed toegang krijgen tot talrijke apps en informatiebronnen, dan kan het misgaan doordat bijvoorbeeld accounts worden overgenomen, hacks worden gezet, of toegang worden verkregen tot privacy gevoelige gegevens (datalekken).

Vraag 3 en 4
Acht u het wenselijk dat AI-systemen zelfstandig handelingen, zoals het doen van aankopen en het aangaan van contracten, kunnen verrichten namens gebruikers?
Zo ja, kunt u aangeven welke toepassingen het kabinet maatschappelijk gezien wenselijk en/of acceptabel vindt, en welke niet?

Antwoord 3 en 4
De wenselijkheid van dit soort autonomere AI-toepassingen hangt samen met de wijze waarop en het domein waarin deze worden ingezet. Bij iedere toepassing is het daarbij in beginsel de vraag in hoeverre de inzet rechtmatig en proportioneel is en in lijn met bestaande wet- en regelgeving. Ook is van belang dat duidelijk is wie verantwoordelijk is voor de handelingen die een AI-systeem namens een gebruiker verricht en dat er voldoende waarborgen bestaan op het gebied van bijvoorbeeld transparantie, controleerbaarheid en de bescherming van (andere) publieke waarden. Uiteraard vind ik het daarbij van belang om de ontwikkelingen op het gebied van AI, zoals agentic AI en autonomere AI-systemen, nauwgezet te volgen en hierin ook in Europees verband op te trekken, bijvoorbeeld via de Europese AI Board en andere relevante EU-gremia.

Vraag 5
In hoeverre is het huidige Nederlandse en Europese toezichtkader (waaronder de AI Act) toereikend om risico's van autonome AI-systemen die zelfstandig taken uitvoeren te ondervangen?

Antwoord 5
De AI-verordening reguleert autonome AI-systemen zoals AI-assistenten op verschillende manieren. Ten eerste zullen alle AI-assistenten die in gevoelige hoog risico toepassingsgebieden of producten worden ingezet aan de strenge eisen uit de AI-verordening moeten voldoen. De AI-verordening noemt expliciet welke gebieden of producten dit zijn. Dit gaat bijvoorbeeld om een AI-assistent die leerlingen in het onderwijs beoordeelt of een AI-assistent die gebruikt wordt als een medisch hulpmiddel. Eén van de eisen aan deze hoog risico AI-systemen is dat risico's beoordeeld en gemitigeerd moeten worden. Als dat niet kan, mag het AI-systeem niet voor die risicovolle toepassing ingezet worden. Hiermee worden de gezondheid, veiligheid en grondrechten van mensen beschermd.
Ten tweede worden er eisen gesteld aan autonome AI-systemen die worden ontworpen om directe interacties met mensen te hebben, bijvoorbeeld als ze zelfstandig e-mails uit kunnen sturen of reacties kunnen plaatsen op het internet. De aanbieder van een dergelijk AI-systeem moet ervoor zorgen dat het duidelijk is dat men contact met een AI-systeem heeft, en niet met een mens.
Ten slotte worden er onder de AI-verordening ook eisen gesteld aan de modellen voor algemene doeleinden die de basis vormen van deze autonome AI-systemen. Als deze modellen capaciteiten met een grote impact hebben, bijvoorbeeld door negatieve effecten op publieke gezondheid, veiligheid, fundamentele rechten of de maatschappij als geheel, dan moeten de aanbieders van deze modellen systeemrisico's in kaart brengen en mitigeren. Capaciteiten zoals vergaande autonomie en het kunnen interacteren met andere hardware en software kunnen mogelijk voor systeemrisico's zorgen. Naast de AI-verordening, is ook de digitaledienstenverordening (DSA) mogelijk relevant. De digitaledienstenverordening verplicht zeer grote online

[3] Autoriteit Persoonsgegevens, 12 februari 2026, AP waarschuwt voor grote beveiligingsrisico's bij AI-agents zoals OpenClaw.

platforms- en zoekmachines om de zogenaamde systeemrisico's die voortvloeien uit het ontwerp of uit de werking van hun dienst en de daaraan verbonden systemen, te beoordelen en te beperken. Dergelijke risico's omvatten onder meer de verspreiding van illegale inhoud en werkelijke of voorzienbare negatieve effecten op grondrechten, de burgerdialoog en verkiezingsprocessen, minderjarigen of het lichamelijke en geestelijke welzijn van personen. Zulke risico's kunnen bijvoorbeeld ontstaan door het niet-authentieke gebruik van de dienst, zoals het aanmaken van nepaccounts, het gebruik van bots en andere geautomatiseerde of gedeeltelijk geautomatiseerde gedragingen. Dit kan leiden tot een snelle en wijdverbreide verspreiding onder het publiek van informatie die illegale inhoud bevat of onverenigbaar is met de algemene voorwaarden van een onlineplatform of onlinezoekmachine, en die bijdraagt aan desinformatiecampagnes. Het is mogelijk dat de risico's van de inzet van autonome AI-systemen binnen zeer grote onlineplatforms langs deze weg moeten worden aangepakt.

Vraag 6
Welke definitie van verantwoorde AI (innovaties) hanteert het kabinet? En in hoeverre passen AI-assistenten daarin?

Antwoord 6
AI-innovaties zijn verantwoord als ze voldoen aan geldende wet- en regelgeving (zoals de AI-verordening, de digitaledienstenverordening en andere wettelijke kaders die van toepassing kunnen zijn) en zijn ontwikkeld op een manier waardoor ze geen voorzienbare negatieve gevolgen hebben voor de gezondheid, veiligheid en grondrechten van mensen.
AI-assistenten kunnen ook op een manier ontwikkeld en gebruikt worden dat ze betrouwbaar en mensgericht zijn. Het is van belang dat zowel ontwikkelaars als gebruikers van AI-assistenten zich bewust zijn van de mogelijke risico's van de technologie in verschillende contexten en hier robuuste maatregelen voor nemen. Aan de ene kant gaat dit om maatregelen om voorzienbare risico's bij de ontwikkeling van het AI-systeem zo veel mogelijk *by design* te ontwikkelen, maar ook om maatregelen om een systeem stop te zetten of aan te passen zodra onvoorziene risico's zich voordoen. Wetgeving zoals de AI-verordening biedt hier kaders voor, onder andere door te bepalen welke toepassingen van AI-assistenten gevolgen kunnen hebben voor de gezondheid, veiligheid en grondrechten van mensen. Toezichthouders, zoals de AP[4] en RDI, monitoren risico's van ontwikkelingen op het gebied van AI en delen *best practices* voor de aanbieders van deze technologie om hun systemen op een verantwoorde manier te ontwikkelen.

Vraag 7
Kunt u aangeven of het naar uw inzicht wenselijk is dat er vanuit de overheid gebruik gemaakt wordt van autonome AI-assistenten? En op welke vlakken gebeurt dit al? Onder welke voorwaarden wordt dit toegestaan en hoe wordt hierop toegezien in de praktijk?

Antwoord 7
De wenselijkheid van autonome AI-assistenten is afhankelijk van de context waarbinnen de AI-systemen worden ingezet. Het is vanzelfsprekend dat de inzet ervan gebeurt in lijn met bestaande regelgeving, zoals de AI-verordening, de AVG in het geval er persoonsgegevens worden verwerkt, en andere relevante (sectorale) wetgeving. Op al deze wettelijke kaders is of wordt toezicht ingericht. Voor overheden geldt daarbij sinds april 2025 het overheidsbrede standpunt generatieve AI, wat ook van toepassing is op autonomere vormen van AI, zoals AI-agenten.[5] Vooraf dienen overheden daarbij altijd een impactassessment (zoals bijvoorbeeld een gegevensbeschermingseffectbeoordeling of een Impact Assessment Mensenrechten en Algoritmes (IAMA)) te hebben uitgevoerd en dient duidelijk te zijn welk maatschappelijk doel precies wordt gediend met de inzet. De in 2025 – in opdracht van BZK – uitgevoerde overheidsbrede monitor generatieve AI laat een duidelijke toename van het aantal generatieve AI-toepassingen zien, maar

[4] Risico's algoritmes & AI: ontwikkelingen in Nederland | Autoriteit Persoonsgegevens.
[5] https://open.overheid.nl/documenten/bc03ce31-0cf1-4946-9c94-e934a62ebe73/file

daarin zijn nog geen tekenen van (veelvuldig) gebruik van autonomere AI-assistenten.[6] Als onderdeel van de AI-prioriteit binnen de Nederlandse Digitaliseringsstrategie (NDS) wordt er voor de Nederlandse overheid gewerkt aan een visie op taalmodellen en aan een herijking van het overheidsbrede standpunt generatieve AI en bijbehorende handreiking.[7] Hierbij worden ook de ontwikkelingen op het gebied van autonome (of agentic) AI nadrukkelijk meegenomen. Deze zullen nog dit jaar met uw Kamer worden gedeeld.

Vraag 8
Hoe wordt op dit moment geborgd dat er in kritieke infrastructuur, in sectoren als defensie, de zorg en de overheid zelf, altijd sprake blijft van «meaningful human control» (ofwel «human in the loop») bij het gebruik van autonome AI-assistenten?

Antwoord 8
In algemene zin hangt de mate van menselijke tussenkomst sterk af van het risico en de mogelijke impact van de toepassing van AI. Onder de AI-verordening worden er eisen gesteld aan onder andere hoog risico AI-systemen die als veiligheidscomponent in de kritieke infrastructuur worden gebruikt, hoog risico AI als medisch product of veiligheidscomponent daarvan en hoog risico AI in gevoelige overheidsgebieden, zoals het toekennen van toeslagen of in de rechtshandhaving. Voor alle hoog risico AI-systemen moet er menselijk toezicht («*human in the loop*») uitgeoefend kunnen worden. Dit houdt onder andere in dat men de capaciteiten van het systeem moet begrijpen en de werking ervan moet kunnen monitoren. Ook moet het systeem stopgezet kunnen worden. De mate van deze menselijke controle is contextafhankelijk: des te groter het risico, des te steviger het menselijk toezicht moet zijn.
Als het specifiek gaat om dergelijke AI-systemen in het Defensiedomein is menselijke controle noodzakelijk om het oordeelsvermogen van de mens voor AI-systemen te behouden, zodat deze conform het (internationaal) recht kunnen worden gebruikt. In Nederland worden nieuwe middelen en strijdmethoden door de Adviescommissie Internationaal Recht en Conventioneel Wapengebruik (AIRCW) getoetst aan het internationaal recht. De Minister van Defensie besluit op basis van het advies al dan niet goedkeuring te verlenen voor gebruik door de krijgsmacht.
Als het specifiek om de zorg gaat rust vanuit de Wet kwaliteit, klachten en geschillen zorg (Wkkgz) de plicht op zorgaanbieders om medische hulpmiddelen, inclusief AI, te gebruiken die veilig en kwalitatief goed zijn. Individuele zorgverleners zijn vanuit de Wet op de geneeskundige behandelingsovereenkomst (WGBO) verplicht om bij hun werkzaamheden de zorg van een goed hulpverlener in acht te nemen. Dit betekent onder meer dat de zorgverlener altijd eindverantwoordelijk blijft voor medische beslissingen en niet blindelings op AI-systemen mag vertrouwen. Bij het gebruik van autonome AI-systemen zal het systeem dus ook ontworpen moeten worden met deze waarborgen als randvoorwaarden. De Minister van VWS is in de brief over innovatie in de beroepsuitoefening[8] van 9 december jl. uitgebreid ingegaan op de acties die worden ondernomen bij het gebruik van AI in de beroepsuitoefening in de zorg.
Wanneer autonome AI-assistenten bijvoorbeeld door overheden worden ingezet in het kader van (algoritmische) besluitvorming, dient daarbij altijd sprake te zijn van betekenisvolle menselijke tussenkomst. Daarbij geldt voor de overheid in het geval van een besluit ook de Algemene Wet Bestuursrecht (Awb). Deze bevat de algemene regels waaraan de overheid moet voldoen bij het nemen van besluiten, onder meer met het oog op transparantie, kenbaarheid en uitlegbaarheid van besluiten. De regels zijn techniekonafhankelijk opgesteld en normeren ook algoritmische besluitvorming en daarmee ook de inzet van AI.
Daarnaast werkt de overheid aan de ontwikkeling van een integraal wegingskader voor de verantwoorde inzet van AI, waarin expliciet aandacht wordt besteed aan het waarborgen van autonomie van AI-systemen, het bevorderen

[6] https://open.overheid.nl/documenten/dafe19ab-8874-42a0-a55f-315ac5825282/file
[7] https://open.overheid.nl/documenten/9c273b71-cebb-4e11-b06f-fa20f7b4b90e/file
[8] Kamerstukken II 2025/2026, 27 529, nr. 354.

van mensgerichte toepassingen, het inrichten van robuuste terugvalopties ten behoeve van continuïteit, en de zorgvuldige afweging of en in welke gevallen AI -bijvoorbeeld in de vorm van AI-assistenten – passend is. Daarbij geldt in het bijzonder dat binnen de publieke dienstverlening het behoud van betekenisvol menselijk contact met burgers en bedrijven leidend is, en dat steeds nauwgezet wordt beoordeeld welke (repetitieve) taken verantwoord door AI kunnen worden ondersteund of overgenomen. Dit integraal wegingskader wordt ontwikkeld in het kader van de NDS, en de conceptversie is voorzien voor dit jaar (na enkele pilots ermee).

Vraag 9
Welke andere waarborgen (vangrails) zijn naar uw verwachting nog nodig om hier goed mee om te gaan, voor zowel overheid als samenleving, en is bijsturing mogelijk?

Antwoord 9
Het is bij nieuwe technologische ontwikkelingen, waaronder autonomie AI-assistenten, belangrijk om goed te kijken naar de waarborgen die nodig zijn om hier goed mee om te gaan. Bijvoorbeeld door het principe van *ethics-by-design* toe te passen zodat publieke waarden zoals privacy, transparantie, autonomie en non-discriminatie al vanaf het begin worden meegenomen. Ik zet mij ervoor in om te zorgen dat publieke waarden gewaarborgd worden, ook bij nieuwe technologische ontwikkelingen zoals quantum en autonome AI-assistenten. Verder merk ik graag op dat binnen de NDS overheden gezamenlijk aan verantwoord en succesvol ontwikkelen van AI-toepassingen werken. Bijkomend doel is om AI breder inzetbaar te maken waardoor de impact en succes van AI groter wordt. Gelet op die bredere inzetbaarheid en de verwachte impact wordt binnen de AI-opschalingsfaciliteit voor overheden een afwegingskader voor het opschalen van AI ontwikkeld, zodat zorgvuldig kan worden gekeken naar de inzet van AI in relatie tot wet- en regelgeving en ethisch verantwoord gebruik. Dit afwegingskader wordt later in 2026 gepubliceerd. Tot slot zou ik ook voorzichtigheid willen bepleiten. Bij experimentele vormen van autonome AI-assistenten zoals OpenClaw is het belangrijk om nu vooral zeer terughoudend te zijn en deze niet te gebruiken op systemen met privacygevoelige of vertrouwelijke gegevens, zoals ook de Autoriteit Persoonsgegevens heeft opgeroepen.[9]

[9] Autoriteit Persoonsgegevens, 12 februari 2026, AP waarschuwt voor grote beveiligingsrisico's bij AI-agents zoals OpenClaw.