Ministerie van Justitie en Veiligheid

> Retouradres Postbus 20301 2500 EH Den Haag

Aan de Voorzitter van de Tweede Kamer
der Staten-Generaal
Postbus 20018
2500 EA DEN HAAG

Turfmarkt 147
2511 DP Den Haag
Postbus 20301
2500 EH Den Haag
www.rijksoverheid.nl/jenv

Uw referentie
7595446

Datum 1 juni 2026
Betreft Antwoorden Kamervragen over het bericht AI-model Mythos geprezen en gevreesd lijkt in handen gevallen van onbevoegden

Hierbij zenden wij u de antwoorden op de Kamervragen van de leden El Boujdaini (D66) en Kathmann (GroenLinks-PvdA) over het bericht 'AI-model Mythos geprezen en gevreesd' lijkt in handen gevallen van onbevoegden met het kenmerk 2026Z08988.

De Minister van Justitie en Veiligheid,

D.M. van Weel

De Staatssecretaris van Digitale Economie en Soevereiniteit,

W.J.M. Aerts

Vragen van de leden El Boujdaini (D66) en Kathmann (GroenLinks-PvdA) aan de minister van Justitie en Veiligheid en de staatssecretaris van Economische Zaken en Klimaat over het bericht AI-model Mythos geprezen en gevreesd lijkt in handen gevallen van onbevoegden
(ingezonden 24 april 2026, 2026Z08988)

Vraag 1
Bent u bekend met het bericht in het NRC over het AI-model Mythos dat mogelijk in handen is gevallen van onbevoegden?

Antwoord op vraag 1
Ja.

Vraag 2
Deelt u de analyse dat ongecontroleerde verspreiding van geavanceerde AI-modellen risico's kan vergroten op cyberaanvallen, geautomatiseerde fraude en andere schadelijke toepassingen? Zo ja, welke risico's acht u het meest urgent? Zo nee, waarom niet?

Antwoord op vraag 2
Het is goed om te benadrukken dat de risico's niet voortkomen uit één specifiek geavanceerd AI-model. Geavanceerde capaciteiten zijn niet enkel meer voorbehouden aan gesloten, zogeheten 'frontier-only' modellen. Ook open, vrij toegankelijke AI-modellen komen steeds dichterbij dergelijke geavanceerde capaciteiten. Zo laat gelekte broncode van Anthropic's Claude Code product zien dat de capaciteiten van een AI-model niet enkel of grotendeels voortkomen uit het model zelf, maar juist uit de technische infrastructuur (het 'harnas') om het AI-model heen.

De ontwikkelingen volgen elkaar in snel tempo op, risico's zijn dan ook zeer situationeel afhankelijk. Eén van de risico's die het kabinet ziet, is dat geavanceerde AI-modellen onder andere kunnen worden ingezet om technische kwetsbaarheden in software op te sporen. Dit kan dankzij dergelijke AI-modellen in veel kortere tijd en op grotere schaal worden gedaan dan eerder mogelijk was. Hierdoor kan de tijd tussen ontdekking van een kwetsbaarheid en het moment dat deze wordt uitgebuit significant worden verkort.

Ook zijn geavanceerde AI-modellen tot dusver afkomstig van commerciële bedrijven en veelal afkomstig uit niet-Europese landen waar andere wet- en regelgeving geldt dan in de EU. Dat kan ertoe leiden dat (a) er weinig tot geen inzicht bestaat in de onderliggende systemen en software van dergelijke systemen en (b) die systemen niet voldoen aan Europese wet- en regelgeving. Het (verder) integreren van AI-modellen in Nederlandse digitale (kern)processen, waarbij de AI-modellen op infrastructuur van niet-Europese aanbieders draait, kan leiden tot toenemende afhankelijkheid.

Ook draaien veel geavanceerde AI-modellen op niet-Europese cloudinfrastructuur. Integratie in Nederlandse digitale processen vergroot daarmee de afhankelijkheid van niet-Europese aanbieders, met bijbehorende risico's op gebied van datasoevereiniteit, continuïteit, vertrouwelijkheid en strategische afhankelijkheid. Aanbieders kunnen

toegang tot specifieke modellen beperken, updates kunnen de uitkomsten veranderen zonder dat de afnemer dat in de hand heeft, en de rekenkracht voor geavanceerde AI is geconcentreerd bij een klein aantal niet-Europese partijen.

Vraag 3
Welke rol spelen geavanceerde AI-modellen op dit moment in het dreigingsbeeld? Welke gevolgen heeft de uitrol van Mythos, binnen afzienbare tijd ook aan het grotere publiek, voor dit dreigingsbeeld?

Antwoord op vraag 3
Geavanceerde AI-modellen spelen in algemene zin op twee manieren een rol in het huidige dreigingsbeeld. Enerzijds kunnen AI-modellen worden ingezet door kwaadwillenden om systemen aan te vallen. Zo kan AI worden ingezet om kwetsbaarheden op te sporen, code te ontwikkelen, phishing te personaliseren of grote hoeveelheden data te verwerken. Dit kan ervoor zorgen dat actoren sneller en op grotere schaal cyberaanvallen kunnen uitvoeren. Anderzijds kunnen AI-modellen worden ingezet ter verdediging. Zo kan AI worden ingezet voor detectie van afwijkend netwerkverkeer, het geautomatiseerd controleren van systemen of het reageren op incidenten. Ook hiervoor geldt dat dergelijke processen sneller en op grotere schaal kunnen worden uitgevoerd. De mogelijke risico's zoals beschreven in het antwoord op vraag 2 gelden voor de bredere trend waarbinnen dergelijke geavanceerde AI-systemen breder beschikbaar worden, waaronder de uitrol van Mythos.

Vraag 4
Bent u van mening dat overheden toegang moeten krijgen tot Mythos zodat zij het kunnen gebruiken om preventief kwetsbaarheden op te sporen en te dichten? Kan dit op een veilige en verantwoorde manier?

Antwoord op vraag 4
Het kabinet bepleit terughoudendheid ten aanzien van toegang tot operationeel gebruik van een niet-Europees leveranciersmodel als oplossingsrichting voor preventieve kwetsbaarheidsdetectie. Drie overwegingen liggen hieraan ten grondslag.

In de eerste plaats laat onafhankelijk onderzoek zien dat toegang tot AI-frontiermodellen in veel gevallen niet noodzakelijk is voor effectieve detectie van kwetsbaarheden. Het verschil tussen frontiermodellen en minder geavanceerde modellen is mogelijk minder groot dan de berichtgeving rond grote modelaankondigingen suggereert. In de tweede plaats zijn er nu aanbieders, van modellen zoals Mythos, die de toegang en voorwaarden bepalen. Toegang via dergelijke constructies kan daarmee ook de afhankelijkheid van die (niet-Europese) partijen vergroten. In de derde plaats kan preventieve kwetsbaarheidsdetectie via AI veilig en verantwoord worden ingericht, mits losgekoppeld van één specifieke leverancier. Het is afhankelijk van de specifieke omstandigheden, voorwaarden en afhankelijkheden en telkens een brede weging in welke mate toegang noodzakelijk is.

Vraag 5
Hoe bereidt u overheidsorganisaties voor op de cyberveiligheidsrisico’s die gepaard gaan met de uitrol van Mythos? Kunt u uiteenzetten welke acties u neemt om de veiligheid van persoonsgegevens van burgers en de ICT-processen van de overheid te garanderen?

Antwoord op vraag 5
Het beschermen van persoonsgegevens van burgers en het waarborgen van de veiligheid van systemen van de overheid vraagt voortdurende aandacht. Zoals beschreven in het antwoord op vraag 2 komen de risico's die van invloed zijn op het waarborgen hiervan niet voort uit één specifiek geavanceerd AI-model zoals Mythos. Om die reden is het noodzakelijk dat overheidsorganisaties rekening houden met bredere ontwikkelingen die mogelijke risico's zijn voor de bescherming van persoonsgegevens of de beveiliging van hun netwerk- en informatiesystemen.

Voor de bescherming van persoonsgegevens geldt dat dit wordt geregeld in de Algemene verordening gegevensbescherming (AVG). De AVG kent onder meer de verplichting tot het treffen van passende technische en organisatorische maatregelen ter beveiliging bij de verwerkingen van persoonsgegevens. Daarbij geldt dat een beoordeling van het passende beveiligingsniveau moet worden uitgevoerd waarbij rekening wordt gehouden met de risico's met betrekking tot de verwerking van persoonsgegevens. Voorafgaand aan een voorgenomen verwerking van persoonsgegevens, die waarschijnlijk een hoog risico inhoudt voor de rechten en vrijheden van betrokkenen, moet voorts een data protection impact assessment (DPIA) uitgevoerd worden om de risico's voor de rechten en vrijheden van betrokkenen in kaart te brengen. Daarbij moet onder meer worden beoordeeld welke maatregelen getroffen worden om de risico's aan te pakken, waaronder veiligheidsmaatregelen en een mechanisme om de bescherming van persoonsgegevens te garanderen. Ook volgt uit de AVG de verplichting om voor het verrichten van verwerkingen van persoonsgegevens uitsluitend gebruik te maken van andere partijen die namens de overheid persoonsgegevens verwerken als verwerkers, indien zij voldoende garanties bieden met betrekking tot het toepassen van passende technische en organisatorische maatregelen, zodat aan de (andere) verplichtingen op grond van de AVG wordt voldaan. Op de naleving van verplichtingen uit de AVG houdt de Autoriteit Persoonsgegevens toezicht.

Vanuit de AI-verordening gelden bovendien regels voor de ontwikkeling en het gebruik van AI-systemen, ook aan de cyberveiligheid van hoog-risico systemen. Bovendien kent de AI-verordening de verplichting tot uitvoering van een Fundamental Rights Impact Assessment (FRIA) bij hoog-risico AI-systemen, om impact van deze systemen op fundamentele rechten van de mens te beoordelen. In de Kamerbrief van 20 april heeft de Staatssecretaris van Digitale Economie en Soevereiniteit u geïnformeerd over de wijze waarop het kabinet voornemens is het toezicht op de naleving van de Europese AI-verordening vorm te geven.[1]

Vanaf het moment van inwerkingtreding van de Cyberbeveiligingswet, waarin de Europese NIS2-richtlijn zal worden geïmplementeerd, gelden op grond van die wet verplichtingen voor de in die wet genoemde organisaties met betrekking tot de beveiliging van hun netwerk- en informatiesystemen. Het voorstel voor deze wet ligt op dit moment ter behandeling voor bij de Eerste Kamer. Deze wet zal gaan gelden voor organisaties uit verschillende sectoren, waaronder ook de overheid. Onderdeel van deze wet zal een zorgplicht zijn die inhoudt dat organisaties die onder de wet vallen verplicht zijn tot het nemen van passende en evenredige, technische, operationele en organisatorische maatregelen, om de risico's met betrekking tot de beveiliging van hun netwerk- en informatiesystemen te beheersen. Daarbij geldt dat organisaties maatregelen moeten nemen die zorgen voor een beveiligingsniveau dat is afgestemd op

[1] Kamerstukken II, 2025/26,

de risico's die voor hun organisatie gelden. Deze risico's kunnen voor die organisaties als gevolg van onder meer veranderende dreigingen of nieuwe technologieën wijzigen, waardoor zij aanvullende maatregelen zullen moeten nemen. Deze zorgplicht wordt in het Cyberbeveiligingsbesluit, de algemene maatregel van bestuur onder de Cyberbeveiligingswet, en in ministeriële regelingen van de voor de sectoren verantwoordelijke vakministers, nader uitgewerkt.

De maatregelen die organisaties, die onder de Cyberbeveiligingswet vallen, moeten nemen betreffen onder meer beleid over risicoanalyses en incidentenbehandeling en de beveiliging van de toeleveringsketen. Krachtens het Cyberbeveiligingsbesluit geldt als maatregel onder meer ook dat organisaties bij attenderingen op voor hun relevante kwetsbaarheden of, dreigingen en beveiligingsadviezen door relevante partijen zoals CSIRTs, moeten beoordelen of op basis hiervan aanpassingen of aanvullingen nodig zijn op hun maatregelen (art. 17 Cbb). Daarnaast is voor overheidsorganisaties die onder de Cbw vallen, voorzien dat in de regeling van de minister van Binnenlandse Zaken en Koninkrijksrelaties als maatregel, die zij in het kader van de zorgplicht moeten nemen, komt te gelden dat zij de normen en overheidsmaatregelen in de Baseline Informatiebeveiliging Overheid (BIO2), het reeds bestaande normenkader voor informatiebeveiliging bij de overheid, toepassen. Op de naleving van de verplichtingen in de Cyberbeveiligingswet en de daar onderliggende regelgeving wordt toezicht gehouden. Voor overheidsorganisaties zal de Rijksinspectie voor Digitale infrastructuur (RDI) als toezichthouder optreden.

Terugkomend in bovengenoemde wet- en regelgeving is dat gegevensbescherming en beveiliging van netwerk- en informatiesystemen een continue en cyclisch proces is. Organisaties dienen voortdurend te toetsen of bestaande maatregelen nog steeds voldoende zijn om risico's te beheersen. Ook risico's als gevolg van het bestaan van AI-modellen, waaronder Mythos, zullen onderdeel zijn van deze continue risico-inschatting door organisaties.

Vraag 6
Welke rol zou een onafhankelijke AI-raad, zoals voorgesteld in de motie-Kathmann/Six Dijkstra (Kamerstuk 26643, nr. 1403), kunnen spelen om de veiligheidsrisico's van geavanceerde AI te monitoren en af te dekken? Hoe wordt deze motie nu uitgevoerd?

Antwoord op vraag 6
Het kabinet onderschrijft het belang van het monitoren en beheersen van veiligheidsrisico's van geavanceerde AI en deelt de opvatting dat dit een aanpak vereist die stevig is verankerd in inhoudelijke expertise. Mede naar aanleiding van de aangenomen motie over een onafhankelijke AI-raad[2], ingediend op 29 september 2025, is gekeken naar een passende structuur om hier invulling aan te geven.

Zoals eerder aan de Kamer gemeld[3], werd in eerste instantie bezien welke rol de NDS-raad hierin kon vervullen. Nu besloten is om de oprichting van de raad niet te formaliseren[4], worden alternatieve invullingen verkend. Zo loopt er tot september 2026 onder andere een interdepartementale verkenning, uitgevoerd door de Digitale Doetank, naar de inrichting van een onafhankelijk AI-veiligheidsinstituut, naar voorbeeld van internationale initiatieven zoals het AI Safety Institute (AISI) in het Verenigd

[2] *Kamerstuk 26643, nr. 1403*
[3] *Kamerstukken II* 2025/26, 26643, nr. 1441
[4] *Kamerstuk 26643, nr. 1510*

Koninkrijk. Het kabinet houdt daarbij rekening met de verschillende bestaande instituten die de overheid en private organisaties reeds voorzien van advies en expertise op het gebied van AI.

Vraag 7
Heeft u voldoende zicht op de risico's van model leakage, model theft en ongeautoriseerde verspreiding van geavanceerde AI-systemen in Nederland en Europa? Zo ja, hoe wordt dit inzicht benut voor beleid en toezicht? Zo nee, welke maatregelen neemt u om dit inzicht te verbeteren?

Antwoord op vraag 7
Aanbieders van bepaalde AI-systemen, zoals systemen met een hoog risico en AI-modellen voor algemene doeleinden, inclusief modellen met een systeemrisico, moeten op grond van de Europese AI-verordening bepaalde informatie over deze systemen openbaar maken. De regels voor AI-modellen voor algemene doeleinden en die met een systeemrisico zijn op 2 augustus 2025 in werking getreden. Voor krachtige AI-modellen met systeemrisico's gelden volgens deze verordening aanvullende verplichtingen om deze risico's in kaart te brengen. Hierdoor hebben lidstaten en de Europese Commissie meer zicht op deze risico's. De AI-verordening kent hiervoor een toezichtstelsel, waarbij zowel nationale toezichthouders als de Europese Commissie nauwlettend toezicht houden op deze risico's. Zo kunnen zij hoog-risico AI-systemen die niet aan de verplichtingen voldoen van de markt halen. Daarnaast vindt er kennisuitwisseling plaats tussen (nationale) toezichthouders, de Europese Commissie en experts uit de lidstaten binnen de door de AI-verordening opgerichte Comité voor Artificiële Intelligentie *(European AI Board)*.

Vraag 8
Hoe beoordeelt u de toereikendheid van bestaande beveiligingsnormen en toezichtmechanismen voor ontwikkelaars en beheerders van krachtige AI-modellen, mede in relatie tot de implementatie van de AIverordening?

Antwoord op vraag 8
De Europese AI-verordening stelt regels aan AI-modellen voor algemene doeleinden waarbij de regels steviger zijn voor dit soort modellen met systeemrisico's, deze regels zijn op 2 augustus 2025 in werking getreden. Hier vallen de meest krachtige AI-modellen onder. Alle AI-modellen voor algemene doeleinden moeten technische documentatie opstellen waarin in ieder geval informatie staat over het beoogde doel van het model, trainingsmethoden en gebruikte data. Aanbieders van modellen met een systeemrisico moeten daarbovenop risico's in kaart brengen en kwetsbaarheden in het model opsporen. Daarnaast moeten zij ernstige incidenten en mogelijk corrigerende maatregelen bijhouden. Het toezicht op al deze modellen is Europees centraal belegd bij de Europese Commissie. Daarnaast kunnen aanbieders aansluiten bij de Europese *General-Purpose AI Code of Practice* die op 10 juli 2025 is gepubliceerd. Deze vrijwillige praktijkcode bevat onder andere informatie over hoe aanbieders kunnen voldoen aan de beveiligingsnormen uit de AI-verordening. Het kabinet is positief over deze vereisten en stimuleert aanbieders om zich aan te sluiten bij de praktijkcode. Het is nog niet mogelijk een sluitend oordeel te vellen over de effectiviteit van deze eisen. Samen met de

Europese Commissie en andere lidstaten houden we de werking van deze vereisten nauwlettend in de gaten en dragen we bij aan de evaluaties hiervan.

Vraag 9
Welke kansen ziet u om via veilige ontwikkeling en deployment van AI de digitale veiligheid te versterken, bijvoorbeeld voor cyberdetectie, opsporing en publieke dienstverlening?

Antwoord op vraag 9
Zie antwoord vraag 3.

Vraag 10
Ziet u in deze casus aanleiding om in Europees verband te pleiten voor versterkte samenwerking rond monitoring van toegangsbeheer, auditing en incidentrespons? Zo ja, op welke wijze?

Antwoord op vraag 10
Organisaties zijn in de eerste plaats zelf verantwoordelijk voor hun eigen digitale weerbaarheid en daarmee onder meer ook voor het voorkomen en detecteren van cyberincidenten. De recente ontwikkelingen rondom AI benadrukken nogmaals het belang van het versterken van de digitale weerbaarheid van organisaties en de samenwerking daartoe bij digitale dreigingen, kwetsbaarheden en incidenten. In Nederland zijn er verschillende organisaties, zoals CSIRTS, die organisaties kunnen ondersteunen door het geven van adviezen en handelingsperspectieven, het verstrekken van informatie over dreigingen en incidenten en het assisteren bij incidentmanagement.

In het geval van grensoverschrijdende incidenten coördineert het Nationaal Cyber Security Centrum (NCSC), vanuit de Nederlandse overheid, de samenwerking met internationale(cyber) partners en werkt daartoe intensief samen met die partijen in onder meer het Europese CSIRT-netwerk. Binnen dit netwerk kunnen CSIRTs snel en effectief operationele informatie met elkaar uitwisselen met ondersteuning door het European Network and Information Security Agency (ENISA). Bovendien wordt binnen dit netwerk en diverse andere EU-samenwerkingsverbanden (zoals de NIS Cooperation Group en het CyCLONe-netwerk) aandacht besteed aan de invloed van de ontwikkeling van AI-modellen op cybersecurity.

Vraag 11
Hoe beoordeelt u de wenselijkheid van meer transparantieverplichtingen voor aanbieders van geavanceerde AI-systemen over beveiligingsmaatregelen, incidenten en misbruikrisico's?

Antwoord op vraag 11
Zoals ook opgenomen in het antwoord op vraag 8 is het nog te vroeg om aanvullende eisen op te stellen. De verplichtingen vanuit de AI-verordening zijn immers pas recent van kracht geworden.

Vraag 12
Kunt u de vragen afzonderlijk beantwoorden en in ieder geval vóór het rondetafelgesprek cyberveiligheid en informatiebeveiliging van 20 mei 2026?

Antwoord op vraag 12
Aangezien niet alle benodigde informatie op tijd was ontvangen is het niet gelukt de beantwoording met uw Kamer te delen voor het rondetafelgesprek cyberveiligheid en informatiebeveiliging van 20 mei 2026.