



AUTORITEIT  
PERSOONSGEGEVENS

# Manipulatieve, misleidende en uitbuitende AI-systemen

## Samenvatting van reacties en vervolgstappen

Autoriteit Persoonsgegevens - Directie Coördinatie Algoritmes (DCA)

December 2024  
DCA-2024-04

---



## Samenvatting en vervolgstappen

Dit document biedt een samenvatting van reacties en benoemt vervolgstappen in relatie tot de eerdere oproep tot input met betrekking tot de eerste twee verboden in artikel 5 van de AI-verordening (zie document DCA-2024-01). Het gaat daarbij om AI-systemen die gebruikmaken van manipulatieve of misleidende technieken, alsmede AI-systemen die gebruikmaken van kwetsbaarheden van mensen. Overkoepelend destilleert de AP drie hoofdobservaties uit de ontvangen reacties.



Hoofdobservatie 1 – Veel respondenten merken op dat de praktijken waarop de verboden betrekking hebben reeds door andere horizontale of sectorspecifieke wet- en regelgeving worden gereguleerd of verboden.

Voor de verboden in kwestie geldt dat er een nauw verband is met regelgeving ter bescherming van consumenten, zoals oneerlijke handelspraktijken, de Digitale dienstenverordening (DSA) en financiële regelgeving. Ook raken de praktijken aan de rechtmatige verwerking van persoonsgegevens zoals geregeld in onder meer de Algemene Verordening Gegevensbescherming (AVG). Volgens de AP hebben de verboden in de AI-verordening een preventief karakter (de AI-systemen mogen niet worden aangeboden of gebruikt) en daarmee vormen de verboden een aanvulling op voornoemde regelgeving. Zij bieden zo een mogelijkheid om effectiever toezicht te houden op gevallen waarin AI-systemen een rol spelen in gevallen van misleiding, manipulatie en uitbuiting van mensen. Daarbij is het belangrijk dat de verboden in de AI-verordening zo veel mogelijk in samenhang wordt uitgelegd en toegepast met deze regelgeving. Dat vergt nauwe samenwerking en afstemming tussen de ACM, AFM en de AP.



Hoofdobservatie 2 – Verschillende respondenten wijzen op de aanvullende mogelijkheden die AI biedt om mensen op overtuigender wijze te misleiden, te manipuleren of uit te buiten.

Een veel genoemd voorbeeld daarbij is het gebruik van *dark patterns*, waarbij interfaces zo worden ontworpen of georganiseerd dat ze mensen misleiden of te manipuleren door hen keuzes of beslissingen te laten maken die ze anders niet zouden nemen. AI-technologie kan manipulerende of misleidende patronen versterken door in de interactie met de gebruiker bijvoorbeeld nog beter in te spelen op eigenschappen of kwetsbaarheden van een persoon.



Hoofdobservatie 3 – Respondenten merken op dat de verbodsbepalingen niet worden uitgewerkt in standaarden en dus in richtsnoeren verder worden verduidelijkt moeten worden.

De verboden in de AI-verordening zijn nieuw en bevatten verbanden en begrippen die – in de AI-verordening zelf – maar beperkt geoperationaliseerd worden. Dat leidt ertoe dat er ook een aantal onderwerpen zijn die verdere verduidelijking behoeven, valt op in de inbreng van respondenten. Dan gaat het bijvoorbeeld om de mate waarin er een verband dient te bestaan tussen het AI-systeem en de misleiding of manipulatie; de omvang van het schadebegrip; wanneer er sprake is van aanzienlijke schade; wat precies kan worden verstaan als een subliminale of misleidende techniek; op welke kwetsbaarheden het verbod betrekking heeft; en of een gedragsverstoring ook geleidelijk plaats kan hebben.

De AP gebruikt deze inzichten ter ondersteuning van de voorbereiding van Nederlandse toezichthouders op het toezicht op de verboden en zal de ontvangen input de komende maanden verder betrekken in de voorlichting en uitleg over de verboden. Hierbij vindt samenwerking en afstemming met andere toezichthouders plaats in onder andere in de AI- en Algoritmekamer van het Samenwerkingsplatform Digitale Toezichthouders (SDT). Daarnaast biedt de input een basis voor de bijdrage van de AP vanuit de praktijk in de discussie over de richtsnoeren over de verboden, waaraan de Europese Commissie werkt.



# I. Achtergrond

## AI-verordening & verboden AI

1. **Sinds 1 augustus 2024 is de Europese AI-verordening (2024/1689) van kracht.** Deze verordening bevat regels over het aanbieden en gebruiken van artificiële (kunstmatige) intelligentie (AI) in de EU. Het uitgangspunt van de AI-verordening is dat er vele nuttige toepassingen van AI zijn, maar dat de technologie ook risico's met zich meebrengt, die beheerst moeten worden. Sommige AI-systemen met een onaanvaardbaar risico worden verboden.

## Oproep tot input

2. **Op 27 september 2024 deed de AP een oproep tot input met betrekking tot de eerste twee verboden in artikel 5 van de AI-verordening.** Het gaat daarbij om AI-systemen die gebruikmaken van manipulatieve of misleidende technieken (hierna: verbod A), alsmede AI-systemen die gebruikmaken van kwetsbaarheden van mensen (hierna: verbod B).
3. **De oproep tot input heeft als doel informatie en inzichten op te halen bij van belanghebbenden (burgers, overheden, bedrijven en andere organisaties) en organisaties die hen vertegenwoordigen.** Hiermee beoogt de AP informatie in te zamelen om een basis te leggen voor de voorbereiding van de verdere uitleg van de verboden in de AI-verordening. Ook bovengenoemde oproep tot input is onderdeel van een reeks aan oproepen over de verschillende verboden in de AI-verordening. Meer informatie over de AI-verordening, de verboden en de rol van de AP in het toezicht is te vinden op de [website van de AP](#) en in de eerder gepubliceerde [oproep tot input](#).
4. **De AP heeft tussen 27 september en 17 november 2024 vijftien reacties ontvangen van verschillende organisaties en individuen met uiteenlopende achtergronden, waaronder ook academici.** De AP is blij dat de oproep zowel burgers, bedrijven, onderzoekers en maatschappelijke organisaties heeft weten te bereiken en dat zij een bijdrage hebben willen leveren met de inzending van input.
5. **Volgend op de oproep tot input is dit samenvattende document opgesteld.** Alle input is door de AP doorgenomen en de hoofdpunten daaruit zijn in dit document opgenomen. In het document wordt enkel in generieke termen verwezen naar de indieners van de ontvangen input. Op een beperkte aantal plekken wordt in dit document de input ook door de AP op waarde geschat. Dat is gedaan waar dat passend is en de inbreng van de AP direct is terug te voeren op de AI-verordening of waar een (aanvullend) gezichtspunt kan worden geboden. Daarmee beoogt de AP bij te dragen aan de discussie over de uitleg van de verboden. Met dit document wordt door de AP niet beoogd al een uitleg te geven aan de verboden of aanverwante regelgeving.

## Toezicht op naleving van de AI-verordening

6. **De AP deed deze oproep tot input vanuit haar rol als coördinerend toezichthouder op algoritmes en AI.** Deze taken zijn binnen de AP belegd bij de directie Coördinatie Algoritmes (DCA). De oproep tot input ligt in het verlengde van het voorbereidende werk dat wordt gedaan ten behoeve van het toezicht op verboden AI-systemen onder de AI-verordening. De regering werkt op dit moment aan de formele aanwijzing van nationale toezichthouders voor de AI-verordening. De AP (vanuit de directie Coördinatie Algoritmes) en de Rijksinspectie Digitale Infrastructuur (RDI) hebben hierover een [advies](#) gegeven in samenwerking en afstemming met andere toezichthouders. Geadviseerd is onder andere om de AP aan te wijzen als markttoezichtautoriteit voor het merendeel van de verboden in artikel 5.



## II. Samenvatting en waardering van ingezonden input

### Algemeen

Verband met andere wet- en regelgeving

Er wordt door verschillende respondenten opgemerkt dat de praktijken waarop de verboden in kwestie betrekking hebben soms reeds door andere horizontale of sectorspecifieke wet- en regelgeving gereguleerd of verboden worden.

7. **Voorbeelden zijn de regels in het consumentenrecht, bijvoorbeeld als het gaat om oneerlijke handelspraktijken of *dark patterns* (o.a. art. 25 DSA), of (specifieke) financiële regelgeving die consumenten beschermt tegen misleiding en manipulatie.** De AP wijst erop dat daarvoor reeds meerdere richtsnoeren beschikbaar zijn, zoals de [Leidraad Bescherming Online Consument](#) van de Autoriteit Consument en Markt (ACM) of de [Beleidsregels](#) van Autoriteit Financiële Markten (AFM) met betrekking tot informatieverstrekking in de financiële sector. Ook worden verwerkingen van persoonsgegevens die ten grondslag liggen aan verboden praktijken gereguleerd door de AVG en zullen verwerkingen van persoonsgegevens ten behoeve van de verboden praktijken ook onrechtmatig zijn.
8. **Volgens de AP is het bijzondere aan de verboden in de AI-verordening dat zij, in de gevallen dat manipulatie, misleiding en uitbuiting een AI-systeem in het spel is, betrekking hebben op zowel het aanbieden als het gebruik van dergelijke systemen.** Daarom kunnen dus zowel aanbieders als gebruiksverantwoordelijken worden aangesproken. Daarmee kan er op verschillende plekken in de AI-waardeketen worden opgetreden tegen schadelijke praktijken met AI. Verder gaat het in het geval van de verboden om veel verschillende soorten schade. Daarnaast hebben de verboden in kwestie ook betrekking op benadeling van *groepen* van personen. Het is daardoor mogelijk minder ingewikkeld om op te treden tegen de verboden AI-praktijken dan op grond van andere juridische kaders, welke zich bijvoorbeeld niet richten op aanbieders, betrekking hebben op psychologische schade of gaan over groepen van personen. Daarom bieden de verboden – in aanvulling op, en uitgelegd in samenhang met andere regels die zien op soortgelijke praktijken – een middel om effectiever toezicht te houden op AI-systemen.
9. **Aansluiting van en tussen regelgeving is belangrijk.** De verboden zullen volgens de AP zo veel mogelijk in overeenstemming en in samenhang met, die bestaande (Europese) regelgeving op onder meer het gebied van consumentenbescherming en financiële dienstverlening, de DSA en de AVG uitgelegd en toegepast moeten worden.
10. **Het toezicht op de verboden in de AI-verordening zou zich daarbij vooral moeten toespitsen op de praktijken waarbij AI-technologie een belangrijke rol speelt.** Dat vergt nauwe samenwerking en afstemming tussen de toezichthouders, waaronder in het bijzonder ACM en AFM, en daarvoor spannen zij zich reeds in.



## Verhouding tot hoog-risico toepassingen en andere verboden

Door respondenten wordt opgemerkt dat de verboden zich ook verhouden tot de hoog-risico toepassingen en andere verboden en dat daarover duidelijkheid moet worden gegeven.

11. **Respondenten werpen de vraag op of de uitzondering voor AI-systemen die geen significant risico op schade voor de gezondheid, veiligheid of de grondrechten van natuurlijke personen inhouden (art. 6 derde lid van de AI-verordening), ook geldt voor de verboden.** Dergelijke systemen worden dan niet als hoog-risicosystemen beschouwd. Volgens de AP is deze uitzondering alleen van toepassing op AI-systemen die anders als hoog risico AI-systemen zouden worden beschouwd (Hoofdstuk III van de AI-verordening). Deze uitzondering is daarom niet van toepassing op de verboden in de AI-verordening. Het zou ook niet passend zijn, omdat de praktijken niet voor niets verboden zijn.
12. **Verskillende respondenten wijzen voor onderhavige verboden ook op het verband met de verboden en hoog risico AI-systemen voor emotieherkenning.** Daarbij gaat het dan met name om de mogelijkheden die emotieherkenning biedt om beter in te kunnen spelen op de kwetsbaarheden van mensen, dan wel hen effectiever te kunnen manipuleren of misleiden.

## Betrekken van professionals

Door een respondent is opgemerkt dat ontwikkelingen op het gebied van AI-technologie elkaar snel opvolgen en dat er tegelijkertijd behoefte is aan duidelijkheid over de toepassing van regelgeving en best practices.

13. **AI-professionals kunnen een bijdrage leveren aan verduidelijking.** De AP vindt die dialoog met stakeholders van belang zodat de AP, samen met andere toezichthouders, een bijdrage kunnen leveren aan de uitleg en guidance over de AI-verordening. De respondent onderstreept ook het belang van educatie en praktijkopleiding. Volgens de AP sluit dat punt goed aan bij het onderwerp van AI-geletterdheid (art. 4 van de AI-verordening) en de stappen die organisaties zouden moeten zetten om te zorgen voor AI-geletterdheid.

## Specifieke punten over de criteria in de verboden

Criterium 1 (verbod A en B): op AI gebaseerde manipulatieve, misleidende of uitbuitingspraktijken

Verskillende respondenten geven aan dat manipulatieve, misleidende of uitbuitingspraktijken steeds vaker worden gebaseerd op AI en dat inzet van AI de risico's vergroot. Volgens respondenten maakt de inzet van AI, in combinatie met het gebruik van beschikbare persoonsgegevens, het mogelijk om mensen te benaderen en met hen te interacteren als zij op hun zwakst of kwetsbaarst zijn. Ook kunnen AI-systemen vanwege hun adaptiviteit ook goed inspelen op iemands omstandigheden of kwetsbaarheden. Verduidelijking lijkt met name nodig op het punt van de samenhang tussen de manipulatieve, misleidende of uitbuitingspraktijk en het AI-systeem.

14. **Respondenten merken vaak de aanvullende mogelijkheden op die AI biedt om *dark patterns* te gebruiken, waarbij interfaces worden ontworpen of georganiseerd om mensen te misleiden of te manipuleren om hen keuzes of beslissingen te laten maken die ze anders niet zouden nemen.** AI-technologie kan dan bijvoorbeeld worden ingezet om een interface en daarin aangeboden informatie verder toe te spitsen op



eigenschappen of kwetsbaarheden van een persoon, waardoor manipulerende of misleidende patronen in de interface worden versterkt.

15. **Verskillende respondenten wijzen ook op de mogelijkheden die generatieve AI biedt om mensen op overtuigender wijze te misleiden of te manipuleren.** Respondenten wijzen daarbij ook op de mogelijkheden om met behulp van AI misinformatie of manipulerende inhoud te verspreiden of aan te bevelen, of mensen bijvoorbeeld ertoe te bewegen hun stemgedrag te veranderen. Een individuele respondent geeft ook aan dat het risico van misleiding, manipulatie of misleiding groter is naarmate AI-technologie dieper geïntegreerd is andere applicaties, dan wanneer de technologie oppervlakkig en zichtbaar aanwezig is in bijvoorbeeld een chatbot. Veel van deze zorgen raken ook aan de aan de regulering van 'AI-modellen voor algemene doeleinden' (*General Purpose AI models*) en de transparantieplichtingen in artikel 50 van de AI-verordening. Ook het toezicht van de ACM en de [Leidraad van de ACM met betrekking tot de Digitaledienstenverordening](#) zijn hier relevant, die voor een belangrijk deel betrekking hebben op online platforms en bijvoorbeeld de verspreiding van mis- en desinformatie.
16. **Sommige respondenten geven aan dat het benodigde verband tussen een AI-systeem en manipulatie, misleiding en uitbuiting verduidelijking behoeft.** Het is volgens respondenten dus de vraag of en wanneer praktijken waarbij sprake is van manipulatie, misleiding of uitbuiting ook altijd een verband hebben met het gebruik van AI. Een respondent merkt ook op dat dit aspect van de verboden cruciaal is voor de reikwijdte van de verboden. Deze respondent merkt ook op dat het gebruik van *andere eenvoudiger technologieën* in combinatie met AI ook kan leiden tot manipulatie, misleiding en uitbuiting. Volgens de AP toont dat aan hoe belangrijk de uitleg van het begrip 'AI-systeem' is voor de toepasselijkheid van de verboden. Ook ondersteunt dat volgens de AP het belang om bij de toekomstige uitleg van dat begrip te kiezen voor een integrale benadering met betrekking tot praktijken waarbij AI-technologie wordt ingezet, om zo de risico's van de inzet van bijvoorbeeld misleidende AI-systemen het hoofd te kunnen bieden.

#### criterium 2 (verbod A): Gebruik maken van subliminale en misleidende technieken

**Respondenten geven aan dat er – in het kader van verbod A – nog veel onduidelijkheid is over wat kan worden verstaan onder subliminale technieken waarvan personen zich niet bewust zijn of van manipulatieve of misleidende technieken. Het is volgens hen met name nodig te verduidelijken wat dergelijke technieken inhouden. Ook de mate manipulatie of misleiding beoogd moet zijn geweest vergt verduidelijking.**

17. **Verskillende respondenten noemen in het kader van de vragen over subliminale technieken voorbeelden als systemen die gebruik maken van niet waarneembare toonhoogten of van beelden die zeer snel worden weergegeven.** Als dergelijke technieken inderdaad kunnen leiden tot gedragsverstoring, dan zou volgens de AP verduidelijkt moeten worden of daarbij inderdaad sprake is van subliminale technieken in de zin van de AI-verordening. Ook noemen respondenten wederom het gebruik van dark patterns in hun antwoorden met betrekking tot subliminale technieken.
18. **Een respondent merkt op dat er in verdere verduidelijking moet worden meegenomen hoe het kan dat dit verbod spreekt over 'doelbewuste' manipulatieve of misleidende technieken.** Dit terwijl in de verbodsbepaling in het onderdeel over de gedragsverstoring wordt gesproken van een verstoring die de doelstelling of het effect is van de inzet van het AI-systeem. Volgens deze respondent zou het daarom mogelijk moeten zijn dat ook manipulatieve of misleidende technieken onder het verbod vallen als die alleen als gevolg hebben (en dus niet *per se* als doel hebben) dat gedrag van personen wezenlijk wordt verstoord. Hoe dit criterium, in het bijzonder de 'doelbewustheid' van het gebruik van manipulatieve of misleidende technieken, moet worden uitgelegd is volgens de AP inderdaad een punt van aandacht.



### criterium 3 (verbod B): Gebruikmaken van kwetsbaarheden

In het kader van verbod B merken respondenten op dat AI-technologieën het steeds makkelijker maken om gebruik te maken van kwetsbaarheden van mensen. Volgens respondenten is het verder nodig om duidelijker te maken om wat voor soort kwetsbaarheden het nu precies gaat en of het daarbij ook kan gaan om het gebruik van *proxies* die een indicatie zijn van de kwetsbaarheid van mensen.

19. **Volgens verschillende respondenten is het onduidelijk wat er precies wordt bedoeld met gebruikmaking van kwetsbaarheden.** Een respondent geeft aan dat het zou helpen als er richtsnoeren komen waarmee kan worden beoordeeld of gebruikmaking van kwetsbaarheden plaatsheeft. Volgens een respondent maakt het gebruik van AI en de (persoons)gegevens die daarbij worden verwerkt het mogelijk om meer inzicht te krijgen in omstandigheden van mensen, zoals hun financiële omstandigheden. Dit zou het makkelijker maken om gebruik te maken van kwetsbaarheden die daarmee samenhangen. Een andere respondent merkt op dat grote verzamelingen van data en het gebruik van AI het ook mogelijk maken om op minder zichtbare en subtiele wijze gebruik te maken van de kwetsbaarheden van mensen. Dan wordt gebruik gemaakt van proxy-gegevens die niet direct duiden op de aanwezigheid van een kwetsbaarheid, maar die daarvan wel een indirecte indicator zijn. Volgens de AP wordt daarmee onderkend dat beschikbaarheid van meer (persoons)gegevens en de toegenomen mogelijkheden om die met AI-technologieën te verwerken de risico's op uitbuiting van kwetsbare mensen alleen maar vergroot. Echter, uiteindelijk is het voor de uitleg van dit verbod volgens de AP van belang of door het AI-systeem gebruik wordt gemaakt van kwetsbaarheden die, zoals beschreven in de AI-verordening, het gevolg zijn van iemands leeftijd, handicap of een specifieke sociale of economische omstandigheid. Bij deze beoordeling is het dus niet per se nodig om te beoordelen hoe en op basis van welke (persoons)gegevens er met het AI-systeem wordt bepaald dat er sprake is van kwetsbaarheid.
20. **Tot slot is het volgens respondenten ook zo dat AI kan worden ingezet ter bescherming of ondersteuning van kwetsbare personen.** Zo geeft een respondent aan dat AI kan worden gebruikt voor de tijdige informatievoorziening of signalering richting kwetsbare gebruikers van bepaalde essentiële dienstverlening, wat soms ook een wettelijke eis is die aan partijen wordt gesteld. Daarnaast biedt de inzet van AI-technologie volgens respondenten ook kansen om bijvoorbeeld mensen met een handicap met ondersteunende AI-technologieën fysiek, zintuigelijk of cognitief te ondersteunen.

### criterium 4 (verbod A en B): Aanzienlijke schade

De respondenten geven aan dat het (in bepaalde gevallen) mogelijk ingewikkeld zal zijn om te bepalen wat voor soort schade kan worden opgelopen bij de verstoring van gedrag, en wanneer deze aanzienlijk is.

21. **Volgens een respondent kan aanzienlijke schade, zoals dat in de overwegingen van de AI-verordening is beschreven als "met name met voldoende belangrijke negatieve gevolgen voor de fysieke of psychologische gezondheid of voor financiële belangen", veel verschillende vormen aannemen.** Deze respondent geeft daarbij ook aan dat de overweging een niet limitatieve opsomming geeft van de soorten negatieve gevolgen door schade.
22. **Respondenten hebben géén eenduidig beeld van de reikwijdte van het schadebegrip.** Volgens een respondent zou het schadebegrip bijvoorbeeld op maatschappelijke schade omvatten, bijvoorbeeld als gevolg van desinformatie, terwijl een andere respondent aangeeft dat dit niet het geval kan zijn. Het is volgens de respondenten van belang dat er verduidelijking komt over op welke soorten schade de verboden betrekking hebben.





23. **Opgemerkt wordt ook dat een causaal verband moet worden bepaald tussen de schade en de verstoring van gedrag en dat dit een onderdeel van dit verbod is dat verduidelijking behoeft.** Wat volgens een respondent zou kunnen helpen is om bepaalde ‘vermoedens’ vast te stellen voor het bepalen van deze schade en verduidelijking te geven over welke denkstappen er nodig zijn bij het bepalen van het causale verband.
24. **Een respondent geeft aan dat om te bepalen of er sprake is van aanzienlijke schade, rekening moet worden gehouden met de ‘relativiteit’ van schade.** Volgens de respondent zou sommige schade voor eenzelfde toepassing van AI voor de ene persoon of groep personen niet aanzienlijk kunnen zijn en voor de ander wel. Daarom zou er bij de beoordeling van de aanzienlijkheid van schade een flexibele benadering moeten worden gekozen. Sommige respondenten merken op dat het vooral ingewikkeld zou kunnen zijn om te beoordelen of er sprake is van aanzienlijke schade indien deze immaterieel is. Alle overziend is er onder respondenten dus behoefte om meer duidelijkheid te verkrijgen over de ‘aanzienlijkheid’ van de schade.

Criterion 5 (verbod A en B): Met als doel of effect/gevolg de wezenlijke verstoring van gedrag

**Uit de reacties van verschillende respondenten volgt het criterium van de “de wezenlijke verstoring van gedrag” vragen oproept.**

25. **Een aantal respondenten geeft aan dat het criterium van een ‘wezenlijke verstoring van gedrag’ ook terug is te vinden in het consumentenrecht.** Dit begrip is volgens sommige respondenten dan ook reeds geoperationaliseerd. Daarbij is dat begrip verbonden aan de ‘gemiddelde consument’. Volgens een respondent is de link met het consumentenrecht duidelijker bij verbod A, waar ook wordt toegelicht dat het criterium ziet op het merkbaar belemmeren van het vermogen van personen om een geïnformeerd besluit te nemen, waardoor zij een besluit nemen dat ze anders niet hadden genomen. Een respondent geeft wel aan dat dit verbod een andere ratio heeft dan bepalingen in het consumentenrecht waarbinnen dit begrip reeds wordt gebruikt. Daarom zou dit criterium niet hetzelfde moeten worden geïnterpreteerd als in het consumentenrecht. Volgens de AP moet het begrip ‘wezenlijke verstoring van gedrag’ zo veel als mogelijk in samenhang met bestaande regelgeving uitgelegd worden. Maar de AP merkt daarbij wel op dat de verboden in de AI-verordening ook betrekking hebben op situaties die niet samenhangen met de afname van producten of diensten en de rol van mensen als consument. Daardoor behoeft dit verbod waarschijnlijk aanvullende operationalisering.
26. **Van belang is verder dat dit criterium ertoe strekt dat mensen ‘merkbaar’ worden belemmerd in hun vermogen om een geïnformeerd besluit te nemen.** Uit de input van respondenten volgt de vraag of het verbod alleen van toepassing is als er een duidelijk besluit wordt genomen, zoals bijvoorbeeld het afnemen van een dienst of het kopen van een product, of dat het verbod ook van toepassing op situaties waarin de beïnvloeding leidt tot een geleidelijke gedragsverandering of verandering van gemoedstoestand die schadelijk is of schadelijke gevolgen heeft voor bijvoorbeeld iemands gezondheid of welzijn. Volgens de AP lijkt dus ook verduidelijking nodig met betrekking tot temporale eigenschappen van mogelijke gedragsverstoringen.



## Criterion 6 (verbod A en B): Individuele en groepsschade

**Volgens verschillende respondenten is het van belang om te benadrukken dat het bij beide verboden kan gaan schade die wordt opgelopen door zowel individuen als door groepen van personen.**

27. **Volgens een respondent zal uit moeten worden gelegd hoe verschillende soorten schade zich verhouden tot het individu of juist tot meerdere personen.** Een respondent geeft aan dat extra belangrijk is om daar ook aandacht te vestigen op schade voor groepen van personen. Bij deze toelichting moet volgens een andere respondent ook aandacht zijn voor schade die 'andere' personen kunnen lijden waarvan het gedrag niet wezenlijk is verstoord. Een respondent geeft aan dat er een verschil zou zijn in de actoren die schade kunnen oplopen per verbod. Volgens deze respondent ziet de schade bij verbod B alleen maar op een individu of individuele personen, terwijl die bij verbod A ook kan zien op groepsschade. Volgens de AP lijkt een tekstuele interpretatie van beide verboden die lezing te ondersteunen, maar ook op dit is meer duidelijkheid en uitsluitel wenselijk.

## Reikwijdte van de verboden

**Volgens respondenten zou er meer duidelijkheid moeten komen over de gevallen waarin een AI-systeem niet onder de verbodsbepalingen zou moeten vallen.**

28. **Een respondent geeft daarbij aan dat bij die verduidelijking aandacht moet worden besteed aan het onderdeel over gangbare en legitieme handelspraktijken, waarbij deze respondent benadrukt dat een gangbare handelspraktijk niet per definitie legitiem hoeft te zijn.** Andere respondenten geven aan dat het behulpzaam zou zijn om meer inzicht te geven in de soorten rechtmatige praktijken in het kader van medische behandelingen waaraan de verboden in deze verordening geen afbreuk mogen doen.



### III. Vervolg

29. **Met de publicatie van dit samenvattende document is het ophalen van informatie via de oproep tot input over de verboden met betrekking tot manipulatieve, misleidende en uitbuitende AI-systemen afgerond.** Er volgt een aantal kernpunten uit de ingezonden input, die van belang zijn voor de uitleg die zal worden gegeven aan de verboden in de AI-verordening:
- a. **Beide verboden bevatten veel begrippen en concepten die in de AI-verordening niet (of zeer beperkt) nader worden uitgelegd.** Invulling en uitleg door (Europese) richtsnoeren, en nadere duiding met concrete voorbeelden, is cruciaal om ervoor te zorgen dat deze verboden bescherming bieden aan burgers en dat er rechtszekerheid wordt geboden aan marktpartijen. Dan gaat het bijvoorbeeld om reikwijdte van de AI-systeem definitie, de technieken om te manipuleren of te misleiden, het gebruikmaken van kwetsbaarheden en de schade die ook door groepen opgelopen kan worden.
  - b. **Met name het verbod met betrekking tot manipulerende en misleidende AI (art. 5, eerste lid, onder a van de AI-verordening) heeft verband met regelgeving op het gebied van consumentenbescherming, financiële dienstverlening en de DSA.** Verduidelijking is nodig over hoe deze verboden zich verhouden tot begrippen in die regelgeving, gegeven het eigenstandige karakter en de ratio van de verboden in de AI-verordening. Het gaat in dit geval bijvoorbeeld om de betekenis van aanzienlijke schade of de wezenlijke verstoring van gedrag. In dit verband behoeven veel genoemde praktijken, zoals het gebruik van dark patterns, bijzondere aandacht.
  - c. **De verboden verhouden zich ook tot andere regels in de AI-verordening, waarop soms ook op EU-niveau toezicht zal worden gehouden.** Over dit verband en de reikwijdte van deze bepalingen moeten de verschillende markttoezichtautoriteiten en de AI Office duidelijkheid verschaffen. Het gaat daarbij om de verhouding met bijvoorbeeld de transparantieverplichtingen (artikel 50), andere verbodsbepalingen of hoog risico toepassingen (zoals bij emotieherkenning) en om de regels voor AI-modellen voor algemene doeleinden.
30. **De AP werkt aan de voorbereiding van het toezicht in Nederland op de verboden en zal de ontvangen input de komende maanden verder betrekken de voorbereiding van de voorlichting en uitleg over de verboden.** Met Nederlandse toezichthouders wordt de opgedane kennis gedeeld in het [Samenwerkingsplatform Digitale Toezichthouders \(SDT\)](#) en wordt in de AI- en Algoritmekamer (AAK) van het SDT besproken welke vervolgacties nodig zijn om de verboden in de AI-verordening te verduidelijken. Die bovenstaande aanpak sluit aan bij de visie van toezichthouders om te streven naar de eenduidige uitleg en toepassing van regelgeving die gelijktijdig van toepassing kan zijn op de verboden AI-praktijken.
31. **De AP zal verzamelde inzichten uit de praktijk ook gebruiken als basis voor bijdrage aan de discussie over de richtsnoeren over de verboden. De Europese Commissie is voornemens om de eerste richtsnoeren begin 2025 publiceren.** Daartoe is de Europese Commissie een consultatie gestart voor belanghebbenden, waaronder aanbieders van AI-systemen, bedrijven, nationale autoriteiten, academici, onderzoeksinstellingen en het maatschappelijk middenveld. Meer informatie over deze richtsnoeren en consultatie zijn te vinden op de [website van de Europese Commissie](#).