Jasper Klopper & Dirk Visser[1]

Vermelding gebruik AI bij AI-gegenereerde content

Bb 2024/51

Wanneer moet bij het gebruik van met AI gegenereerde content (tekst, afbeeldingen, audio en video) het gebruik van AI worden vermeld? Wat zegt de AI Act daarover en wat vloeit voort uit andere regelgeving? Dit artikel geeft antwoord op deze vragen.

1. Inleiding

Met de komst van AI-toepassingen als ChatGPT, DALL-E en Midjourney en tal van andere programma's, is het mogelijk om gemakkelijk, snel en goedkoop content (tekst, afbeeldingen, audio en video) te genereren. Die door AI gegenereerde content (ook wel 'output' genoemd) kan ook commercieel worden ingezet. Dat gebeurt dan ook al op grote schaal, en de verwachting is dat het gebruik van AI-gegenereerde content alleen maar zal toenemen.[2] In een eerdere bijdrage in dit tijdschrift hebben wij een overzicht gegeven van waar de gebruiker van dergelijke content op moet letten.[3] Daarbij kwamen onder meer de algemene voorwaarden van de gebruikte AI-tool, auteursrechten en portretrechten aan de orde. Inmiddels is per 1 augustus 2024 de Europese AI Act in werking getreden.[4] Deze verordening bevat verschillende verplichtingen voor gebruikers van AI-gegenereerde content, waaronder transparantieverplichtingen.[5] In dit artikel gaan wij verder in op die transparantieverplichtingen. Daarbij behandelen wij niet alleen de transparantieverplichtingen zoals die in de AI Act zijn opgenomen. Ook de algemene regels van het reclamerecht en oneerlijke handelspraktijken komen aan bod. Die regels zijn immers nauw verwant aan de transparantieverplichtingen waarin de AI Act voorziet én zijn ook relevant voor commercieel gebruik van AI-output.

2. De AI Act

Het doel van de AI Act is het bevorderen van de ontwikkeling en toepassing van veilige en betrouwbare AI-systemen door zowel publieke als private partijen. Daarbij gaat de verordening uit van een risicogebaseerde aanpak: hoe groter de risico's (op schade voor de samenleving) van een bepaald AI-systeem, hoe strenger de regels waaraan dat systeem en het gebruik daarvan is gebonden.[6]

Welke soorten AI-systemen in de AI Act worden geïdentificeerd en welke verplichtingen daarbij horen, gaat het bestek van deze bijdrage te buiten. Wij richten ons op de transparantieverplichtingen die gelden voor AI-systemen die content kunnen genereren. Deze verplichtingen zijn te vinden in artikel 50 van de AI Act, dat zich richt op zowel aanbieders als gebruikers van AI-systemen. Het beoogde doel van artikel 50 is dat degene die met AI in aanraking komt, – dat kan een klant zijn die contact heeft met een chatbot van de klantenservice van zijn bank, maar ook iemand die op YouTube een advertentie ziet die door AI is gecreëerd – steeds weet dat hij of zij met AI te maken heeft. In de kern gaat het er dus om dat het publiek niet wordt misleid.
Voor het (commercieel) gebruik van door AI gegenereerde output is met name lid 4 van artikel 50 relevant. Dit luidt als volgt:

> "Gebruiksverantwoordelijken van een AI-systeem dat beeld-, audio- of videocontent genereert of bewerkt die een deepfake vormt, maken bekend dat de content kunstmatig is gegenereerd of gemanipuleerd. Deze verplichting geldt niet wanneer het gebruik bij wet is toegestaan om strafbare feiten op te sporen, te voorkomen, te onderzoeken of te vervolgen. Wanneer de content deel uitmaakt van een kennelijk artistiek, creatief, satirisch, fictief of analoog werk of programma, zijn de transparantieverplichtingen van dit lid beperkt tot de openbaarmaking van het bestaan van dergelijke gegenereerde of bewerkte content op een passende wijze die de weergave of het genot van het werk niet belemmert.
> Gebruiksverantwoordelijken van een AI-systeem dat tekst genereert of bewerkt die wordt gepubliceerd om het publiek te informeren over aangelegenheden van algemeen belang, maken bekend dat de tekst kunstmatig is gegenereerd of bewerkt. Deze verplichting is echter niet van toepassing wanneer het gebruik bij wet is toegestaan om strafbare feiten op te sporen, te voorkomen, te onderzoeken of te vervolgen of wanneer de door AI gegenereerde content een proces van menselijke toetsing of redactionele controle heeft ondergaan en wanneer een natuurlijke of rechtspersoon redactionele verantwoordelijkheid draagt voor de bekendmaking van de content."

Artikel 50 lid 4 richt zich dus op twee soorten gebruik: (i) wanneer AI wordt gebruikt om zogeheten 'deepfakes' te creëren en (ii) wanneer AI wordt ingezet om nieuwsberichten op te stellen. Beide soorten gebruik worden hieronder verder besproken.

3. Deepfakes

De AI Act definieert een deepfake als "door AI gegenereerd of gemanipuleerd beeld-, audio- of videomateriaal dat een gelijkenis vertoont met bestaande personen, voorwerpen, plaatsen, entiteiten of gebeurtenissen, en door een persoon

---

1 Jasper Klopper & Dirk Visser zijn advocaten in Amsterdam. Dirk Visser is daarnaast hoogleraar IE in Leiden.

2 Er komen immers steeds nieuwe applicaties bij, en bestaande applicaties worden in rap tempo intelligenter en dus breder inzetbaar.

3 'Commercieel gebruik van beeldmateriaal gemaakt met AI', *Bb* 2023/42.

4 Verordening (EU) 2024/1689 tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie.

5 De meeste bepalingen zullen vanaf 2 augustus 2026 worden gehandhaafd. Voor die tijd kunnen de bepalingen van de AI Act al wel worden toegepast ter invulling van reeds bestaande zorgvuldigheidsnormen.

6 Bepaalde toepassingen van AI (zoals cognitieve gedragsmanipulatie) worden zelfs helemaal verboden, omdat de risico's voor de samenleving onaanvaardbaar worden geacht.

ten onrechte voor authentiek of waarheidsgetrouw zou worden aangezien".[7] Een deepfake is dus *iedere* door AI gegenereerde afbeelding, audio of video die voor *echt* kan doorgaan. Hoewel de term 'deepfake' vooral bekend is als aanduiding voor gemanipuleerde beelden van bekende personen, is de definitie in de AI Act dus veel ruimer. Deze ziet immers op 'bestaande personen' ongeacht of ze bekend zijn, en daarnaast ook op 'voorwerpen, plaatsen, entiteiten of gebeurtenissen'. Dus ook een afbeelding van een dier, een plant, een landschap of een gebouw kan eronder vallen. Het is o.i. overigens goed verdedigbaar dat de definitie zo moet worden uitgelegd dat ook fotorealistische afbeeldingen van *niet bestaande* personen, situaties of gebeurtenissen onder de definitie van deepfake vallen, wanneer zij de indruk wekken afbeeldingen te zijn van personen etc. die wél bestaan. Het lijkt dan te gaan om 'bestaande personen' en de situatie waarin zij zich bevinden lijken 'bestaande gebeurtenissen'. Ook in die gevallen is er een risico op misleiding van het publiek, en dat is nu juist wat de AI Act wil voorkomen. Een dergelijk ruime uitleg heeft wel tot gevolg dat relatief onschuldige situaties er ook onder vallen. Een afbeelding van een niet bestaand persoon in een reclame voor tandpasta of pindakaas die de indruk wekt dat het om een bestaand persoon gaat, is dan een deepfake in de zin van de AI Act. Afbeeldingen die niet de indruk wekken dat ze 'echt' zijn, zoals cartoons, tekeningen en andere afbeeldingen en die niet 'fotorealistisch' zijn, vallen niet onder de definitie. Dergelijke afbeeldingen kunnen overigens wel onder het portretrecht vallen en als persoonsgegeven zijn aan te merken, maar het zijn geen deepfakes in de zin van de AI Act.

Deepfakes kunnen op uiterst schadelijke manieren worden toegepast. Zo zijn deepfakes een middel om desinformatie te verspreiden (denk aan een deepfake van een president die onwaarheden verkondigt). Ook bij oplichting en porno kunnen deepfakes worden ingezet.[8] Tegen deze achtergrond heeft de Europese wetgever deepfakes als zodanig weliswaar niet verboden, maar wel een transparantieplicht aan gebruikers van deepfakes opgelegd.[9] Deze transparantieplicht geldt overigens niet alleen voor de evident schadelijke toepassingen die zojuist werden genoemd, maar voor alle gebruik van deepfakes. Het enige criterium is dat de afbeelding 'door een (gemiddeld) persoon ten onrechte voor authentiek of waarheidsgetrouw zou worden aangezien'. Het gebruik van het woord 'zou' betekent dat deze onjuiste indruk moet worden weggenomen door de transparantieplicht in kwestie. Op grond van artikel 50 lid 4 AI Act mag een deepfake alleen worden gebruikt als het voor degene die met die deepfake wordt geconfronteerd duidelijk is dat het om een deepfake gaat (dus: een disclaimer dat het om content gaat die kunstmatig is gegenereerd of gemanipuleerd).

Lid 5 van artikel 50 geeft nadere informatie over de wijze waarop die disclaimer moet worden gepresenteerd:

> "De in de leden 1 tot en met 4 bedoelde informatie wordt uiterlijk op het moment van de eerste interactie of blootstelling op duidelijke en te onderscheiden wijze aan de betrokken natuurlijke personen verstrekt. De informatie moet aan de toepasselijke toegankelijkheidseisen voldoen."

Deze instructies roepen de nodige vragen op. Wat valt er onder "de eerste interactie of blootstelling" en wie zijn de "betrokken natuurlijke personen"? Moet de betreffende informatie pas worden verstrekt op het moment dat een deepfake wordt gepubliceerd, of moet dat al eerder worden gedaan? Het is waarschijnlijk dat het gaat om iedere 'interactie' met een persoon die niet zelf de deepfake heeft vervaardigd en dus mogelijk niet weet dat het een deepfake is. Dat geldt zeker voor extern gebruik buiten een onderneming, bijvoorbeeld via internet of andere media, gericht op consumenten of zakelijke klanten. Maar het geldt vermoedelijk ook voor intern gebruik. Ook bij gebruik van AI-gegenereerde content in interne stukken en communicatie moet transparantie worden betracht.

Gezien de ruime definitie zullen deze transparantieverplichtingen gelden voor veel content die door AI wordt gegenereerd. Het is voor de toepasselijkheid van de transparantieverplichtingen verder ook niet relevant of die content nog enige redactionele controle heeft ondergaan (zoals dat bij nieuwsberichten wel relevant is, zie verder hieronder).

Ten aanzien van deepfakes die onderdeel zijn van content die deel uitmaakt van een 'kennelijk artistiek, creatief, satirisch, fictief of analoog werk of programma', zijn de transparantieverplichtingen van artikel 50 lid 4 beperkt tot de openbaarmaking van het bestaan van dergelijke gegenereerde of bewerkte content op een passende wijze die de weergave of het genot van het werk niet belemmert. Dat betekent dat niet op storende wijze in, op of bij iedere (bewegende) afbeelding het deepfake karakter hoeft te worden vermeld, omdat dit het 'genot van het werk' zou belemmeren. Het gebruik van AI moet echter wel in de aankondiging, aftiteling of in het colofon worden vermeld. De aanduiding 'analoog' vormt hier vermoedelijk een tegenstelling tot de 'digitale' output van het AI-systeem die op grond van artikel 50 lid 2 van AI Act moet 'worden gemarkeerd in een machineleesbaar formaat en detecteerbaar zijn als kunstmatig gegenereerd of gemanipuleerd'. Bij 'analoge' content is 'machineleesbare' transparantie immers niet mogelijk en voor de menselijke gebruiker is die sowieso niet kenbaar. Het aanbrengen van die 'machineleesbare' markeringen is de verantwoordelijkheid van de aanbieder van het AI-systeem, niet van de gebruiker ervan. Maar het is aannemelijk dat de gebruiker die deze output verspreidt deze machineleesbare markering niet mag verwijderen. Daarnaast heeft hij dus zijn eigen transparantieplicht ten aanzien van de ruime categorie deepfakes.

7 Artikel 3 lid 60 AI Act.

8 Uit een onderzoek (2019) volgde dat maar liefst 96% van de deepfakes pornografisch van aard zijn. Zie Sensity (2019). *The State of Deepfakes: Landscape, Threats, and Impact*. Medium.

9 Het lijkt overigens niet waarschijnlijk dat partijen met kwade intenties die gebruikmaken van deepfakes zich iets zullen aantrekken van de transparantieverplichtingen van de AI Act.

4. Nieuwsberichten en andere tekst

Artikel 50 lid 4 legt ook transparantieverplichtingen op aan gebruikers van AI-systemen die *tekst* genereren of bewerken die wordt gepubliceerd om "het publiek te informeren over aangelegenheden van algemeen belang". Dit betekent dat AI niet mag worden ingezet om nieuwsberichten op te stellen of zelfs maar te bewerken zonder dat dit wordt vermeld. Anders dan bij deepfakes, heeft de wetgever hier wél een belangrijke uitzondering gemaakt. De transparantieverplichting geldt niet als de door AI gegenereerde tekst "een proces van menselijke toetsing of redactionele controle heeft ondergaan en wanneer een natuurlijke of rechtspersoon redactionele verantwoordelijkheid draagt voor de bekendmaking van die content". Op grond van de formulering van de uitzondering lijkt het erop dat deze vereisten van menselijke controle en redactionele verantwoordelijkheid cumulatief zijn. Op grond van deze 'redactionele uitzondering' zullen nieuwsredacties in de praktijk gebruik kunnen maken van AI zonder dat zij dat specifiek hoeven te melden: de meeste nieuwsberichten zullen immers onderhevig zijn aan ten minste enige mate van redactionele controle.

Deze in de AI Act opgenomen regel en de redactionele uitzondering gelden naar de letter alleen voor het gebruik van AI bij nieuwsberichten, en niet voor andere vormen van gebruik van AI-gegenereerde tekst. Het is duidelijk dat het risico van misleiding bij nieuwsinformatie groot is. O.i. is echter goed verdedigbaar dat een dergelijke hoofdregel en uitzondering ook zouden moeten gelden, of op grond van de maatschappelijke zorgvuldigheid al gelden, voor andere vormen van gebruik van tekst.[10] Bij wetenschappelijke teksten zal transparantie over gebruik van AI naar alle waarschijnlijkheid voortvloeien uit de regels betreffende wetenschappelijke integriteit. Maar bij zakelijke en commerciële teksten voor intern en extern gebruik ligt o.i. misleiding en aansprakelijkheid op de loer wanneer met AI gegenereerde teksten worden gebruikt die niet "een proces van menselijke toetsing of redactionele controle" hebben ondergaan en waarbij geen "redactionele verantwoordelijkheid" wordt genomen. Met AI vertaalde reclames en aanbiedingen op websites kunnen immers fouten bevatten, en die fouten kunnen in concrete gevallen misleiding tot gevolg hebben. Als het gebruik van een AI-vertaalmachine daarbij niet wordt vermeld, is de kans groot dat de aanbieder aansprakelijk is voor schade die voortvloeit uit de misleiding die het gevolg is van dergelijke (vertaal)fouten.[11] Het is dus zaak om bij ieder gebruik van AI-gegenereerde teksten waarbij niet "een proces van menselijke toetsing of redactionele controle" (waar nodig door een 'native speaker') heeft plaatsgevonden, het feit dat de tekst of de vertaling met AI is gegenereerd te vermelden. Het is overigens nog beter om, naast de vermelding van AI-gebruik, ook altijd enige mate van redactionele controle uit te voeren. Daarmee kunnen fouten – en de daaruit voortvloeiende aansprakelijkheid – worden voorkomen.

5. Het gebruik van overige AI-content

Buiten de hiervoor behandelde categorieën bevat de AI Act geen transparantieverplichtingen voor algemeen commercieel gebruik van content die door AI is gegenereerd. Dit betekent dat door AI gegenereerd beeld-, audio- of videomateriaal dat niet als deepfake valt aan te merken (omdat dat materiaal niet ten onrechte voor waar kan worden aangezien) commercieel mag worden ingezet, zonder dat daarbij hoeft te worden vermeld dat dat materiaal met AI is gemaakt. Ook het (commerciële) gebruik van door AI gegenereerde of bewerkte tekst is op grond van de AI Act niet aan specifieke transparantieverplichtingen gebonden.

Evengoed hebben adverteerders zich te houden aan de algemene regels van het reclamerecht en oneerlijke handelspraktijken die van toepassing zijn op alle vormen van commerciële communicatie.[12] Die regels beogen, net als de transparantieverplichtingen van de AI Act, te voorkomen dat de consument wordt misleid. De enkele inzet van door AI gegenereerde content in commerciële communicatie is niet zonder meer misleidend en dus niet zonder meer in strijd met het reclamerecht. Afhankelijk van de omstandigheden *kan* het echter wel misleidend zijn om door AI gegenereerde content in reclame te gebruiken, of dat te doen zonder dat expliciet te vermelden. Dat kan bijvoorbeeld het geval zijn wanneer door AI gegenereerd beeld wordt gebruikt om de werking van cosmetische producten te onderschrijven (waarbij dan een al te rooskleurige voorstelling van zaken wordt gegeven).[13] Het plaatsen van een disclaimer is in het laatste geval mogelijk onvoldoende om de misleiding weg te nemen.

Een duidelijke vermelding van het gebruik van door AI gegenereerde content kan een verkeerde indruk bij de consument voorkomen en is dus aan te raden. O.i. kan de vermelding dat AI is gebruikt onder omstandigheden zelfs essentiële informatie zijn die aan de consument *moet* worden verstrekt voordat hij tot zijn aankoopbeslissing komt.[14] Denk bijvoorbeeld aan een bioscoopfilm die volledig door AI is gemaakt: er zijn mensen die zo'n film om die reden niet hoeven te zien en daar dus ook geen geld aan willen uitgeven.

10 Overigens zal ook het begrip 'nieuws' vrij ruim uitgelegd moeten worden, vergelijkbaar met het begrip 'journalistieke activiteiten': "alle activiteiten die bekendmaking aan het publiek van informatie, meningen of ideeën tot doel hebben". Het staat aan de nationale rechter om te beoordelen of dit het geval is (HvJ EG 16 december 2008, ECLI:EU:C:2008:727).

11 De enkele vermelding van gebruik van AI maakt overigens nog niet dat de gebruiker niet aansprakelijk is. De vermelding is echter wel een relevante factor in dit verband. Zonder vermelding is de kans dat de gebruiker volledig aansprakelijk is voor een fout in door AI gegenereerde tekst in ieder geval groter dan met vermelding.

12 Artikel 50 lid 6 van de AI Act maakt ook duidelijk dat de transparantieverplichtingen uit het artikel geen afbreuk doen aan andere transparantieverplichtingen die zijn vastgelegd in het Unierecht of het toepasselijke nationale recht. Bovendien dient de gebruiker van door AI gegenereerde content ook rekening te houden met de gebruikersvoorwaarden van de tool die is gebruikt. Die voorwaarden kunnen de gebruiker verplichten om te vermelden dat content is gemaakt door of met behulp van AI.

13 www.reclamecode.nl/nrc/reclamecode-cosmetische-producten-rcp/.

14 Het niet tijdig vermelden van essentiële informatie aan de consument of zakelijke klant kwalificeert als een 'misleidende omissie' en is een misleidende oneerlijke handelspraktijk op grond van artikel 6:193d BW en misleidende reclame op grond van artikel 6:194 lid 2 BW.

6. Conclusie en aanbeveling

Bij met AI gegenereerde *afbeeldingen, filmpjes en geluiden* moet het gebruik van AI *altijd, direct en duidelijk* worden gemeld wanneer bij een gebruiker de indruk zou kunnen ontstaan dat het om *echte* foto's, filmpjes of geluidsopnames gaat van *echte* personen, objecten, situaties of gebeurtenissen. Bij met AI gegenereerde *tekst* moet het gebruik van AI *ook altijd, direct en duidelijk* worden gemeld, *tenzij* de tekst "een proces van menselijke toetsing of redactionele controle" heeft ondergaan én "een natuurlijke of rechtspersoon redactionele verantwoordelijkheid draagt voor de bekendmaking van die content". Voor alle duidelijkheid: het vermelden van gebruik van AI bij AI-gegenereerde content neemt niet de aansprakelijkheid weg voor eventuele inbreuken op het auteursrecht, portretrecht, de privacy (AVG) of andere rechten van derden. De gebruiker blijft verantwoordelijk voor de inhoud van de content, ook bij vermelding van het gebruik van AI. De transparantieplicht is een (minimum)verplichting, geen vrijbrief ten aanzien van de inhoud van de content. De menselijke controle is bedoeld om inbreuken en andere onrechtmatigheid te voorkomen, maar als die controle tekortschiet, is er (redactionele) verantwoordelijkheid en (mogelijk) aansprakelijkheid. En ook als het gebruik van AI staat vermeld, kan de aanbieder nog altijd (risico)aansprakelijk zijn, zowel voor inbreuken op rechten van derden als voor misleiding.