



UvA-DARE (Digital Academic Repository)

GenAI ingebed op sociale media: AI-recht versus platformrecht

Leerssen, P.

Publication date

2024

Document Version

Final published version

Published in

Mediaforum

License

Unspecified

[Link to publication](#)

Citation for published version (APA):

Leerssen, P. (2024). GenAI ingebed op sociale media: AI-recht versus platformrecht. *Mediaforum*, 36(6), 189-190. https://www.mediaforum.nl/art/90-6310_GenAI-ingebed-op-sociale-media-AI-recht-versus-platformrecht

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

GenAI ingebed op sociale media: AI-recht versus platformrecht

Opinie

Paddy Leerssen*

Op 26 september 2024 hebben Instagram en Facebook een nieuwe AI-tool gelanceerd, 'Imagine'. Hiermee kunnen gebruikers afbeeldingen genereren op basis van hun posts, direct vanuit de gebruiksomgeving van het platform. *Imagine* is inmiddels in wel 22 landen beschikbaar (hoewel nog niet in Europa), en geldt als het nieuwste en belangrijkste voorbeeld van een bredere trend: sociale media integreren steeds meer generatieve AI-toepassingen in hun diensten, zowel voor adverteerders als voor eindgebruikers.¹ Een verse schep AI-drek de publieke sfeer inwerpen gaat nu met slechts één druk op de knop.²

Juridisch gezien leidt deze integratie van AI en platformdiensten tot markante interacties tussen de AI Act en de Digital Services Act (DSA).³ Hieronder een korte beschouwing over aansprakelijkheid en digitale watermerken.

Wat aansprakelijkheid betreft, zullen lezers van dit blad bekend zijn met het *safe harbour*-beginsel, dat online platforms (zijnde een vorm van 'hosting provider') vrijwaart van aansprakelijkheid voor illegale gebruikersinhoud.⁴ Een GenAI-dienst zoals ChatGPT of Midjourney kan zich klaarblijkelijk niet op deze vrijwaring beroepen voor zijn outputs, want het gaat hier niet om informatie verstrekt door de gebruiker – zo betoogde ook Daniël de Weerd onlangs in ons recente themanummer over AI⁵ (De *inputs* of prompts van gebruikers vallen daarentegen mogelijk wel onder de *safe harbour* wanneer de dienst deze opslaat.) Hiermee is nog niet gezegd dat GenAI-diensten altijd aansprakelijk zijn voor hun outputs; dit valt nog onder nationaal recht te bezien. Maar de AI-aanbieder geniet in ieder geval niet dezelfde privileges als platforms en andere hosting providers.

Bij ingebedde GenAI ligt het nu iets ingewikkelder. Hoe meer platforms zich mengen in het *genereren* van inhoud, des te dichter zoeken zij de grens van hun *safe harbour* op. Kunnen we nog wezenlijk zeggen dat een *Imagine* afbeelding door de gebruiker *verstrekt* is aan het platform? Is een *Imagine* afbeelding nu gebruikersinhoud, AI-inhoud, of allebei?

Hierover lijkt het laatste woord nog niet gezegd. Uit bestaande rechtspraak weten we dat een platform zijn beschermde status kan verliezen, als zijn verwerking verder gaat dan "een louter technische en automatische verwerking" en het geen "actieve rol heeft waardoor het kennis heeft van of controle

heeft over die gegevens".⁶ Enerzijds is het de vraag of ingebedde GenAI-content wel "verstrekt" is door de eindgebruiker en heeft de tussenpersoon een verdergaande controle dan bij de meeste hostingdiensten. Anderzijds behoudt de gebruiker in de meeste gevallen nog wel de technische mogelijkheid om van publicatie af te zien; hij behoudt als het ware de eindredactie en daarmee een belangrijke vorm van controle. Het voert daarom wellicht te ver om nu te concluderen dat platforms geen bescherming genieten voor *Imagine*-inhoud. Maar voor andere ingebedde AI-functionaliteiten ligt dat wellicht nog anders. Denk bijvoorbeeld aan real-time livestreaming contexten, waar de gebruiker geen kans heeft om de outputs voor publicatie te controleren. Waar ligt de grens?

Naast aansprakelijkheid regelt het EU-recht nu ook bepaalde markeringsverplichtingen voor AI-inhoud. Aanbieders van systemen voor *generative AI* zijn verplicht hun AI-gegenereerde inhoud als zodanig herkenbaar te maken door middel van een digitale markering – bijvoorbeeld een visueel watermerk of een label in de metadata. De vereisten voor deze markering zijn flexibel en houden rekening met "de specifieke kenmerken en beperkingen van de verschillende soorten content, de uitvoeringskosten en de algemeen erkende stand van de techniek".⁷ Ook de *gebruiksverantwoordelijke* van het systeem heeft een verplichting herkenbaarheid te waarborgen (als het gaat om misleidende 'deepfakes').⁸

Wederom zijn ingebedde platformtoepassingen een lastig geval. De herkenbaarheidsregel voor aanbieders biedt namelijk een uitzondering "voor zover de AI-systemen een ondersteunende functie voor standaardbewerking vervullen of de door de gebruiksverantwoordelijke of de semantiek daarvan verstrekte inputdata niet substantieel wijzigen".⁹ Een ingebedde platformtoepassing kan hier mogelijk sneller aanspraak op maken dan een zelfstandig systeem. Denk bijvoorbeeld aan schrijfstijlcoaches of selfiefilters die je helpen om je posts aantrekkelijker te maken. Van een louter 'ondersteunende functie' lijkt allicht géén sprake bij *Imagine*, waar de gebruikersdata wel degelijk "substantieel gewijzigd" is, namelijk van tekst naar beeld.

Ook is er weer overlap met de DSA. De lijst van mogelijke risicobeperkingsmaatregelen onder Artikel 35 DSA nodigt platforms ook uit misleidende GenAI-inhoud "door opvallende markeringen [te] onderscheiden". In de zelfregulerende prak-

* Mr. P.J. Leerssen is postdoctoraal onderzoeker aan de UvA, gelieerd aan het IViR en redacteur van dit blad.

1 Voor Facebookadverteerders bestaat al enkele maanden een 'AI sandbox': <https://www.facebook.com/business/news/introducing-ai-sandbox-and-expanding-meta-advantage-suite>.

2 Vrij vertaald naar het Engelse 'AI-slop': een recent fenomeen waarbij commerciële partijen op volledig geautomatiseerde wijze AI-afbeeldingen delen met het oog om zoveel mogelijk clicks te winnen (en daarmee reclameinkomsten van Facebook). Dit levert de meest absurde en verontrustende afbeeldingen op, zoals de inmiddels beruchte 'Shrimp Jesus'. <https://www.404media.co/facebook-algorithm-is-boosting-ai-spam-that-links-to-ai-generated-ad-laden-click-farms/>

3 Verordening (EU) 2022/2065 betreffende een eengemaakte markt voor digitale diensten en tot wijziging van Richtlijn 2000/31/EG, *PbEU L 277/1* ('Digital Services Act'). Verordening (EU) 2024/1689 tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie, *PbEU L 2024/1689* ('AI Act').

4 DSA, Art. 6.

5 D. De Weerd, 'Aansprakelijkheid van aanbieders van generatieve AI-toepassingen onder de Digital Services Act', *Mediaforum* 2024-5.

6 HvJ EU 23 maart 2010, ECLI:EU:C:2010:159 (*Google France/Louis Vuitton*), par. 114.

7 AI Act, Art 50(2).

8 AI Act, Art. 50(4).

9 AI Act, Art. 50(2).

tijk hebben platforms als Facebook ook soortgelijke maatregelen getroffen (hoewel hun markeringen niet altijd even 'opvallend' zijn).¹⁰ Voorziet de wet dan twee watermerken per post, één van het platform en één van het model? Mogelijk kan Meta in zo'n geval gebruikmaken van de voorrangsregel uit overweging 118 van de Considerans van de AIA, die voorschrijft dat "overeenkomstige verplichtingen van deze verordening als vervuld worden beschouwd" wanneer zij overlappen met systeemrisicoverplichtingen uit de DSA. Men zou zich bovendien kunnen afvragen of niet-herkenbare deepfakes gelden als een vorm van *illegale* inhoud, en als zodanig onder het regime voor inhoudsmoderatie vallen. Kortom, het platform blijft een belangrijk aanspreekpunt voor de regulering van generatieve AI.

Het voorgaande illustreert enkele raakvlakken tussen AIA en DSA als het gaat om ingebedde AI op grote platforms. Naast

aansprakelijkheid en herkenbaarheidsverplichtingen spelen nog meer vragen over de verhouding tussen de twee systeemrisicokaders in deze wetten (de eerder genoemde voorrangsregel uit de considerans biedt wellicht enig houvast, maar schept ook zijn eigen grensgevallen).

Het moet nog maar blijken of ingebedde AI-toepassingen werkelijk een vlucht zullen nemen. Hoe nuttig is zo'n functie nu echt? Sterker nog, de gehele AI-industrie bouwt nog steeds op los zand; de commerciële vooruitzichten blijven twijfelachtig terwijl de milieuschade rampzalige proporties aanneemt. Maar als AI zich voorgoed inbedt op sociale media, dan brengt het complexe juridische vraagstukken mee. Slechts een paar jaar geleden leek een onderscheid tussen platformregulering en AI-regulering nog voor de hand te liggen, maar in de nieuwe plannen van Silicon Valley staan ze opeens op ramkoers.

¹⁰ <https://www.theverge.com/2024/9/12/24242998/facebook-instagram-ai-label-update-edited-content>.