



Universiteit
Leiden
The Netherlands

De AI-verordening beschermt onvoldoende tegen deepfakes

Custers, B.H.M.

Citation

Custers, B. H. M. (2025). De AI-verordening beschermt onvoldoende tegen deepfakes. *Nederlands Juristenblad*, 100(27), 2235-2241. Retrieved from <https://hdl.handle.net/1887/4291016>

Version: Publisher's Version

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/4291016>

Note: To cite this publication please use the final published version (if applicable).

De AI-verordening beschermt onvoldoende tegen deepfakes

Bart Custers¹

Vorig jaar trad de AI-verordening in werking, waarmee de EU probeert de ontwikkeling en het gebruik van AI in goede banen te leiden. Maar door de snelle technologische ontwikkelingen lijkt de wet nu al barsten te vertonen. Met name voor generatieve AI (kunstmatig aangemaakte tekst, foto's en video's) lijken de risico-inschattingen niet adequaat te zijn. Deepfakes en deepnudes worden als een beperkt risico gezien, terwijl er alle reden is om die als een hoog risico te beschouwen.

1. Inleiding

De AI-verordening is uniek in de wereld, geen enkele andere jurisdictie kent tot op heden een vergelijkbaar omvangrijke wet die de ontwikkeling en het gebruik van AI reguleert. Om te voorkomen dat de AI-verordening snel gedateerd zou raken, is gekozen voor een risicogerichte aanpak: als de risico's van een bepaalde vorm van AI laag zijn, zijn er weinig of geen regels van toepassing, maar als de risico's hoog zijn, legt de AI-verordening allerlei restricties en waarborgen op. Bij de allerhoogste risico's legt de AI-verordening zelfs een verbod op.

De AI-verordening werd in 2021 voorgesteld door de Europese Commissie als wettelijk kader voor AI-systemen en de daaraan verbonden risico's. Betrekkelijk snel voor EU-wetgeving werd de wet na enkele aanpassingen van kracht in 2024. Echter, gedurende deze periode vond ook een explosieve ontwikkeling van AI plaats, met name van generatieve AI. Met gratis of goedkope tools zoals ChatGPT, Dall-E en Midjourney kunnen gebruikers tekst, foto's en video's aanmaken.²

Dat zorgt voor leuke, grappige en creatieve beelden, maar tegelijkertijd zijn er ook allerlei vormen van misbruik. Deze bijdrage gaat in het bijzonder in op deepfakes en deepnudes. Deepfakes zijn door AI gegenereerde beelden die (zeer) realistisch lijken en waarmee mensen eenvoudig op het verkeerde been kunnen worden gezet. Zulke beelden kunnen worden gebruikt voor desinformatie en nepnieuws en kunnen besluitvorming en gedrag van mensen vergaand beïnvloeden.³ Deepnudes zijn een bijzondere variant van deepfakes waarin beelden van slachtoffers worden gecombineerd met pornografisch materiaal.⁴ Onderzoek laat zien dat slachtoffers hiervan

Deepfakes zijn door AI gegenereerde beelden die (zeer) realistisch lijken en waarmee mensen eenvoudig op het verkeerde been kunnen worden gezet

langdurige en ernstige psychologische klachten ondervinden.⁵

Deepfakes worden in de AI-verordening gezien als AI-technologie met een beperkt risico. Voor deze categorie

Auteur

1. Prof. mr. dr. ir. B.H.M. Custers is hoogle-raar Law & Data Science aan de Universiteit Leiden en directeur van eLaw, het Centrum voor Recht en Digitale Technologie.

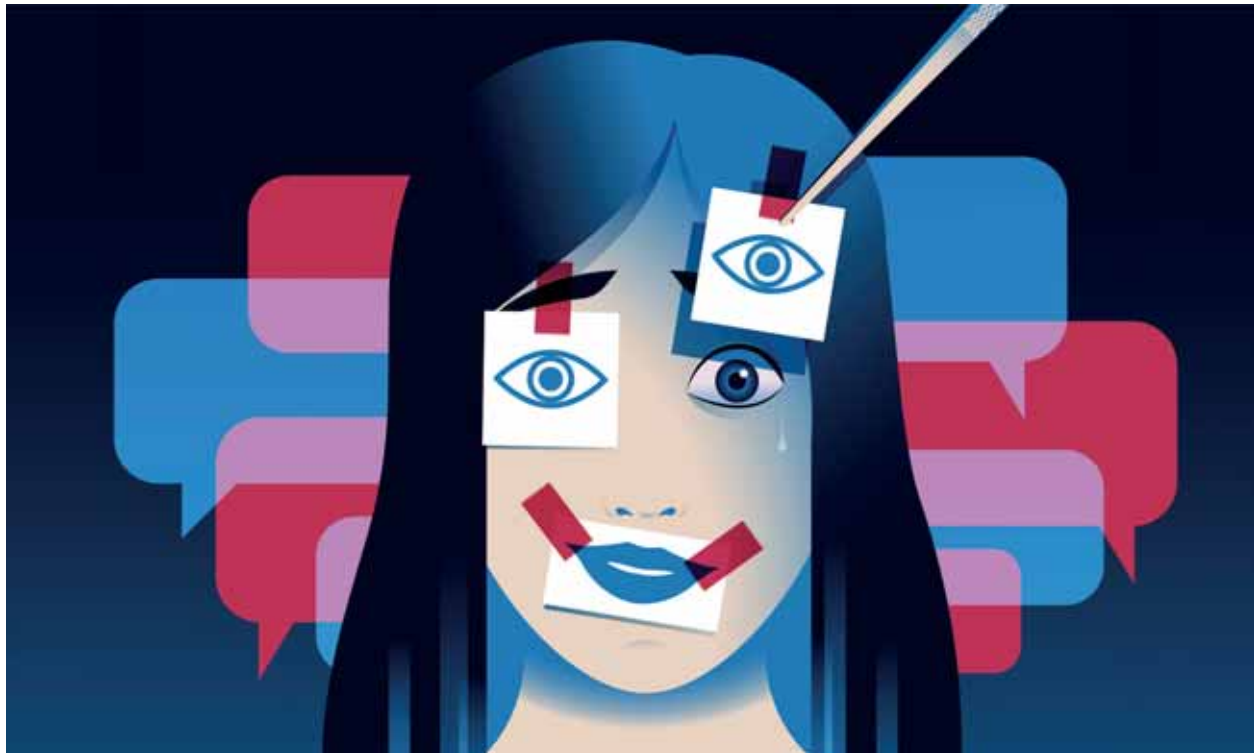
Noten

2. A. Bandi, P.V.S.R. Adapa & Y.E.V.P.K. Kuchi, 'The power of generative ai: A review of requirements, models, input-output formats, evaluation metrics, and challenges', *Future Internet* 2023, 15(8), 260; S. Feuerriegel, J. Hartmann, C. Janiesch & P. Zschech, 'Generative ai', *Business & Information Systems Engineering* 2024, 66(1), p. 111-126.

3. R. Chesney & D. Citron, 'Deepfakes and the new disinformation war: The coming age of post-truth geopolitics', *Foreign Affairs* 2019, 98, 147.

4. A. Gilen, C. van de Heyning & M. Walrave, 'The Scourge of Deepnude Applications: A Fundamental Rights Perspective', in: E. Kosta, D. Hallinan, P. de Hert & S. Nusselder (eds.), *Data protection, privacy and artificial intelligence: to govern or to be governed, that is the question*, New York: Bloomsbury 2025, p. 11-32.

5. S. Royer, 'Belaging in tijden van doxing en deep nudes', *Computerrecht: Tijdschrift voor Informatica, Telecommunicatie en Recht* 2022, p. 287-290.



© Moor Studio/Getty images

geldt alleen een transparantieplichting: voor gebruikers moet duidelijk zijn dat het om nepbeelden gaat. Als gebruikers worden gefaciliteerd om echt van nep te onderscheiden, zou dat voldoende moeten zijn, zo is de gedachte.⁶ Wanneer gebruikers weten dat bepaalde beelden nep zijn, zou de invloed op hun besluitvorming en gedrag niet problematisch zijn.

In deze bijdrage wordt betoogd dat er twee problemen zijn met deze redenering. Ten eerste is er inmiddels veel onderzoek beschikbaar dat laat zien dat deepfakes en deepnudes niet slechts een 'beperkt risico' vormen. Dit gebruik van generatieve AI kan grote maatschappelijke en persoonlijke schade berokkenen en kan de gezondheid, veiligheid en grondrechten van burgers aanzienlijk bedreigen.

Ten tweede wordt betoogd dat de transparantieplichting onvoldoende bescherming biedt tegen deze risico's. Onderzoek laat zien dat ook beelden waarvan duidelijk is dat ze nep zijn niettemin ernstige schade kunnen berokkenen. Daar komt nog bij dat transparantie zeer lastig is te handhaven, omdat bij het verspreiden van nepbeelden de labels (al dan niet opzettelijk) snel verloren kunnen gaan.

Deze bijdrage is als volgt opgebouwd. Paragraaf 2 biedt relevante achtergrond over de AI-verordening, in het bijzonder de risicogerichte aanpak en de classificatie van deepfakes. Paragraaf 3 beschrijft deepfake- en deepnude-technologie en wat daarmee allemaal mogelijk is. Paragraaf 4 biedt een overzicht van de risico's van deepfakes en deepnudes en laat zien dat die niet 'beperkt' zijn. Paragraaf 5 bespreekt de transparantieplichting en laat zien dat deze onvoldoende is om de problemen te adresseren. Paragraaf 6 sluit af met conclusies.

2. De AI-verordening

De AI-verordening reguleert AI door middel van een risicogerichte aanpak.⁷ Er worden vier categorieën risico's

onderscheiden: onaanvaardbaar risico, hoog risico, beperkt risico en minimaal risico.⁸ AI-systemen met een onaanvaardbaar risico, zoals systemen die gedrag subliminaal manipuleren of de sociaal-economische status van mensen vergaren, zijn verboden.⁹ AI-systemen die de gezondheid, veiligheid of grondrechten van mensen bedreigen, worden beschouwd als een hoog risico en hiervoor gelden strenge regels, zoals evaluaties voordat ze op de markt komen en extra toezicht daarna.¹⁰ Meer algemene AI-toepassingen zoals chatbots en emotieherkenningsystemen worden gezien als een beperkt risico, waarvoor alleen een transparantieplichting geldt: het moet voor gebruikers duidelijk zijn dat het om een AI-tool gaat.¹¹ Voor alle overige vormen van AI, zoals videogames en spamfilters, wordt het risico als minimaal ingeschat en legt de AI-verordening geen regels op. Ook slimme oplossingen voor bestandsbeheer, zoals functies voor indexering, opzoeking, tekst- en spraakverwerking of vertalingen vallen in deze categorie van zeer lage risico's.¹²

Om ervoor te zorgen dat de AI-verordening niet te snel gedateerd zou raken, is gekozen voor technologie-neutrale wetgeving,¹³ waarbij niet zozeer de (werking van de) technologie wordt beschreven, maar waarbij de nadruk ligt op de gevolgen en risico's van de verschillende vormen van AI. Als bepaalde vormen van AI verder worden doorontwikkeld, blijft de AI-verordening van toepassing, omdat niet de onderliggende technologie bepalend is voor de reikwijdte. Nu zijn de meeste vormen van AI nog bereikbaar via computers en het internet, maar het is niet ondenkbaar dat AI straks in autonome robots wordt geïmplementeerd. Die zouden dan nog steeds binnen het bereik van de AI-verordening vallen.

Artikel 3 lid 1 AI-verordening geeft een lange en ingewikkelde definitie van wat een AI-systeem is. De kern van deze definitie is dat een AI-systeem een op een

machine gebaseerd systeem is met een bepaalde mate van autonomie en aanpassingsvermogen. Autonomie houdt in dat menselijke tussenkomst beperkt of afwezig is. Er is dan geen (of geen significante) menselijke schakel in het genereren van output of het nemen van beslissingen.¹⁴ Aanpassingsvermogen houdt in dat het AI-systeem zelflerend is. Algoritmen leveren bij dezelfde input altijd dezelfde output en zijn zodoende niet zelflerend. AI-systemen zijn uitgebreide netwerken van algoritmen waaraan verschillende wegingsfactoren kunnen worden toegekend.¹⁵ Daardoor kan na verloop van tijd dezelfde input toch leiden tot veranderde output, omdat het systeem 'bijgeleerd' heeft.

De definitie in artikel 3 AI-verordening is bewust ruim en open geformuleerd met het oog op toekomstige ontwikkelingen. Om tegelijkertijd toch verdere verduidelijking aan te brengen in welke technologieën bedoeld zijn, bevat Bijlage III een gedetailleerde beschrijving van AI-systemen met een hoog risico.¹⁶ Door technologiebeschrijvingen niet in de wetsartikelen op te nemen, blijft er voor rechters ruimte voor rechtsontwikkeling en voor de wetgever is het eenvoudiger een bijlage aan te passen dan wetsartikelen te herzien, mocht dat in de toekomst nodig zijn.

3. Deepfakes en deepnudes

Eerdere vormen van AI waren vooral gericht op patroonherkenning, zoals gezichtsherkenning, spraakherkenning, tekstherkenning en het classificeren van beeldmateriaal.¹⁷ Op basis van grote hoeveelheden data worden systemen dan getraind (of trainen zichzelf) om bijvoorbeeld te herkennen op welke plaatjes een vlinder staat. De meest recente technologie is in staat dit proces ook omgekeerd te doorlopen en aan de hand van beschrijvingen tekst en beeld op te bouwen.¹⁸ Dit heet generatieve AI en hiermee kan kunstmatige content worden aangemaakt, bijvoorbeeld gezichten van niet-bestaande mensen.¹⁹

In de afgelopen jaren zijn er online steeds meer gratis en goedkope tools beschikbaar gekomen die zulke AI

Door technologiebeschrijvingen niet in de wetsartikelen op te nemen, blijft er voor rechters ruimte voor rechtsontwikkeling

aanbieden. Wellicht het bekendste voorbeeld is ChatGPT, een gratis toepassing waarmee gebruikers teksten kunnen aanmaken, vertalen of redigeren.²⁰ Scholieren en studenten gebruiken ChatGPT voor hun opdrachten en scripties, advocaten en rechters gebruiken AI voor hun pleidooien en vonnissen en journalisten gebruiken het voor hun artikelen. De nieuwste versie biedt ook de mogelijkheid om foto's te genereren. ChatGPT is niet de enige aanbieder, ook andere online tools zoals Dall-E, Midjourney en Microsoft Copilot kunnen tekst, foto's en video's aanmaken.²¹

Het beeldmateriaal dat hiermee kan worden gemaakt, kan zeer tot de verbeelding spreken. Het zorgt voor grappige, mooie en creatieve foto's en video's. Zo zijn er nepbeelden verschenen van de paus in een dik dons-jack,²² tennissende wasbeertjes en president Trumps perikelen in de gevangenis.²³ Mensen kunnen een Barbie,²⁴ Muppet²⁵ of andere animatie van zichzelf maken. Met tools als Fotor²⁶ en Picwand²⁷ kunnen gebruikers beelden genereren die voorspellen hoe ze eruit zullen zien op latere leeftijd.

Ook de filmindustrie maakt volop gebruik van AI, in Hollywood waren er al stakingen en protestacties van acteurs en scenarioschrijvers die bang waren hun baan te verliezen.²⁸ In de filmwereld bestaan inmiddels voorbeelden van AI die overleden acteurs weer tot leven wekt met behulp van archiefmateriaal en oude scènes. Een bekend voorbeeld betreft de acteur Paul Walker die om het leven kwam bij een auto-ongeluk tijdens de opnames van de film

6. Cf. overweging 134 AI-verordening.

7. Cf. R. Baldwin, M. Cave & M. Lodge, *Understanding regulation: theory, strategy, and practice*, Oxford university press, 2011; J. Black, 'Risk based regulation', *Risk and regulatory policy: improving the governance of risk*, OECD, 2010, p. 185-224, regulatoryreform.com/wp-content/uploads/2015/02/OECD-Risk-and-Regulatory-Policy-2010.pdf.

8. M. Ebers, 'Truly risk-based regulation of artificial intelligence how to implement the EU's AI Act', *European Journal of Risk Regulation* 2024, p. 1-20. doi.org/10.1017/err.2024.78.

9. Art. 5 AI-verordening.

10. Hoofdstuk III AI-verordening.

11. Art. 50 AI-verordening.

12. Overweging 53 AI-verordening.

13. M. Hildebrandt & L. Tielemans, 'Data protection by design and technology neutral law', *Computer Law & Security Review*

2013, 29(5), p. 509-521.

14. S. Robbins, 'The many meanings of meaningful human control', *AI and Ethics* 2024, 4(4), p. 1377-1388; G. Mecacci, D. Amoroso, L.C. Siebert, D. Abbink, J. van den Hoven & F.S. de Sio, *Research handbook on meaningful human control of artificial intelligence systems*, Edward Elgar Publishing, 2024.

15. Vergelijkbaar met het menselijke brein: neuronen geven elektrische en chemische signalen door, altijd op dezelfde manier. Maar het uitgebreide netwerk van neuronen in de hersenen biedt de mogelijkheid bepaalde routeringen te versterken na verloop van tijd. Daardoor leert iemand nieuwe dingen bij.

16. Voor de categorie verboden AI-systemen met onaanvaardbaar hoge risico's is artikel 5 in de AI-verordening opgenomen. Deze categorie is verder uitgewerkt in richtsnoeren, zie: European Commission, *Com-*

mission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), C(2025) 884 final, Brussel, 4 februari 2025.

17. A. Häuselmann, 'Disciplines of AI: An Overview of Approaches and Techniques', in: B. Custers & E. Fosch-Villaronga (eds.) *Law and Artificial Intelligence Regulating AI and Applying AI in Legal Practice*, Heidelberg: Springer 2022, p. 43-70.

18. A. Narayanan & S. Kapoor, *AI Snake Oil*, Princeton: Princeton University Press 2024.

19. Zie bijvoorbeeld thispersondoesnotexist.com.

20. openai.com/index/chatgpt/.

21. A. Bandi, P.V.S.R. Adapa & Y.E.V.P.K. Kuchi, 'The power of generative ai: A review of requirements, models, input-output formats, evaluation metrics, and challenges', *Future Internet* 2023, 15(8), 260; S. Feuerriegel, J. Hartmann, C. Janiesch &

P. Zschech, 'Generative ai', *Business & Information Systems Engineering* 2024, 66(1), p. 111-126.

22. J. Golby, 'I thought I was immune to being fooled online. Then I saw the pope in a coat', *The Guardian* 27 maart 2023, theguardian.com/commentisfree/2023/mar/27/pope-coat-ai-image-baby-boomers.

23. K. Devlin & J. Cheetham, 'Fake Trump arrest photos: How to spot an AI-generated image', *BBC News* 24 maart 2023, bbc.com/news/world-us-canada-65069316.

24. barbie selfie.ai/intl/

25. fotor.com/ai-image-generator/muppet/

26. fotor.com

27. picwand.ai

28. B. Rutten, 'Tarantino hoeft nog niet te vrezzen, maar AI kan veel filmmakers vervangen', *Nieuwsuur* 15 mei 2023, nos.nl/nieuwsuur/artikel/2475176-tarantino-hoeft-nog-niet-te-vrezzen-maar-ai-kan-veel-filmmakers-vervangen.

Fast & Furious deel 7. Voor de film was pas twee-derde van de scenes waarin de acteur meespeelde opgenomen en de makers konden opeens niet verder. Uiteindelijk heeft producent Universal ervoor gekozen oude beelden van de acteur te combineren met opnames van zijn twee broers, een 3D-reconstructie en een acteur die zijn bewegingen imiteerde.²⁹ Daarmee kon de film toch voltooid worden. Een ander opzienbarend voorbeeld was de Star Wars film Rogue One, waarin acteur Peter Cushing een rol speelde met echte acteurs terwijl hij al 22 jaar dood was.³⁰

Veel van deze beelden zijn duidelijk nep, maar sommige lijken (zeer) realistisch. Dit zijn de zogeheten deepfakes.³¹ Deze kunnen voor problemen zorgen wanneer mensen denken dat de beelden echt zijn en daarop hun beeldvorming en besluitvorming baseren. Met deepfakes

De risico's kunnen zeer groot zijn, bijvoorbeeld wanneer het om politieke content gaat

kunnen identiteiten van bekende personen worden misbruikt, bijvoorbeeld door filmsterren bepaalde producten te laten aanbevelen of politici bepaalde woorden in de mond te leggen.³² De Tweede Kamer werd zo misleid tijdens een videogesprek met een medewerker van de Russische oppositieleider Navalny.³³ Het bleek totaal iemand anders te zijn waarvan het gezicht door AI-technologie was aangepast.

De risico's kunnen zeer groot zijn, bijvoorbeeld wanneer het om politieke content gaat.³⁴ Deepfakes kunnen worden gebruikt om democratische verkiezingen te manipuleren, iets dat inmiddels veelvuldig plaatsvindt. Nepbeelden kunnen mensen aanzetten tot demonstraties, haat of criminele handelingen.³⁵ Het kan leiden tot maatschappelijke polarisatie en zelfs tot staatsgrepen en oorlog. Dat bleek toen in 2019 de president van Gabon in een deepfake video tot het publiek sprak. Hij was na een medische ingreep in het buitenland maandenlang niet gezien. De video leidde tot speculaties en kort daarna tot een coup.³⁶

Een andere populaire toepassing is het genereren van deepnudes, deepfake pornografie, waarbij beelden van bijvoorbeeld beroemdheden worden gecombineerd met pornografisch materiaal.³⁷ Maar liefst 98% van alle deepfakes die online circuleren bestaat uit deepnudes.³⁸ Hoe deepnudes werken is goed gedocumenteerd door presentatrice Welmoed Sijsma die ontdekte dat er van haar online een nep-pornofilm circuleerde.³⁹ Ze maakte de persoonlijke documentaire waarin ze de maker van haar nep-filmpje opspoorde.⁴⁰ Het maken van deepnudes beperkt zich niet tot beroemdheden, ook andere vrouwen kunnen hiervan slachtoffer worden. Onderzoek laat zien dat 70% van de deepnudes is gemaakt van niet-bekende personen.⁴¹ Als dat gebeurt, is dat regelmatig door mensen uit hun directe omgeving.⁴² Slachtoffers zijn vrijwel altijd vrouwen: 98% van de deepnudes wordt van vrouwen

gemaakt.⁴³ Deepnudes worden soms gemaakt door mensen die dat grappig of leuk vinden, in tegenstelling tot de slachtoffers. Deepnudes worden ook gebruikt voor wraakporno, afpersing en treiteren.⁴⁴ Onder jongeren spelen macht, status en populariteit ook een belangrijke rol.⁴⁵

Verschillende eenvoudig toegankelijke AI-tools voor het maken van deepfake pornografie zijn inmiddels van het internet gehaald. De makers van Deepnude haalden hun tool van het internet na aanhoudende kritiek.⁴⁶ De website Mr. Deepfakes verdween dit jaar omdat de service-provider niet langer deze toepassing wilde ondersteunen.⁴⁷ Niettemin zijn er nog tal van andere websites over – voldoende reden voor de Amerikaanse overheid om onlangs stevige wetgeving aan te nemen die deepnudes verbiedt.⁴⁸ Apps als *Nudify*, *Undress* en *ClothesOff* worden miljoenen keren per maand bezocht.⁴⁹

4. De risico's zijn niet 'beperkt'

Artikel 3 lid 60 AI-verordening definieert deepfakes 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 ten onrechte voor authentiek of waarheidsgetrouw zou worden aangezien. Hieronder vallen ook deepnudes. In artikel 50 lid 4 AI-verordening worden deepfakes, inclusief deepnudes, geclassificeerd als 'beperkt risico'. Echter, de risico's van deepfakes en deepnudes zijn niet 'beperkt'.

Er is inmiddels veel onderzoek beschikbaar dat laat zien dat deepfakes en deepnudes serieuze schade kunnen berokkenen aan slachtoffers en aan de maatschappij.⁵⁰ Slachtoffers van deepfakes kunnen ernstige reputatieschade lijden en deepfakes kunnen de rechtsstaat en democratie ondermijnen. Deepfakes kunnen aanzetten tot haat, geweld en andere vormen van criminaliteit.⁵¹ De psychologische gevolgen van deepfakes komen het duidelijkst naar voren bij deepnudes. Slachtoffers van deepnudes kunnen naast reputatieschade langdurige psychologische klachten ondervinden van deepnudes, waaronder angststoornissen, spanningsklachten, depressie, PTSS en suïcidale gedachten.⁵²

Dit is serieuze schade en de kans dat deze schade zich voordoet is groot. Zoals in de volgende paragraaf wordt besproken, zijn deze risico's grotendeels ook aanwezig als de beelden duidelijk nep zijn. In veel gevallen maakt dat niet uit voor slachtoffers en voor evallen van het maatschappelijk vertrouwen in online informatie. Kortom, dat sprake zou zijn van een beperkt risico is niet in lijn met de connotatie die hoort bij de term 'beperkt risico'.

Het begrip 'beperkt risico' is echter een juridisch begrip dat door de AI-verordening wordt geïntroduceerd. De betekenis van een juridisch begrip kan afwijken van wat we in maatschappelijke context onder die term verstaan. Maar ook binnen de context van het juridische begrip 'beperkt risico' dat de AI-verordening introduceert zijn de risico's van deepfakes en deepnudes niet helemaal passend binnen de wetssystematiek.

Ten eerste gaat het mis bij de beoogde bescherming. In de AI-verordening zijn beperkte risico's gekoppeld aan transparantie. Weten of beeldmateriaal echt of nep is zou voldoende moeten zijn om schade te voorkomen. Zoals in de volgende paragraaf wordt getoond, zijn veel deepfakes

en deepnudes duidelijk nep, maar kunnen ze niettemin aanzienlijke schade veroorzaken. Dus bezien vanuit de belangen die de AI-verordening probeert te beschermen (*Schutznorm*), slaat de categorisering van deepfakes en deepnudes als beperkt risico de plank mis. Er wordt vrijwel geen bescherming geboden tegen de hierboven beschreven risico's.

Ten tweede gaat het mis bij de onderbouwing. Volgens de wetssystematiek van de AI-verordening is sprake van een hoog risico als de gezondheid, veiligheid en grondrechten van burgers worden bedreigd.⁵³ Deepfakes kunnen een ernstige bedreiging vormen voor de veiligheid en grondrechten van burgers zodra ze de rechtsstaat en democratie ondermijnen.⁵⁴ Het recht op privacy en het recht op bescherming van persoonsgegevens zijn grondrechten die onder druk komen te staan als personen zonder toestemming op ongewenste manieren worden

geportretteerd, want in beginsel zijn deepfakes alleen toegestaan als de betrokkene daarvoor toestemming heeft gegeven.⁵⁵ Vrijheid van meningsuiting en vrijheid van religie kunnen ingeperkt worden wanneer mensen woorden in de mond worden gelegd, met name wanneer dat uitspraken zijn die haaks staan op hun denkbeelden en levensovertuigingen.⁵⁶ Wanneer nepbeelden in rechtszaken terechtkomen, kan het recht op een eerlijk proces worden geschonden.⁵⁷ Kortom, allerlei rechten en vrijheden kunnen onder druk komen te staan. Deepnudes kunnen bijzonder grote schade aan de gezondheid van slachtoffers aanrichten. Dat kan bij psychosomatische klachten en zelfmoordpogingen omslaan in problemen ten aanzien van fysieke gezondheid en welzijn.⁵⁸ Kortom, de risico's van deepfakes en deepnudes passen qua onderbouwing veel beter in de categorie 'hoog risico' die de AI-verordening biedt.

29. BBC, 'Fast & Furious 7 to 'go ahead' after Walker's death', *BBC* 2 december 2013, [bbc.com/news/entertainment-arts-25182688](https://www.bbc.com/news/entertainment-arts-25182688).

30. A. Pulver, 'Rogue One VFX head: 'We didn't do anything Peter Cushing would've objected to'', *The Guardian*, 16 januari 2017, [theguardian.com/film/2017/jan/16/rogue-one-vfx-jon-knoll-peter-cushing-ethics-of-digital-resurrections](https://www.theguardian.com/film/2017/jan/16/rogue-one-vfx-jon-knoll-peter-cushing-ethics-of-digital-resurrections).

31. Á.F. Gambín, A. Yazidi, A. Vasilakos, H. Haugerud & Y. Djenouri, 'Deepfakes: Current and future trends', *Artificial Intelligence Review* 2024, 57(3), 64.

32. S. Suwajanakorn, S.M. Seitz & I. Kemelmacher-Shlizerman, 'Synthesizing Obama: learning lip sync from audio' *ACM Transactions on Graphics (TOG)* 2017, 36(4), p. 1-13.

33. NOS, 'Veel onduidelijk over 'deepfake-gesprek' van Kamerleden met medewerker Navalny', *NOS* 24 april 2021, nos.nl/artikel/2378060-veel-onduidelijk-over-deepfake-gesprek-van-kamerleden-met-medewerker-navalny.

34. B. Chesney & D. Citron, 'Deep fakes: A looming challenge for privacy, democracy, and national security', *California Law Review* 2019, 107, 1753; M. Pawelec, 'Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions', *Digital society* 2022, 1(2), 19.

35. J. Langa, 'Deepfakes, real consequences: Crafting legislation to combat threats posed by deepfakes', *Boston University Law Review* 2021, 101, 761.

36. S. Cahlan, 'How misinformation helped spark an attempted coup in Gabon', *The Washington Post* 13 februari 2020, [washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-](https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/)

[sparked-an-attempted-coup-gabon/](https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/).

37. R. Metz, 'A terrifying app for making any woman appear naked was killed off by its creator', *CNN* 28 juni 2019, edition.cnn.com/2019/06/27/tech/deepnude-naked-women-ai.

38. Home Security Heroes, *State of Deepfakes: Realities, Threats and Impact*, 2023 securityhero.io/state-of-deepfakes/
39. D. Jansen, 'Welmoed Sijtsma maakt docuseries over deepfake porno. Mijn hoofd is gestolen voor een pornofilm', *Algemeen Dagblad* 4 november 2022, ad.nl/show/welmoed-sijtsma-maakt-docuserie-over-deepfake-porno-mijn-hoofd-is-gestolen-voor-een-pornofilmje-a07f3c92/.

40. npa.nl/start/serie/welmoed-en-de-sex-fakes/seizoen-1/welmoed-en-de-sex-fakes.

41. Sensity, 'Automating Image Abuse', juli 2023, sensity.ai.

42. K. West, 'I was deepfaked by my best friend', *BBC* 2 april 2024, [bbc.com/news/uk-68673390](https://www.bbc.com/news/uk-68673390).

43. Z. Holmes, '97% of new deepfakes are pornographic, experts believe revenge porn will spike', *The Big Smoke* 7 april 2021, [usa.thebigsmoke.com.au/2019/10/19/97-of-new-deepfakes-are-pornographic-experts-believe-revenge-porn-will-spike/](https://www.thebigsmoke.com.au/2019/10/19/97-of-new-deepfakes-are-pornographic-experts-believe-revenge-porn-will-spike/).

44. E. Blancaflor, J.I. Garcia, F.D. Magno & M.J. Vilar, 'Deepfake Blackmailing on the Rise: The Burgeoning Posterity of Revenge Pornography in the Philippines', in: *Proceedings of the 2024 9th International Conference on Intelligent Information Technology* 2024, p. 295-301.

45. N.D. Szyf, A. Gilen & M. Giacometti, *Deepnudes onder Belgische jongeren*, Brussel: Instituut voor gelijkheid voor de gelijkheid van vrouwen en mannen 2023, ijgm-iefh.belgium.be/sites/default/files/deepnudes_under_belgische_jongeren.pdf.

46. BBC, 'App that can remove women's clothes from images shut down', *BBC* 28 juni 2019, [bbc.com/news/technology-48799045](https://www.bbc.com/news/technology-48799045).

47. L. Ferris, 'AI-generated porn site Mr. Deepfakes shuts down after service provider pulls support', *CBS News* 5 mei 2025, [cbsnews.com/news/ai-generated-porn-site-mr-deepfakes-shuts-down/](https://www.cbsnews.com/news/ai-generated-porn-site-mr-deepfakes-shuts-down/).

48. NOS, 'Nieuwe Amerikaanse wet biedt meer bescherming tegen deepfakes en wraakporno', *NOS* 20 mei 2025, nos.nl/artikel/2568001-nieuwe-amerikaanse-wet-biedt-meer-bescherming-tegen-deepfakes-en-wraakporno.

49. A. Olson, 'The Double-Side of Deepfakes: Obstacles and Assets in the Fight Against Child Pornography' *Georgia Law Review* 2021, 56, 865; S. Caramela, 'Nudify Deepfake Bots on Telegram Are Up to 4 Million Monthly Users', *Vice* 16 oktober 2024, [vice.com/en/article/nudify-deepfake-bots-telegram/?utm_source=dlvr.it&utm_medium=mastodon](https://www.vice.com/en/article/nudify-deepfake-bots-telegram/?utm_source=dlvr.it&utm_medium=mastodon).

50. A. de Rancourt-Raymond & N. Smaili, 'The unethical use of deepfakes', *Journal of Financial Crime* 2023, 30(4), p. 1066-1077; I. Kalpokas & J. Kalpokiene, *Deepfakes: a realistic assessment of potentials, risks, and policy regulation*, Heidelberg: Springer Nature 2022; P.S. Chauhan, 'Deepfake: Risks and Opportunities', *Computer* 2024, 57(06), p. 141-144.

51. R. Rini & L. Cohen, 'Deepfakes, deep harms', *Journal of Ethics and Social Philosophy* 2022, 22, 143.

52. K. Hao, 'Deepfake porn is ruining women's lives. Now the law may finally ban it', *MIT Technology Review* 2021, [technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/](https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/).

Yadav, M.Z. Sadique, S. Kumar, R. Sharma, M. Sharma, R. Sharma & T. Rattan, 'Psy-

chological Trauma and Legal Challenges of Deep fake Technology', *Sciences of Conservation and Archaeology* 2025, 37(1), p. 143-150.

53. Zie onder andere overwegingen 7, 53 en 65-66 AI-verordening.

54. F.R. Moreno, 'Generative AI and deepfakes: a human rights approach to tackling harmful content', *International Review of Law, Computers & Technology* 2024, 38(3), p. 297-326.

55. N. Lussier, 'Nonconsensual deepfakes: detecting and regulating this rising threat to privacy', *Idaho Law Review* 2022, 58, 353; B. Custers, F. Dechesne, W. Pieters, B. Schermer & S. van der Hof, 'Consent and Privacy', in: A. Müller & P. Schaber (eds.), *Handbook of the Ethics of Consent*, London: Routledge 2018, p. 247-258. In de strijd tegen deepfakes overweegt Denemarken om burgers copyright te geven op hun eigen gezicht en stem, zie 'In de strijd tegen deepfakes krijgen Denen copyright op eigen gezicht en stem', *NOS* 13 juli 2025, nos.nl/nieuwsuur/artikel/2574932.

56. A. Barber, 'Freedom of expression meets deepfakes', *Synthese* 2023, 202(2), 40; B. van der Sloot, Y. Wagenveld, 'Deepfakes: regulatory challenges for the synthetic society', *Computer Law & Security Review* 2022, 46, 105716; J. Ice, 'Defamatory political deepfakes and the first amendment', *Case Western Reserve Law Review* 2019, 70, 417.

57. R. Pfefferkorn, 'Deepfakes in the Courtroom', *Boston University Public Interest Law Journal* 2019, 29, 245; A.E. Venema & Z.J. Gerads, 'Digital forensics, deepfakes, and the legal process', *The SciTech Lawyer* 2020, 16(4), p. 14-23.

58. J.L. Hatch, *Deepfake: Cause for Sui-cide... Or Prelude to Murder*, 2024.

De risico's van deepfakes en deepnudes passen qua onderbouwing veel beter in de categorie 'hoog risico' die de AI-verordening biedt

Alles bij elkaar zijn de risico's van deepfakes niet 'beperkt', in maatschappelijke noch in juridische betekenis. Vanuit de systematiek van de AI-verordening passen de risico's van deepfakes juist goed in de categorie hoog risico (vanwege de dreiging voor gezondheid, veiligheid en grondrechten van burgers) en duidelijk niet in de categorie 'beperkt risico' (vanwege de Schutznorm).

5. Alleen een transparantieplichting is onvoldoende

De categorisering van deepfakes en deepnudes als 'beperkt risico' in de AI-verordening heeft tot gevolg dat slechts de transparantieplichting van toepassing is (artikel 50 AI-verordening). De kern van deze verplichting is dat voor gebruikers duidelijk moet zijn dat het beeldmateriaal door AI is vervaardigd. Dat kan bijvoorbeeld door de beelden te voorzien van een label of een introductie. De vorm waarop kenbaar wordt gemaakt dat het om nepbeelden gaat is niet voorgeschreven, zolang het maar duidelijk is voor 'normaal geïnformeerde en redelijk voorzichtige en oplettende' gebruikers.⁵⁹ Dit moet kenbaar zijn vanaf de eerste interactie met of blootstelling aan de beelden.⁶⁰

Er zijn twee problemen met de transparantieplichting. Het eerste probleem is al genoemd in de vorige paragraaf: de transparantieplichting biedt geen of vrijwel geen bescherming tegen de risico's van deepfakes en de schade die ze kunnen aanrichten.

Deepfakes hoeven niet altijd overtuigend te zijn. In veel gevallen zijn ze dat ook niet. Niettemin kunnen de beelden zodanig veel onzekerheid, verwarring en wantrouwen opbouwen, dat vertrouwen in alle berichtgeving erodeert.⁶¹ Ook als de beelden duidelijk nep zijn, daalt het vertrouwen in overheidsinstellingen, nieuwsmedia en verkiezingen.⁶² Omgekeerd kan echte content sneller worden afgedaan als nep. Actoren die de rechtsstaat of democratie willen ondermijnen kunnen dat ook bereiken door het creëren van verwarring en wantrouwen, het is niet noodzakelijk dat de boodschap in het nepnieuws daadwerkelijk overkomt. De schade zit niet altijd in de inhoud van valse berichtgeving via deepfakes, maar in het punt dat onduidelijk is welke informatie wel en niet betrouwbaar is. Dat is op zichzelf al ondermijnend.

Ook voor deepnudes is de schade voor slachtoffers niet minder als de beelden worden gelabeld als nepbeelden. Een slachtoffer dat de beelden te zien krijgt of erover hoort weet namelijk al lang dat de beelden nep zijn. Veel anderen zullen dat ook weten of vermoeden. De makers van deepnudes pretenderen vaak niet eens dat het echte beelden zijn. De hierboven al genoemde reputatieschade en psychologische schade voor slachtoffers is er niet anders door.⁶³

Ten tweede is er een serieus handhavingsprobleem met de transparantieverlichting. Actoren die deepfakes maken of verspreiden kunnen dat van buiten de EU doen, waardoor ze zich niet verplicht voelen de informatie te labelen. Als ze de informatie wel labelen, maar bijvoorbeeld in de aankondiging, dan kan die meta-informatie snel verloren gaan als de deepfakes toegankelijk worden gemaakt via embedding of deeplinking. Als de labels in het beeldmateriaal zijn verwerkt, bijvoorbeeld een aankondiging in de eerste paar seconden, kan dat er eenvoudig worden afgeknipt. Wellicht de beste manier is om de beelden van een watermerk te voorzien of voortdurend in een hoekje het label te tonen dat het om een deepfake gaat. Maar het voortdurend in beeld brengen doet af aan de inhoud, waardoor makers en gebruikers van deepfakes minder geneigd zullen zijn deze aanpak te kiezen. Artikel 50 AI-verordening legt de transparantieplichting enkel op aan de aanbieders van de AI-systemen die deepfakes maken, niet aan degenen die de beelden daarna bekijken en verspreiden. Als zij de labels verwijderen, dan lijkt dat niet in strijd met de AI-verordening.

Uiteraard kunnen toezichthouders proberen om ongelabelde deepfakes offline te halen, maar populaire content verspreidt zich razendsnel. Het is vrijwel onmogelijk om al het beeldmateriaal volledig van het internet te verwijderen, zeker op plaatsen die buiten hun jurisdicties vallen. Hetzelfde geldt voor het uitdelen van boetes. De AI-verordening biedt weliswaar de mogelijkheid tot het opleggen van stevige boetes,⁶⁴ maar dat is lastig uitvoerbaar als de deepfakes worden gemaakt en aangeboden buiten de EU.

Toezichthouders kunnen internetserviceproviders aanspreken om de meest problematische deepfakes offline te halen en te weren, maar dat kan niet op basis van de AI-verordening. Daarvoor moet de Digital Service Act (DSA) worden ingezet.⁶⁵ Met behulp van de DSA kan illegale content offline worden gehaald, maar alleen onder bepaalde voorwaarden, bijvoorbeeld wanneer het om beelden gaat die aanzetten tot geweld en criminaliteit, beelden die beschermd zijn via intellectueel eigendom, of pornografische beelden. Deepnudes vallen hier wel onder, maar niet zonder meer alle vormen van problematische deepfakes. Maar ook via de DSA kan het lastig zijn al deze content tegen te houden.⁶⁶

6. Conclusie

Deepfakes kunnen grote maatschappelijke en persoonlijke schade berokkenen. Bij deepnudes ligt de nadruk (nog meer) op persoonlijk leed. Het is duidelijk dat de risico's van deepfakes niet 'beperkt' zijn. Zowel feitelijk als in juridische zin zijn de risico's serieus. De gezondheid, veiligheid en grondrechten van burgers worden bedreigd. Daar komt bij dat transparantie alleen onvoldoende is: ook beelden waarvan duidelijk is dat ze nep zijn, kunnen ernstige schade berokkenen. Transparantie is bovendien lastig te handhaven, omdat bij het verspreiden van nepbeelden de labels (al dan niet opzettelijk) snel verloren kunnen gaan, aangenomen dat de beelden überhaupt worden gelabeld. Daarmee is de huidige risico-inschatting niet adequaat. De EU zou deepfakes daarom als een hoog risico moeten bestempelen in de AI-verordening.

Voor deepnudes zou nog een stap verder kunnen worden gegaan door deze technologie op te nemen in de lijst met verboden AI-technologie

Voor deepnudes zou, vanwege de zeer schadelijke gevolgen, nog een stap verder kunnen worden gegaan door deepnude technologie op te nemen in de lijst met verboden AI-technologie in artikel 5 AI-verordening. In deze lijst staan onder meer het gebruik van subliminale beïnvloedingstechnieken, social scoring systemen, predictive policing en real-time biometrische identificatie op

afstand. Zulke AI-systemen kunnen zeer ernstige inbreuken maken op onder meer het recht op privacy, het recht op non-discriminatie en het recht op menselijke waardigheid. Omdat voor deepnudes hetzelfde geldt, kan worden gesteld dat deepnude AI-technologie niet als 'hoog risico' moet worden gezien, maar als 'onaanvaardbaar risico'.

Omdat ook beelden die duidelijk als nep zijn gelabeld serieuze schade kunnen aanrichten, lijkt het beter om ook de definitie van deepfakes die de AI-verordeningen geeft in artikel 3 lid 60 aan te passen. In de definitie zou de zinsnede 'en door een persoon ten onrechte voor authentiek of waarheidsgetrouw zou worden aangezien' moeten worden verwijderd. Het maakt namelijk niet uit of de deepfakes pretenderen authentiek of waarheidsgetrouw te zijn. Het maakt ook niet uit of de deepfakes door gebruikers of slachtoffers worden aangezien als authentiek of waarheidsgetrouw. De schade wordt immers veroorzaakt door de inhoud van de beelden, niet door het feit dat ze als echt worden gepercipieerd. ●

59. Art. 50 lid 1 AI-verordening.

60. Art. 50 lid 5 AI-verordening.

61. C. Vaccari & A. Chadwick, 'Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news', *Social media+ society*, 2020, 6(1); H. Etienne, 'The future of online trust (and why Deepfake is advancing it)', *AI and Ethics* 2021, 1(4), p. 553-562.

62. P. Dixit, M. Kaur & L. Mistrean, 'The Impact of Deepfake Technology on Public Trust in Economic Policies, in: *Mastering Deepfake Technology: Strategies for Ethical Management and Security*, Gistrup: River Publishers 2025, p. 189-210.

63. K. Hao, 'Deepfake porn is ruining women's lives. Now the law may finally ban it', *MIT Technology Review* 2021, technologyreview.com/2021/02/12/1018222/

deepfake-revenge-porn-coming-ban/. G. Yadav, M.Z. Sadique, S. Kumar, R. Sharma, M. Sharma, R. Sharma & T. Rattan, 'Psychological Trauma and Legal Challenges of Deep fake Technology', *Sciences of Conservation and Archaeology* 2025, 37(1), p. 143-150.

64. Art. 99-101 AI-verordening.

65. C. Cauffman & C. Goanta, 'A new order: The digital services act and consumer

protection', *European Journal of Risk Regulation*, 2021 12(4), p. 758-774; T. Enarsson, 'Navigating hate speech and content moderation under the DSA: insights from ECtHR case law', *Information & Communications Technology Law* 2024, 33(3), p. 384-401.

66. A. Reyna, 'DMA and DSA Effective Enforcement - Key to Success', *Journal of Antitrust Enforcement* 2024, 12(2), p. 320-324.