

AI en kinderpornografisch materiaal: Stand van zaken en een blik op de toekomst

samenvatting

Analiv Hasert, Andreia Costa Marchante, Terezia Kucarova
Erasmus Universiteit Rotterdam

Introductie

Kinderpornografisch materiaal vormt een snel veranderende en dynamische bedreiging en is een van de schadelijkste vormen van online uitbuiting met kinderen als doelwit. Recente ontwikkelingen op het gebied van kunstmatige intelligentie (AI) hebben dit landschap ingrijpend veranderd: dankzij generatieve AI-technologie kunnen daders hyperrealistisch en eenvoudig schaalbaar kinderpornografisch materiaal produceren, zonder dat zij daarvoor directe toegang tot een kind nodig hebben. Zowel de omvang als de ernst van AI-gegenereerd kinderpornografisch materiaal nemen snel toe; de combinatie van steeds makkelijker beschikbare hulpmiddelen zorgt ervoor dat de technologische drempels voor het maken en verspreiden van dergelijke inhoud aanzienlijk worden verlaagd. Aangezien het moeite kost om met wettelijke kaders en opsporingsmethoden gelijke tred te houden met de technologische ontwikkelingen, is er dringend behoefte aan evidence-based kennis die als basis kan dienen voor effectieve preventieve en reactieve maatregelen.

AI-ontwerpprogramma's

De belangrijkste technologie achter AI-gegenereerd kinderpornografisch materiaal is het diffusiemodel: een vorm van generatieve AI die afbeeldingen genereert op basis van ingevoerde tekst. Er zijn twee brede categorieën te onderscheiden binnen AI-gegenereerd kinderpornografisch materiaal, namelijk: AI-gegenereerde beelden en door AI bewerkte inhoud. AI-gegenereerd kinderpornografisch materiaal omvat fictieve afbeeldingen van minderjarigen, die grotendeels worden gemaakt met behulp van open-source diffusiemodellen, die daders minder bescherming van buitenaf bieden dan de closed-source alternatieven.

Dankzij de opkomst van precisietechnieken binnen deze categorie kunnen modellen verder worden gespecialiseerd om gerichte resultaten te genereren. Bij door AI gemanipuleerd kinderpornografisch materiaal wordt gebruikgemaakt van een echte afbeelding van een persoon, die vervolgens met behulp van AI verder wordt bewerkt, waardoor de betrokkene rechtstreeks het slachtoffer wordt. Nieuwe technieken zorgen ervoor dat de dreiging zich in hoog tempo uitbreidt, aangezien synthetische bewegende beelden en audio-deepfakes de productie van multimodaal misbruikmateriaal mogelijk maken.

Motivatie: Makers

Uit onderzoek blijkt dat AI-gegenereerd kinderpornografisch materiaal om uiteenlopende redenen wordt geproduceerd en dat deze verder gaan dan louter seksuele bevrediging. Sommige makers worden gedreven door financieel gewin en maken gebruik van platforms en gemeenschappen die specifieke aanvragen of ecosystemen voor het delen van modellen ondersteunen. Anderen streven naar status, technisch aanzien of acceptatie binnen gemeenschappen van makers. AI maakt ook dwangconstructies mogelijk, zoals sextortion. Het is een eigenschap van AI dat de drempel voor belediging wordt verlaagd. Voor het maken van dit materiaal is directe toegang tot een kind niet langer nodig, en zelfs onervaren gebruikers kunnen met behulp van deze tools zeer realistische inhoud produceren. Hierdoor wordt de groep potentiële makers groter en ontwikkelt AI-gegenereerd kinderpornografisch materiaal zich tot een goedkope, op grote schaal voorkomende en zeer flexibele vorm van misbruik. Ook jongere gebruikers spelen een rol, aangezien er bewijs is dat sommige jongeren dit materiaal maken uit nieuwsgierigheid, groepsdruk, wraak of om te experimenteren, vaak zonder zich bewust te zijn van de schade die dit op de lange termijn kan veroorzaken.

Motivation: Consumenten

Uit onderzoek blijkt dat de drijfveren voor consumenten van kinderpornografie onder meer seksuele bevrediging, verslaving, parafiele seksuele interesses – met name pedofilie – en normalisering kunnen omvatten. Uit onderzoek is gebleken dat mensen verslaafd kunnen raken aan pornografie, waarbij sprake

is van vergelijkbare basismechanismen en onderliggende neurologische processen als bij verslavende stoffen. Ook komt het voor dat mensen die herhaaldelijk kinderpornografisch materiaal bekijken, dwangmatig gedrag vertonen en aangeven dat zij dit als zeer bevredigend ervaren. Bovendien blijkt uit onderzoek dat pedofilie een sterke voorspellende factor is voor het verbruik van kinderpornografisch materiaal, vooral wanneer er sprake is van hyperseksualiteit. Bovendien kan normalisering een rol spelen bij al deze drijfveren, aangezien consumenten van AI-gegenereerd kinderpornografisch materiaal een gevoel van straffeloosheid en een gebrek aan empathie voor de slachtoffers kunnen koesteren door het gebruik van dit materiaal, omdat zij het gevoel hebben dat er minder of geen echte schade wordt aangericht. Uit onderzoek is echter ook gebleken dat het bekijken, bezitten of vervaardigen van kinderpornografisch materiaal afwijkende gedachten, fantasieën en driften versterkt en het risico op escalatie kan vergroten.

Psychologische gevolgen en schade: Slachtoffers

Door AI aangestuurde seksuele uitbuiting van kinderen brengt de slachtoffers ernstige psychologische schade toe. Slachtoffers hebben aangegeven dat zij angst, schaamte, vernedering en zelfverwijt hebben ervaren, alsook gevoelens van hulpeloosheid en verlies van controle over hun identiteit, wat wijst op symptomen die overeenkomen met angst, depressie en PTSS. Het specifieke karakter van AI-gegenereerd kinderpornografisch materiaal vormt een unieke uitdaging: het verwijderen van dit materiaal wordt bemoeilijkt door de eindeloze replicatie en snelle verspreiding ervan online, waardoor het trauma in stand wordt gehouden en de waardigheid van het slachtoffer voortdurend wordt geschonden. Het is voor slachtoffers steeds moeilijker om juridische stappen te ondernemen, aangezien onlineplatforms verzoeken tot verwijdering vaak vertragen en de versnipperde wettelijke regelingen de toegang tot de rechter bemoeilijken. Er is weinig onderzoek gedaan naar slachtoffers van AI-gegenereerd kinderpornografisch materiaal, en met de bestaande interventies lukt het niet om adequaat rekening te houden met deze complexe aspecten.

Maatregelen ter beperking en preventie

Wereldwijd zijn er diverse wettelijke en technologische kaders opgezet om kinderpornografisch materiaal tegen te gaan en/of wordt hier momenteel aan gewerkt. In de Europese Unie richten de inspanningen zich momenteel op Verordening 2011/93/EU en de herziening van 2024 die nog in behandeling is. Dit wordt aangevuld door de EU AI-verordening. In het Verenigd Koninkrijk wordt in wetgeving zoals de Protection of Children Act, Coroners and Justice Act en Crime and Policing Bill ook aandacht besteed aan de toename van AI-gegenereerd kinderpornografisch materiaal. In de Verenigde Staten richten de PROTECT Act en, meer recentelijk, de STOP CSAM Act en de ENFORCE Act zich eveneens op de aanpak van het maken en bezitten van AI-gegenereerd kinderpornografisch materiaal. Bovendien zijn er, hoewel deze ondergeschikt zijn aan wettelijke handhaving, ook technologische en medische maatregelen die het vermijden waard zijn, waarbij enkele preventiestrategieën bestaan uit door het bedrijfsleven geïnitieerde digitale watermerken, het samenstellen van veiligere trainingsdatasets (bijvoorbeeld LAION-5B) en medische behandelingen voor seksueel afwijkend gedrag die mogelijk een zekere preventieve werking bieden.

Gaten in de wetgeving en uitdagingen op het gebied van opsporing

De huidige wetgeving inzake materiaal met seksueel misbruik van kinderen is nog niet volledig aangepast aan AI-gegenereerde inhoud. Een belangrijk probleem is dat er bij AI-gegenereerd kinderpornografisch materiaal mogelijk geen echt kind op de afbeelding zelf te zien is, wat de juridische kwalificatie en de kwestie omtrent strafbaarheid bemoeilijkt. In verschillende rechtsgebieden bestaat er onduidelijkheid over wie verantwoordelijk is: de eindgebruiker, degene die het model heeft getraind, het platform of de aanbieder van het systeem dat de inhoud mogelijk heeft gemaakt. In de EU is het huidige wettelijke kader opgesteld voordat moderne generatieve AI bestond en het biedt daarom niet altijd duidelijkheid over synthetisch materiaal. Deze kloof is vooral zichtbaar in de periode tussen de oudere wetgeving en de nog in behandeling zijnde regels. Ook het opsporen wordt steeds moeilijker.

Traditionele hash-based methoden zijn effectief voor bekend materiaal, maar minder bruikbaar wanneer elke AI-gegenereerde afbeelding uniek is. Watermerken en AI-detectiesystemen zijn weliswaar nuttig, maar overtreders kunnen deze vaak verwijderen, wijzigen of omzeilen. Tegelijkertijd wordt AI-gegenereerd kinderpornografisch materiaal steeds fotorealistischer en komt het steeds vaker voor.

Toekomstperspectief: Onderzoeksaanbevelingen

Na het verzamelen van informatie in een evidence-based rapport over de huidige stand van kennis en wetgeving op het gebied van door AI gemanipuleerd en AI-gegenereerd kinderpornografisch materiaal sluiten wij af met een overzicht van onderzoeksprioriteiten die als leidraad kunnen dienen voor toekomstig beleid en handelingsperspectief op dit snel evoluerende terrein. Tot de onderzoeksprioriteiten behoren onder meer nader onderzoek naar systematisch misbruik van open-source AI-modellen, de integriteit van trainingsdatasets en de implicaties van opkomende technologieën zoals het verenigen van verschillende tools of agentische AI. Het blijft van cruciaal belang om de cognitieve en morele kaders van makers en consumenten te bestuderen om doelgroepen waarbij actie vereist is in kaart te brengen; er is verder onderzoek nodig naar de sociale en ontwikkelingsgerelateerde drijfveren achter het maken van dergelijk materiaal, zoals online gemeenschappen en in contact komen met jongeren, alsook naar het identificeren van consumentenprofielen en het verduidelijken van het verband tussen de consumptie van (AI-gegenereerd) kinderpornografisch materiaal en verslaving, een verband dat buiten de context van volwassenenpornografie nog onvoldoende is onderzocht. Bovendien is er nog steeds weinig onderzoek gedaan naar de specifieke psychologische gevolgen van slachtofferschap van AI-gegenereerd kinderpornografisch materiaal, met name wat betreft de gendergerelateerde aspecten van de schade, secundaire victimisatie en de toereikendheid van bestaande ondersteuningsregelingen voor slachtoffers.