Navigeren in een nieuw risicodomein

Een kwalitatief, verkennend onderzoek naar de rol van generatieve kunstmatige intelligentie (GenAI) bij risicovol of strafbaar gedrag onder jongeren

Senna Stoltink – 2867253

Masterthesis Criminologie

Afstudeerrichting Interventiecriminologie

Faculteit der Rechtsgeleerdheid

Vrije Universiteit Amsterdam

Begeleider: dr. J. Hanrath

Tweede lezer: dr. J. van der Kemp

In opdracht van het RIEC Midden-Nederland

25 juni 2026

VU VRIJE UNIVERSITEIT AMSTERDAM

RIEC Midden-Nederland

Inhoudsopgave

Samenvatting

Dit onderzoek verkent wat de opkomst van generatieve kunstmatige intelligentie (GenAI) betekent voor de betrokkenheid van jongeren bij risicovol of strafbaar gedrag. Om dit te verkennen, zijn vijf informanten geïnterviewd over hun zicht op AI-gerelateerde criminaliteit en hun zorgen daarover met betrekking tot jongeren. Vervolgens zijn drie focusgroepen gehouden met dertien jongeren tussen de 14 en 23 jaar over hun ervaringen met en opvattingen over GenAI-gebruik. De resultaten laten zien dat het zicht op AI-gerelateerde criminaliteit momenteel beperkt is doordat AI-gebruik bij delicten niet systematisch wordt geregistreerd en niet altijd wordt herkend door de politie of door slachtoffers. De informanten signaleren dat de brede beschikbaarheid van deepfake- en deepnude-tools kan bijdragen aan een lagere drempel voor GenAI-gerelateerd delictgedrag onder jongeren. Tegelijkertijd blijkt dat de deelnemende jongeren zelf normatieve opvattingen hebben over het gebruik van GenAI, voornamelijk met betrekking tot het maken van deepfakes. Deepnudes worden daarbij unaniem afgewezen, terwijl andere vormen van deepfakes afhankelijk van de context worden beoordeeld. Buiten de schoolcontext ervaren de jongeren weinig concrete richtlijnen voor GenAI-gebruik. De feitelijke betrokkenheid van de deelnemende jongeren bij risicovol of strafbaar GenAI-gebruik komt in dit onderzoek beperkt naar voren. Slechts een kleine groep ICT-vaardige jongeren experimenteert met het omzeilen van ingebouwde veiligheidsmechanismen (*guardrails*) van GenAI. Daarnaast geven verschillende jongeren aan dat *guardrails* hen helpen om de grenzen van acceptabel GenAI-gebruik te herkennen. De bevindingen suggereren dat GenAI vooral een nieuw risicodomein creëert waarin jongeren moeten navigeren tussen wat technisch mogelijk en maatschappelijk aanvaardbaar is. Het onderzoek draagt bij aan een nog beperkt kennisveld door inzicht te bieden in de wijze waarop jongeren omgaan met de mogelijkheden en risico's van GenAI.

1. Inleiding

1.1 Aanleiding en probleemstelling

De opkomst van generatieve kunstmatige intelligentie (GenAI) heeft in korte tijd geleid tot ingrijpende veranderingen in de manier waarop mensen informatie zoeken, communiceren en digitale content creëren. Generatieve AI verwijst naar AI-systemen die op basis van instructies van gebruikers (*prompts*) nieuwe content kunnen genereren, zoals tekst, afbeeldingen, video en audio (Ciubotaru, 2025). Toepassingen zoals ChatGPT, Claude en Grok zijn sinds eind 2022 toegankelijk voor een breed publiek en worden gebruikt voor uiteenlopende doeleinden, variërend van onderwijs en werk tot entertainment (Du, 2024).

Tegelijkertijd waarschuwen veiligheidsdiensten en onderzoekers dat GenAI, ondanks de *guardrails* (ingebouwde veiligheidsmechanismen), ook kan worden ingezet voor risicovol of strafbaar gedrag (Europol, 2025; Hayward & Maas, 2020; NCTV, 2025). Zo rapporteert AI-ontwikkelaar Anthropic (2025) dat haar taalmodel Claude wordt gebruikt als hulpmiddel bij verschillende vormen van cybercriminaliteit, variërend van het stelen van gegevens tot het ontwikkelen van malware. GenAI heeft hierbij een drempelverlagende werking, doordat het gemakkelijk toegankelijk is via het internet en het gebruik ervan geen specifieke technische kennis vereist (Europol, 2025; NCTV, 2025; Schuilenburg & Soudijn, 2024).

Hoewel GenAI kan worden ingezet voor uiteenlopende vormen van criminaliteit, krijgen deepfakes bijzondere aandacht door het gemak waarmee zij kunnen worden gemaakt en hun brede toepasbaarheid. Deepfakes, een combinatie van de woorden *deep learning* en *fake*, zijn synthetische beelden of audio die worden gemaakt met behulp van GenAI (Chadha et al., 2021). Dit materiaal is vaak hyperrealistisch, waardoor het lastig is om onderscheid te maken tussen deepfakes en niet-gemanipuleerd materiaal (Chadha et al., 2021). Deepfakes worden gebruikt voor uiteenlopende vormen van criminaliteit, waaronder fraude, identiteitsmisbruik en het genereren van seksueel misbruikmateriaal (Hayward & Maas, 2020; Schuilenburg & Soudijn, 2024; Kopecký & Voráč, 2025).

Deze ontwikkelingen zijn bijzonder relevant met het oog op jongeren. Zij groeien op in een sterk gedigitaliseerde samenleving en brengen een aanzienlijk deel van hun tijd online door (Van Der Meijs, z.d.). Hun vertrouwdheid met digitale technologie vertaalt zich dan ook naar intensief gebruik van GenAI (Verma & Lall, 2025). Zo blijkt uit cijfers van het Centraal Bureau voor de Statistiek (2024) dat bijna een kwart van de Nederlanders van twaalf jaar en ouder GenAI gebruikt om teksten, video's of afbeeldingen te genereren. Tegelijkertijd bevinden zij zich in een levensfase waarin experimenteren, identiteitsvorming en het verkennen van sociale grenzen een belangrijke rol spelen (Goldsmith & Wall, 2019; Yu et al., 2025). De combinatie van deze ontwikkelingsfase met de laagdrempelige toegang tot GenAI roept vragen op over de rol die GenAI kan spelen bij risicovol of strafbaar gedrag onder hen.

Ook buiten expliciet criminele contexten blijkt dat het gebruik van GenAI door jongeren tot problematische situaties kan leiden. Zo waarschuwde de politie recent voor een trend waarbij jongeren met behulp van GenAI deepfakes maken die de indruk wekken dat zich een dakloos persoon in hun woning bevindt, om hun ouders te laten schrikken (NOS, 2025). Dit leidde in verschillende gevallen tot valse meldingen en onnodige inzet van politiecapaciteit, waaronder de inzet van een politiehelikopter (NOS, 2025).

Daarnaast bestaan er in relatie tot jongeren zorgen over het gebruik van deepfake-technologie voor het maken van deepnudes. Een deepnude is “een nep-naakfoto of nep-naaktvideo die met deepfake-technologie gemaakt is” (Offlimits, 2026). De toegankelijkheid van dergelijke toepassingen heeft ertoe geleid dat het aantal meldingen van deepnudes de afgelopen jaren is toegenomen (Offlimits, 2025). Ook bestaande generatieve AI-systemen, zoals Grok, worden bekritiseerd omdat het daarmee relatief eenvoudig is om dit soort beelden te genereren. Jongeren komen online met deze toepassingen in aanraking. Zo geeft ruim één op de vijf Belgische jongeren aan wel eens een deepnude te hebben gezien via sociale media, zoals Snapchat, Instagram of X (Szyf et al., 2023, p. 38). Daarnaast heeft 60,5 procent van de jongeren die met deze toepassingen bekend zijn zelf geprobeerd om een deepnude te maken (Szyf et al., 2023, p. 42). Daarbij variëren de motieven van nieuwsgierigheid tot het vernederen van de ander (Szyf et al., 2023, p. 43). Bovendien zijn de afgelopen jaren in verschillende landen zaken aan het licht gekomen waarbij door minderjarigen deepnudes zijn gemaakt van klasgenoten, zoals in de Verenigde Staten (Sharp, 2024), Spanje (Jones, 2024), Engeland (Weale, 2025) en Australië (Beazley & Touma, 2024).

Ondanks deze ontwikkelingen is nog weinig bekend over de manieren waarop GenAI mogelijk een rol speelt bij de betrokkenheid van Nederlandse jongeren bij risicovol of strafbaar gedrag. Bestaand onderzoek naar GenAI-gebruik door jongeren richt zich voornamelijk op legitieme toepassingen, zoals voor onderwijs, creativiteit en productiviteit terwijl risicovol of strafbaar gebruik grotendeels buiten beschouwing blijft. Tegelijkertijd bevinden jongeren zich in een leeftijdsfase waarin het verkennen van grenzen een belangrijke rol speelt (Goldsmith & Wall, 2019; Yu et al., 2025). Bovendien blijken online omgevingen, door de ervaren anonimiteit, toegankelijkheid en lossere gedragsnormen, een omgeving waarin jongeren gemakkelijker experimenteren met normoverschrijdend gedrag (Goldsmith & Wall, 2019; Suler, 2004). Hoewel in toenemende mate wordt gewezen op de risico’s van GenAI, ontbreekt vooralsnog inzicht in de mate waarin jongeren deze mogelijkheden daadwerkelijk verkennen en benutten.

Het is relevant om hier oog voor te hebben, omdat GenAI zich in hoog tempo ontwikkelt en steeds sterker aanwezig is in de online leefwereld van jongeren. Tegelijkertijd worden professionals binnen het zorg- en veiligheidsdomein in toenemende mate geconfronteerd met nieuwe digitale verschijningsvormen van risicovol gedrag, waaronder deepfakes en deepnudes. Om hier adequaat op te kunnen reageren is meer inzicht nodig in de manieren

waarop jongeren GenAI gebruiken, welke grenzen zij daarbij ervaren en in hoeverre zij in aanraking komen met risicovolle of strafbare toepassingen van GenAI. Het huidige onderzoek beoogt een bijdrage te leveren aan het verkleinen van dit kennishiaat.

1.2 Onderzoeksvraag en deelvragen

Dit onderzoek heeft als doel te verkennen wat de opkomst van generatieve kunstmatige intelligentie (GenAI) betekent voor de betrokkenheid van de deelnemende jongeren bij risicovol of strafbaar gedrag. Daarom staat in deze bijdrage de volgende onderzoeksvraag centraal: 'Wat betekent de opkomst van generatieve kunstmatige intelligentie (GenAI) voor de betrokkenheid van jongeren bij risicovol of strafbaar gedrag?'

Gezien het verkennende karakter van het onderzoek wordt zowel aandacht besteed aan de perspectieven van professionals die vanuit hun expertise zicht hebben op ontwikkelingen rondom GenAI en/of jongeren als aan de ervaringen en opvattingen van jongeren zelf. Ter beantwoording van de hoofdvraag zijn daarom vijf deelvragen opgesteld. De eerste twee deelvragen worden beantwoord op basis van interviews met vijf informanten:

1. In hoeverre is er zicht op criminaliteit waarbij generatieve AI een rol speelt volgens de informanten?
2. Welke zorgen hebben de informanten over generatieve AI in relatie tot jongeren?

Vervolgens worden de overige drie deelvragen beantwoord op basis van drie focusgroepen met de deelnemende jongeren:

3. Hoe beoordelen de deelnemende jongeren wat al dan niet verantwoord is bij het gebruik van generatieve AI?
4. Welke risicovolle of criminele toepassingen van generatieve AI komen de deelnemende jongeren tegen in hun omgeving?
5. Hoe gaan de deelnemende jongeren om met de ingebouwde veiligheidsmechanismen (*guardrails*) bij hun gebruik van generatieve AI?

1.3 Leeswijzer

Het huidige onderzoek is opgebouwd uit meerdere onderdelen. Allereerst wordt in het theoretisch kader de wetenschappelijke achtergrond geschetst die er tot op heden over GenAI en criminaliteit, en het gebruik van GenAI door jongeren is. Ook worden enkele theoretische concepten besproken die worden gebruikt voor het duiden van de bevindingen. Daarna wordt in de methode uiteengezet op welke wijze dit onderzoek is uitgevoerd. Vervolgens worden de resultaten van de interviews en de focusgroepen besproken aan de hand van de vijf deelvragen. Daarna wordt een conclusie van de hoofdbevindingen gepresenteerd, en wordt besproken hoe de resultaten zich verhouden tot de bestaande literatuur. Tot slot wordt kritisch

gereflecteerd op het huidige onderzoek en wordt afgesloten met enkele aanbevelingen voor vervolgonderzoek en de praktijk.

2. Theoretisch kader

Hoewel steeds meer veiligheidsdiensten erop wijzen dat generatieve AI (GenAI) wordt ingezet voor het plegen van criminaliteit, is nog nauwelijks onderzocht hoe dit mogelijk speelt bij jongeren. In dit hoofdstuk wordt die kwestie verkend aan de hand van de wetenschappelijke kennis die voorhanden is. Eerst wordt ingegaan op GenAI als technologie die zowel bestaande delicten kan versterken als nieuwe vormen van online delictgedrag kan faciliteren. Daarbij worden verschillende verschijningsvormen besproken. Vervolgens wordt besproken hoe de online omgeving gelegenheid biedt tot het plegen van criminaliteit en de mogelijke invloed van GenAI op die gelegenheid. Daarna wordt ingegaan op de online omgeving als digitale risicocontext voor jongeren. Hier wordt het daderschap van online criminaliteit onder jongeren besproken, evenals de factoren van het internet die ertoe kunnen leiden dat zij online sneller de grens opzoeken. Tot slot wordt het gebruik van GenAI door jongeren besproken, en wordt beschouwd hoe dit een rol kan spelen bij hun betrokkenheid bij risicovol of strafbaar gedrag.

2.1 GenAI en criminaliteit

Verschillende veiligheidsdiensten waarschuwen dat GenAI in toenemende mate een rol speelt in de modus operandi van daders (Europol, 2025; Nationaal Coördinator Terrorismebestrijding en Veiligheid, 2024). Tegelijkertijd is nog onduidelijk in hoeverre dit werkelijk al gebeurt. Binnen de literatuur wordt vooral gesproken over criminaliteit waarbij GenAI wordt ingezet in termen van dreigingen, zorgen en voorspellingen. Daarnaast zijn er enkele studies waarin het gebruik van GenAI voor criminaliteit empirisch is onderzocht. In dit hoofdstuk wordt daarom onderscheid gemaakt tussen empirische en toekomstverkennende studies.

De relatie tussen GenAI en criminaliteit is niet eenduidig. Binnen de literatuur worden verschillende indelingen gehanteerd om dit te duiden. Doorgaans wordt daarbij onderscheid gemaakt tussen AI als hulpmiddel bij bestaande vormen van criminaliteit en AI als zowel middel als doelwit van criminaliteit (Hayward & Maas, 2020; Jeong, 2020). Dit sluit aan bij het klassieke onderscheid binnen de cybercrime-literatuur, tussen cybercriminaliteit in enge zin (waarbij ICT zowel het middel als doelwit is) en gedigitaliseerde criminaliteit (waarbij ICT wordt ingezet voor de uitvoering van bestaande delicten) (Jeong, 2020; Van der Wagen et al., 2020). Het huidige onderzoek zal zich richten op het gebruik van GenAI als hulpmiddel bij risicovol of strafbaar gedrag.

2.1.1 Nieuwe criminele werkwijzen

Enerzijds heeft de komst van GenAI gezorgd voor het ontstaan van nieuwe criminele werkwijzen (Europol, 2025). Europol (2025, p. 21) stelt dat GenAI hierdoor een “catalyst for crime” is, waarmee wordt bedoeld dat het een factor is die criminaliteit sneller kan laten

plaatsvinden. Dit lijkt vooral te worden beïnvloed door de komst van AI-tools voor het genereren van deepfakes.

Ten eerste kunnen daders met deepfakes verschillende manieren van oplichting professionaliseren (Hayward & Maas, 2020; Van der Wagen et al., 2020). Deepfaketechnologie maakt het mogelijk om afbeeldingen, video's en audiofragmenten te genereren die lastig van echt te onderscheiden zijn (Chadha et al., 2021). Hiermee kunnen personen worden nagebootst om geloofwaardiger over te komen op het slachtoffer (Dzuba, 2025; Schuilenburg & Soudijn, 2024). Hoewel deepfaketechnologie geen nieuw fenomeen is, heeft GenAI wel geleid tot een bredere beschikbaarheid van tools om dergelijke content te creëren. Daarnaast hebben deze tools de drempel verlaagd om deepfakes te maken, omdat het gebruik ervan geen specifieke technische kennis vereist (Europol, 2025; Kopecký & Voráč, 2025; Nationaal Coördinator Terrorismebestrijding en Veiligheid, 2024; Schuilenburg & Soudijn, 2024).

Doordat deepfakes de mogelijkheid bieden om anderen te imiteren, kunnen ze worden ingezet voor uiteenlopende vormen van criminaliteit. Ten eerste kan het worden ingezet voor verschillende vormen van fraude, zoals CEO-fraude of datingfraude, waarbij criminelen zich voordoen als iemand anders om het slachtoffer geld afhandig te maken (Hayward & Maas, 2020; Van der Wagen et al., 2020).

Daarnaast wordt deepfake-technologie in toenemende mate ingezet om deepnudes te genereren (Kopecký & Voráč, 2025; Offlimits, 2025; Szyf et al., 2023). Een deepnude is "een nep-naaktfoto of nep-naaktvideo die met deepfake-technologie gemaakt is" (Offlimits, 2026, p.8). Expertisecentrum Offlimits nam in 2024 en 2025 een stijging van het aantal meldingen van deepnudes waar ten opzichte van het jaar ervoor (Offlimits, 2025). Tegelijkertijd signaleren zij dat de kwaliteit van de beelden verbetert en daardoor moeilijker als gemanipuleerd te herkennen is (Offlimits, 2025). Bovendien gaat Offlimits (2025) uit van een onderschatting van de werkelijke omvang van deepnudes, omdat slachtoffers vaak niet weten dat er beelden van hen circuleren of hier geen melding van maken. Deepnudes vallen onder de definitie van seksueel beeldmateriaal in de Wet seksuele misdrijven. Het maken en bezitten van deepnudes is strafbaar op grond van artikel 254ba van het Wetboek van Strafrecht. Naast dat zij op zichzelf een strafbaar feit vormen, kunnen deepnudes worden ingezet voor afpersing (sextortion) (Kopecký & Voráč, 2025).

Deepnudes kwalificeren als kindermisbruikmateriaal (CSAM[1]) wanneer een minderjarige wordt afgebeeld (Offlimits, 2026). Uit onderzoek door Protect Children (2026) blijkt dat deefake-technologie daarvoor wordt ingezet. De omvang van AI-gegenereerde CSAM is aanzienlijk toegenomen sinds de komst van GenAI (Protect Children, 2026). Verder blijkt dat

[1] CSAM staat voor Child Sexual Abuse Material.

niet alleen volwassen, maar ook jongeren deze beelden maken (Kopecký & Voráč, 2025; Szyf et al., 2023).

2.1.2 Efficiëntere uitvoering van bestaande werkwijzen

Anderzijds kan GenAI worden ingezet voor een efficiëntere modus operandi van bestaande delicten (Alahmed et al., 2024; Brundage et al., 2018; Europol, 2025). Ten eerste kan GenAI een ondersteunende rol hebben bij het uitvoeren van phishing. Daders kunnen GenAI geloofwaardige phishingberichten laten schrijven die zijn afgestemd op de ontvanger, wat ook wel wordt aangeduid als “spear phishing” (Hayward & Maas, 2020, p. 7; Schmitt & Flechais, 2023). Daarnaast kunnen ze de schrijfstijl van een ander imiteren om geloofwaardiger over te komen op het slachtoffer (Falade, 2023; Gupta et al., 2023; Van der Wagen et al., 2020). Door die overtuigende formulering zijn AI-gegenereerde phishingberichten lastiger als zodanig te herkennen, doordat ze nauwelijks te onderscheiden zijn van legitieme communicatie (Dzuba, 2025).

Verder stelt GenAI daders in staat om deze berichten op een geloofwaardige manier te vertalen zonder zelf kennis te hebben van de taal (Brundage et al., 2018). Hieruit blijkt dat daders hun bereik kunnen uitbreiden zonder zelf te beschikken over de kennis die daarvoor nodig is (Blauth et al., 2022). Deze studies over de inzet van GenAI voor phishing zijn overwegend toekomstverkennend van aard. Echter is uit een recent intelligence-rapport van Anthropic (2025) gebleken dat hun taalmodel Claude wordt ingezet voor het maken van phishinglinks en het genereren van berichten voor datingfraude (p.19-p.24).

Daarnaast kan GenAI worden gebruikt op verschillende momenten in het uitvoeringsproces van cyberaanvallen (Guembe et al., 2022). Zo kan het worden ingezet om malware te maken, kwetsbaarheden in beveiligingssystemen te ontdekken en wachtwoorden te raden (Guembe et al., 2022; Thanh & Zelinka, 2019; Trieu & Yang, 2018). Trieu & Yang (2018) voerden een experimenteel onderzoek uit waarin ze een AI-model ontwikkelden om wachtwoorden te raden. Hieruit concludeerden zij dat hun AI-model dit effectiever uitvoerde dan hacken op basis van traditionele methoden (Trieu & Yang, 2018). Daarnaast is uit onderzoek door Anthropic (2025) gebleken dat Claude wordt gebruikt voor verschillende cyberdelicten zoals het stelen van data en het ontwikkelen en verspreiden van malware (p. 4-20). Hoewel dit onderzoek zicht enkel richtte op Claude, is het volgens Anthropic (2025) aannemelijk dat ook andere frontier-modellen worden misbruikt ter ondersteuning van deze delicten.

De besproken literatuur vormt een eerste beeld van de uiteenlopende verschijningsvormen van criminaliteit waarbij GenAI een rol kan spelen. GenAI fungeert niet als een volledig nieuw crimineel fenomeen, maar als een technologie die zowel bestaande delicten kan versterken, automatiseren of opschalen als nieuwe vormen van online delictgedrag kan faciliteren.

Daarnaast verlaagt GenAI de vereiste kennis voor het plegen van bepaalde delicten zoals phishing, wat kan zorgen voor een verbreding van de daderpopulatie. Dit vormt een aanwijzing dat GenAI invloed heeft op de gelegenheid tot het plegen van (online) criminaliteit.

2.2 Gelegenheid tot criminaliteit in de online omgeving

De klassieke criminologische gelegenheidstheorieën richten zich op criminaliteit in de fysieke omgeving. Echter stellen Van der Wagen et al. (2020, p. 256) dat deze ook “goed toepasbaar zijn op cybercriminaliteit”. In de volgende sectie zal daarom worden ingegaan op de routine activiteitentheorie en hoe die wordt toegepast op criminaliteit in digitale omgevingen.

2.2.1 De routine activiteitentheorie

De routine activiteitentheorie (RAT) stelt dat criminaliteit kan ontstaan wanneer een gemotiveerde dader en aantrekkelijk doelwit samenkomen in tijd en ruimte, zonder dat er geschikt toezicht aanwezig is (Cohen & Felson, 1979). Hierbij ligt de focus op de manier waarop gelegenheid de keuzes van een potentiële dader beïnvloedt (Newburn, 2017). Een gemotiveerde dader is iemand die de kennis, vaardigheden en motivatie heeft om criminaliteit te plegen (Weulen Kranenbarg & Weerman, 2022). Daarnaast kan een aantrekkelijk doelwit zowel een object als persoon zijn (Newburn, 2017; Weulen Kranenbarg & Weerman, 2022). Verder is geschikt toezicht iets of iemand dat in de regel afschrikkend werkt op de dader (Newburn, 2017).

2.2.2 Routine activiteiten in de online omgeving

Verschillende onderzoekers stellen dat de RAT ook online toepasbaar is, mits de elementen worden beschouwd in hun digitale context (Ahmad & Thurasamy, 2022; Vakhitova et al., 2021; Weerman, 2019). Ten eerste zijn er meer gemotiveerde daders, doordat mensen online eerder geneigd kunnen zijn om de grens over te gaan dan offline (Suler, 2004). Bepaalde eigenschappen van het internet kunnen hiervoor zorgen. Suler (2004) stelt dat online een ontremming van het gedrag kan optreden door bepaalde factoren, zoals anonimiteit en de onzichtbaarheid van het slachtoffer. Dit wordt aangeduid als het ‘online disinhibitie effect’ (Suler, 2004). Daarnaast worden online criminele tools en services aangeboden, ook wel *cybercrime-as-a-service* (Manky, 2013). Dit zorgt ervoor dat daders niet te hoeven beschikken over bepaalde kennis of vaardigheden voor het uitvoeren van een online delict. Bovendien bleek uit onderzoek door Brewer et al. (2018) dat de beschikbaarheid van dergelijke tools en services zorgt voor meer gelegenheid voor het plegen van online criminaliteit[2] door jongeren.

Ten tweede zijn online meer geschikte slachtoffers aanwezig. Naast personen kunnen namelijk ook ICT-systemen het doelwit van criminaliteit vormen (Van der Wagen et al., 2020). Daarnaast zijn online meer slachtoffers te bereiken. Verschillende onderzoekers stellen dat het besteden van meer tijd online samenhangt met een verhoogde kans op slachtofferschap van cyber- of gedigitaliseerde criminaliteit (Akgül, 2021; Bossler et al., 2011; Vale et al., 2022).

[2] Online criminaliteit wordt binnen dit onderzoek gebruikt als verzamelterm voor zowel cyber- als gedigitaliseerde vormen van criminaliteit, naar voorbeeld van het CBS en het WODC.

Uit onderzoek van Vale et al. (2022) bleek bijvoorbeeld dat jongeren die intensiever gebruik maken van digitale apparaten zoals een smartphone of laptop, een hogere kans hadden om slachtoffer te worden van online intimidatie.

Hoewel over de betekenis van gemotiveerde daders en geschikte slachtoffers in de online omgeving enige overeenstemming bestaat, ontbreekt dat tot dusver over bekwaam toezicht in de online omgeving. Enerzijds stellen Vakhitova et al. (2021) dat mensen die fysiek aanwezig zijn tijdens het internetgebruik, zoals gezinsleden of huisgenoten, kunnen optreden als informele toezichthouders voor een ander. Echter bleek uit onderzoek door Reyns et al. (2015) dat dergelijk toezicht het risico op slachtofferschap van cyberstalking niet verlaagde en soms juist liet toenemen. Volgens hen komt dat doordat fysieke toezichthouders geen zicht hebben op wat er online gebeurt en daardoor niet adequaat kunnen ingrijpen (Reyns et al., 2015). Andere onderzoekers interpreteren bekwaam toezicht als vormen van digitale bescherming. Voorbeelden hiervan zijn de aanwezigheid van firewall software en antivirussoftware waarmee digitale bedreigingen worden afgewend (Bossler et al., 2011; Dzuba, 2025). Bij onderzoek naar online criminaliteit onder jongeren wordt ook gekeken naar ouders als fysieke toezichthouders (Wissink et al., 2023).

2.2.3 Invloed van GenAI op gelegenheid

De routine activiteitentheorie (RAT) wordt in toenemende mate toegepast op diverse vormen van online criminaliteit. Echter staat de toepassing ervan op delicten waarbij GenAI een ondersteunende rol inneemt vrijwel volledig in de kinderschoenen. Voor zover bekend werd binnen één recent onderzoek door Dzuba (2025) de RAT toegepast op een vorm van AI-criminaliteit, namelijk AI-gefaciliteerde phishing. Daaruit komt naar voren dat GenAI op verschillende manieren invloed heeft op de elementen van de theorie.

Ten eerste gaat de RAT uit van een dader die actief strafbaar gedrag vertoont. Dzuba (2025) stelt dat dit mogelijk is veranderd door de automatiseringsmogelijkheden van GenAI. Doordat GenAI een deel van het uitvoeringsproces van een delict over kan nemen, roept dat vragen op over wie de dader is (Campedelli, 2025; Dzuba, 2025; Schuilenburg & Soudijn, 2024). Zo stellen Guembe et al. (2022) dat bij de uitvoering van DDoS-aanvallen met behulp van GenAI nog slechts beperkte menselijke tussenkomst nodig is, omdat het systeem in staat is zijn aanvalsstrategie tussentijds zelfstandig bij te sturen.

Daarnaast is het mogelijk dat GenAI de drempel heeft verlaagd om delicten te plegen. Dat komt doordat het een deel van de (technische) kennis die nodig is om een delict te plegen, weg kan nemen. Hoewel generieke taalmodellen ethische barrières hebben, zijn er ook taalmodellen zonder barrières die bewust zijn ontwikkeld voor criminele doeleinden (Fire et al., 2025; Van der Wagen et al., 2020). Deze modellen worden aangeduid als dark Large

Language Models (dark LLMs), en zijn in staat om gedetailleerde instructies voor criminele handelingen te geven (Fire et al., 2025).

Bovendien bestaan er AI-agents die autonoom kunnen handelen en daardoor ook zelfstandig illegale handelingen kunnen verrichten (Hayward & Maas, 2020; Nerantzi & Sartor, 2024).Dit roept vragen op over strafrechtelijke aansprakelijkheid. Nerantzi & Sartor (2024, p. 674) gebruiken hiervoor de term “AI hard crimes”: situaties waarin AI-gedreven delicten plaatsvinden zonder dat een persoon strafrechtelijk verantwoordelijk kan worden gehouden. Volgens hen ontstaat in dergelijke gevallen een kloof in de strafrechtelijke aansprakelijkheid, omdat onduidelijk blijft of de mens of de AI-agent verantwoordelijk moet worden gehouden voor het delict (Nerantzi & Sartor, 2024). Deze visie wordt ondersteund door Campedelli (2025), die stelt dat criminologen daardoor verder moeten kijken dan GenAI als hulpmiddel bij het uitvoeren van criminaliteit. Het is dus nog de vraag in hoeverre GenAI zelf kan worden gezien als dader, of slechts als een verlengstuk van iemand met criminele intenties (Dzuba, 2025).

Daarnaast lijkt GenAI volgens Dzuba (2025) ook invloed te hebben op het slachtofferbereik van daders. Dat komt doordat GenAI kan worden ingezet om phishingberichten te vertalen en deze op grote schaal te verspreiden (Brundage et al., 2018; Dzuba, 2025). Hierdoor kunnen veel meer slachtoffers tegelijk worden bereikt door één dader (Brundage et al., 2018; Dzuba, 2025). Daarnaast worden taalmodellen zoals Claude gebruikt voor *victim profiling*, oftewel het analyseren van potentiële slachtoffers (Anthropic, 2025, p. 22). Ook dit zorgt voor een uitbreiding van het slachtofferbereik van daders.

Verder stelt Dzuba (2025) dat de komst van GenAI heeft gezorgd voor nieuwe uitdagingen op het gebied van toezicht op online content. Dat komt doordat daders met behulp van GenAI realistische deepfakes kunnen creëren die nauwelijks van echt te onderscheiden zijn (Park & Nan, 2025; Schuilenburg & Soudijn, 2024). Daardoor kan het door zowel beveiligingssystemen als mensen niet altijd als gemanipuleerd worden herkend (Park & Nan, 2025; Schmitt & Flechais, 2023). De meeste Large Language Models (LLMs) zoals ChatGPT bevatten ethische veiligheidswaarborgen (*guardrails*) om misbruik, zoals het genereren van misleidende content en het produceren van schadelijke of ongepaste output, tegen te gaan. Echter zijn deze modellen kwetsbaar voor *jailbreaking*, waarbij gebruikers een opdracht op een bepaalde manier formuleren om deze ingebouwde beperkingen te omzeilen (Fire et al., 2025, p.2). Bovendien bevatten de eerder besproken dark LLMs geen *guardrails* (Fire et al., 2025).

Op basis van het voorgaande stelt Dzuba (2025) dat de RAT moet worden herzien, waarbij rekening wordt gehouden met de door GenAI veroorzaakte verschuivingen. Dit gaat over de drempelverlaging voor daders, de uitbreiding van het slachtofferbereik en de bemoeilijkte detecteerbaarheid van AI-gefaciliteerde criminaliteit. Daarnaast is er sprake van

onduidelijkheid over daderschap bij delicten waarbij GenAI wordt ingezet. Hoewel het onderzoek van Dzuba (2025) zich specifiek richt op GenAI-gefaciliteerde phishing, is het voorstelbaar dat deze verschuivingen zich ook voordoen bij andere vormen van criminaliteit waarbij GenAI als hulpmiddel wordt aangewend.

2.3 Jongeren in een digitale risicocontext

Deze veranderingen in online gelegenheid en daderschap zijn relevant met het oog op jongeren, omdat zij veel tijd online doorbrengen. Het merendeel van hen is dagelijks online (Kanne et al., 2019) en hun online en offline leefwereld raken steeds meer met elkaar verweven (Van der Wagen et al., 2020; Weulen Kranenbarg & Van 't Hoff-de Goede, 2023). Dit wordt ook wel aangeduid als de "hybride leefwereld" van jongeren, waarbij hun digitale en fysieke ervaringen voortdurend door elkaar heen lopen (De Jong & De Wit, 2024). Jongeren communiceren met leeftijdsgenoten, vinden ontspanning en vormen hun identiteit zowel online en offline waarbij beide domeinen elkaar beïnvloeden (Scott et al., 2022). In de volgende sectie zal daarom worden ingegaan op de online omgeving als risicocontext voor jongeren.

2.3.1 Daderschap van online criminaliteit onder jongeren

Door hun uitgesproken online aanwezigheid lopen jongeren de kans om online in aanraking te komen met criminaliteit, als slachtoffer maar ook als dader (Weijers, 2024; Weulen Kranenbarg & van 't Hoff-de Goede, 2023). Volgens Weijers (2024, p. 85) heeft de snelle digitalisering ervoor gezorgd dat niet alleen ICT-vaardige jongeren maar ook "doorsneejongeren" zich met online criminaliteit zijn gaan bezighouden. Jongeren laten daarbij steeds vaker een hybride vorm van delictgedrag zien, waarbij online en offline criminaliteit naast elkaar voorkomen. Zo is uit onderzoek door Roks et al. (2021) gebleken dat jongeren die voorheen vooral betrokken waren bij straatgerichte criminaliteit, zich ook bezig zijn gaan houden met vormen van online criminaliteit zoals fraude en phishing. Deze ontwikkeling wordt door de auteurs aangeduid als "the hybridization of street offending" (Roks et al., 2021, p. 932).

Deze ontwikkeling lijkt te worden ondersteund door cijfers over online daderschap onder jongeren. Hoewel de algemene jeugdcriminaliteit in Nederland al jaren afneemt en op korte termijn stabiliseert, geldt die afname niet direct wanneer specifiek wordt gekeken naar registratiecijfers over online criminaliteit onder jongeren (Tollenaar et al., 2024). Uit het Cybercrimebeeld blijkt namelijk dat de helft van de Nederlandse cybercrimeverdachten 25 jaar of jonger is (Openbaar Ministerie, 2024, p. 3).

Tegelijkertijd vormen politie- en justitieregistraties een beperkte basis voor het vaststellen van de omvang van online daderschap onder jongeren. Dat komt doordat de politie maar beperkt zicht heeft op dit fenomeen, onder andere door de onzichtbaarheid van daders en de lage aangiftebereidheid van slachtoffers (Weijers, 2024). Zo is uit onderzoek door het Centraal Bureau voor de Statistiek (2025) gebleken dat maar 18 procent van de slachtoffers van online criminaliteit aangifte doet bij de politie.

Door deze beperkingen wordt daderschap van online criminaliteit onder jongeren gemeten aan de hand van zelfrapportage-enquêtes (Tollenaar et al., 2024; Van der Wagen et al., 2020). De resultaten van de Monitor Zelfgerapporteerde Jeugdcriminaliteit tonen dat cyber-

en gedigitaliseerde delicten behoren tot de meest voorkomende zelfgerapporteerde delicten onder jongeren (Tollenaar et al., 2024). Hieruit blijkt dat voor zover daderschap onder jongeren zichtbaar is, vooral hacken, “diefstal van virtuele goederen” en het “online verspreiden van seksueel beeldmateriaal van minderjarigen” naar voren komen (Tollenaar et al., 2024, p.33).

Er kan dus gesteld worden dat er slechts beperkt zicht is op zowel online daderschap als slachtofferschap onder jongeren. Dit komt doordat politie- en registratiedata slechts een deel van de werkelijkheid weerspiegelen. Hoewel de exacte omvang lastig te bepalen is, lijken de beschikbare cijfers er wel op te wijzen dat online daderschap en slachtofferschap onder jongeren reëel aanwezig zijn.

2.3.2 Affordances: faciliterende kenmerken van het internet

Jongeren zijn een relatief oververtegenwoordigde groep binnen het online daderschap in Nederland (Openbaar Ministerie, 2024). Verschillende onderzoekers stellen dan ook dat de digitale omgeving een gelegenheidscontext is voor jongeren om criminele activiteiten te plegen (Weijers, 2024; Weulen Kranenbarg et al., 2022). Deze visie wordt gedeeld door Goldsmith & Wall (2019), die stellen dat het internet bepaalde kenmerken bezit die het voor jongeren verleidelijk kunnen maken om online grenzen op te zoeken en risicovol of strafbaar gedrag te ondernemen. Goldsmith & Wall (2019) noemden dit *affordances*, wat verwijst naar “de mogelijkheid (en uitnodiging) tot een bepaalde handeling” (Van der Wagen et al., 2020, p. 278). Het affordance-perspectief richt zich niet alleen op de gebruiker, maar op de wederzijdse beïnvloeding tussen de mens en de technologie (Goldsmith & Wall, 2019; Van der Wagen et al., 2020).

Goldsmith & Wall (2019) stellen dat het internet verschillende *affordances* heeft die relevant zijn voor jongeren. Ten eerste is er sprake van *accessibility* (toegankelijkheid), waarmee wordt bedoeld dat het voor de jongere makkelijk en snel is om de technologie te gebruiken en dat diegene zich bekwaam voelt in het gebruik ervan (Goldsmith & Wall, 2019, p. 104). Online omgevingen zoals games zijn “virtual convergence settings” waar jongeren op een laagdrempelige manier in aanraking kunnen komen met criminaliteit (Goldsmith & Wall, 2019, p. 104). Bovendien kan het gevoel van bekwaamheid ervoor zorgen dat de bereidheid van een jongere tot het ondernemen van risicovol gedrag toeneemt (Goldsmith & Wall, 2019, p. 104).

Ten tweede zorgt *anonymity* (anonimiteit) ervoor dat een jongere zich online veiliger voelt dan offline om bepaald gedrag te vertonen, zichzelf te uiten en om grenzen op te zoeken (Goldsmith & Wall, 2019, p. 104-105). Volgens Suler (2004) speelt anonimiteit ook een rol bij online disinhibitie, waarbij online een ontremming van gedrag kan optreden. Ten derde verwijst *affordability* (betaalbaarheid) naar de lage kosten van internetgebruik wat volgens Goldsmith & Wall (2019, p. 105) onzorgvuldig gebruik ervan kan aanmoedigen.

Ten vierde verwijst *abundance* (overvloed) naar de onafgebroken aanvoer van online content (Goldsmith & Wall, 2019, p. 105). Ten vijfde duidt *ambivalence* (ambivalentie) op gedragsnormen die online losser zijn dan offline, wat volgens Goldsmith & Wall (2019, p. 105-106) kan leiden tot minder zelfregulatie waardoor een jongere eerder normoverschrijdend gedrag kan vertonen. Daarnaast duidt ambivalentie op een “sense of unreality”, waardoor de risico’s van ongewenst gedrag lager worden ingeschat (Goldsmith & Wall, 2019, p. 106).

Ten zesde wordt met *arousal* (opwinding) bedoeld dat jongeren online worden blootgesteld aan nieuwe en spannende content, wat volgens Goldsmith & Wall (2019, p. 106) de opstap kan vormen naar risicogedrag. Tot slot verwijst *asymmetry* (asymmetrie) ernaar dat jongeren nooit volledige controle hebben over de manier waarop technologie werkt (Goldsmith & Wall, 2019, p. 106). Hierdoor kunnen zij onbedoeld in aanraking komen met content waar ze oorspronkelijk niet naar op zoek waren, bijvoorbeeld door de werking van bepaalde algoritmes (Goldsmith & Wall, 2019, p. 106). Gezamenlijk laten deze *affordances* zien dat de digitale omgeving specifieke omstandigheden creëert die online delictgedrag aantrekkelijker, toegankelijker of eenvoudiger kunnen maken voor jongeren.

De toepassing van het affordance-perspectief op online criminaliteit staat nog in de kinderschoenen. Echter zijn er enkele voorbeelden waarbij de theorie is toegepast binnen cybercriminologisch onderzoek. Zo is uit onderzoek door Chan et al. (2019) gebleken dat de toegankelijkheid van social mediaplatformen invloed heeft op de mate waarin zij gebruikt worden voor cyberpesten. Uit onderzoek door Tanrikulu (2019) is gebleken dat deze toegankelijkheid ook voor jongeren een reden kan zijn om digitale platformen te gebruiken voor het cyberpesten van leeftijdsgenoten. Ook speelt anonimiteit een rol bij cyberpesten onder jongeren, omdat daders zich online onzichtbaarder en zekerder voelen om een ander te pesten (Tanrikulu, 2019). Binnen het huidige onderzoek zou het affordance-perspectief een raamwerk kunnen bieden voor de redenen waarom jongeren GenAI mogelijk inzetten voor risicovol of strafbaar gedrag.

2.4 Gebruik van GenAI door jongeren

De invloed van GenAI op criminaliteit en de aanwezigheid van jongeren in een digitale risicocontext roept vragen op over de invloed van GenAI op hun mogelijke betrokkenheid bij risicovol of strafbaar gedrag. Echter is de wetenschappelijke kennis over de manier waarop deze twee ontwikkelingen mogelijk samenkomen nog beperkt. Daarom zal in de volgende sectie eerst worden ingegaan op de manieren waarop jongeren GenAI gebruiken. Vervolgens zullen de risico's van het gebruik en misbruik van GenAI door jongeren worden besproken.

2.4.1 GenAI als hulpmiddel

Jongeren maken als vroege gebruikers van nieuwe technologie in toenemende mate gebruik van GenAI, zoals ChatGPT (Liu & Wang, 2025). Onderzoek naar dit gebruik richt zich veelal op de toepassing ervan binnen de schoolcontext. Uit een omvangrijk onderzoek door Abdaljaleel et al. (2024) onder studenten uit verschillende landen is gebleken dat de meerderheid van hen taalmodellen zoals ChatGPT gebruikt ter ondersteuning bij het leren. Zo gebruiken jongeren GenAI om vragen te stellen over opdrachten en om feedback te krijgen op hun eigen werk (Dikilitas et al., 2024; Sah et al., 2025). Zij gebruiken hiervoor GenAI, omdat ze het ervaren als toegankelijk en snel in gebruik (Abdaljaleel et al., 2024; Brandtzaeg et al., 2025). Daardoor gaven jongeren binnen onderzoek door Dikilitas et al. (2024) de voorkeur voor feedback door GenAI boven feedback door een docent. Bovendien gaven jongeren aan zich gemotiveerder en productiever voelen wanneer zij aan een schoolopdracht werken met behulp van GenAI (Dikilitas et al., 2024). Die ervaren efficiëntie zorgt ervoor dat vooral jongeren die veel tijdsdruk ervaren tijdens hun studie positief zijn over het gebruik van GenAI (Sah et al., 2025). Deze ervaren efficiëntie van GenAI door jongeren sluit aan bij de *affordance accessibility* van Goldsmith & Wall (2019, p. 104).

2.4.2 GenAI als steun- en inspiratiebron

Daarnaast gebruiken jongeren GenAI voor emotionele steun. Hierbij kan worden gedacht aan het vragen om steun of advies bij mentale problemen of zorgen (Brandtzaeg et al., 2025; Chaudhuri, 2024). Daarbij verkiezen sommige jongeren steun door GenAI boven steun door een mens (Brandtzaeg et al., 2025; Chaudhuri, 2024). Zo gaven jongeren binnen onderzoek door Brandtzaeg et al. (2025) aan sneller geneigd te zijn om zich kwetsbaar op te stellen tijdens een gesprek met een AI-taalmodel, doordat zij zich anoniemer voelen. Door die anonimiteit ervaren zij minder oordeel in vergelijking met interacties met mensen (Brandtzaeg et al., 2025; Chaudhuri, 2024). Volgens Goldsmith & Wall (2019) kan die anonimiteit ertoe leiden dat jongeren in online omgevingen eerder de grens over gaan. Een andere toepassing van GenAI door jongeren is voor vermaak of als inspiratiebron (Brandtzaeg et al., 2025; Chaudhuri, 2024). Zo gaven enkele jongeren binnen onderzoek door Brandtzaeg et al. (2025,

p. 9-10) aan dat zij GenAI gebruiken om grappige verhalen over vrienden te schrijven. Ook worden AI-chatbots door jongeren ingezet als raadgever voor alledaagse kwesties of dilemma's (Brandtzaeg et al., 2025; Chaudhuri, 2024).

2.4.3 Risico's van GenAI-gebruik voor jongeren

Deze toename in gebruik van GenAI door jongeren roept echter ook vragen op over de mogelijke negatieve consequenties ervan. Yu et al. (2025) hebben een grootschalig onderzoek verricht naar de risico's van het gebruik van GenAI door jongeren. Daarvoor bestudeerden zij Reddit[3]-discussies tussen jongeren, de AI Incident database[4] en AI-chatlogs van jongeren. Op basis hiervan brachten zij de meest voorkomende risico's van AI-gebruik door jongeren in kaart.

Ten eerste kan het gebruik van GenAI een negatieve impact kan hebben op "de psychologische, emotionele en cognitieve gezondheid van jongeren" (Yu et al., 2025, p. 5). Doordat GenAI is ontworpen om het gesprek met de gebruiker zo lang mogelijk te laten duren, zijn taalmodellen erop getraind om bevestigend te reageren. Hierdoor ontstaat het risico dat schadelijke gedachten of gedrag onbedoeld wordt bevestigd door een AI-taalmodel (Yu et al., 2025).

Daarnaast kan deze constante bevestiging leiden tot overmatige afhankelijkheid van GenAI (Brandtzaeg et al., 2025; Chaudhuri, 2024; Yu et al., 2025). Dit risico is bijzonder relevant met het oog op jongeren. Zij zijn gevoelig voor onmiddellijke behoeftebevrediging doordat de beloningscentra van hun hersenen nog volop in ontwikkeling zijn (Romer et al., 2010). Deze bevestiging kan GenAI bieden door direct en bevestigend te antwoorden (Yu et al., 2025). Daarnaast kan overmatige afhankelijkheid van GenAI een negatieve impact hebben op het kritisch denkvermogen en de zelfredzaamheid van jongeren (Brandtzaeg et al., 2025; Gerlich, 2025; Yu et al., 2025). Zo is uit onderzoek door Gerlich (2025) gebleken dat overmatig AI-gebruik samenhangt met een afname van het kritisch denkvermogen van jongeren.

Een ander risico is de verstoring van de sociale en morele ontwikkeling van jongeren (Yu et al. (2025, p. 7). Zij bevinden zich in een leeftijdsfase waarin deze sociale en morele normen in interactie met anderen worden gevormd, en waarin het verkennen van grenzen centraal staat (Yu et al., 2025). Echter kunnen zich problemen voordoen wanneer jongeren veel interacteren met GenAI. Dat komt doordat GenAI uit kunstmatige intelligentie bestaat en geen sociaal bewustzijn heeft. Daardoor is het niet om staat om op een genuanceerde manier te reageren, met "wederzijds respect, empathie en grenzen" zoals bij menselijk contact wel het

[3] Reddit is een internetforum waar gebruikers discussies kunnen starten en op elkaar kunnen reageren.
[4] De AI Incident database is een openbaar toegankelijke database die is opgezet om de schade die is ontstaan door de inzet van AI in kaart te brengen.

geval is (Yu et al., 2025, p. 7). Dit kan ertoe leiden dat grensoverschrijdend of schadelijk gedrag van jongeren wordt bekrachtigd door een taalmodel (Yu et al., 2025).

Tot slot kunnen jongeren in aanraking komen met aanstootgevende content door het gebruik van GenAI (Yu et al., 2025). Dit betreft voornamelijk seksuele en geweldsgerelateerde content (Yu et al. (2025). Door het ontbreken van kritisch tegengeluid wordt dergelijke content tijdens de interactie met een taalmodel sneller genormaliseerd (Yu et al., 2025). Bovendien zijn er enkele gevallen bekend waarbij een AI-chatbot jongeren aanzette tot zelfbeschadiging, zonder dat de vragen van de jongeren hiertoe aanleiding gaven (Yu et al., 2025). Dit is een voorbeeld van de *affordance asymmetry* van Goldsmith & Wall (2019, p. 106), waarbij technologie op een onverwachte en ongewenste manier kan werken zonder dat de gebruiker dit heeft geïnitieerd.

2.4.4 Misbruik van GenAI door jongeren

Naast de risico's die bestaan voor jongeren, zijn er ook verschillende manieren waarop zij GenAI kunnen misbruiken. Yu et al. (2025, p. 9-10) verwijzen hiernaar als het "misuse and exploitation risk", waarbij zij onderscheid maken tussen onbedoeld en opzettelijk misbruik van GenAI. Onbedoeld misbruik kan optreden wanneer gebruikers blindelings vertrouwen op de output van GenAI, zonder stil te staan bij mogelijke onnauwkeurigheden (Yu et al., 2025, p. 9). Volgens hen komt dit risico voort uit de neiging van GenAI om informatie en adviezen te hallucineren (Yu et al., 2025, p. 9-10). Hallucineren houdt in dat GenAI output geeft die feitelijk onjuist is (Ciubotaru, 2025). Dit kan het vertrouwen van jongeren in een taalmodel schaden (Brandtzaeg et al., 2025; Ciubotaru, 2025), en kan leiden tot het onbedoeld verspreiden van desinformatie (Ciubotaru, 2025). Hoewel het verspreiden van desinformatie op zichzelf niet strafbaar is, kan het een risico vormen wanneer de misleidende informatie wordt ingezet om gedrag te sturen of schadelijk gedrag uit te lokken (Blauth et al., 2022; Park & Nan, 2025). In zulke gevallen kan desinformatie leiden tot strafbare feiten zoals opruiing, bedreiging of het beïnvloeden van verkiezingen.

Opzettelijk misbruik wordt door Yu et al. (2025, p. 10) aangeduid als "deliberate actions by users who manipulate GAI to create, disseminate, or facilitate harmful behaviors". Deze categorie omvat veel verschillende gedragingen. Zo kwamen Yu et al. (2025, p. 10) voorbeelden tegen waarbij jongeren GenAI gebruikten om gewelddadige verhalen te genereren. Ook zagen zij voorbeelden van chatbots om extremistisch gedachtegoed te verspreiden. Dit kan ertoe leiden dat jongeren in een tunnelvisie van hun eigen gedachtengoed terechtkomen. Binnen de literatuur wordt dit aangeduid als een echokamer. Dit houdt in dat iemand zich in een omgeving bevindt waarin diens mening of overtuigingen continu worden bevestigd zonder kritisch tegengeluid (Hartmann et al., 2025, p. 2). Hoewel het hierbij

doorgaans gaat om bevestiging door andere personen, is het voorstelbaar dat een AI-chatbot een vergelijkbare functie kan vervullen en daarmee ook als echokamer kan fungeren.

Daarnaast verwijst “GAI-facilitated interpersonal harm” naar situaties waarin jongeren GenAI misbruiken om anderen gericht schade toe te brengen (Yu et al., 2025, p. 11). Voorbeelden hiervan zijn het gebruik van GenAI voor phishing (Schmitt & Flechais, 2023), cyberstalking (Yu et al., 2025) en het maken van deepnudes (Kopecký & Voráč, 2025; Szyf et al., 2023). Vooral het maken van deepnudes door jongeren heeft de afgelopen jaren in toenemende mate aandacht gekregen binnen de media.

Hoewel wetenschappelijk onderzoek naar het onderling creëren van deepnudes door jongeren tot op heden beperkt is, wijzen de beschikbare onderzoeken er wel op dat dit plaatsvindt. Zo is uit onderzoek door Kopecký & Voráč (2025) onder jongeren tussen de 12 en 17 jaar gebleken dat zij onderling deepnudes van elkaar maken. Daarnaast heeft onderzoek onder Belgische jongeren tussen de 15 en 25 jaar uitgewezen dat 12,8 procent van hen bekend was met deepnude-apps, en dat 60,5 procent van hen die apps wel eens heeft gebruikt (Szyf, et al., 2023, p. 11). Uit beide onderzoeken is naar voren gekomen dat jongens vaker deepnudes maken dan meisjes (Kopecký & Voráč, 2025, p. 551; Szyf et al., 2023, p. 62). Ook zijn de platforms vaker gericht op het uitkleden van vrouwen dan mannen (Szyf et al., 2023, p. 26). Verder is regulatie van deze platformen zeer beperkt. Szyf et al. (2023) analyseerden de 25 meest populaire deepnude-apps en concludeerden dat deze apps misbruik in theorie verbieden, maar dat er in de praktijk weinig gebeurt om dit te voorkomen. Bovendien blijken de apps het gebruik in veel gevallen juist aan te moedigen (Szyf et al., 2023).

2.5 Synthese

Concluderend laat de literatuur zien dat GenAI een breed inzetbare technologie is die uiteenlopende mogelijkheden biedt voor leren, creativiteit en persoonlijke ondersteuning, maar ook kan worden ingezet voor risicovol of strafbaar gedrag. Deze ontwikkeling is bijzonder relevant in relatie tot jongeren, aangezien zij opgroeien in een tijd waarin GenAI een steeds prominentere rol speelt in hun leefwereld. Binnen dit kader biedt het affordance-perspectief inzicht in de manier waarop technologie handelingsmogelijkheden creëert die online delictgedrag voor jongeren aantrekkelijker, toegankelijker of eenvoudiger kunnen maken. Kenmerken zoals *accessibility* en *anonymity* zijn daarbij ook zichtbaar in het gebruik van GenAI door jongeren. Daarnaast wijzen eerdere studies naar GenAI-gebruik door jongeren op verschillende risico’s die kunnen bijdragen aan risicovol of strafbaar gedrag.

3. Methodologie

Om de onderzoeksvraag te beantwoorden, is in deze studie empirisch onderzocht wat de opkomst van generatieve AI (GenAI) mogelijk betekent voor de betrokkenheid van jongeren bij risicovol of strafbaar gedrag. Dit hoofdstuk zal een toelichting bieden van de wijze waarop het onderzoek is uitgevoerd.

3.1 Onderzoeksdesign

Ter beantwoording van de onderzoeksvraag is gekozen voor een kwalitatief, verkennend onderzoek. Het doel hiervan is om te verkennen wat de opkomst van GenAI mogelijk betekent voor de betrokkenheid van jongeren bij risicovol of strafbaar gedrag. Door de beperkte beschikbare kennis over het onderwerp is gekozen voor een onderzoeksopzet die bestond uit twee fases. Naar voorbeeld van Spradley (1980, p. 77-78) vond eerst een verkennende "'Grand Tour" van het onderwerp plaats. In deze eerste fase zijn oriënterende gesprekken gevoerd met personen met expertise op het gebied van digitalisering, criminaliteit en het online gedrag van jongeren. Het doel van deze fase was het verkennen en afbakenen van relevante thema's die in de tweede fase verder uitgediept konden worden. Vervolgens zijn in de tweede, verdiepende fase focusgroepen met jongeren gehouden. Deze fases zullen nader worden toegelicht in dit hoofdstuk.

De kwalitatieve dataverzameling verliep via een cyclisch iteratief proces. Dat houdt in dat nieuwe inzichten die ontstonden tijdens de dataverzameling werden gebruikt om de daaropvolgende stappen aan te scherpen (Decorte & Zaitch, 2016, p. 474; Hennink et al., 2020). Door deze terugkerende afstemming tussen de verzamelde data en de vervolgstappen kon gaandeweg verdieping worden gezocht (Hennink et al., 2020). Daarnaast was het onderzoek inductief van aard. Dit houdt in dat de inzichten voortkwamen uit de verzamelde data en dat het perspectief van de respondenten centraal stond (Evers, 2015). Door de inzichten van de experts en de jongeren met elkaar te combineren, kunnen de resultaten vanuit meerdere perspectieven worden belicht.

3.2 Oriënterende fase

Tijdens de oriënterende fase zijn gesprekken gevoerd met informanten die deskundig zijn op minimaal één van de twee relevante domeinen: de impact digitalisering op criminaliteit of het online gedrag van jongeren. Deze informanten zijn geselecteerd vanwege hun professionele ervaring en inhoudelijke expertise op deze onderwerpen. Het doel van deze gesprekken was om kennis op te doen over het onderwerp, en de belangrijkste thema's en vraagstukken te identificeren die in de volgende fase verder uitgediept konden worden.

3.2.1 Zoekstrategie literatuur

In de beginfase van de Grand Tour is een verkennend literatuuronderzoek uitgevoerd om zicht te krijgen op het onderzoeksveld, welke kwesties opvallen en hoe er door onderzoekers over de onderwerpen wordt gesproken. Er is gezocht naar relevante wetenschappelijke artikelen via Google Scholar en Scopus. Hiervoor is gekozen omdat het gebruik van GenAI voor criminaliteit op het snijvlak ligt van technologiestudies en criminologie. Deze zoekmachines hebben het mogelijk gemaakt om een breed en interdisciplinair zoekbereik te realiseren. De zoektermen bestonden uit combinaties van de termen '(generative) artificial intelligence', 'crime' en verwante zoektermen. Daarnaast is gericht gezocht naar literatuur over jongeren met betrekking tot dit thema met termen gerelateerd aan '(generative) artificial intelligence' en 'adolescents'. Er is gezocht naar relevante studies in zowel het Engels als het Nederlands.

Naast wetenschappelijke zoekmachines is ook gezocht op het internet. Dit leverde grijze literatuur op, zoals inlichtingenrapporten van Europol en Anthropic en een onderzoeksrapport over deepnudes onder jongeren van het Instituut voor de Gelijkheid van Vrouwen en Mannen (IGVM).

Het beeld dat tijdens de beginfase van de Grand Tour naar voren is gekomen, heeft op meerdere manieren de literatuurstudie en verdere richting van het onderzoek beïnvloedt. Ten eerste is op basis hiervan de focus aangescherpt van AI naar GenAI, aangezien dit een thema was dat vaak terugkwam in relatie tot zorgen over jongeren. Daarnaast is naar voren gekomen dat binnen de wetenschap veel werd gesproken over het onderzoeksonderwerp in termen van dreigingen en zorgen, in plaats van concrete casussen. Daarom is in de literatuurstudie explicieter onderscheid gemaakt tussen empirisch en toekomstverkennend onderzoek.

3.2.2 Werving van informanten

Tijdens de oriënterende fase zijn informanten geworven met expertise op het gebied van AI-criminaliteit, digitalisering en criminaliteit of het online gedrag van jongeren. Zij zijn geworven via het internet of aangedragen vanuit het netwerk van Straatwaarden van het RIEC Midden-Nederland. Straatwaarden richt zich op het voorkomen en terugdringen van jonge aanwas in de georganiseerde criminaliteit. Haar netwerk bestaat uit verschillende partners met uiteenlopende expertise, variërend van de politie tot verschillende jongerenwerkorganisaties. Daarnaast is een oproep gedeeld in het netwerk Jonge Aanwas en het netwerk Digitaal Wijkagenten van de politie. De wervingsstrategie is opgenomen in bijlage A.

Bij de benadering werden personen geïnformeerd over het doel van het onderzoek en werd met hen besproken of hun expertise bij het onderzoek aansloot. Deze strategie heeft vijf

respondenten opgeleverd, waarmee een individueel interview is afgenomen. Zij beschikken over uiteenlopende expertise[5], waarvan tabel 1 een overzicht biedt.

Tabel 1

Kenmerken van de informanten

	Expertise
Informant 1	Strategisch Specialist Digitaal bij de politie met expertise op het gebied van GenAI.
Informant 2	Ondersteunt cybervaardige jongeren in hun talentontwikkeling bij stichting Cyberbrein en heeft zicht op het online gedrag van jongeren.
Informant 3	Operationeel Specialist Publiek Private Samenwerking Cybercrime bij de politie en houdt zich bezig met de impact van AI op criminaliteit.
Informant 4	Online jongerenwerker bij stichting Epic Youth en heeft zicht op het online gedrag van jongeren.
Informant 5	Senior onderzoeker cybercrime bij het Nederlands Studiecentrum voor Criminaliteit en Rechtshandhaving (NSCR) en doet onderzoek naar digitalisering en cybercriminaliteit, waar AI onderdeel van uitmaakt.

3.2.3 Interviews

De individuele interviews vonden plaats tussen eind februari en eind maart 2026. Er zijn vijf interviews afgenomen, waarvan twee online en drie fysiek plaatsvonden. Voorafgaand aan elk interview is het toestemmingsformulier ondertekend door de informant. Hierin werden onder meer de aard en het doel van het onderzoek toegelicht. Indien het interview online plaatsvond, is het toestemmingsformulier op voorhand naar de informant gestuurd en ondertekend geretourneerd. De opzet van dit formulier is opgenomen in bijlage C1.

De interviews hadden een semigestructureerde opzet, en zijn afgenomen aan de hand van een topiclijst. Deze opzet bood ruimte om op bepaalde punten dieper in te gaan, afhankelijk van de expertise van de informant. Dit betekent dat de hoofdonderwerpen overeenkwamen, maar dat bij enkele informanten dieper is ingegaan op de invloed van AI op criminaliteit en bij de anderen op AI-gebruik door jongeren. De topiclijst is opgenomen in bijlage D1.

[5] Enkel de expertise van de informanten die door de onderzoeker als het meest relevant wordt beschouwd voor dit onderzoek, wordt weergegeven in tabel 1.

3.2.4 Aanvullende informatieverzameling en het logboek

Tijdens deze oriënterende fase heeft de onderzoeker, naast de literatuurstudie en de interviews, op verschillende manieren aanvullende informatie verzameld. Deze aanvullende informatiebronnen waren behulpzaam bij het verder afbakenen van het onderzoeksonderwerp. Dit proces is bijgehouden in een logboek (bijlage B). Zo werd een intern intelligencebeeld verkregen van de Intelcel Cyber Midden-Nederland over de invloed van AI op criminaliteit. Daarnaast werd een overleg bijgewoond tussen het Europol Innovation Lab en het team Cybercrime van de politie over AI en criminaliteit. Het Europol Innovation Lab onderzoekt onder andere hoe nieuwe en opkomende technologieën, waaronder kunstmatige intelligentie, kunnen zorgen voor nieuwe dreigingen voor de internationale veiligheid. Tijdens dit overleg werd de onderzoeker in de gelegenheid gesteld om enkele vragen te stellen over de mate waarin er aandacht is voor AI-criminaliteit onder jongeren bij Europol. Verder werd een kennisbijeenkomst bijgewoond over AI in het veiligheidsdomein, waar Offlimits[6] vertelde over de impact van GenAI op het maken van deepnudes van en door jongeren. De onderzoeker heeft aantekeningen gemaakt van deze aanvullende informatie in de vorm van veldnotities, om deze informatie mee te kunnen nemen in de analyse (Evers, 2015, p. 19). Per activiteit is een korte samenvatting van de veldnotities opgenomen in het logboek (bijlage B).

3.3 Verdiepende fase

Tijdens de verdiepende fase zijn focusgroepen gehouden met jongeren om te onderzoeken hoe zij denken over de thema's die in de oriënterende fase naar voren kwamen. Deze fase was gericht op het verdiepen van de inzichten uit de oriënterende fase door het verkrijgen van een rijker beeld van het onderwerp vanuit het perspectief van de doelgroep, namelijk de jongeren zelf. Daarnaast heeft dit het mogelijk gemaakt om inzichten van de informanten te verbinden aan de ervaringen en meningen van de deelnemende jongeren.

3.3.1 Werving van jongeren

De werving van de jongeren verliep op meerdere manieren. Ten eerste is via verschillende jongerenwerkers geprobeerd om jongeren te werven voor het onderzoek. Zij zijn benaderd omdat zij direct in contact staan met jongeren, en mogelijk zicht hebben op jongeren die veel online zijn. Ook hebben zij doorgaans een vertrouwensband met de jongeren waarmee zij in contact staan, waardoor zij in een positie zijn om hen op een laagdrempelige manier te benaderen voor deelname. Echter leverde dit geen respondenten op waardoor de wervingsstrategie werd aangepast. Vervolgens werd Halt benaderd met informatie over het onderzoek, en de vraag of via hen jongeren konden worden benaderd. Vanuit Halt is een

[6] Offlimits is een expertisecentrum online misbruik waar melding kan worden gemaakt van misbruikbeelden en waar mensen terecht kunnen voor hulp als zij zelf slachtoffer zijn geworden van online misbruik.

groepje jongeren aangedragen die deel wilden nemen aan het onderzoek. Dit leverde drie respondenten op, waarmee de eerste focusgroep is gehouden. Vervolgens werd een mbo-school benaderd. Hierbij is de vraag voor deelname aan het onderzoek uitgezet onder leerlingen van twee ICT-opleidingen. Dit resulteerde in tien respondenten, waarmee de tweede en derde focusgroep zijn afgenomen.

De inclusiecriteria voor de jongeren waren dat zij tussen de 12 en 23 jaar oud zijn, woonachtig in Nederland en wel eens gebruik maken of hebben gemaakt van GenAI. Hun deelname aan dit onderzoek was volledig vrijwillig. De enige vereiste voor hun deelname was het ondertekenen van een toestemmingsformulier (bijlage C2). Een aanvullende eis voor deelname van jongeren onder de 16 jaar was dat ook hun ouder of voogd het toestemmingsformulier ondertekende (bijlage C3). In totaal hebben dertien jongeren deelgenomen aan het onderzoek, variërend in leeftijd van 14 tot 23 jaar. De onderzoeksgroep bestond uit zowel vrouwen als mannen. Zij waren afkomstig van verschillende opleidingen waaronder de middelbare school, de universiteit en diverse mbo-ICT-opleidingen. De kenmerken van de respondenten worden weergegeven in bijlage E.

3.3.2 Focusgroepen

Tijdens de verdiepende fase werden focusgroepen met jongeren gehouden over GenAI. Er is gekozen voor het organiseren van focusgroepen omdat het de onderzoeker in staat stelde rijke informatie te verzamelen over het onderzoeksonderwerp waar nog relatief weinig over bekend is (Hennink et al., 2020). Tijdens een focusgroep konden jongeren in een groepssetting worden bevraagd over verschillende thema's die relevant zijn voor het beantwoorden van de onderzoeksvragen (Hennink et al., 2020). Hierdoor kon ook worden verkend hoe hun meningen zich tot elkaar verhouden.

De focusgroepen vonden zowel online als fysiek plaats. De fysiek gehouden focusgroepen vonden plaats op de school van de respondenten. Daarbij was ook een secondant van de onderzoeker aanwezig om non-verbale signalen van de respondenten vast te leggen (Evers, 2015). Tevens waren bij twee focusgroepen ook een begeleider of docent aanwezig. Bijlage E geeft een overzicht van de verdeling van de respondenten en andere aanwezigen tijdens de focusgroepen.

De focusgroepen hadden een semigestructureerde opzet en werden afgenomen aan de hand van een topiclijst (bijlage D2). Deze opzet bestond uit drie onderdelen. Ten eerste is ingegaan op wat GenAI volgens de jongeren inhoudt. Aangezien dit een breed begrip is, kon hierdoor op voorhand helder worden gemaakt waar het onderzoek precies over ging. Vervolgens werd met de jongeren besproken hoe zij GenAI gebruiken, bijvoorbeeld welke tools en platforms zij gebruiken en waarom. Daarnaast werd in dit deel besproken hoe zij over het algemeen over GenAI denken. In het tweede deel is verkend wat de jongeren wel en niet

toelaatbaar vinden bij het gebruik van GenAI. Ook werd in dit deel besproken welke risico's het gebruik van GenAI volgens hen kan hebben voor jongeren, en in hoeverre zij deze risico's zelf tegenkomen. Vervolgens is in het derde deel verkend hoe de jongeren denken over het gebruik van GenAI voor verschillende vormen van risicovol of strafbaar gedrag. Dit werd gedaan aan de hand van stellingen die telkens begonnen met "Het is oké om AI te gebruiken om…", zoals "Het is oké om AI te gebruiken om *deepnudes* van iemand anders te maken".

Bij de afname van de focusgroepen is een flexibele aanpak gehanteerd. Dit houdt in dat de onderzoeker ruimte heeft gelaten om in te spelen op de inbreng van de jongeren, en om onderwerpen verder te verkennen wanneer deze spontaan naar voren kwamen. Deze aanpak sluit aan bij het verkennende karakter van het onderzoek, aangezien op voorhand niet geheel duidelijk was in hoeverre de jongeren al met de besproken thema's bezig waren. De topiclijst fungeerde daarom als leidraad voor de gesprekken, maar er is vooral doorgevraagd op thema's die de jongeren zelf aandroegen. Hiermee is ook rekening gehouden bij de opbouw van de topiclijst, door eerst te vragen naar risico's die jongeren zelf ervaren voordat de stellingen over criminele handelingen met GenAI als hulpmiddel aan bod kwamen.

De stellingen zijn bewust abstract geformuleerd. Dit is gedaan om te voorkomen dat jongeren onbedoeld werden geïnformeerd over malafide gebruik van GenAI. Het doel van de stellingen was namelijk om morele oordelen uit te lokken, en niet om jongeren technische instructies te geven over het gebruik van GenAI voor diverse strafbare handelingen. Daarnaast is benadrukt dat jongeren niet verplicht waren om persoonlijke ervaringen te delen. Dit is gedaan om een veilige en comfortabele sfeer te creëren, en de jongeren het gevoel te geven dat zij vrijuit konden spreken zonder druk om gevoelige of belastende informatie te delen in de groep en met de onderzoeker.

3.4 Data-analyse

De verworven data zijn geanalyseerd volgens een aantal stappen. Ten eerste zijn de audio-opnames van zowel de interviews als de focusgroepen verbatim getranscribeerd. Hierbij zijn de antwoorden van de respondenten letterlijk uitgewerkt (Decorte & Zaitch, 2016). Deze manier van transcriberen sluit aan bij het emic perspectief, waarin de mening en ervaringen van de respondenten centraal staan (Hennink et al., 2020, p. 208). Dit heeft de onderzoeker in staat gesteld zo dicht mogelijk bij de standpunten van de respondenten te blijven. Ook heeft de onderzoeker hiermee beoogd om de kans op misinterpretatie van de data te verkleinen (Evers, 2015; Hennink et al., 2020). Om die nauwkeurigheid verder te bevorderen, zijn bij de focusgroepen ook non-verbale signalen opgenomen in het transcript. Deze signalen hebben bijgedragen aan een vollediger begrip van de antwoorden. Bovendien bieden zij waardevolle informatie over de sfeer van de focusgroepen, die van invloed kan zijn geweest op de mate waarin jongeren bereid waren om hun mening en ervaringen te delen.

De transcriptie heeft telkens direct plaatsgevonden na de afname van het interview of de focusgroep. Hierdoor konden nieuwe, relevante en onverwachte thema's snel worden opgemerkt en meegenomen in de verdere dataverzameling, waardoor latere gesprekken meer diepgang kregen (Hennink et al., 2020). Zo gaven de jongeren tijdens de eerste focusgroep aan dat zij zelf wel eens een vraag hebben gesteld waarop ChatGPT weigerde antwoord te geven. Dit leverde waardevolle anekdotes op, waardoor de onderzoeker deze vraag heeft toegevoegd aan de topiclijst voor de daaropvolgende focusgroepen. De transcripten van de focusgroepen zijn na de uitwerking geanonimiseerd, door namen en andere herleidbare details weg te halen.

De volgende stap in het proces bestond uit het analyseren van de verzamelde data. Hiervoor is het kwalitatieve analyseprogramma ATLAS.ti 26 gebruikt. Met behulp van dit programma is de onderzoeker op zoek gegaan naar thema's in de data. Dit werd gedaan in meerdere stappen. Ten eerste zijn de transcripten, veldnotities en secundaire databronnen opnieuw doorgelezen om een helder overzicht te krijgen van de inhoud. Daarna werd elk document gecodeerd door middel van de analysetechniek open coderen. Dit houdt in dat tekstfragmenten zijn voorzien van een open code die het fragment in enkele woorden beschrijft (Decorte & Zaitch, 2016). De codenaam is zo dicht mogelijk aangesloten op de tekst, om de nabijheid tot de data te behouden en de kans op misinterpretatie door de onderzoeker te beperken (Decorte & Zaitch, 2016; Hennink et al., 2020). Dit sluit aan bij de inductieve benadering, waarbij de codes voortkomen uit de bijdragen van de respondenten (Hennink et al., 2020). Het open coderen vormde de eerste stap van datareductie (Decorte & Zaitch, 2016).

Vervolgens is axiaal gecodeerd. Hierbij zijn de open codes opnieuw bekeken en waar nodig hernoemd. Ook zijn codes met een overeenkomstige betekenis samengevoegd tot overkoepelende codes. Deze analysestap heeft ervoor gezorgd dat overzicht ontstond in de data (Decorte & Zaitch, 2016). Op basis van de axiale codes zijn de onderzoeksvragen verder afgestemd op de informatie die uit de dataverzameling naar voren is gekomen. Tot slot is in de derde fase van het analyseproces selectief gecodeerd. Hierbij is bestudeerd hoe de hoofdcategorieën in relatie staan tot elkaar (Decorte & Zaitch, 2016). Dit is gedaan door per deelvraag een overzicht (*network view*) te maken in ATLAS.ti van de axiale codes die behulpzaam zijn voor het beantwoorden ervan. Deze netwerken hebben de leidraad gevormd voor het schrijven van de resultatenparagraaf. Een overzicht hiervan is opgenomen als codeboom onder bijlage F.

Naast het coderen en organiseren van de codes, zijn tijdens het analyseproces memo's aangemaakt in ATLAS.ti. Er zijn zowel empirische als methodologische memo's bijgehouden. Het bijhouden van memo's heeft ervoor gezorgd dat waardevolle inzichten die zijn ontstaan tijdens het analyseproces tussentijds konden worden vastgelegd, zoals ideeën voor het axiaal coderen en mogelijk bruikbare citaten (Decorte & Zaitch, 2016; Hennink et al., 2020).

Daarnaast hebben methodologische memo's ruimte geboden om te reflecteren op de keuzes die tijdens het analyseproces werden gemaakt, en de mogelijke impact van die keuzes op de kwaliteit van de analyse (Decorte & Zaitch, 2016; Hennink et al., 2020).

3.5 Ethiek

Tijdens het gehele onderzoeksproces is rekening gehouden met het ethiekreglement van de Faculteit der Rechtsgeleerdheid van de Vrije Universiteit Amsterdam. In dit reglement zijn de ethische richtlijnen van de uitvoering van criminologisch onderzoek vastgelegd. Om deze richtlijnen te volgen zijn verschillende maatregelen genomen tijdens de voorbereiding en uitvoering van het onderzoek.

3.5.1 Informed consent

Een belangrijk onderdeel van het uitvoeren van ethisch verantwoord onderzoek is het zorgen voor informed consent van de deelnemers. Dit bestond uit meerdere onderdelen. Ten eerste hebben alle respondenten voorafgaand aan hun deelname het toestemmingsformulier doorgelezen en ondertekend. Met dit formulier werden zij geïnformeerd over het doel en de aard van het onderzoek, wat hun deelname aan het onderzoek inhield en eventuele risico's die daaraan verbonden waren. Bij de jongeren werd het toestemmingsformulier (bijlage C2) telkens aan het begin van de focusgroepen besproken, om te verzekeren dat zij begrepen wat hun deelname inhield. Ook is ervoor gekozen om de risico's van hun deelname mondeling toe te lichten aan het begin van de focusgroep. Bij deelnemers die jonger waren dan 16 jaar heeft zowel de jongere als hun ouder of voogd het formulier ondertekend (bijlage C3).

3.5.2 Vertrouwelijkheid en datamanagement

Tijdens het onderzoek is bewust omgegaan met het beschermen van persoonsgegevens van de respondenten. Volledige anonimiteit kon niet worden gegarandeerd aan de jongeren, aangezien de gesprekken plaatsvonden in een groepssetting en er een audio-opname is gemaakt (Swanborn, 2015). In plaats daarvan is gezorgd voor vertrouwelijkheid door de transcripten van de focusgroepen te anonimiseren en daarbij herleidbare details weg te laten (Swanborn, 2015). Daarnaast zijn de audio-opnames opgeslagen op een beveiligde server van het RIEC, waar alleen de onderzoeker toegang tot had. Na transcriptie zijn deze audio-opnames op zorgvuldige wijze verwijderd. Verder is in overleg met de informanten besloten hun identiteit niet geheel te anonimiseren, omdat hun functie geloofwaardigheid toevoegt aan hun uitspraken als informant (Swanborn, 2015).

3.5.3. Ethische waarborgen voor jongeren

Binnen sociaalwetenschappelijk onderzoek met jongeren is het belangrijk om oog te hebben voor hun kwetsbare positie. Hierbij moet een balans worden gevonden tussen het meenemen van hun stem in het onderzoek en het beschermen van hun welzijn. Hiervoor zijn enkele maatregelen genomen. Ten eerste zijn de vrijwilligheid en vertrouwelijkheid van hun deelname op meerdere momenten benadrukt door de onderzoeker, zowel voorafgaand als tijdens de focusgroepen. Daarnaast zijn tijdens de focusgroepen enkele strafbare handelingen besproken die werden uitgevoerd met GenAI, zoals het maken van *deepnudes*. Hierbij bestond het risico dat de jongeren hier zelf slachtoffer of dader van zijn geweest. Daarom is voorafgaand aan de focusgroepen benadrukt dat de keuze over wat de jongeren hierover wilden delen te alle tijden bij henzelf lag. Ook is aan het eind van elke focusgroep tijd ingebouwd voor een passende afsluiting. Daarbij werd ten eerste benadrukt dat het gebruik van GenAI voor strafbaar gedrag niet afdoet aan de strafbaarheid daarvan. Vervolgens is in het kader van nazorg een korte uitleg gegeven over het meldpunt en de hulplijn van Offlimits. Ook hebben de jongeren de mogelijkheid gekregen om na te praten over de onderwerpen die aan bod kwamen, als zij daar behoefte aan hadden.

3.6 Validiteit en betrouwbaarheid

Gedurende het onderzoek zijn verschillende maatregelen genomen om de validiteit en betrouwbaarheid te versterken. Ten eerste is aan het begin van ieder interview en elke focusgroep met de respondent(en) besproken wat zij onder GenAI verstaan. GenAI is namelijk een breed begrip, waardoor de definities die de respondenten ervan hanteren uiteen kunnen lopen. Door dit aan het begin te bespreken, kon worden nagegaan of hun definitie ervan overeenkwam met dat van de onderzoeker. Tijdens de eerste focusgroep heeft de onderzoeker het begrip kort toegelicht aan de jongeren, omdat er wat verwarring over bestond. Hiermee is beoogd de constructvaliditeit van het onderzoek te versterken (Fischer & Julsing, 2023, p. 86).

Daarnaast is bij de werving van informanten vooraf met hen afgestemd of hun expertise aansloot bij het onderzoek. Hiermee is geprobeerd om op voorhand vast te stellen dat de zij beschikten over kennis die relevant was voor het beantwoorden van de onderzoeksvragen. Deze afstemming had als doel de interne validiteit van het onderzoek te bevorderen, oftewel dat met de interviews werd onderzocht wat de onderzoeker beoogde te onderzoeken (Matthijsse, 2023, p. 97).

Verder zijn binnen dit onderzoek de inzichten van informanten en jongeren met elkaar gecombineerd. Hierdoor konden de bevindingen vanuit verschillende perspectieven worden belicht. Het combineren van bronnen is een manier van datatriangulatie, waarmee is getracht te voorkomen dat conclusies afhankelijk waren van slechts één bron of interpretatie

(Matthijsse, 2023, p. 96). Daarmee is beoogd om de interne betrouwbaarheid van het onderzoek te versterken (Matthijsse, 2023, p. 96).

Tot slot is tijdens het onderzoek een logboek van zowel de wervingsstrategie van de respondenten (bijlage A) als de aanvullende informatieverzameling (bijlage B) bijgehouden, zoals werd besproken in sectie 3.2.2 en 3.2.4. Deze overzichten vormen een *audit trail* van het onderzoeksproces; een gedetailleerd spoor van beslissingen, contacten en opgeleverde informatie (Swanborn, 2015). Dit heeft gezorgd voor transparantie over het onderzoeksproces, door de navolgbaarheid van de gemaakte keuzes te vergroten (Swanborn, 2015). Ook is een overzicht gemaakt van de codegroepen (bijlage F), die de leidraad hebben gevormd voor het resultatenhoofdstuk. Met deze acties heeft de onderzoeker beoogd om de betrouwbaarheid van het onderzoek te versterken (Fischer & Julsing, 2023, p. 83).

4. Resultaten

In dit hoofdstuk worden de onderzoeksresultaten gepresenteerd aan de hand van de vijf deelvragen. Allereerst worden in sectie 4.1 en 4.2 de bevindingen uit de Grand Tour besproken. In sectie 4.1 wordt ingegaan op de mate waarin de informanten zicht hebben op criminaliteit waarbij AI een rol speelt. In sectie 4.2 wordt beschreven welke zorgen zij uiten over GenAI in relatie tot jongeren. Vervolgens worden in sectie 4.3 tot en met sectie 4.5 de bevindingen uit de focusgroepen met jongeren besproken. In sectie 4.3 wordt uiteengezet hoe de deelnemende jongeren GenAI gebruiken en wat zij daarbij verantwoord vinden. Vervolgens wordt in sectie 4.4 wordt ingegaan op de mate waarin jongeren risicovol of strafbaar gebruik van GenAI tegenkomen in hun omgeving. Tot slot wordt in sectie 4.5 beschreven hoe de deelnemende jongeren omgaan met de ingebouwde veiligheidsmechanismen (*guardrails*) bij hun gebruik van GenAI.

4.1 Zicht op criminaliteit gepleegd met AI

In deze sectie wordt besproken in hoeverre er zicht is op AI-gerelateerde criminaliteit volgens de informanten. Eerst zal worden ingegaan op de mate waarin hier zicht op is volgens de informanten. Vervolgens zal worden ingegaan op de factoren die daaraan ten grondslag liggen volgens hen. Tot slot wordt besproken in hoeverre het hebben van zicht op AI-gerelateerde criminaliteit gewenst is volgens de informanten.

4.1.1 Beperkt zicht

Uit de interviews met de informanten blijkt dat het zicht op AI-gerelateerde criminaliteit zeer beperkt is. De meesten van hen geven aan nauwelijks concrete casussen tegen te komen waarin AI een aantoonbare rol speelt bij de uitvoering van een delict en ook geen zicht te hebben op de mogelijke omvang daarvan. Ook vanuit wetenschappelijk perspectief, stelt respondent 5, is dit tot op heden een “totaal onderbelicht gebied”. Dit beperkte zicht hangt volgens de informanten samen met verschillende factoren.

4.1.2 “Zolang je er niet naar zoekt, bestaat het ook niet”

Een eerste reden die informanten noemen, is dat AI-gebruik bij het plegen van strafbare feiten niet actief wordt opgespoord. Bij het opnemen van aangiftes wordt niet standaard doorgevraagd naar de mogelijkheid dat AI een rol heeft gespeeld bij de uitvoering van een delict. Volgens de informanten komt dit doordat vaak geen aandacht is bij de politie voor de mogelijkheid daarvan. Daarover vertelt informant 3:

> Zolang je er niet naar zoekt, bestaat het ook niet, ga je het ook niet vinden zeg maar. En dat is het probleem waar we nu tegenaan lopen. Uh, bij de aangiftes wordt er

misschien ook niet altijd doorgevraagd over de mogelijkheid van AI, en dat is al aan de voorkant van wat we bij de politie zien.

Wanneer AI-gebruik niet wordt uitgevraagd of benoemd door de aangever verschijnt het ook niet in registraties. Daardoor blijft mogelijk AI-gebruik buiten beeld bij het uitvoeren van zoekslagen in politiesystemen. Dit wordt ook benoemd in het Intelbeeld over AI door de politie Midden-Nederland. Binnen dat onderzoek is gezocht op basis van trefwoorden zoals AI in politiesystemen. Daaruit blijkt volgens de onderzoekers dat “veel registraties en meldingen gerelateerd aan AI niet door de Landelijke Cybercrime Query eruit gefilterd worden.”

Daarnaast speelt mee dat GenAI kan worden ingezet bij een breed scala aan delicten. Voorbeelden die de informanten noemen variëren van phishing en identiteitsfraude tot afpersing, documentvervalsing, drugsgerelateerde delicten en cybercriminaliteit. Doordat GenAI in veel verschillende modus operandi kan worden ingezet, voelt niemand zich volgens de informanten geroepen om hier regie over te nemen. Informant 3 omschrijft dit als volgt: “Dus het probleem is dat het een probleem van ons allemaal is waardoor het eigenlijk een probleem van niemand is hè, omdat iedereen denkt van oh, iemand anders komt het oplossen”. Deze diffuse verantwoordelijkheid, zo stelt de informant, zorgt ervoor dat er geen structurele aandacht is voor AI-gebruik waardoor het zicht van opsporingsinstanties verder versplintert. Hierdoor ontstaat een situatie waarin AI-gebruik voor criminaliteit mogelijk wel voorkomt, maar niet altijd als zodanig wordt herkend of vastgelegd.

Verder is het volgens de informanten lastig om een representatief beeld te vormen van de omvang van AI-gebruik voor criminaliteit op basis van bestaande aangiftes. Zij wijzen op de lage aangiftebereidheid van cybercriminaliteit in het algemeen, waardoor veel slachtoffers überhaupt geen aangifte doen. Bovendien hebben slachtoffers volgens hen vaak geen zicht op en aandacht voor de manier waarop het delict waarvan zij slachtoffer zijn geworden is uitgevoerd, en dus ook niet of daarbij AI is gebruikt. Volgens de informanten is de kans daardoor groot dat een deel van de mogelijke AI-gerelateerde criminaliteit niet in politiesystemen wordt vastgelegd.

Tegelijkertijd blijkt uit het AI-Intelbeeld van de politie Midden-Nederland dat de onderzoekers tijdens hun analyse van registraties regelmatig meldingen aantroffen waarin AI een rol werd toegeschreven, zonder dat die betrokkenheid kon worden vastgesteld. Deze zijn vaak gebaseerd op vermoedens van de aangever, die bijvoorbeeld voort kunnen komen uit verward gedrag. Voorbeelden hiervan die door de informanten worden benoemd zijn meldingen van verwarde personen die vermoeden dat AI hen manipuleert, of dat hun familieleden deepfakes zijn. Informant 1, die een vergelijkbare zoekstrategie hanteert bij haar onderzoek, vertelt daarover:

> Ik ben nu toevallig bezig, maar ik heb 600 hits, nou ja de eerste 30 heb ik nu bekeken, een is er naaktbeelden maar de rest is allemaal uh, fake anders. Of verwarde personen die zeggen ja ik weet niet, mijn zoon kan ook een deepfake zijn, jullie moeten kijken of hij wel echt mijn zoon is, maar dan gaat het om psychische patiënten.

Hierdoor ontstaat een dataset waarin meldingen met en zonder daadwerkelijke AI-betrokkenheid door elkaar lopen, wat het moeilijker maakt om daadwerkelijk AI-gerelateerde delicten te identificeren. Dat maakt het uitvoeren van zoekslagen met trefwoorden zoals “AI” in politiesystemen volgens verschillende informanten geen effectieve zoekstrategie. Daardoor moet in de praktijk elke melding handmatig worden beoordeeld. Dit maakt het in kaart brengen van daadwerkelijk AI-gerelateerde criminaliteit op basis van bestaande meldingen volgens hen een arbeidsintensief proces.

4.1.3 “AI is gewoon een middel”

Daarnaast geven de informanten aan dat het vastleggen van AI-gebruik bij de uitvoering van een delict niet per definitie wenselijk is. Zij vertellen AI in veel gevallen slechts te zien als een middel bij de uitvoering van een bestaand delict. Informant 1 maakt daarbij de vergelijking met traditionele middelen die worden gebruikt door daders: “Maar er is dus geen enkele standaard van vastleggen want AI is gewoon een middel. We leggen ook niet elke hamer vast of elke breekijzer dat een inbreker gebruikt”. Daarnaast wordt het toevoegen van een aparte registratiecode voor AI-gebruik bij een delict door sommige informanten als onpraktisch gezien, omdat het tot op heden nog onduidelijk is wanneer AI-gebruik evident genoeg is om vast te leggen.

Bovendien wijst informant 3 erop dat het juridisch ingewikkeld kan zijn om vast te stellen of een verdachte AI heeft geraadpleegd ter voorbereiding of uitvoering van een delict. Het gericht doorzoeken van digitaal beslag om dit te achterhalen valt namelijk niet altijd binnen de toegestane onderzoeksvragen van opsporingsonderzoek. Hierdoor bestaat de kans dat dergelijke informatie niet kan worden verzameld. Dat obstakel licht informant 3 nader toe: “We hebben een Landeck-arrest waardoor we alleen maar nu in het digitaal beslag met een specifieke vraag mogen kijken wat dus dat onderzoek dient, en dit is meer een soort beschouwende vraag.” Hierdoor blijft ook dit potentiële zichtpunt tot op heden beperkt.

Concluderend is het zicht op AI-gerelateerde criminaliteit volgens de informanten beperkt. Dat zicht wordt belemmerd door een combinatie van organisatorische, praktische en juridische factoren. AI-gebruik wordt door de politie niet actief uitgevraagd of herkend en bovenal niet op systematische wijze geregistreerd. Tegelijkertijd zorgen de lage aangiftebereidheid, onwetendheid bij slachtoffers en meldingen waarin betrokkenheid van AI

niet eenduidig kan worden vastgesteld ervoor dat bestaande data slecht beperkt inzicht biedt in hoeverre er sprake is van AI-gerelateerde criminaliteit.

4.2 Zorgen over GenAI in relatie tot jongeren

Tijdens de gesprekken uiten de informanten verschillende zorgen over GenAI in relatie tot jongeren. In het volgende deel worden de drie zorgen besproken die daaruit het sterkst naar voren komen. Eerst wordt ingegaan op zorgen met betrekking tot de onduidelijke grens tussen grappige en grensoverschrijdende deepfakes. Daarna wordt de vaak genoemde zorg over uitkleedapps en de toegankelijkheid daarvan besproken. Tot slot wordt ingegaan op zorgen over de afhankelijkheid van jongeren van GenAI.

4.2.1 Deepfakes: onduidelijke grens tussen grappig en grensoverschrijdend

De informanten geven aan dat het maken van deepfakes onder jongeren vaak gaat over iets onschuldigs en “speels”. Zij zien dat jongeren dergelijke tools bijvoorbeeld gebruiken om stemmen te vervormen of afbeeldingen van anderen te bewerken voor de grap. Zo vertelt informant 4 over een situatie waarin jongeren beelden van hem bewerkten met AI:

> Maar nu hebben ze, van een van die foto’s van mij, hebben ze mij op een Jeu de Boulesbaan gezet bij een bejaardenbingo. Maar goed in die zin is dat lollig en speels en hè, kan het in deze context bij ons in die groepen helemaal geen kwaad.

Tegelijkertijd signaleren meerdere informanten dat deze speelsheid kan verschuiven richting gebruik dat voor anderen als grensoverschrijdend wordt ervaren. Volgens hen is het voor jongeren vaak niet duidelijk waar die grens precies ligt. De informant uit het voorgaande voorbeeld benoemt dit zelf ook: “Maar dat is wel natuurlijk hoever gaat die grens daarin, want je bent natuurlijk wel met foto's van mij andere beelden aan het creëren, daar proberen we wel het gesprek over aan te gaan.” De informanten benadrukken dat dezelfde tools die worden ingezet voor grappige bewerkingen, ook kunnen worden gebruikt voor het maken van deepfakes die schadelijke gevolgen voor anderen hebben. Volgens informant 4 kan dit een “afglijdende schaal” zijn, waarbij jongeren al dan niet bewust steeds verder de grens opzoeken bij het maken van deepfakes.

Dit blijkt ook uit een voorbeeld van informant 1 over een recente zaak waarbij een wijkagent foto’s ontving van mensen in een lokale pizzeria. De afgebeelde personen droegen een politie-uniform, hadden wapens vast en brachten de Hitlergroet. Dit leidde tot verwarring over de echtheid van de beelden en tot vragen over mogelijke strafbare feiten doordat wapens werden afgebeeld. Uit nader onderzoek bleek dat de afbeelding met AI gegenereerd was door een jongere. Dat benadrukt de zorg van de informanten over de onduidelijke grens tussen onschuldig experimenteren met deepfakes voor de grap en het creëren van content die schadelijk of zelfs strafbaar is.

4.2.2 Toegankelijkheid van uitkleedapps

De meeste zorgen die de informanten uiten, hebben betrekking op AI-tools voor het genereren van deepnudes (AI-gegenereerde naaktbeelden). Ten eerste benadrukken zij dat deze tools zeer toegankelijk zijn voor jongeren, doordat ze in grote aantallen beschikbaar zijn in app-stores. Tegelijkertijd wordt er volgens de informanten weinig gedaan om die apps offline te halen. Over dit gebrek aan regulering vertelt informant 1:

> Er zijn gewoon heel veel nudify-apps, die staan gewoon in de app stores. En er wordt niks aan gedaan door Android en Apple totdat journalisten of anderen erover gaan klagen, maar die zijn niet actief bezig om uitkleedapps uit hun stores te verwijderen.

Daarnaast is volgens de informanten niet altijd een aparte app nodig om deepnudes te maken. Ook GenAI-toepassingen die via het internet toegankelijk zijn, zoals Grok, kunnen worden gebruikt om seksuele deepfakes van anderen te creëren. Verder signaleren de informanten dat jongeren via online platforms zoals Telegram en Discord in aanraking komen met AI-tools die ontwikkeld zijn om deepnudes te genereren. Zij zien dat deze tools onderling worden gedeeld in Discord-groepen en in sommige gevallen zelfs actief worden gepromoot richting jongeren. Over die zorg vertelt informant 4:

> Het wordt veel makkelijker gemaakt dus daar zijn wel zorgen over ook naar de jongeren en ook qua promoties. In Telegramgroepen worden steeds meer bots en van dat soort AI-tools gepromoot richting de doelgroep en dat is ook gewoon zorgwekkend. Ook inderdaad een gewone foto van iemand van TikTok of Instagram kan je zo omzetten naar naaktbeelden en klaar. Dat zijn wel zorgelijke signalen.

Volgens de informanten die met jongeren werken sluiten deze ontwikkelingen aan bij meldingen die zij vanuit scholen ontvangen. Daar zijn signalen over jongeren die deepnudes maken van leeftijdsgenoten of docenten op basis van bestaande foto's. Dergelijke meldingen versterken hun zorgen over de mate waarin jongeren in aanraking komen met deepnudes, als dader of slachtoffer.

4.2.3 Afhankelijkheid van AI-taalmodellen

De informanten geven aan dat GenAI een steeds groter onderdeel wordt van de leefwereld van de jongeren waar zij mee in contact staan. Zij merken dat AI-taalmodellen zoals ChatGPT niet alleen gebruikt worden voor schoolopdrachten, maar ook voor alledaagse vragen, meningsvorming en het uitvoeren van uiteenlopende taken. Volgens de informanten lijkt het gebruik van AI-taalmodellen voor deze jongeren in de meeste gevallen

vanzelfsprekend geworden. Hierdoor staan jongeren volgens hen lang niet altijd stil bij de mogelijke risico's of beperkingen van AI-taalmodellen.

De voornaamste zorg die informanten hierbij uiten gaat over de groeiende afhankelijkheid van AI-taalmodellen onder jongeren. Zij merken dat sommige jongeren moeite hebben om taken zelfstandig uit te voeren, zonder eerst een taalmodel te raadplegen. Informant 4 ziet dit ook terug in zijn communicatie met jongeren:

> Het wordt voor ons ook steeds duidelijker in de communicatie die we krijgen, zeker als het gaat om stagiaires die net begonnen zijn, dat ze zelfs tot het niveau van appen naar ons, dat ze daar echt al elke zin al door AI laten schrijven voordat ze het eruit gooien.

Deze afhankelijkheid vinden informanten zorgelijk, omdat zij zien dat jongeren geneigd zijn om de antwoorden van een taalmodel voor waar aan te nemen zonder hier zelf kritisch bij na te denken. Dit blijkt ook uit de observatie van informant 4, die stelt dat jongeren "blind vertrouwen" op AI doordat zij "überhaupt alles erin gooien en wat ze terugkrijgen aannemen als waarheid." Dit is zorgelijk, omdat taalmodellen regelmatig onnauwkeurige of misleidende informatie genereren.

Ook informant 2 merkt dat jongeren geneigd zijn de antwoorden van een taalmodel voor waar aan te nemen: "Je ziet dat dat ze steeds minder grip krijgen op wat technologie kan en doet en steeds meer maar vertrouwen op dat het wel goed zal zijn." Volgens hem hangt dit samen met het feit dat jongeren over het algemeen weinig inzicht hebben in de manier waarop AI-systemen werken, oftewel "wat er onder de motorkap zit". Daardoor zijn zij zich er volgens hem in mindere mate van bewust dat de antwoorden van taalmodellen zoals ChatGPT onwaarheden kunnen bevatten.

Daarnaast wijzen de informanten op het risico dat AI-taalmodellen sturend kunnen zijn in de manier waarop zij jongeren adviseren. Zij merken dat jongeren steeds vaker keuzes of persoonlijke kwesties met een taalmodel bespreken, waarbij jongeren een taalmodel in sommige gevallen behandelen als een "online vriend" volgens informant 3. De informanten stellen dat dit problematisch kan zijn, omdat taalmodellen gesprekken niet inhoudelijk begrijpen terwijl dat wel zo over kan komen op jongeren. Bovendien bestaat daardoor het risico dat er niet wordt ingegrepen wanneer gesprekken geleidelijk een schadelijke richting op gaan. Dit kan vooral risicovol zijn voor jongeren die worstelen met hun mentale gezondheid, doordat een taalmodel onbedoeld negatieve gedachten of schadelijke ideeën kan bevestigen. Informant 4 benadrukt dit risico:

> Het is natuurlijk heel sturend. Ook als het gaat om mentale gezondheid moet je je er bewust van zijn dat het naar je toe kan praten en dat het je ook op die manier dingen kan aanpraten die er misschien helemaal niet zijn.

Volgens de informanten kan deze combinatie van afhankelijkheid, het kritiekloos vertrouwen op antwoorden en de sturende manier waarop taalmodellen reageren ertoe leiden dat jongeren door AI worden aangezet tot risicovol gedrag. Zij benadrukken daarom het belang van bewustwording van de beperkingen van taalmodellen, en het ontwikkelen van een kritische houding ten aanzien van de informatie afkomstig van taalmodellen.

Samenvatten benoemen de informanten drie ontwikkelingen die zij zorgelijk vinden. Ten eerste hebben jongeren volgens hen soms moeite om de grens tussen speels en grensoverschrijdend gebruik van deepfakes te herkennen, waardoor onschuldig bedoelde bewerkingen kunnen overgaan in risicovolle of strafbare content. Daarnaast komen jongeren door de brede beschikbaarheid van uitkleedapps in aanraking met toepassingen die misbruik met behulp van GenAI faciliteren, wat volgens de informanten wordt bevestigd door signalen vanuit scholen. Tot slot uiten de informanten zorgen over de potentieel sturende werking van taalmodellen, waardoor jongeren mogelijk worden aangezet tot risicovol of strafbaar gedrag.

4.3 Verantwoord gebruik van GenAI volgens jongeren

Tijdens de focusgroepen is met de deelnemende jongeren besproken op welke manieren zij GenAI gebruiken en wat volgens hen daarbij verantwoord is. In de volgende sectie wordt hierop ingegaan. Allereerst volgt een kort overzicht van de manieren waarop deze jongeren GenAI in hun dagelijks leven gebruiken. Vervolgens wordt besproken hoe zij de grenzen van verantwoord GenAI-gebruik beoordelen en welke uitdagingen zij daarbij ervaren. Tot slot wordt ingegaan op hun behoefte aan duidelijke richtlijnen voor verantwoord AI-gebruik.

4.3.1 Manieren van gebruik

Tijdens de focusgroepen beschrijven deze jongeren dat zij GenAI op uiteenlopende manieren gebruiken in hun dagelijks leven. Hun gebruik vormt een breed spectrum: sommige jongeren gebruiken GenAI nagenoeg voor alles, terwijl anderen het bewust beperken of zelfs vermijden.

Voor een deel van de jongeren is GenAI een praktische hulpbron voor schoolwerk. Zij gebruiken taalmodellen om teksten te controleren, extra uitleg te krijgen of informatie sneller te vinden dan via reguliere zoekmachines. Bovendien geven sommige jongeren aan schoolopdrachten, zoals presentaties, volledig door GenAI te laten maken.

Aan de andere kant zijn er jongeren die GenAI bewust zo min mogelijk gebruiken. Dit doen zij vanwege zorgen over de klimaatimpact van AI-datacentra, verlies van creativiteit of omdat zij niet afhankelijk willen zijn van taalmodellen. Vooral jongeren van de mediavormgeving-opleiding hechten waarde aan het zelf maken van hun projecten. Zij zien kunst gemaakt met GenAI als gestolen kunst, omdat het wordt gegenereerd op basis van het werk van anderen. Die jongeren vinden het daarom jammer dat hun opleiding AI-gebruik steeds vaker aanmoedigt. Deze uiteenlopende manieren van AI-gebruik, van intensief inzetten tot bewust vermijden, vormen het vertrekpunt voor de manier waarop deze jongeren beoordelen wat zij al dan niet verantwoord vinden bij het gebruik van AI.

4.3.2 Grenzen van verantwoord gebruik

In hun beoordeling van wat al dan niet verantwoord is bij het gebruik van GenAI, gaan de uitspraken van de deelnemende jongeren vooral over deepfakes met GenAI. Hierbij maken zij onderscheid tussen verschillende categorieën deepfakes. Enerzijds zijn er toepassingen die jongeren in geen enkel geval acceptabel vinden. Anderzijds zijn er jongeren die zeggen het “eigenlijk wel voor alles” verantwoord vinden om GenAI te gebruiken. Ook benoemen zij vormen van deepfake-gebruik die zij als grappig of ogenschijnlijk onschuldig ervaren.

Ten eerste zeggen alle deelnemende jongeren dat het maken van deepnudes van anderen in geen enkel geval verantwoord is. Daarvoor geven zij verschillende redenen. Allereerst stellen zij dat deepnudes hetzelfde effect kunnen hebben op het slachtoffer als het

maken of verspreiden van niet met AI-gegenereerde naaktbeelden. Daardoor doet het feit dat deepnudes gemanipuleerde beelden zijn volgens hen niet af aan de ernst daarvan. Daarnaast kunnen deepnudes volgens hen een grote impact hebben, doordat het slachtoffer naar anderen moet bewijzen dat de beelden niet echt zijn. Echter is het volgens de jongeren soms ook afhankelijk van de situatie in welke mate die beelden serieus worden genomen. Zo vertelt respondent 5 te hebben meegemaakt dat er deepnudes werden gemaakt van iemand uit haar omgeving. De echtheid van die beelden werd door anderen volgens haar direct in twijfel getrokken, wat het volgens haar minder erg maakte:

> Een beste vriendin van mijn vriendje die heeft wel eens dat een ex van haar uh video's had ge-uhm, haar kleding een soort van had weggehaald op video's, en daarmee dan een soort van kon zeggen van 'oh mijn ex is echt een slet want ze heeft dit soort video's online gepost' et cetera. Ja, ja, dat was in principe al heel snel dat iedereen wist dat dat niet waar was, maar dat is ook wel omdat het best een populaire meid is die heel veel mensen mogen. Maar ja, als je wat minder gemogen wordt of zo, dan denk ik niet dat dat heel prettig is, uh sowieso niet, maar er zijn wel omstandigheden dat het minder erg is dan, ja.

Toch kunnen de beelden volgens de jongeren een blijvende impact hebben op de reputatie van het slachtoffer, zelfs wanneer duidelijk is dat de beelden niet echt zijn. Over die impact die deepnudes kunnen hebben, vertelt respondent 7: "Maar zoiets heeft wel gewoon directe impact op een persoons reputatie. Ook al is het niet waar, de schade blijft."

Hoewel deepnudes volgens deze jongeren het meest eenduidige voorbeeld van onverantwoord AI-gebruik is, komt tegelijkertijd naar voren dat hun oordeel over andere vormen van deepfakes minder eenduidig is. Zowel bij (niet-seksuele) deepfakes als met AI-gemanipuleerde audio variëren de meningen van de jongeren over verantwoord AI-gebruik.

Ten eerste vinden sommige jongeren het namaken of aanpassen van andermans stem met GenAI in geen enkel geval verantwoord. Zij voelen zich er niet fijn bij dat een ander hen dingen kan laten zeggen die niet hebben plaatsgevonden. Ook kan met AI-gemanipuleerde audio volgens hen snel de grens over gaan doordat ze gebruikt kunnen worden voor misleiding, oplichting of het verspreiden van desinformatie. Zij noemen voorbeelden zoals iemand schadelijke uitspraken laten zeggen die nooit hebben plaatsgevonden, of het gebruik van een met AI-gemanipuleerde stem om een familielid op te lichten.

Daarnaast vinden zij het gebruik van een deepfake-stem in sommige gevallen respectloos, bijvoorbeeld wanneer de stem van een overleden artiest wordt gebruikt om een nieuw lied te maken. Verder wijzen zij op het risico dat iemands stem in een AI-database terechtkomt en er vervolgens geen toezicht is op wat met die data wordt gedaan. Dat maakt

het volgens hen ook niet verantwoord om een stem te manipuleren met GenAI voor de grap. Respondent 5 zegt daarover:

> Nou ja ook voor een grapje denk ik ja weet je, je stem, diegenes stem staat nu wel in een soort database waarschijnlijk waar die nooit meer uit komt. Dus ja ik zou zeggen gewoon ook niet als, ook niet voor de leuk of zo weet je.

Andere jongeren vinden het maken van stemmen of niet-seksuele deepfakes met GenAI wel verantwoord, mits het gaat om onschuldige grappen en de intentie duidelijk speels is. Zo zegt respondent 3 er wel om te kunnen lachen wanneer GenAI wordt gebruikt door vrienden onderling om elkaar een ander accent te geven.

Ook deepfake-beelden kunnen volgens sommige jongeren grappig zijn, bijvoorbeeld wanneer vrienden elkaar onderling iets grappigs laten doen. Uit de gesprekken blijkt dat de deelnemende jongeren drie factoren meewegen bij de beoordeling of het maken van een deepfake verantwoord is: (1) de relatie tussen de maker en de afgebeelde persoon, (2) de intentie van de maker en (3) de mening van de afgebeelde persoon over de deepfake. Deze drie factoren komen tijdens alle focusgroepen herhaaldelijk naar voren.

Allereerst geven jongeren aan belang te hechten aan wie de deepfake maakt. Doorgaans worden deepfakes die door vrienden worden gemaakt door deze jongeren sneller geaccepteerd, omdat zij ervan uitgaan dat vrienden hun grenzen kennen en geen kwaad in de zin hebben. Wanneer de maker een onbekende is, ervaren de jongeren dit al snel als ongemakkelijk omdat zij niet kunnen inschatten wat diegene met die beelden of audio zal doen. Zo vertelt respondent 1: “Nou ja vrienden of vriendinnen die gaan er niet zo heel snel wat mee doen. En zeg maar iemand waar je niet heel veel mee om gaat, ja je weet niet wat diegene er mee wilt gaan doen.”

Daarnaast nemen de jongeren de intentie van de maker van de deepfake in acht. Zij maken een duidelijk onderscheid tussen deepfakes die zijn bedoeld als grap en deepfakes die worden ingezet om iemand te schaden, belachelijk te maken of te misleiden. De jongeren benadrukken dat hun mening over een deepfake afhankelijk is van de bedoeling erachter. Zo zegt respondent 2 over met AI-gegenereerde foto’s:

> Tuurlijk kan dat wel extremer worden gedaan, uh foto’s die je zelf nooit zou hebben gemaakt of andere zouden maken, die zouden dan kunnen worden verspreid, maar dat is ook weer een beetje zo van ja wat is de intentie. Is het haat, wil je echt diegene pijn doen met die foto’s of is het gewoon een grappig iets tussen vrienden.

Tot slot overwegen de jongeren ook wat de afgebeelde persoon zelf van de deepfake vindt of zou vinden. Zonder toestemming van de afgebeelde is een deepfake volgens hen al snel ongepast. Volgens sommige jongeren zou daarom in theorie expliciet toestemming nodig zijn van de afgebeelde persoon. Echter gaat dit er in de praktijk volgens deze jongeren vaak anders aan toe. Volgens hen is die toestemming binnen vriendschappen vaak implicieter, omdat zij elkaars grenzen kennen. Hierover vertelt respondent 3: “Ja ik denk dat dat weer zo’n dingetje is van officieel moet je toestemming vragen om foto’s te maken van mensen uh maar ja inderdaad als je vrienden bent dan ken je elkaars grenzen”. Respondent 2 voegt daaraan toe dat zij dit vaak zelf wel weet in te schatten:

> Ik denk dat uh dat als je gewoon goede vrienden bent dat dat een beetje verondersteld is. Als je al een bepaalde tijd vrienden bent bijvoorbeeld dat je stickers, iedereen heeft denk ik wel eens een Whatsapp-sticker van vrienden gemaakt of heeft van zichzelf een sticker gekregen en dat je dat gewoon als grapje stuurt. En daar heb je geen toestemming voor gevraagd maar ja niemand die zegt dit is niet normaal, maar je weet ook zelf de grens. Je gaat niet je vriend te kijk zetten of zo of uitkleden met AI om dat rond te sturen, ja.

Ondanks deze ‘beoordelingscriteria’ is de grens tussen grappige en grensoverschrijdende deepfakes volgens enkele jongeren soms niet helemaal helder. Een voorbeeld dat zij zelf aanhalen, betreft een situatie op school waarin een klasgenoot deepnudes van docenten had gemaakt door hun gezichten op vrouwelijke lichamen te zetten met een deepnude AI-tool. Over wat hij ervan vond toen dat hem dat ter ore kwam, vertelt respondent 12:

> Uh ja, dom. Ja ik hield me er niet echt mee bezig eigenlijk. Ik zag het wel voorbijkomen dat die jongen apart werd genomen en toen hoorde ik dat het was omdat hij AI-filmpjes maakte. Toen dacht ik van ja, ja [lacht] best wel dom toch?

Andere jongeren geven toe dat zij de beelden eerst ook grappig vonden, maar erkennen dat het ook een grens over ging. Respondent 10 zegt hierover: “Ja dat is natuurlijk wel grappig aan de ene kant. Maar ja het is natuurlijk niet de bedoeling”. Volgens de jongeren zag hun klasgenoot de ernst hiervan pas in nadat hij daarop door docenten werd aangesproken. Dit voorbeeld laat volgens hen zien dat de grens tussen humor en schadelijk gebruik van deepfakes niet voor elke jongere vanzelfsprekend is en daarom regelmatig soms moet worden verduidelijkt. Tevens bevestigt hun docent dat dit soort situaties elk jaar voorkomen op de school. Respondent 12 reageert daarop: “Maar het is gewoon lol trappen

dat te ver gaat. Het zijn gewoon uh onbesproken regels die dan weer duidelijk gemaakt moeten worden. Want je gaat uiteindelijk wel over mensen hun grens heen."

4.3.3 Behoefte aan richtlijnen

Deze behoefte aan regels over gebruik van GenAI komt terug in meerdere gesprekken met de jongeren. Zij denken dat het voor anderen soms lastig kan zijn om in te schatten wanneer een deepfake binnen de grenzen van verantwoord gebruik valt. Zij vinden het opvallend dat op school vaak wél afspraken bestaan over het gebruik van GenAI voor opdrachten of toetsing, maar dat daarbuiten geen expliciete richtlijnen zijn over wat acceptabel is. Volgens respondent 3 zouden dergelijke richtlijnen op korte termijn moeten worden gerealiseerd, doordat GenAI zich op dit moment snel ontwikkelt:

> Uh maar ik denk als je dat wil bevorderen dan moet je ook wel snel komen met nieuwe regels. Want zoals [respondent 2] zei, het wordt gebruikt voor goed en voor slecht en daar moet je denk ik gewoon echt heel duidelijk in zijn, en niet zo van oké we maken nu een reglement en we geven er vijf jaar geen aandacht aan. Want dit is juist iets wat nu heel snel ontwikkelt.

Volgens meerdere jongeren zouden richtlijnen over GenAI-gebruik het voor jongeren duidelijker maken waar de grens ligt, omdat dit niet voor iedereen vanzelfsprekend blijkt. Op die manier stellen zij dat situaties zoals het voorgaande voorbeeld in de toekomst wellicht voorkomen kunnen worden.

Samenvattend laten deze resultaten zien dat jongeren verschillende grenzen hanteren bij hun gebruik van GenAI. Vooral deepfakes staan centraal bij hun beoordeling van verantwoord gebruik van GenAI. Deepnudes worden door alle jongeren als onacceptabel beschouwd, terwijl andere vormen van deepfakes, zoals niet-seksuele beelden of AI-gegenereerde stemmen, situationeel worden beoordeeld. Bij hun oordeel wegen zij voornamelijk drie factoren mee: (1) de relatie tussen de maker en de afgebeelde persoon, (2) de intentie van de maker en (3) de mening van de afgebeelde persoon over de deepfake. Hoewel de deelnemende jongeren binnen vriendschappen vaak vertrouwen op impliciete grenzen en er in de schoolcontext regels bestaan over AI-gebruik, ervaren zij daarbuiten een gebrek aan richtlijnen. Volgens hen kan dat helpen om het grijze gebied rondom verantwoord AI-gebruik voor jongeren op te helderen.

4.4 Ervaringen van jongeren met risicovol of strafbaar GenAI-gebruik in hun omgeving

Tijdens de focusgroepen is met de deelnemende jongeren besproken welke vormen van risicovol of strafbaar gebruik van GenAI zij in hun omgeving tegenkomen. In de volgende sectie wordt dit toegelicht. Ten eerste wordt beschreven welke vormen van risicovol of strafbaar AI-gebruik deze jongeren zelf waarnemen of vermoeden. Vervolgens wordt besproken hoe zij aankijken tegen de vraag wie verantwoordelijk is wanneer GenAI wordt ingezet voor strafbare handelingen.

4.4.1 Beperkte ervaringen

De meerderheid van de jongeren geeft aan geen risicovol of strafbaar gebruik van GenAI tegen te komen in hun omgeving. Sommige van hen vertellen dergelijk gebruik wel te kennen uit nieuwsberichten. Een kleiner deel van de jongeren noemt wel voorbeelden, maar deze worden meestal slechts door één jongere benoemd. Bovendien is de aard van deze voorbeelden vaak speculatief, omdat de jongeren niet altijd zeker weten of daarbij daadwerkelijk GenAI is gebruikt. Ter illustratie van de aard daarvan worden die voorbeelden in deze sectie besproken.

4.4.2 Ervaringen met deepnudes en uitkleedapps

Tijdens alle focusgroepen komen de onderwerpen deepnudes en uitkleedapps ter sprake. Op een enkeling na hebben alle jongeren hier wel eens van gehoord, meestal via online nieuwsbronnen. Echter heeft de meerderheid van hen geen ervaring met deepnudes in hun omgeving. Toch komen tijdens de focusgroepen twee situaties naar voren waarin deepnudes zijn gemaakt door mensen uit de omgeving van deze jongeren. De eerste situatie betreft respondent 5, die vertelt dat er deepnudes zijn rondgegaan van een kennis van haar, gemaakt door het ex-vriendje van die persoon. Daarnaast vertellen jongeren uit de derde focusgroep dat er onlangs deepnudes zijn rondgegaan van twee mannelijke docenten. Beide voorbeelden zijn reeds besproken in sectie 4.3.2.

4.4.3 Ervaringen met AI-advertenties met bekende personen

Enkele jongeren vertellen dat zij online wel eens nepadvertenties tegenkomen waarin een bekend persoon een product lijkt te promoten. Doorgaans gaat het hierbij om een nieuwe online game of een nieuw product uit de supermarkt. Zij vermoeden dat daarbij GenAI is gebruikt om stemmen of beelden te manipuleren. Respondent 6 vertelt hoe hij dat online tegenkomt:

> Nou AI kan bijvoorbeeld ook je stem opnemen en die kan ook bijvoorbeeld advertenties daarmee maken, met jouw stem eronder. Dat heb ik ook een paar keer voorbij zien

komen, bijvoorbeeld op YouTube of Instagram. Bijvoorbeeld dat een bekende Twitch-streamer of YouTuber opeens een AI-advertentie voorbij ziet komen met zijn of haar stem eronder, en dat is natuurlijk strafbaar.

Verder noemen de jongeren enkele andere voorbeelden of vermoedens van risicovol of strafbaar AI-gebruik, maar deze worden telkens slechts één keer genoemd. Ten eerste stelt respondent 3 wel eens te hebben gehoord van oplichting met GenAI via het kledingverkoopplatform Vinted, waarbij een koper een productfoto met AI zou bewerken zodat het lijkt alsof het kapot is om geld terug te krijgen. Echter heeft hij dit niet zelf gezien maar van vrienden gehoord. Daarnaast vertelt respondent 5 advertenties te hebben gezien waarmee zogenoemde snuff-films worden gepromoot. Dit zijn films waarin de moord van een bestaand persoon wordt gefilmd met financieel gewin als doel. Zij vermoedt dat deze advertenties met GenAI zijn gemaakt, maar dit is volgens haar niet met zekerheid te zeggen. Beide voorbeelden lijken dus te gaan om speculatief gebruik van GenAI voor risicovolle of strafbare doeleinden.

4.4.4 Opvattingen over verantwoordelijkheid bij strafbare handelingen met GenAI

De deelnemende jongeren hebben verschillende opvattingen over wie verantwoordelijk is wanneer AI wordt ingezet voor strafbare handelingen. Hun opvattingen laten drie duidelijke lijnen zien. Ten eerste vindt de meerderheid van de jongeren dat de persoon die GenAI hiervoor inzet verantwoordelijk blijft. Zij benadrukken dat iemand bewust kiest om AI te gebruiken voor een bepaald doel en dat de gebruiker degene is die AI aanstuurt op het uitvoeren van de strafbare handelingen. Voor hen speelt de intentie van de gebruiker bij dit verantwoordelijkheidsvraagstuk wederom een belangrijke rol.

Daarnaast zijn de jongeren van mening dat ontwikkelaars van AI-systemen eveneens een mate van verantwoordelijkheid dragen. Zij vinden dat die ontwikkelaars betere *guardrails* moeten inbouwen om misbruik van AI-modellen te voorkomen. Zo vertelt respondent 8 dat die *guardrails* soms nog te makkelijk te omzeilen zijn bij ChatGPT:

> Beide, bedrijven en personen zijn verantwoordelijk. Ik bedoel, een bedrijf kan altijd betere *guardrails* maken. Uh, ik weet niet of je het gehoord hebt, maar pff een paar maanden geleden was er iemand die gewoon in een gedicht aan ChatGPT vroeg, gewoon uh, rijmend, hoe hij cocaïne moest maken en ChatGPT antwoordde gewoon. Omdat ze daar geen *guardrails* voor gemaakt hebben, want dat hadden ze niet verwacht. Dus ja, ik vind dat die persoon die dat vroeg verantwoordelijk is maar ik vind ook dat OpenAI daarvoor verantwoordelijk is. Want misschien hebben hun er niet over gedacht, maar die *guardrails* moeten er gewoon zijn soms.

Tegelijkertijd verwachten verschillende jongeren dat het in de praktijk vaak onduidelijk is wie verantwoordelijk is. Zo is het bij deepfake-advertenties of stemmanipulatie met GenAI volgens hen moeilijk vast te stellen wie de dader is. Daarover zegt respondent 6: “Mensen weten dan niet wie ze moeten aanwijzen van dat is hun schuld, omdat je niet kan weten van wie of wat, ja hoe ze ertoe zijn gekomen om zo ver te komen.”

Samenvattend geeft de meerderheid van de deelnemende jongeren aan weinig tot geen risicovolle of strafbare toepassing van GenAI tegen te komen in hun omgeving. De voorbeelden die worden genoemd zijn incidenteel en soms speculatief. Echter hebben enkele jongeren wel een indirecte ervaring met deepnudes. Wat betreft verantwoordelijkheid bij strafbare handelingen waarbij GenAI wordt ingezet wijzen de deelnemende jongeren vooral naar de gebruiker. Daarnaast vinden zij dat AI-bedrijven hun *guardrails* moeten verbeteren om misbruik tegen te gaan. Tegelijkertijd vermoeden zij dat het in de praktijk lastig kan zijn om vast te stellen wie in dergelijke situaties ‘de dader’ is en daarom ook bij wie de schuld van de strafbare handeling ligt.

4.5 Ervaringen van jongeren met *guardrails* van GenAI

In dit hoofdstuk wordt toegelicht welke ervaringen de jongeren hebben de ingebouwde veiligheidsmechanismen (*guardrails*) van GenAI. Eerst wordt beschreven hoe zij in aanraking komen met *guardrails*. Vervolgens wordt ingegaan op vragen die zij hebben gesteld aan taalmodellen waarbij de *guardrails* in werking traden. Daarna wordt aandacht besteed aan manieren waarop jongeren bewust proberen om *guardrails* te omzeilen

Uit de focusgroepen blijkt dat de deelnemende jongeren in uiteenlopende mate experimenteren met de *guardrails* van GenAI. Hoewel de meeste jongeren deze grenzen niet actief opzoeken, beschrijven zij hier wel mee in aanraking te zijn gekomen in diverse situaties. Daarbij ontstaan twee duidelijke patronen. Enerzijds zijn er jongeren die incidenteel of uit nieuwsgierigheid op *guardrails* van AI-modellen stuiten, anderzijds probeert een kleinere groep deze bewust te omzeilen. Deze situaties zullen in de volgende secties worden toegelicht.

4.5.1 Bewustzijn van guardrails

In alle gesprekken benoemen jongeren dat taalmodellen, zoals ChatGPT en Grok, zijn uitgerust met *guardrails* die voorkomen dat ongepaste of schadelijke informatie wordt gedeeld met de gebruiker. Deze *guardrails* worden door hen omschreven als een “muur” die bepaalt welke vragen al dan niet beantwoord mogen worden door GenAI. De deelnemende jongeren lijken zich daarmee bewust van het bestaan van deze begrenzingen.

4.5.2 Confrontatie met guardrails

Voor een deel van de jongeren, met name degenen zonder technische ICT-achtergrond, vormt deze begrenzing een eerste confrontatie met het feit dat bepaalde vragen niet zijn toegestaan. Meerdere jongeren geven aan wel eens een vraag te hebben gesteld aan een AI-taalmodel die werd tegengehouden door *guardrails*. Hierbij gaat het om situaties waarin zij niet bewust op zoek zijn naar manieren om *guardrails* te omzeilen, maar er onverwacht mee worden geconfronteerd wanneer zij een vraag stellen. De vragen die zij stelden variëren van het opzoeken van manieren om illegale films te streamen, tot het achterhalen van de prijs van softdrugs, de waarde van medicatie en het zoeken naar informatie over de werking van zuur op een lichaam. Deze jongeren stellen dat hun vragen vooral voortkomen uit nieuwsgierigheid, niet uit de intentie om iets strafbaars te doen. Zo vertelt respondent 5 over haar vraag aan ChatGPT over de waarde van ADHD-medicatie:

> Ja dat was het ook, want het was meer uit een soort nieuwsgierigheidje. Het was niet dat ik nu opeens onder een bruggetje m’n ADHD-medicatie ga zitten dealen, maar [lacht] het was gewoon meer dat, hoe dat werkt. Daar kreeg ik geen antwoord op.

De jongeren geven aan daardoor geen actie te ondernemen om toch aan de gevraagde informatie te komen. Daarnaast merkt respondent 5 op dat ChatGPT in dergelijke gevallen niet alleen de vraag weigert, maar soms ook een alternatief advies geeft. Zo adviseerde het taalmodel in haar geval om de medicijnen weg te gooien of terug te brengen naar de apotheek.

Verder wordt voor sommige jongeren pas duidelijk dat hun vraag ongepast kan zijn door de confrontatie met *guardrails*. Zo legt respondent 6 uit dat hij uit nieuwsgierigheid door een true crime film aan ChatGPT vroeg wat een bepaald zuur met een menselijk lichaam doet. Door de weigering van het taalmodel besefte hij dat de vraag buiten de context argwaan kan opwekken:

> […], dat ik vroeg van ja wat doet dit bepaald zuur dan met het menselijk lichaam als het met de huid in contact komt en dat soort dingen. Maar toen dacht ChatGPT gelijk dat ik het wilde gebruiken. En ik dacht van nee, ik wil gewoon weten van wat zijn de effecten daarvan of iets in die richting, ik weet niet meer precies. Maar toen zij 'ie ook van nee dat ga ik niet beantwoorden, en toen dacht ik oh... [lacht].

Hoewel deze jongeren de *guardrails* respecteren, wijzen zij erop dat een onbeantwoorde vraag in theorie kan leiden tot risicovol gedrag onder jongeren doordat iemand zelf gaat experimenteren.

4.5.3 Bewust omzeilen van guardrails

Naast jongeren die onbedoeld op *guardrails* stuiten, is er een kleinere groep jongeren die aangeeft deze beperkingen wel eens bewust te omzeilen. Dit betreft voornamelijk de jongeren met een technische ICT-achtergrond (van de opleidingen ICT System Engineer en ICT Support). Zij beschrijven dat dit mogelijk is door *prompts* zodanig te formuleren dat het taalmodel alsnog antwoord geeft op een vraag die normaliter door *guardrails* wordt geblokkeerd.

Een voorbeeld hiervan is het gebruiken van AI-taalmodellen om op een illegale manier toegang te krijgen tot content zoals films of online games. De jongeren leggen uit dat dit mogelijk is door de *prompt* zo te richten dat het model tóch een link of instructie geeft. Volgens respondent 9 is dat mogelijk door een *prompt* "een beetje te richten zodat het toch net die grens overgaat." Respondent 10 illustreert dit met een concreet voorbeeld waarin hij ChatGPT gebruikt om illegaal games te downloaden naar zijn PC:

> Ik heb ook wel eens uh, pas heb ik een Switch Emulator gedownload op m'n PC. En via ChatGPT heb ik een website gekregen van hem waar ik gratis games kon

downloaden voor die Switch Emulator. Waar eigenlijk ja, ik denk wat in totaal eigenlijk als je het goed doet vierhonderd euro kost, heb ik gewoon gratis gekregen.

De jongeren zeggen zich ervan bewust te zijn dat dit illegaal is. Daarom nemen zij verschillende voorzorgsmaatregelen om hun identiteit te beschermen. Zo verbergen zij hun IP-adres via een VPN (Virtual Private Network), gebruiken zij een niet-persoonlijk e-mailadres voor hun account en delen zij bewust geen persoonlijke informatie in de chat.

Daarnaast stellen zij dat het omzeilen van *guardrails* niet voor alle jongeren even toegankelijk is. Volgens respondent 8 kan dit makkelijk zijn “zo lang je weet wat er gebeurt en wat je aan het doen bent”, maar wordt het lastig wanneer iemand niet weet hoe *prompts* werken en hoe je die op een bepaalde manier moet “richten”. Volgens deze jongeren bepalen technische vaardigheden en kennis over de werking van taalmodellen in belangrijke mate of een jongere in staat is om GenAI te misbruiken.

4.5.4 Experimenteren met uncensored AI-modellen

Enkele jongeren met een technische ICT-achtergrond gaan verder dan het omzeilen van *guardrails* via *prompts* en experimenteren met uncensored AI-modellen. Dit zijn taalmodellen die bewust zijn ontworpen zonder *guardrails*. De jongeren vertellen dat dergelijke modellen beschikbaar zijn via de desktopapplicatie LM Studio. Door een taalmodel via LM Studio te downloaden, draait het lokaal op hun eigen PC. Hierdoor blijft data, zoals de *prompts* die zij invoeren, privé.

Hoewel uncensored AI-modellen in theorie alle vragen beantwoorden door het ontbreken van *guardrails*, ervaren deze jongeren dat de output vaak onbetrouwbaar is doordat de modellen sterk hallucineren. Zij leggen uit dat het verwijderen van de *guardrails* ertoe leidt dat delen van de onderliggende code van het model worden weggehaald. Respondent 8 stelt dat hierdoor ook stukjes van de “slimheid van het model” worden verwijderd, waardoor het minder accuraat antwoord. Volgens hem heeft een jongere die misbruik wil maken van GenAI dan ook weinig aan de antwoorden van een uncensored AI-model: “Dus ja als je vraagt hoe maak ik een atomische bom dan geeft hij wel antwoord en als je dan vraagt hé waar kan ik hem kopen? Ga naar atomischebom.nl, maar dat bestaat niet [lacht].”

Daarnaast vertelt respondent 8 wel eens zelf een uncensored AI-model te hebben gehost. Dit houdt in dat hij het model handmatig heeft opgezet en uitgevoerd op zijn eigen PC, waardoor hij eveneens beschikt over een taalmodel zonder *guardrails*. Volgens respondent 8 is het “verbazend hoe makkelijk het is” om dat zelf te doen. Tegelijkertijd benadrukken jongeren dat deze toegankelijkheid vooral geldt voor technisch vaardige gebruikers. Volgens hen moet iemand daarvoor namelijk beschikken over een krachtige PC en bovendien over voldoende technische kennis.

Concluderend blijkt dat de deelnemende jongeren in wisselende mate de ingebouwde veiligheidsmechanismen (*guardrails*) van GenAI verkennen. De meeste jongeren stuiten slechts incidenteel op deze *guardrails* en ondernemen geen verdere stappen wanneer hun vraag wordt geweigerd door een AI-taalmodel. Enkele ICT-vaardige jongeren geven aan te experimenteren met het omzeilen van *guardrails* of uncensored AI-modellen zonder *guardrails*. Echter stellen zij dat de output van uncensored AI-modellen vaak onbetrouwbaar is, waardoor zij het risico op daadwerkelijk misbruik door jongeren beperkt achten. Gezamenlijk laten deze bevindingen zien dat enkele jongeren in aanraking komen met risicovolle of strafbare toepassingen van GenAI, maar dat de mate waarin zij deze actief verkennen uiteenloopt en in de meeste gevallen wordt beperkt door *guardrails*.

5. Conclusie

In dit onderzoek is verkend wat de opkomst van generatieve kunstmatige intelligentie (GenAI) mogelijk betekent voor de betrokkenheid van jongeren bij risicovol of strafbaar gedrag. Om hiervan een beeld te schetsen, zijn in eerste aanleg vijf informanten gesproken over hun zicht op AI-gerelateerde criminaliteit en hun zorgen hierover in relatie tot jongeren. Vervolgens is tijdens drie focusgroepen met dertien jongeren besproken hoe zij GenAI gebruiken en in hoeverre zij risicovol of strafbaar gebruik van GenAI bij zichzelf of anderen ervaren. Over het geheel genomen laten de bevindingen zien dat GenAI vooral een nieuw risicodomein creëert waarin jongeren moeten navigeren tussen wat technisch mogelijk en maatschappelijk aanvaardbaar is. Tegelijkertijd lijkt de feitelijke betrokkenheid van de deelnemende jongeren bij risicovol of strafbaar gebruik van GenAI beperkt.

De bevindingen wijzen op vijf deelconclusies. Ten eerste blijkt het zicht op AI-gerelateerde criminaliteit volgens de informanten momenteel beperkt. Dat komt doordat AI-gebruik bij delicten niet systematisch wordt geregistreerd door de politie en de rol van AI niet altijd wordt herkend door politie of slachtoffers. Daarnaast zorgen meldingen waarin AI-betrokkenheid niet eenduidig kan worden vastgesteld en beperkingen in het opsporingsonderzoek ervoor dat bestaande registraties weinig inzicht bieden in de omvang van AI-gerelateerde criminaliteit. Door dit gebrek aan zicht blijft de daadwerkelijke omvang van AI-gerelateerde criminaliteit onduidelijk, evenals de mate waarin jongeren daarbij betrokken zijn.

Ten tweede signaleren de informanten ontwikkelingen die de kans op risicovol of strafbaar gedrag met GenAI door jongeren kunnen vergroten. Daarbij wordt voornamelijk gewezen op de brede beschikbaarheid en gebruiksvriendelijkheid van deepfake- en deepnudetools. Dit kan volgens de informanten de drempel voor jongeren verlagen om met dergelijke toepassingen te experimenteren. Daarbij speelt op de achtergrond dat zij meemaken dat jongeren het soms lastig vinden om te grens tussen speels en grensoverschrijdend gebruik van GenAI te herkennen.

Ten derde blijkt dat de deelnemende jongeren zelf normatieve afwegingen maken bij hun gebruik van GenAI. Voornamelijk bij het beoordelen van deepfakes nemen zij verschillende punten in acht, namelijk (1) de relatie tussen de maker van de deepfake en de afgebeelde persoon, (2) de intentie van de maker en (3) de mening van de afgebeelde persoon over de deepfake. Tegelijkertijd blijkt dat jongeren buiten de schoolcontext weinig concrete richtlijnen hebben voor GenAI-gebruik. Hierdoor moeten zij zelf interpreteren waar de grens tussen verantwoord en onverantwoord GenAI-gebruik ligt, wat kan leiden tot een grijs gebied waarin normen niet altijd duidelijk zijn.

Ten vierde komen de deelnemende jongeren in beperkte mate risicovol of strafbaar GenAI-gebruik tegen in hun eigen omgeving. De meeste jongeren benoemen slechts

incidentele of speculatieve voorbeelden. Daarentegen hebben enkele jongeren meegemaakt dat iemand uit hun omgeving slachtoffer is geworden van deepnudes, zoals een kennis of een docent.

Ten vijfde lijkt op basis van de bevindingen in het huidige onderzoek ook de feitelijke betrokkenheid van de deelnemende jongeren bij risicovol of strafbaar GenAI-gebruik beperkt. Een belangrijke verklaring hiervoor lijkt de aanwezigheid van ingebouwde veiligheidsmechanismen (*guardrails*) van GenAI. Verschillende jongeren geven aan dat deze ingebouwde veiligheidsmechanismen niet alleen bepaalde verzoeken blokkeren, maar hen ook helpen te herkennen wanneer een vraag of handeling mogelijk grensoverschrijdend is. Enkel een kleine groep technisch vaardige jongeren experimenteert bewust met het omzeilen van *guardrails*, zoals voor het illegaal downloaden van games of door het gebruik van uncensored AI-modellen zonder *guardrails*. Echter achten zij het misbruik daarvan door andere jongeren laag, vanwege de vereiste kennis om *guardrails* te omzeilen en de beperkte betrouwbaarheid van uncensored AI-modellen.

Gezamenlijk suggereren de bevindingen dat GenAI vooral een nieuw risicodomein creëert waarin jongeren moeten navigeren tussen wat technisch mogelijk en maatschappelijk aanvaardbaar is. Hoewel GenAI nieuwe mogelijkheden biedt voor normoverschrijdend gedrag, wijzen de resultaten van het huidige onderzoek niet op een substantiële betrokkenheid van de deelnemende jongeren bij risicovol of strafbaar gedrag waarbij GenAI een rol speelt.

6. Discussie

6.1 Bevindingen in perspectief

De bevindingen van dit onderzoek moeten worden geplaatst tegen de achtergrond van een nog beperkte wetenschappelijke basis over de rol van generatieve AI (GenAI) in het online daderschap van jongeren. Hoewel GenAI binnen de literatuur wordt benoemd als technologie die nieuwe mogelijkheden biedt voor risicovol of strafbaar gedrag (Blauth et al., 2022; Europol, 2025; Hayward & Maas, 2020), ontbreekt het tot op heden aan empirisch onderzoek dat de daadwerkelijke omvang hiervan in kaart brengt. De bevindingen van het huidige onderzoek laten eveneens zien dat het zicht op AI-gerelateerde criminaliteit beperkt is. De informanten geven aan dat AI-gebruik bij delicten niet systematisch wordt vastgelegd door de politie en wellicht niet altijd wordt herkend. Daarnaast spelen meldingen die enkel gebaseerd zijn op vermoedens van AI-gebruik en juridische obstakels een rol bij dit beperkte zicht. Het huidige onderzoek vormt daarmee een aanvulling op de bestaande literatuur, door inzicht te bieden in enkele onderliggende factoren die het zicht op AI-gerelateerde criminaliteit in de praktijk bemoeilijken.

Daarnaast sluiten de bevindingen aan bij de literatuur die GenAI beschrijft als technologie die bestaande delicten kan versterken, automatiseren of opschalen doordat de benodigde technische kennis van verschillende delicten afneemt (Europol, 2025; Kopecký & Voráč, 2025; Nationaal Coördinator Terrorismebestrijding en Veiligheid, 2024; Schuilenburg & Soudijn, 2024). De resultaten laten zien dat deze drempelverlaging niet enkel theoretisch is, maar zich in de praktijk ook manifesteert in de leefwereld van jongeren. Zo blijkt uit de resultaten dat deepfake- en deepnude-tools laagdrempelig beschikbaar zijn via app-stores en communicatieplatformen, en in sommige gevallen volgens de informanten zelfs actief worden gepromoot richting de doelgroep. Waar de literatuur dus vooral waarschuwt voor deze risico's, blijkt uit de bevindingen dat jongeren daadwerkelijk in aanraking komen met deze toepassingen.

Verder kunnen de resultaten van dit onderzoek worden geplaatst binnen het affordance-perspectief van Goldsmith & Wall (2019), maar laten daarbij zowel bevestiging als nuancering zien. Dit perspectief stelt dat digitale technologieën specifieke handelingsmogelijkheden (*affordances*) bieden die risicovol online gedrag aantrekkelijker, toegankelijker of eenvoudiger kunnen maken voor jongeren. Binnen dat perspectief zou GenAI kunnen fungeren als een faciliterende of 'verleidende' omgeving, die risicovol gedrag bij jongeren waarschijnlijker maakt. De toegankelijkheid (*accessibility*) van GenAI, de anonimiteit (*anonymity*) waarmee jongeren met GenAI-tools kunnen experimenteren en de losser ervaren gedragsnormen daarbij (*ambivalence*) kunnen verklaren waarom zij eerder risicovol of strafbaar gedrag vertonen tijdens hun gebruik van GenAI.

De resultaten sluiten deels aan bij deze theoretische aannames. Enerzijds stellen enkele deelnemende jongeren uit nieuwsgierigheid risicovolle vragen te stellen aan GenAI, waarbij zij zich pas door de werking van *guardrails* (ingebouwde ethische beperkingen van AI-taalmodellen) beseffen dat de vraag buiten de context argwaan kan opwekken. Dit kan een aanwijzing zijn voor de losser beschouwde gedragsnormen en anonimiteit die jongeren ervaren in interactie met GenAI, waardoor zij sneller risicovolle vragen stellen aan een AI-taalmodel. Ook de toegankelijkheid van tools die risicovol of strafbaar GenAI-gebruik stimuleren, zoals deepnude-tools, blijken onderdeel te zijn van de leefwereld van de deelnemende jongeren.

Anderzijds laten de resultaten zien dat de deelnemende jongeren zelf normatieve afwegingen maken bij het maken van deepfakes. Hierbij nemen zij voornamelijk drie factoren in acht, namelijk (1) de relatie tussen de maker van de deepfake en de afgebeelde persoon, (2) de intentie van de maker en (3) de mening van de afgebeelde persoon over de deepfake. Daarmee bieden de bevindingen een aanvulling op het affordance-perspectief. Waar Goldsmith & Wall (2019) beschrijven hoe technologische *affordances* de drempel tot normoverschrijdend gedrag kunnen verlagen, laten de resultaten zien dat de deelnemende jongeren ondanks deze verlaagde drempel zelf afwegingen maken over wat zij al dan niet acceptabel vinden bij het gebruik van GenAI. Tegelijkertijd illustreren de uiteenlopende opvattingen van de deelnemende jongeren over het maken van deepfakes dat dergelijke afwegingen niet voor elke jongere op dezelfde manier worden gemaakt. Deze bevindingen suggereren dat *affordances* van GenAI niet direct leiden tot gedrag, maar bestaan naast de normatieve afwegingen van jongeren.

Een aanvullende observatie betreft de rol van ingebouwde veiligheidsmechanismen (*guardrails*) bij AI-taalmodellen. Verschillende jongeren gaven aan dat zij door een afwijzing door deze *guardrails* pas beseften dat hun verzoek of vraag aan een AI-taalmodel mogelijk grensoverschrijdend is. Dit suggereert dat de deelnemende jongeren grenzen bij AI-gebruik niet altijd vooraf herkennen, maar deze soms pas tijdens hun interactie met GenAI ontdekken. Deze bevindingen nuanceren het uitgangspunt van het affordance-perspectief, waarin technologie primair wordt beschouwd als bron van nieuwe mogelijkheden voor risicovol gedrag. GenAI blijkt namelijk niet alleen dergelijke handelingsmogelijkheden te creëren, maar kan door de werking van *guardrails* juist ook een begrenzende rol spelen bij potentieel risicovol of strafbaar AI-gebruik door jongeren.

6.2 Sterke punten en beperkingen

Het huidige onderzoek levert waardevolle inzichten over een onderwerp dat zich nog in een vroeg stadium van wetenschappelijke ontwikkeling bevindt. Doordat het onderzoeksveld nog in ontwikkeling is, bood dit ruimte om het onderzoek open en exploratief vorm te geven zonder vooraf sterk afgebakende verwachtingen. Deze benadering sluit aan bij de uitgangspunten van kwalitatief onderzoek, waarin ruimte voor onverwachte inzichten centraal staat (Hennink et al., 2020). Door deze open houding konden nieuwe perspectieven en nuances worden blootgelegd die anders mogelijk buiten beeld zouden zijn gebleven.

Tegelijkertijd kent het onderzoek enkele beperkingen die de interpretatie van de resultaten beïnvloeden. Ten eerste zijn de bevindingen niet generaliseerbaar naar alle Nederlandse jongeren, aangezien dit een kleine steekproef betreft van jongeren die bovendien bereid waren deel te nemen. Hierdoor kan sprake zijn van selectiebias, waarbij jongeren met sterkere meningen over of meer kennis van GenAI oververtegenwoordigd kunnen zijn. Deze beperking heeft gevolgen voor de externe validiteit van het onderzoek.

Daarnaast zijn de bevindingen met betrekking tot de jongeren gebaseerd op zelfrapportage van hun gedrag, waardoor sociaal-wenselijke antwoorden niet kunnen worden uitgesloten. Zo bestaat de kans dat zij hun eigen gedrag onderrapporteren of hun kritische houding ten opzichte van verantwoord GenAI-gebruik overschatten. Deze afhankelijkheid van zelfrapportage van gedrag kan daardoor de constructvaliditeit beïnvloeden.

Verder kan de groepssetting waarin de gesprekken met hen plaatsvonden van invloed zijn geweest op de resultaten. Enerzijds kan dit ertoe hebben geleid dat deelnemers zich konden herkennen in de ervaringen van anderen, wat mogelijk heeft bijgedragen aan rijkere data dan individuele interviews zouden hebben opgeleverd. Anderzijds kan de groepssetting van invloed zijn geweest op de mate waarin de jongeren bereid waren om hun eigen ervaringen met risicovol of strafbaar gebruik van GenAI te delen.

Tot slot zijn de resultaten van het huidige onderzoek sterk tijdgebonden zijn. GenAI-technologieën, de toegankelijkheid daarvan en wetgeving daaromheen zijn tijdens de uitvoering van dit onderzoek nog volop in ontwikkeling. Hierdoor zijn de percepties en ervaringen van zowel de informanten als de jongeren hierover eveneens dynamisch. De bevindingen moeten daarom worden beschouwd als een momentopname binnen de huidige technologische en maatschappelijke context, wat de generaliseerbaarheid naar toekomstige situaties beperkt.

6.3 Aanbevelingen voor vervolgonderzoek en praktijk

Het huidige onderzoek biedt enkele aanknopingspunten voor vervolgonderzoek. Ten eerste is het, gezien de kleine steekproef binnen dit onderzoek, waardevol om grootschaliger kwantitatief onderzoek uit te voeren naar de prevalentie van risicovol of strafbaar gebruik van GenAI onder jongeren. Dit kan bijvoorbeeld worden benaderd vanuit een affordance-perspectief, waarbij wordt onderzocht in welke mate de toegankelijkheid en gebruiksvriendelijkheid van GenAI samenhangen met verschillende vormen van grensoverschrijdend gebruik. Daarmee kan een beeld worden gevormd van de mate waarin dit gedrag daadwerkelijk voorkomt onder jongeren.

Ten tweede komen deepnudes in dit onderzoek naar voren als de meest toegankelijke strafbare toepassing van GenAI onder jongeren. Opvallend is dat zij deepnudes unaniem afkeuren, maar dat er ook situaties naar voren komen waarin dergelijke beelden aanvankelijk als grappig worden beschouwd. Toekomstig onderzoek kan zich daarom richten op de sociale mechanismen die deze opvattingen van jongeren beïnvloeden. Wellicht spelen groepsdynamiek, normaliserende humor en de onduidelijke grens tussen speels en schadelijk gebruik hierbij een rol.

Naast deze punten uitten de deelnemende jongeren tijdens de gesprekken hun behoefte aan concrete en toegankelijke richtlijnen voor verantwoord GenAI-gebruik voor jongeren. Een aanvullende aanbeveling is daarom om richtlijnen te ontwikkelen die jongeren houvast bieden voor verantwoord gebruik van GenAI. Verder blijkt uit de gesprekken dat het voor de deelnemende jongeren niet altijd duidelijk is wanneer GenAI-gebruik strafbaar is. Daarom zou in deze richtlijnen ook duidelijk moeten worden opgenomen welke vormen van GenAI-gebruik altijd strafbaar zijn, zoals het maken van deepnudes.

Literatuurlijst

Abdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N. A., Abazid, H., Malaeb, D., Mohammed, A. H., Hassan, B. A. R., Wayyes, A. M., Farhan, S. S., Khatib, S. E., Rahal, M., Sahban, A., Abdelaziz, D. H., Mansour, N. O., AlZayer, R., Khalil, R., Fekih Romdhane, F., Hallit, R., . . . Sallam, M. (2024). A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Scientific Reports, 14*(1983), 1-14. https://doi.org/10.1038/s41598-024-52549-8

Ahmad, R., & Thurasamy, R. (2022). A systematic literature review of routine activity theory's applicability in cybercrimes. *Journal Of Cyber Security And Mobility, 11*(3), 405-432. https://doi.org/10.13052/jcsm2245-1439.1133

Akgül, G. (2021). Routine activities theory in cyber victimization and cyberbullying experiences of Turkish adolescents. *International Journal Of School & Educational Psychology, 11*(2), 135-144. https://doi.org/10.1080/21683603.2021.1980475

Anthropic. (2025). *Threat Intelligence Report: August 2025*. https://www.anthropic.com/news/detecting-countering-misuse-aug-2025

Beazley, J., & Touma, R. (2024, 12 juni). Bacchus Marsh Grammar: schoolboy arrested after 50 female students allegedly targeted in fake explicit AI photos scandal. *The Guardian*. https://www.theguardian.com/australia-news/article/2024/jun/12/schoolboy-arrested-after-allegedly-posting-fake-explicit-images-of-female-students-ntwnfb

Blauth, T. F., Gstrein, O. J., & Zwitter, A. (2022). Artificial intelligence crime: An overview of malicious use and abuse of AI. *IEEE Access*, *10*, 77110-77122. https://doi.org/10.1109/access.2022.3191790

Bossler, A. M., Holt, T. J., & May, D. C. (2011). Predicting online harassment victimization among a juvenile population. *Youth & Society, 44*(4), 500–523. https://doi.org/10.1177/0044118x11407525

Brandtzaeg, P. B., Følstad, A., & Skjuve, M. (2025). Emerging AI individualism: How young people integrate social AI into everyday life. *Communication And Change, 1*(11), 1-22. https://doi.org/10.1007/s44382-025-00011-2

Brewer, R., Cale, J., Goldsmith, A., & Holt, T. (2018). Young people, the internet, and emerging pathways into criminality: A study of Australian adolescents. *Adelaide Research & Scholarship, 12*(1), 115–132. https://doi.org/10.5281/zenodo.1467853

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., Héigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., . . . Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. arXiv. https://doi.org/10.48550/arxiv.1802.07228

Campedelli, G. M. (2025). *A Criminology of machines*. arXiv. https://doi.org/10.48550/arxiv.2511.02895

Centraal Bureau voor de Statistiek. (2024). *Bijna kwart Nederlanders gebruikt kunstmatige intelligentie zoals ChatGPT.* https://www.cbs.nl/nl-nl/nieuws/2024/36/bijna-kwart-nederlanders-gebruikt-kunstmatige-intelligentie-zoals-chatgpt

Centraal Bureau voor de Statistiek. (2025). *Online veiligheid en criminaliteit 2024.* https://www.cbs.nl/nl-nl/longread/rapportages/2025/online-veiligheid-en-criminaliteit-2024?onepage=true

Chadha, A., Kumar, V., Kashyap, S., & Gupta, M. (2021). Deepfake: An overview. In *Lecture notes in networks and systems* (pp. 557–566). https://doi.org/10.1007/978-981-16-0733-2_39

Chan, T. K. H., Cheung, C. M. K., & Wong, R. Y. M. (2019). Cyberbullying on social networking sites: The crime opportunity and affordance perspectives. *Journal Of Management Information Systems*, *36*(2), 574-609. https://doi.org/10.1080/07421222.2019.1599500

Chaudhuri, A. (2024). The psychosocial reasons for the surge of chatbot usage among young adults: A review. *International Journal of Interdisciplinary Approaches in Psychology, 2*(9), 1-18. https://www.psychopediajournals.com/index.php/ijiap/article/view/534

Ciubotaru, B.-I. (2025). The hallucination problem in generative artificial intelligence: Accuracy and trust in digital learning. *Proceedings of the International Conference on Virtual Learning, 20*, 35-45. https://doi.org/10.58503/icvl-v20y202503

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44*(4), 589-591. https://doi.org/10.2307/2094589

Decorte, T. & Zaitch, D. (2016). *Kwalitatieve methoden en technieken in de criminologie*. Acco.

De Jong, J. D., & De Wit, N. (2024). Plus-preventie voor de 2 procent. *Vakblad Sociaal Werk*, *25*(1), 5-9. https://doi.org/10.1007/s12459-024-1962-5

Dikilitas, K., Klippen, M. I. F., & Keles, S. (2024). A systematic rapid review of empirical research on students' use of ChatGPT in higher education. *Nordic Journal of Systematic Reviews in Education, 2*(1), 103-125. https://doi.org/10.23865/njsre.v2.6227

Du, Y. (2024). The impact of artificial intelligence on people's daily life. *The Frontiers of Society, Science and Technology*, *6*(6), 12-18. https://doi.org/10.25236/FSST.2024.060603

Dzuba, C. (2025). Artificial intelligence in social engineering: A literature review through the lens of routine activity theory. *Issues in Information Systems, 26*(2), 452-465. https://doi.org/10.48009/2_iis_2025_134

Europol. (2025). *European Union Serious and Organised Crime Threat Assessment 2025: The changing DNA of serious and organised crime*. Publications Office of the European Union. https://www.europol.europa.eu/publications-events/main-reports/socta-report

Evers, J. (2015). *Kwalitatief interviewen: Kunst én kunde* (2e druk). Boom Lemma uitgevers.

Falade, P. V. (2023). *Decoding the threat landscape: ChatGPT, FraudGPT, and WormGPT in social engineering attacks*. arXiv. https://doi.org/10.48550/arxiv.2310.05595

Fire, M., Elbazis, Y., Wasenstein, A., & Rokach, L. (2025). *Dark LLMs: The growing threat of unaligned AI models*. arXiv. https://doi.org/10.48550/arXiv.2505.10066

Fischer, T., & Julsing, M. (2023). *Onderzoek doen!: Kwantitatief en kwalitatief onderzoek* (4e druk). Noordhoff uitgevers.

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 1-28. https://doi.org/10.3390/soc15010006

Goldsmith, A., & Wall, D. S. (2019). The seductions of cybercrime: Adolescence and the thrills of digital transgression. *European Journal Of Criminology, 19*(1), 98-117. https://doi.org/10.1177/1477370819887305

Guembe, B., Azeta, A., Misra, S., Osamor, V. C., Fernandez-Sanz, L., & Pospelova, V. (2022). The emerging threat of AI-driven cyber attacks: A review. *Applied Artificial Intelligence, 36*(1), 1-34. https://doi.org/10.1080/08839514.2022.2037254

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. *IEEE Access, 11*(1), 80218–80245.https://doi.org/10.1109/access.2023.3300381

Hartmann, D., Wang, S. M., Pohlmann, L., & Berendt, B. (2025). A systematic review of echo chamber research: Comparative analysis of conceptualizations, operationalizations, and varying outcomes. *Journal Of Computational Social Science*, *8*(52), 1-59. https://doi.org/10.1007/s42001-025-00381-z

Hayward, K. J., & Maas, M. M. (2020). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture: An International Journal, 17*(2), 209-233. https://doi.org/10.1177/1741659020917434

Hennink, M., Hutter, I., & Bailey, A. (2020). *Qualitative Research Methods*. Sage publications.

Jeong, D. (2020). Artificial intelligence security threat, crime, and forensics: Taxonomy and open issues. *IEEE Access, 8*(1), 184560-184574. https://doi.org/10.1109/access.2020.3029280

Jones, S. (2024, 9 juli). Spain sentences 15 schoolchildren over AI-generated naked images. *The Guardian*. https://www.theguardian.com/world/article/2024/jul/09/spain-sentences-15-school-children-over-ai-generated-naked-images

Kanne, P., Nannes, L. & Andringa, W. (2019). *Opvoeden. De balans tussen vrijheid geven en verantwoordelijkheid nemen.* I&O Research.

Kopecký, K., & Voráč, D. (2025). The phenomenon of deep nudes – a new threat to children and adults. *AI & Society, 41*(1), 545-556. https://doi.org/10.1007/s00146-025-02425-4

Liu, Y., & Wang, H. (2025). Who on earth is using generative AI? *World Development, 199*(1),1-46. https://doi.org/10.1016/j.worlddev.2025.107260

Manky, D. (2013). Cybercrime as a service: A very modern business. *Computer Fraud & Security*, *2013*(6), 9-13. https://doi.org/10.1016/s1361-3723(13)70053-8

Matthijsse, M. (2023). *Interviewen in kwalitatief onderzoek*. Uitgeverij SWP.

Nationaal Coördinator Terrorismebestrijding en Veiligheid (2024, 9 december). *Versterkte dreigingen in een wereld vol kunstmatige intelligentie.* https://www.nctv.nl/documenten/2024/12/09/versterkte-dreigingen-in-een-wereld-vol-kunstmatige-intelligentie

Nerantzi, E., & Sartor, G. (2024). 'Hard AI Crime': The deterrence turn. *Oxford Journal of Legal Studies, 44*(3), 673-701. https://doi.org/10.1093/ojls/gqae018

Newburn, T. (2017). *Criminology*. Routledge.

NOS. (2025, 6 oktober). Politie waarschuwt ouders voor geintje met AI-zwerver. *NOS Nieuws.* https://nos.nl/artikel/2585443-politie-waarschuwt-ouders-voor-geintje-met-ai-zwerver

Offlimits. (2025). *Jaarverslag 2025*. Offlimits. https://data.maglr.com//3878/issues/66984/788967/downloads/offlimits-jaarverslag2025.pdf

Offlimits. (2026). *Terminologiegids 2026*. Offlimits. https://offlimits.nl/assets/offlimits_nl/downloadable_files/terminologiegids26digi.v4.pdf

Openbaar Ministerie. (2024, 11 juni). *Samenvatting CCBN 2024 Nederlands.* https://www.om.nl/documenten/cybercrime/2024/ccbn/samenvatting-ccbn-nederlands

Park, S. & Nan, X. (2025). Generative AI and misinformation: A scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation. *AI & Society*, *41*(2), 1501-1515. https://doi.org/10.1007/s00146-025-02620-3

Protect Children. (2026). *CSAM perpetrator research report: Findings from a survey of CSAM perpetrators on digital platform use and design.* https://www.protectchildren.fi/en/post/tmat-csam-perpetrator-research-report

Reyns, B. W., Henson, B., & Fisher, B. S. (2015). Guardians of the cyber galaxy: An empirical and theoretical analysis of the guardianship concept from routine activity

theory as it applies to online forms of victimization. *Journal Of Contemporary Criminal Justice, 32*(2), 148-168. https://doi.org/10.1177/1043986215621378

Roks, R., A., Leukfeldt, E. R., & Densley, J. A. (2021). The hybridization of street offending in the Netherlands. *British Journal of Criminology, 61*(4), 926-945. https://doi.org/10.1093/bjc/azaa091

Romer, D., Duckworth, A. L., Sznitman, S., & Park, S. (2010). Can adolescents learn self-control? Delay of gratification in the development of control over risk taking. *Prevention Science*, *11*(3), 319-330. https://doi.org/10.1007/s11121-010-0171-8

Sah, R., Hagemaster, C., Adhikari, A., Lee, A., Sun, N. (2025). Generative AI in higher education: Student and faculty perspectives on use, ethics, and impact. *Issues in Information Systems, 26*(2), 373-386. https://doi.org/10.48009/2_iis_129

Schmitt, M., & Flechais, I. (2023). Digital deception: Generative artificial intelligence in social engineering and phishing. *Artificial Intelligence Review, 57*(12), 1-18. https://doi.org/10.1007/s10462-024-10973-2

Schuilenburg, M., & Soudijn, M. (2024). AI-criminaliteit: Een verkenning van actuele verschijningsvormen. *Justitiële Verkenningen, 50*(1), 12-29. https://doi.org/10.5553/jv/016758502024050001002

Scott, R. A., Stuart, J., & Barber, B. L. (2022). Connecting with close friends online: A qualitative analysis of young adults' perceptions of online and offline social interactions with friends. *Computers in Human Behavior Reports*, *7*, 1-11. https://doi.org/10.1016/j.chbr.2022.100217

Sharp, J. (2024, 8 maart). 5 Beverly Hills middle school students expelled following AI generated nude photo scandal. *CBS News.* https://www.cbsnews.com/news/5-beverly-hills-middle-school-students-expelled-following-ai-generated-nude-photo-scandal/?linkId=353624654

Spradley, J. P. (1980). *Participant observation*. Cengage learning.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*(3), 321-326. https://doi.org/10.1089/1094931041291295

Swanborn, P. G. (2015). *Basisboek sociaal onderzoek* (6e druk). Boom Lemma uitgevers.

Szyf, N. D., Gilen, A., & Giacometti, M. (2023). *Deepnudes among young people in Belgium: The numbers, the market, the impact*. Instituut voor de Gelijkheid van Vrouwen en Mannen (IGVM/IEFH). https://igvm-iefh.belgium.be/sites/default/files/deepnudes_among_young_people_in_belgium.pdf

Tanrikulu, I. (2019). How do school children learn cyberbullying perpetration? *Journal of Theoretical Educational Sciences, 12*(1), 16-27. https://doi.org/10.30831/akukeg.512556

Thanh, C. T., & Zelinka, I. (2019). A survey on artificial intelligence in malware as next-generation threats. *MENDEL, 25*(2), 27-34. https://doi.org/10.13164/mendel.2019.2.027

Tollenaar, N., Beijers, J., & Van der Laan, A. M. (2024). *Monitor zelfgerapporteerde jeugddelinquentie 2023*. Wetenschappelijk Onderzoek- en Datacentrum. https://repository.wodc.nl/bitstream/handle/20.500.12832/3385/Cahier-2024-14-volledige-tekst.pdf?sequence=9&isAllowed=y

Trieu, K., & Yang, Y. (2018). Artificial intelligence-based password brute force attacks. *Proceedings of the 13th Midwest Association for Information Systems Conference (MWAIS 2018).* AIS Electronic Library (AISeL). https://aisel.aisnet.org/mwais2018/39/

Vakhitova, Z. I., Go, A., & Alston-Knox, C. L. (2021). Guardians against cyber abuse: Who are they and why do they intervene? *American Journal Of Criminal Justice, 48*(1), 96 -122. https://doi.org/10.1007/s12103-021-09641-w

Vale, M., Pereira, F., Spitzberg, B. H., & Matos, M. (2022). Cyber-harassment victimization of Portuguese adolescents: A lifestyle-routine activities theory approach. *Behavioral Sciences & The Law, 40*(5), 604-618. https://doi.org/10.1002/bsl.2596

Van Der Meijs, F. (z.d.). *Feiten en cijfers over schermgebruik*. Trimbos-instituut. https://www.trimbos.nl/kennis/digitale-media-gokken/expertisecentrum-digitalisering-en-welzijn/feiten-en-cijfers-digitale-media/schermgebruik/

Van Der Wagen, W., Oerlemans, J., & Kranenbarg, M. W. (2020). *Basisboek Cybercriminaliteit: Een criminologisch overzicht voor studie en praktijk*. Boom criminologie.

Verma, S., & Lall, A. J. (2025). Youth in the age of AI: Examining the role of digitalization in shaping daily life. *International Journal For Research in Applied Science And Engineering Technology*, *13*(4), 350–354. https://doi.org/10.22214/ijraset.2025.68289

Weale, S. (2025, 2 december). The rise of deepfake pornography in schools: ‘One girl was so horrified she vomited’. *The Guardian.* https://www.theguardian.com/society/ng-interactive/2025/dec/02/the-rise-of-deepfake-pornography-in-schools

Weerman, F. (2019). Criminaliteit, digitalisering en de online sociale wereld: Dezelfde processen in een nieuwe sociale context? *Tijdschrift Voor Criminologie, 61*(4), 395-404.https://doi.org/10.5553/tvc/0165182x2019061004008

Weijers, I. (2024). Een halve eeuw jeugdcriminaliteit. *Justitiële Verkenningen, 50*(4), 80-93. https://doi.org/10.5553/JV/016758502024050004006

Weulen Kranenbarg, M., & van 't Hoff-de Goede, S. (2023). Online criminaliteit in criminologisch perspectief: Recente ontwikkelingen in het onderzoek naar daders en

slachtoffers van online criminaliteit. *Tijdschrift voor Criminologie, 65*(4), 400-419. https://doi.org/10.5553/TvC/0165182X2023065004002

Weulen Kranenbarg, M., & Weerman, F. (2022). Online jeugdcriminaliteit in coronatijd: Ontwikkelingen in zelfgerapporteerd daderschap, activiteiten en contacten met vrienden bij een hoogrisicogroep. *Tijdschrift voor Criminologie, 64*(4), 396-426. https://doi.org/10.5553/tvc/0165182x2022064004004

Wissink, I. B., Asscher, J. J., & Stams, G. (2023). Online delinquent behaviors of adolescents: Parents as potential “influencers”? *International Journal Of Offender Therapy And Comparative Criminology, 69*(8), 898-920. https://doi.org/10.1177/0306624x231206521

Yu, Y., Liu, Y., Zhang, J., Qian, Z., Yun, H., & Yang, W. (2025). Understanding generative AI risks for youth: An empirical taxonomy for safer digital futures. *Proceedings Of The Association For Information Science And Technology, 62*(1), 1-17. https://doi.org/10.1002/pra2.1526

Bijlagen

Bijlage A: Wervingsstrategie informanten en jongeren

Datum	Persoon/organisatie	Actie	Opgeleverde informatie
20-11-2025	Gedrags- en systeemonderzoeker TNO	Benaderd voor mogelijke deelname onderzoek.	Gaf aan geen relevante kennis over het onderwerp te hebben.
25-11-2025	Diverse personen waaronder van de politie en het RIEC.	Deelgenomen aan overleg van het Netwerk Jonge Aanwas. Hierin verteld over het onderzoek en oproep gedaan voor informanten.	-
1-12-2025	Netwerk Jonge Aanwas	Oproep geplaatst in het netwerk Jonge Aanwas op het intranet van de politie.	-
2-12-2025	Jongerenwerk Utrecht	Benaderd voor werving van jongeren voor het onderzoek.	Overleg gepland.
3-12-2025	Henk van Ee, stichting Cyberbrein	Benaderd voor mogelijke deelname onderzoek.	Interview gepland op 27-2-2026.
3-12-2025	Thomas Vermeij, stichting Epic Youth	Benaderd voor mogelijke deelname onderzoek.	Interview gepland op 3-3-2026.
3-12-2025	Onderzoeker TNO	Benaderd voor mogelijke deelname onderzoek.	-
3-12-2025	Manon den Dunnen, Strategisch Specialist Digitaal politie	Benaderd voor mogelijke deelname onderzoek.	Interview gepland op 24-2-2026.
10-12-2025	Jongerenwerk Utrecht	Overleg over mogelijke deelname jongeren onderzoek.	Vraag uitgezet onder jongerenwerkers. Geen resultaat.
10-12-2025	Keerpunt	Benaderd voor werving van jongeren voor het onderzoek.	Vraag uitgezet bij collega's maar geen resultaat.
16-12-2025	Online jongerenwerker	Benaderd voor werving van jongeren voor het onderzoek en afnemen interview.	-
18-12-2025	Halt	Benaderd voor werving van jongeren voor het onderzoek.	Vraag uitgezet bij enkele jongeren. Uiteindelijk leverde dit drie respondenten op.
13-1-2026	Rutger Leukfeldt, onderzoeker digitalisering en cybercrime NSCR	Benaderd voor mogelijke deelname onderzoek.	Interview gepland op 5-3-2025.
21-1-2026	Halt	Overleg over afnemen focusgroep met jongeren.	

22-1-2026	Joost Willemsen, politie	Belafspraak over AI-intelbeeld dat hij heeft opgesteld over AI-criminaliteit voor de politie.	Artificial intelligence Intelbeeld
17-3-2026	Halt	Laatste afstemming focusgroep.	Toestemmingsformulieren verstuurd naar ouders van kinderen <16 jaar.
1-4-2026	Mbo-school	Benaderd voor werving van jongeren voor het onderzoek.	Vraag uitgezet onder jongeren van twee ict-opleidingen. Vervolgens twee focusgroepen gepland op 10-4-2026.

Bijlage B: Logboek aanvullende informatieverzameling

Datum	Persoon/organisatie	Activiteit	Samenvatting veldnotitie
13-11-2025	Henrique van Huisstede, team cybercrime politie Midden-Nederland	Online voorgesprek over de impact van AI op criminaliteit.	- AI kan op verschillende manieren worden ingezet in modus operandi daders. - AI-systemen zijn kwetsbaar voor 'jailbreaking': technieken om ethische barrières te omzeilen. - Er bestaan dark Large Language Models die deze ethische barrières niet hebben. - AI-agents kunnen autonoom handelen en in de toekomst onderling communiceren.
20-1-2026	Netwerk Digitale Wijkagenten politie	Oproep geplaatst in netwerk.	Twee reacties van digitaal wijkagenten: - Wijkagent 1: "Vanuit mijn functie als digitaal wijkagent in Lelystad-Zeewolde signaleer ik geen AI gebruik door jongeren met een crimineel oogmerk.". Heeft dit ook besproken met jeugdagent, die tot dezelfde conclusie kwam. - Wijkagent 2 heeft situatie meegemaakt waarin minderjarige vrouw slachtoffer was geworden van verspreiden van deepnudes.
5-2-2026	Team cybercrime politie Midden-Nederland & Europol Innovation Lab	Online meeting naar aanleiding van EMPACT-bijeenkomst (European Multidisciplinary Platform Against Criminal Threats).	- Bij Europol Innovation Lab nog weinig specifieke aandacht voor AI-gebruik door jongeren. - Krijgen steeds meer jongeren in beeld die online extremistisch gedachtengoed uiten. Denken aan risico's voor jongeren in richting van ingezogen raken in eigen denkbeelden en foutieve informatieverspreiding met AI.
2-4-2026	Het Regionaal Samenwerkingsverband Integrale Veiligheid (RSIV)	Kennisbijeenkomst 'AI in het veiligheidsdomein'. Sprekers:	- AI speelt steeds grotere rol bij alle drie de onderdelen van Offlimits: meldpunt, hulplijn en preventielijn.

		- Kees Verhoeven - DataExpert - Burgemeester Tjarda Struik en Offlimits	- 50% stijging in meldingen van AI-gegenereerd beeldmateriaal: 1472 (2024) naar 2214 (2025). - Offlimits signaleert dat steeds vaker AI wordt gebruikt voor sextortion. Dit is de grootste categorie van het totale aantal meldingen. - Steeds meer tieners zijn slachtoffer van of gebruiken uitkleedapps. - Rapport uitgebracht over nudify-websites. - Aanpak: No To Nudify. Er is op Europees niveau een verbod op nudify-tools aanstaande. - Offlimits zag met komst van Grok een grote stijging in het aantal tieners dat slachtoffer was van deepnudes. Spande daarom een kort geding aan tegen Grok en heeft gelijk gekregen van de rechter. Betekent dat rechter Grok en X verbiedt om uitkleedbeelden en CSAM te genereren en verspreiden (ECLI:RBAMS:2026:3106).

Bijlage C: Toestemmingsformulieren

Bijlage C1: Toestemmingsformulier informant

Mijn naam is Senna en momenteel ben ik bezig met mijn masterscriptie naar de invloed van kunstmatige intelligentie (AI) op jeugdcriminaliteit. In dit onderzoek richt ik mij op hoe jongeren omgaan met AI en welke rol het bij hen speelt in relatie tot criminaliteit of daaraan grenzend gedrag.

Deelname aan dit onderzoek bestaat uit een interview. Aan je deelname aan dit onderzoek zijn geen bijzondere risico's verbonden. De verzamelende gegevens worden uitsluitend gebruikt voor dit onderzoek en voor wetenschappelijke doeleinden.

Toestemmingsverklaring:

- Ik heb genoeg informatie gekregen over het interview en het onderzoek;
- Ik ben in de gelegenheid gesteld om vragen te stellen over het onderzoek en mijn deelname;
- Ik neem vrijwillig deel aan het onderzoek;
- Ik weet dat ik mijn deelname op elk moment kan beëindigen zonder hiervoor een reden op te geven;
- Ik geef toestemming voor het maken van een geluidsopname van het interview voor de analyse van het onderzoek. Deze geluidsopname zal worden getranscribeerd en uitsluitend door de onderzoeker worden gebruikt. Na afronding van het onderzoek zal deze opname worden verwijderd;
- Ik geef toestemming dat de onderzoeker de resultaten mag gebruiken voor het onderzoek en voor wetenschappelijke doeleinden.

Ik stem toe met deelname aan het onderzoek.

Naam deelnemer:……

Handtekening: ……

Datum: ……

Bijlage C2: Toestemmingsformulier jongeren

Het onderzoek
Het doel van dit onderzoek is om meer informatie te verzamelen over hoe jongeren generatieve AI gebruiken en de rol die AI kan spelen bij risicovol of strafbaar gedrag. Ik ben benieuwd naar jouw mening en ervaringen en wil het daar graag met je over hebben tijdens deze bijeenkomst.

Wat vraag ik van jou?
Ik ben benieuwd hoe jij over dit onderwerp denkt. Tijdens deze bijeenkomst gaan we samen praten over de volgende dingen: hoe gebruik je zelf AI? Wat zie je om je heen en online gebeuren met AI door jongeren? Wat vind je dat je wel/niet mag doen met AI? Welke risico's zitten er aan het gebruik van AI door jongeren?

Wat gebeurt er met de gegevens?
Tijdens de bijeenkomst wordt een audio-opname gemaakt. Die wordt opgeslagen op een beveiligde server en verwijderd zodra het onderzoek klaar is. Alleen ik (Senna) heb toegang tot de opname. De antwoorden die je geeft, worden alleen gebruikt voor dit onderzoek. Je blijft anoniem: alle informatie die mogelijk naar jou te herleiden is, laat ik weg. Er zitten geen bijzondere risico's aan je deelname aan dit onderzoek.

Toestemmingsverklaring:
- Ik snap wat meedoen aan het onderzoek inhoudt;
- Ik heb vragen kunnen stellen over het onderzoek en nagedacht of ik wil meedoen;
- Als ik wil stoppen met het onderzoek, dan kan dat zonder dat ik hoef te zeggen waarom;
- Ik begrijp dat mijn deelname anoniem is, zodat niemand erachter kan komen welke antwoorden ik heb gegeven.

Ik doe mee aan het onderzoek:
- ☐ Ja
- ☐ Nee

Bijlage C3: Toestemmingsformulier ouders/verzorgers

Veiligheidsinstanties zien dat generatieve kunstmatige intelligentie (GenAI) wordt gebruikt bij het uitvoeren van verschillende criminele activiteiten. Er is nog weinig duidelijk over hoe dit mogelijk speelt in de (online) omgeving van jongeren. Het RIEC Midden-Nederland is daar benieuwd naar. Daarom wil ik (Senna) het daar graag over hebben met [redacted] op **donderdag 26 maart**.

Het onderzoek
Het doel van dit onderzoek is om meer informatie te verzamelen over hoe jongeren generatieve AI gebruiken en wat zij hier in hun (online) omgeving mogelijk over zien in relatie tot risicovol of strafbaar gedrag. Dit onderzoek voer ik uit in opdracht van het RIEC Midden-Nederland en voor mijn afstudeerscriptie aan de Vrije Universiteit Amsterdam.

Wat vragen we van u?
Ik wil de jongeren, waaronder uw kind, graag een aantal vragen stellen over dit onderwerp. Tijdens de bijeenkomst gaan we samen praten over verschillende thema's rondom generatieve AI. Voor onderzoek met jongeren onder de 16 is hiervoor toestemming van een ouder of verzorger nodig. Daarom vragen we van u of u toestemming wilt geven voor deelname van uw kind via dit formulier.

Wat gebeurt er met de gegevens?
De antwoorden die uw kind geeft worden alleen gebruikt voor dit onderzoek. Tijdens de bijeenkomst wordt een audio-opname gemaakt. Deze wordt opgeslagen op een beveiligde server en verwijderd zodra het onderzoek is afgerond. Alleen de onderzoeker heeft toegang tot de opname.

Toestemming
Als uw kind mee mag doen dan kunt u dat hieronder aangeven. Deelname aan het onderzoek brengt geen bijzondere risico's met zich mee. Als u vragen heeft over het onderzoek mag u contact opnemen met de onderzoeker Senna Stoltink (senna.stoltink@riec-mn.nl).

Toestemmingsverklaring:
Ik heb genoeg informatie gekregen over de focusgroep en het onderzoek. Ik ben in de gelegenheid gesteld om vragen te stellen over het onderzoek en de deelname. Mijn kind neemt vrijwillig deel aan het onderzoek. Als mijn kind wil stoppen met het onderzoek dan kan dat zonder opgave van reden. Ik begrijp dat de antwoorden die mijn kind geeft anoniem verwerkt zullen worden en dat de onderzoeker deze mag gebruiken voor het onderzoek.

Mijn kind mag meedoen aan het onderzoek:
- ☐ Ja
- ☐ Nee

Bijlage D: Topiclijsten

Bijlage D1: Topiclijst interview informant

Introductie
- Kennismaking
- Praktische en ethische zaken: inhoud toestemmingsformulier + ondertekenen
- Kennismaking respondent:
    - Kan je wat vertellen over jezelf?
    - Kan je me meenemen in je expertise?

Operationalisering AI
- AI, waar hebben we het dan over?

AI-criminaliteit algemeen
- Hoe kan AI worden ingezet voor het plegen van een delict?
    - Vormen van criminaliteit
    - Platformen/tools
    - Toegankelijkheid voor daders
    - Benodigde kennis of vaardigheden
    - Wanneer sprake van inzet AI voor plegen delict
    - Grijze gebieden
- Kan je wat vertellen over de omvang van criminaliteit waarbij AI wordt gebruikt?
    - Registratie
    - Specifiek voor de verschillende vormen
    - Kenmerken daders/slachtoffers
    - Wat blijft onder de radar

Jongeren en AI-criminaliteit
- In hoeverre is er zicht op AI-criminaliteit door jongeren?
    - Concrete signalen/trends
    - Blinde vlekken
- Welke vormen van AI-criminaliteit kunnen jongeren plegen?
    - Platformen/tools
    - Toegankelijkheid
    - Benodigde kennis of vaardigheden
    - Onbekende vormen om te verwachten
        - Aanleiding verwachting
- Welke factoren zie je die verklaren waarom jongeren AI zouden inzetten voor criminaliteit?
    - Doorvragen op punten die respondent benoemt
- In hoeverre is er aandacht voor AI-criminaliteit door jongeren naar jouw inzicht?
    - Waar blijkt dat uit
    - Voorbeelden
    - Verhouding tot AI-criminaliteit in het algemeen

Toekomstbeeld

- Hoe zal AI-criminaliteit (door jongeren) zich in de komende jaren ontwikkelen?
    - Signalen
    - Risico's/scenario's om rekening mee te houden
    - Grootste onzekerheden in voorspelling
    - Waar nog geen zicht op maar wel belangrijk

Aanpak en aanbevelingen

- Is er al een aanpak voor (een vorm van) AI-criminaliteit (door jongeren)?
    - Hoe ziet aanpak eruit
    - Goede punten/verbeterpunten
    - Effectiviteit
    - Zo nee: wat zou belangrijk zijn bij ontwikkeling aanpak?
    - Risico's of blinde vlekken specifiek voor jongeren
- Wat zou je anderen willen meegeven over dit onderwerp?

Afronding

- Zijn er nog andere dingen die je kwijt wilt over dit onderwerp?
    - Vragen over onderzoek
    - Terugkoppeling
    - Andere mensen om te benaderen

Bijlage D2: Topiclijst focusgroep jongeren

Introductie

- Kennismaking: wie wij zijn, wat onderzoek inhoudt, wat we gaan doen tijdens focusgroep.
- Ethiek: uitleggen inhoud toestemmingsformulier + ondertekenen
- Praktische dingen:
    - Reageer op elkaar maar laat elkaar wel uitpraten
    - Geen goede of foute antwoorden, ben benieuwd naar jullie mening en ervaringen
    - Risico toelichten
- Kennismaking respondenten: naam, leeftijd en in welk schooljaar zit je?

Deel 1: gebruik van generatieve AI door jongeren

- Waar denken jullie aan bij generatieve AI?
    - Evt. toelichten: met generatieve AI kan je teksten, video's of audio maken.
- Welke AI-tools/platformen gebruiken jullie?
    - Waarvoor?
    - Hoe vaak?
    - Waarom die tools/platformen?
    - Waarom gebruik je daarvoor AI?
- (Wat vinden jullie van AI?)

Deel 2: wat mag (niet) met AI?

- Wanneer vinden jullie het oké om AI te gebruiken?
    - Waarom?
- Wanneer niet?
    - Waarom?
    - Zie je dat wel eens om je heen?
- Welke risico's kunnen er zitten aan het gebruik van AI door jongeren?
    - Waarom is dit een risico?
    - Voorbeeld van situatie waarin dit risico kan ontstaan?
    - Zie je dit wel eens om je heen?
- Zien jullie ook risico's die te maken hebben met misbruik of criminaliteit?
    - Dit wel eens gezien in je omgeving (online/offline)?
    - Zelf wel eens opdracht gegeven die AI niet wilde uitvoeren?
- Wanneer is het gebruik van generatieve AI volgens jullie strafbaar?
    - Hebben jullie dat wel eens gezien in je omgeving of online?
    - Wat gebeurde er toen?
    - Hoe makkelijk/moeilijk is het voor jongeren om AI daarvoor te gebruiken denk je?
- Wat zou je doen als je ziet dat iemand AI gebruikt waarvan je denkt dat het niet oké of zelfs strafbaar is?
    - Waarom?
    - Wat als het met jou persoonlijk te maken heeft?

Deel 3: grenzen via stellingen

Werkvorm stellingen eens/oneens: Het is oké om AI te gebruiken om…

- Iemands stem na te maken (voice cloning)
    - Waarom?
    - Heb je dit wel eens gedaan/iemand anders zien doen?
    - Zijn er situaties waarin het juist wel/niet oké is?
- Een neppe foto of video van iemand uit je omgeving te maken (deepfakes)
    - Waarom vind je dat?
    - Heb je dit wel eens gedaan/iemand anders zien doen?
    - Aan iemand bij “oneens”: zou je antwoord veranderen als het gaat om een bekend persoon?
- Naaktbeelden van iemand te maken (deepnudes)
    - Waarom vind je dat?
    - (Heb je dit wel eens gedaan/iemand anders zien doen?)
- Toegang te krijgen tot iemands account (hacken)
    - Waarom vind je dat?
    - Heb je dit wel eens gedaan/iemand anders zien doen?
- Documenten zoals een ID na te maken of aan te passen (fraude)
    - Waarom vind je dat?
    - Heb je dit wel eens gedaan/iemand anders zien doen?
- Bij genoeg tijd: als iemand AI gebruikt voor iets crimineels, is diegene daar niet verantwoordelijk voor.
    - Waarom vind je dat?

Afronding

- Is er nog iets anders dat jullie kwijt willen over dit onderwerp?
- Hebben jullie nog vragen?
- Bedanken
- Nazorg: Offlimits: gratis anonieme hulplijn voor als jij of iemand uit je omgeving online iets vervelends heeft meegemaakt en je erover wil praten.

Bijlage E: Kenmerken van de jongeren en focusgroepen

Respondent	Leeftijd	Geslacht	Opleiding	Focusgroep
R1	14	Vrouw	Middelbare school	Focusgroep 1
R2	19	Man	WO-bachelor	Focusgroep 1
R3	16	Vrouw	Middelbare school	Focusgroep 1
R4	17	Vrouw	Mbo Mediavormgeving	Focusgroep 2
R5	16	Vrouw	Mbo Mediavormgeving	Focusgroep 2
R6	23	Man	Mbo Mediavormgeving	Focusgroep 2
R7	22	Man	Mbo Mediavormgeving	Focusgroep 2
R8	16	Man	Mbo System Engineer	Focusgroep 3
R9	17	Man	Mbo System Engineer	Focusgroep 3
R10	17	Man	Mbo System Engineer	Focusgroep 3
R11	18	Man	Mbo Medewerker ICT support	Focusgroep 3
R12	18	Man	Mbo Medewerker ICT support	Focusgroep 3
R13	17	Man	Mbo System Engineer	Focusgroep 3
Andere aanwezigen				
Twee Halt-begeleiders				Focusgroep 1
Secondant van de onderzoeker				Focusgroep 2 en 3
Docent				Focusgroep 3

Bijlage F: Codeboom

Zicht op omvang

- Concrete signalen AI-crim jongeren
- Beperkt zicht op omvang
    - Verklaringen beperkt zicht op omvang
    - Registratie omvang geen prioriteit
- Obstakels zicht krijgen op omvang

Zorgen van informanten

- Zorgen m.b.t. taalmodellen
    - Zorg onbewuste omgang met privacygegevens
    - Zorg afhankelijkheid van taalmodellen
        - Zorg impact op beslissingen
    - Zorg mentale problemen verergeren
- Zorgen m.b.t. deepfakes
    - Zorg deepfakes beïnvloeden perceptie
    - Zorg toegankelijkheid deepnudes/uitkleedapps

Gebruik van GenAI door jongeren

- Gebruik school/leren/zoeken
    - Richtlijnen GenAI vanuit school
- Creatief gebruik
- Indirect gebruik
- Motivatie gebruik

Grenzen verantwoord gebruik jongeren

- Manieren van verantwoord gebruik
- Manieren van onverantwoord gebruik
    - Deepnudes niet ok
    - Deepfake stem niet ok
- Grenzen gebruik afhankelijk van…
    - Afhankelijk van relatie maker - afgebeelde
    - Afhankelijk van intentie maker
    - Afhankelijk van mening afgebeelde
    - Weet zelf grenzen vrienden

Ervaring met AI-criminaliteit jongeren

- Geen ervaring AI-crim omgeving
- Voorbeelden ervaring AI-crim omgeving
    - Ervaringen deepnudes
    - Ervaringen deepfake-advertentie
- Verantwoordelijkheid AI-crim
    - Persoon blijft verantwoordelijk
    - AI-ontwikkelaars deels verantwoordelijk
    - Verantwoordelijkheid onduidelijk

Zelf verkennen grenzen GenAI jongeren

- Beschrijving van *guardrails*
- Ervaringen met *guardrails*
- Bewust omzeilen *guardrails*
- Ervaringen uncensored taalmodel