Tekoälyn hyödyntäminen lapsiin kohdistuvassa kyberrikostutkinnassa

Turun yliopisto
Tietotekniikan laitos
TkK-tutkielma
Tieto- ja viestintätekniikka
Huhtikuu 2024
Lauris Seebeck

TURUN YLIOPISTO
Tietotekniikan laitos

Lauris Seebeck: Tekoälyn hyödyntäminen lapsiin kohdistuvassa kyberrikostutkinnassa

TkK-tutkielma, 33 s.
Tieto- ja viestintätekniikka
Huhtikuu 2024

Digitalisaation myötä yhä useampi lapsi on yhteydessä internettiin yleistyneiden mobiililaitteiden seurauksena. Tämän takia on tärkeää pyrkiä turvaamaan lapsille mahdollisimman turvallinen internetkokemus, ehkäisemällä lapsiin kohdistuvia seksuaalirikoksia. Tutkielmassa käsitellään käsitteitä digitaalisuus, kyberturvallisuus ja tekoäly. Tutkielmassa käsitellään, millaisia vaaroja lapsia vaanii internetissä, miten voidaan suojella lapsia internetissä, ja voidaanko tekoälyllä tukea lapsiin kohdistuvan seksuaalirikosten torjuntaa. Tutkielmassa syvennytään tekoälyn menetelmiin, kuten tukivektorikoneeseen (engl. Support Vector Machine, SVM) ja konvoluutioneuroverkkoon (engl. Convolution Neural Network, CNN). Pohditaan myös tekoälyn kykyä toimia tukena lastensuojelutyössä ottamalla tukea muista tekoälytutkimuksista. Lopputuloksena on arvioitu, että tekoäly ei vielä ole tarpeeksi luotettava toimiakseen rikostutkinnassa, mutta nykytutkimusten mukaan, tekoäly voi olla potentiaalinen työkalu mm. suurten data määrien läpikäynnissä, keskustelualustojen läpikäynnissä sen kuva- ja tekstianalyyttisten taitojen ansiosta, sekä epäilyttävien ihmisten identifioinnissa. Tekoäly on potentiaalinen työkalu sekä rikostutkinnassa, mutta myös lastensuojelussa yleisellä tasolla.

Asiasanat: tekoäly, lastensuojelu, tukivektorikone, konvoluutioneuroverkko

Sisällys

1 Johdanto

Vuosi vuodelta 1990-luvulta lähtien internetistä on tullut yhä yleisempi työkalu [1] kaikissa ikäryhmissä [2]. Digitalisaatio on mahdollistanut sen, että suurella osalla maailman väestöstä on jonkinlainen yhteys internettiin. Älypuhelimien kasvavan suosion takia monilla on kellon ympäri suora yhteys internettiin. Tämän vuoksi myös verkkoturvallisuudesta on tullut vielä tärkeämpää ja ajankohtaisempaa. Ja koska internet on vielä monille puolituntematon, tulee tällaisia käyttäjiä suojella erityisesti internetissä tapahtuvista rikoksista. Yksi tällainen ryhmä on lapset ja nuoret.

On tärkeää tarjota lapsille ja nuorille kattava mediakasvatus, [3] koska se edistää heidän turvallisuuttaan internetissä. Lisäksi lasten internetin käytön valvonta voi auttaa estämään heitä altistumasta haitalliselle sisällölle. Tällaista valvontaa laajalla skaalalla voi tukea poliisi ja kyberturvallisuuteen erikoistuneet yksiköt [4]. Lisäksi Covid-19 aikana lasten näyttöaika on huomattavasti kasvanut [5], ja yhä useampi lapsi on löytänyt tiensä internettiin. Tämän tutkielman teemaksi valikoitui, kuinka tekoälyn erilaisia menetelmiä voidaan hyödyntää internetin monitoroinnissa ja siten suojella lapsia haitalliselta sisällöltä.

Tämän tutkielman tutkimuskysymykset ovat:

TK1 Miten tekoälyä voitaisiin hyödyntää lastensuojelussa ja rikostutkinnassa, ja mitkä ovat sen mahdollisuudet ja rajoitukset?

TK2 Miten tekoälyn tekniikoita, kuten tukivektorikonetta tai konvoluutioneuro-

verkkoja, voitaisiin käyttää rikostutkijoiden avuksi epäilyttävän aineiston löytämisessä ja lapsiin kohdistuvan hyväksikäytön estämisessä?

Hakukantoina on käytetty Google Scholar -tietopankkia ja IEEE Xplore -sivustoa. Tutkielman hakulauseina on käytetty mm. *(AI OR Artificial Intelligence) AND (Child protection OR (child* AND Internet)) sekä (AI OR Artificial Intelligence) AND (Child* OR Internet) AND (("Child protection"OR Child*) AND protection).* Hakulauseilla löytyi IEEE Xplore sivustolta noin 286 artikkelia, joista valikoitui 30. Tutkielmaan valikoitui artikkeleja, joissa pohdittiin tekoälyn hyödyntämistä lastensuojeluun internetissä tai verkkolaitteilla. Lisäksi artikkeleista pyrittiin löytämään tutkimusta tekoälystä poliisityössä. Näistä valikoitui vielä noin 10 tutkielmaa lopputarkasteluun. Lopputarkasteluun päätyivät tutkimukset, joissa käydään läpi haluttuja tekoälyn menetelmiä, eli tukivektorikonetta ja konvoluutioneuroverkkoja. Lopputarkastuksen vanhin tutkimus julkaistiin vuonna 2012 ja uusin vuonna 2021. Tutkielmassa on käytetty myös Google-hakukonetta muuhun tiedonhakuun, kuten pohjatietojen etsimiseen digitalisaatiosta, lasten verkkokäyttäytymisestä, kyberrikostorjunnasta ja tekoälymenetelmistä.

Tutkielmassa käsitellään digitalisaatiota, tekoälyä sekä kyberturvallisuutta. Tutkielman johdanto on ensimmäisessä kappaleessa. Johdannossa käydään läpi tutkielman aineistohakua, rajausta sekä tutkimuskysymykset. Toisessa kappaleessa määritellään digitaalisuus, ja pohditaan lasten digitoimintaa. Kappaleessa pohditaan myös lapsiin kohdistuvista riskeistä ja mahdollisuuksista internetissä. Kolmannessa luvussa käydään läpi kyberrikostorjuntaa, kyberrikollisuutta ja tietoturvaa. Luvussa lisäksi pohditaan, kuinka tietoisia vanhemmat ja lapset ovat tietoturvallisuudesta internetissä, sekä mitkä osa-alueet vaikuttavat lapsen verkkoturvallisuuteen. Neljännessä kappaleessa määritellään tekoäly, ja pohditaan tämän mahdollisuuksia ja heikkouksia. Tutkielma käsittelee myös neljännessä kappaleessa erilaisia tekoälyn sekä menetelmiä että algoritmeja, joita voidaan hyödyntää internetin monitoroinnissa.

Tutkielmassa keskitytään vain teksti- ja kuva-analyyttisiin menetelmiin. Algoritmeja valittiin kaksi kappaletta, joita käsitellään monipuolisesti. Nämä algoritmit ovat tukivektorikone (engl. Support Vector Machine, SVM) ja konvoluutioneuroverkko (engl. Convolution Neural Network, CNN). Lisäksi tutkielmassa tutustutaan myös PrevBOT nimiseen robottiohjelmaan eli bottiin, joka on konsepti keskustelupalstojen monitorointiohjelmistosta, jota suunnitellaan kyberrikostutkintaan käytettäväksi. Kyseiset algoritmit valikoituivat sen mukaan, mitä esiintyi eniten valituissa aineistoissa. Viimeisestä kappaleesta löytyy tutkielman yhteenveto sekä vastataan tutkimuskysymyksiin.

2 Digitalisaatio, internet ja lasten digitoiminta

2.1 Digitaalinen muutos

Digitalisaatio tarkoittaa ilmiötä, jossa teknologian avulla pyritään helpottamaan ja parantamaan ihmisten jokapäiväistä elämää. Digitalisaatio vaikuttaa lähes kaikkeen yhteiskunnallisiin osa-alueisiin, jos ei suoraan, niin välillisesti. Tämän myötä toivotaan, että arki helpottuu ja sujuvoituu. Kuitenkin digitalisaatio aiheuttaa myös paljon huolta ja ylimääräistä murhetta. Tavoitteena on helpottaa elämää tuomalla vaihtoehtoisia tapoja hoitaa asioita, eikä korvata kasvokkain tapahtuvaa asiointia. [6][7]

Digitalisaatiossa on tärkeää huomioida kyberturvallisuuden tärkeys. Kyberturvallisuus toimii digitalisaation kanssa yhteistyössä turvaamalla tämän kehityksen [8]. Kyberturvallisuudella tarkoitetaan tietoteknillisen toiminnan turvallisuutta. Kyberturvalla pyritään ehkäisemään kyberrikollisuutta sekä ihmis- ja valtiotasolla, ja mahdollistamaan mahdollisimman monelle turvallinen internetkokemus. [9]

Kun puhutaan digitalisaatiosta, on tärkeää ottaa esille tekoäly. 1950-luvulta lähtenyt idea koneesta, joka olisi yhtä älykäs kuin ihminen [10] [11], on vahvasti yhteydessä digitalisaatioon. Tekoälyä nykypäivänä hyödynnetään hyvinkin monipuolisesti erilaisissa tehtävissä [12]. Tekoälyn koetaan myös olevan tärkeä osa digitaalista

muutosta sen innovatiivisuudensa ja produktiivisuudensa vuoksi [13], mutta myös tehokkaan ongelmaratkaisukyvynsä takia organisaatiotasolla[14].

2.2 Lasten digitoiminta

Teknologiasta on tullut normi monien lasten kotona. Nykyajan lapset ovat yhä enemmän yhteydessä internettiin, ja lapsilla yleistyvät älypuhelimet mahdollistavat nopean yhteyden internettiin. [15] Noin 33 % maapallon lapsilla on yhteys internettiin [16], jota käytetään enimmäkseen oppimiseen, kommunikointiin, pelaamiseen ja luovaan työhön. Internet kuitenkin altistaa lapsia myös esimerkiksi kiusaamiselle ja sopimattomalle sisällölle. Lisäksi näistä altistavista tekijöistä on tullut kokoaikaisesti läsnäolevia, minkä takia esimerkiksi koulukiusaaminen seuraa lapsia kotiin kyberkiusaamisen muodossa [17]. [18]

Marc Prenskyn [19] kovasti kritisoitu diginatiivisuus yhdistetään usein nykyajan lapsiin. Marc Presnky viittaa diginatiiveilla ihmisiin, jotka ovat syntyneet ja kasvaneet digitaalisen teknologian ympärillä, mikä alkoi noin 1980-luvulla. Näiltä henkilöiltä oletetaan erinomaista digiosaamista. Tilanne ei usein ole näin. Vaikka nykyajan lapsilla on mahdollisuus käyttää erilaisia teknologisia työkaluja, ei tämä suoraan viittaa siihen, että nykyajan lapset olisivat erinomaisia teknologian tai internetin käyttäjiä. Myöhemmät tutkimukset ovat osoittaneet, että syntymävuodella ei ole selkeää yhteyttä siihen, kuka on diginatiivi ja kuka ei. [20] Vuosien varrella on huomattu medialukutaidon osaamisen tärkeys, sillä moni tutkimus on todennut, että monella nuorella on vaikeuksia etsiä ja arvioida internetistä löydettyä tietoa [21] [22]. Tämä on huomioitu korostamalla ICT-taitojen tärkeyttä myös lasten koulutusjärjestelmässä. [18]

Staksrud, Livingstone, Haddon ja Ólafssonin raportissa [23] *What do we know about children's use of online technologies?: a report on data availability and research gaps in Europe [2nd edition]* viitataan tutkimuksen ensimmäiseen painokseen, jos-

		Content: Lapsi on vastaanottavassa roolissa	Contact Lapsi on osallisena	Conduct: Lapsi on käyttäjän roolissa
Mahdollisuudet	**Pedagoginen oppiminen ja digitaalinen lukutaito**	Opetusmateriaali	Mielenkiinnoista jakaminen	Yhdessä oppiminen
	Identiteetti ja yhteisö	Henkilökohtaisista asioista tuen etsiminen (terveys/seksuaalisuus jne.)	Sosialinen verkostointi, jakaa yhteisiä kokemuksia	Identiteetin ilmaiseminen
	Luovuus ja itsensä ilmaisu	Resurssien moninaisuus	Inspiraation saaminen luovaan tekoon	Käyttäjien sisällöllinen luominen
Riskit	**Agressiivisuus**	Väkivaltaista ja vihamielistä sisältöä	Kiusaamisen, vainonnan tai ahdistelun kohteeksi tuleminen	Kiusaaminen tai toisen ahdistelu
	Seksuaalisuus	Pornografisen/haitallista seksuaalista sisältöä	Tuntemattomien tapaaminen, groomaamisen kohteeksi joutuminen	Pornografisen materiaalin luominen/julkaiseminen
	Arvot	Rasistisia, puolueellista tieto/neuvoa (esim. huumeet)	Itsensä vahingoittaminen, ei-toivottu suostuttelu	Neuvon antaminen esim. itsemurhaan, syömishäiriöön
	Mainonta	Mainonta, roskaposti ja sponsorointi	Seurauksen tai henkilötietojen keräämisen kohteeksi joutuminen	Uhkapelaus, laiton tiedostojen lataaminen, hakkerointi

Kuva 2.1: Luokittelu lapsiin kohdistuvista riskeistä ja mahdollisuuksista [23] [18]

sa lapset internetin käyttäjinä on jaettu seuraaviin luokkiin: content, contact ja conduct. EU Kids Online kehitti tämän luokituksen riskityypeistä lapsen roolin ja koetun riskin tyypin perusteella.

Nämä luokitukset tarkoittavat [23][18] (ks. talukko 2.1):

- Lapsi on vastaanottavassa roolissa. Tilanteet, jossa lapsi voi altistua haitalliselle tai laittomalle materiaalille *(Content)*
- Lapsi on osallisena. Tilanteessa lapsi voi joutua osallistumaan tuntemattomien ihmisten kanssa keskusteluun, lapsen yksityisyysrajojen rikkomiseen, tai joutuu internetin kautta vaaninnan tai kiusaamisen kohteeksi *(Contact)*
- Lapsi on käyttäjän roolissa. Lapsi osallistuu toisten käyttäjien kiusaamiseen, pornograafisen materiaalin luontiin ja lähettämiseen *(Conduct)*

3 Tietoturva ja kyberrikollisuus

3.1 Kyberrikostorjunta

Covid-19 pandemian huipulla huomattiin myös jyrkkä kasvu lasten seksuaalisen hyväksikäyttöä koskevassa verkkomateriaalissa [24]. Vuonna 2020 lapsiin kohdistuva seksuaalisen hyväksikäyttöä koskeva materiaali oli vuoteen 2019 verrattuna nelinkertaistunut, kertoo ITU:n toimintapoliittinen asiakirja [25]. Lasten suojelu digitaalisissa ympäristöissä on globaali haaste. Lastensuojeluun liittyviä haasteita ovat mm. olemassa olemattomat kansainväliset ja kansalliset lainsäädännöt, suunnitelmat, strategiat, rahoitukset ja instituutiot. Lisäksi yksi lastensuojelun haasteista verkossa on sen rajaton luonteisuus. Rajattomalla tarkoitetaan tässä tapauksessa sitä, että rikos, rikolliset, uhrit ja tekninen infrastruktuuri eivät ole sijoittuneet yhteen lainkäytäntöalueeseen, ja vaatii siksi usein kansainvälistä yhteistyötä [26] [27]. Tästä syystä kyberrikosten tutkinta lainvalvontaviranomaisilla vaikeutuu. Kyberrikosten tutkinta vaatii useamman alueen sekä julkisten että yksityisten viranomaisten välistä yhteistyötä. [25, 27, 26]

Tekoälyn yleistyessä on noussut esiin kysymys, voisiko sitä käyttää poliisityössä. Tekoälyä voitaisiin hyödyntää erilaisissa osa-alueissa kuten, epäiltyjen profiloinnissa, pimeän verkon rahan kulun analysoinnissa ja lapsiin kohdistuvan pornografisen materiaalin havaitsemisessa. Mutta ennen kuin tekoälyä voidaan hyödyntää poliisityössä, tulee pitää huoli siitä, että tekoäly pystyy suojelemaan perusoikeuksia, kuten

henkilötietojen ja yksityisyyden suojaa ja syrjimättömyyttä [28]. [29]

3.2 Kyberrikollisuus

Yleistyneen internetin käytön takia myös kyberrikollisuus on kasvanut vuosien aikana [30]. Varsinkin COVID-19 pandemian aikana on huomattu nousu kyberrikollisuudessa, josta yleisimmät olivat verkkopetokset, huijaukset ja tietojen kalastus [24][31]. Kyberuhat voidaan jakaa kolmeen eri kategoriaan [32]:

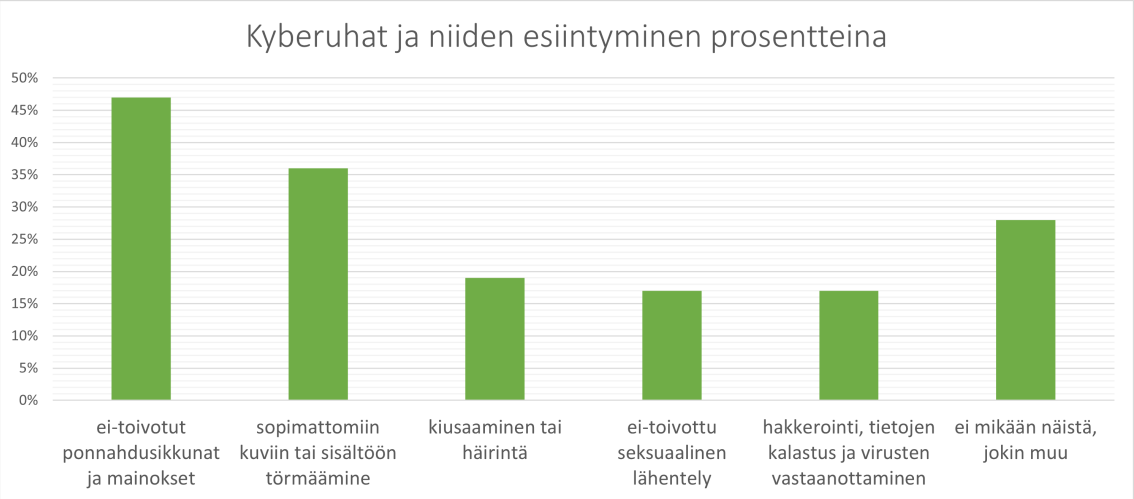
- *Kyberrikollisuus* – yksi tai useampi toimija pyrkii hyökkäämään järjestelmiin, ja siten hyötymään hyökkäyksestä rahallisesti, tai aiheuttamalla häiriötä.
- *Kyberhyökkäys* – usein poliittisesti motivoitunut tiedonkeruu.
- *Kyberterrorismi* – suunniteltu hyökkäys sähköisiä tietojärjestelmiä, ohjelmia ja tietoa vastaan, millä pyritään aiheuttaa pelkoa ja huolta. Toisinaan määritelmää laajennetaan kattamaan kaikkia kyberhyökkäyksiä, joissa uhataan väkivallalla ja joiden tarkoituksena on pelotella kohdeyleisöä [33].

LexiNexis Risk Solutions kyberrikollisuuden raportti vuodelta 2020 [34] raportoi, että suurin osa ihmisiin kohdistuvasta kyberrikollisuudesta kohdistui nuoriin aikuisiin ja yli 75 vuotiaisiin. Kyberrikollisuus oli myös raportin mukaan kasvanut huomattavasti 2019 ja 2020 välillä. Tämän uskotaan johtuvan uusien käyttäjien määrän noususta Covid-19 pandemian aikana vuonna 2020 [35]. Myös Atlas VPN:n raportti on päätynyt samaan lopputulemiin. Sen lisäksi, että nuoret aikuiset ovat eniten ihmisryhmistä alttiita kyberrikollisuudelle, niin he myös epätodennäköisimmin ilmoittavat heihin kohdistuvasta kyberrikollisuudesta [36].

Covid-19 pandemian aikana moni lapsi joutui nopealla aikataululla siirtymään kotiin opiskelemaan. Toukokuussa 2021 noin 80 % yhdysvaltalaisista opiskelijoista opiskelivat vain etäyhteyden avulla. Pandemian aikana lasten altistuminen erilaisille kyberuhille kasvoi huomattavasti, sillä yhä useampi lapsi alkoi käyttämään

verkkoa päivittäisenä opiskelun kanavana. [37] Myös ransomware-hyökkäykset oppilaitoksia vastaan lisääntyivät pandemian alkamisen jälkeen, sekä moni oppilaitos joutui verkkovakoilun kohteeksi. Vakoilun kohteiksi joutui myös opiskelijat ja etäopintoissa käytetyt alustat, kuten Zoom. Opiskelijat olivat myös joutuneet välillisesti kyberuhkien kohteeksi. Human Rights Watch havaitsi [38], että 145 koulutusteknologian (EdTech) etäopiskeluvälineisä 163:sta olivat jakaneet opiskelijat tietoja kolmansille osapuolille. Tämä johti US Federal Trade Commission (FTC) lausuntoon, jossa painotettiin, että Edtech tulee suojella opiskelijoiden tietoja, eikä voi vaatia opiskelijoita luopumaan yksityisyydestään vastineeksi verkko-opiskelusta. [39]

Boston Consulting Group (BCG) [37] järjesti yhteistyössä tutkimusyrityksen Dynatan kanssa tutkimuksen, jossa kyseltiin noin 41 000 vanhemmalta ja lapselta lasten verkkokäyttäytymisestä, ja näiden verkkouhkien reagoinnista. Tutkimus pidettiin 2021 helmikuun ja maaliskuun välisenä aikana, johon ottivat osaa lapset ja vanhemmat 24 eri maasta. Tutkimuksen mukaan noin 72 % lapsista 8–17 vuoden ikäisistä ovat kokeneet ainakin kerran jonkinlaisen kyberuhan verkossa, ja vain noin 52 % lapsista tuntee olonsa turvalliseksi verkossa. Nämä kyberuhat ja niiden esiintyminen prosentteina on esitetty kuvassa 3.1.



Kuva 3.1: Kyberuhat ja niiden esiintyminen prosentteina [37]

Ja vain noin 40 % vanhemmista on kertonut, että heidän lapsensa ovat ilmoittaneet törmänneensä näihin kyberuhkiin. [37] Nazilah Ahmadin ja hänen tiiminsä mainitsivat tutkimuksessaan [40], että vanhempien kyberturvatietoisuus on vielä hyvin rajallista. Tutkimuksessa myös vahvasti ilmaistiin, että viranomaisten ja kansalaisjärjestöjen tulisi olla isommassa roolissa tietoturvan lisäämisessä vanhemmille [40].

3.3 Tietoturva

Digitalisaation myötä, kun informaatiotekniikasta tulee entistä yleisempää, tarvitaan tehokkaampia ja monipuolisempia keinoja turvautua kyberhyökkäyksiltä. Tietoturva viittaa teknologiaan, toimenpiteeseen tai käytäntöön, jolla ehkäistään kyberuhkia, tai vähennetään niiden vaikutusta. Tietoturvaa tarvitaan mm. kaikissa verkkoon liitetyissä laitteissa, sensitiivistä tietoa sisältävissä laitteissa, verkossa ja elektronisissa laitteissa.

Millaisia keinoja on turvata lapsia verkossa? Boston Consult Group (BCG) on havainnoinut, että lapsia suojeleva ekosysteemi voidaan jakaa viiteen eri osalliseen [37]:

- Teknologiayritykset ja yksityiset sektorit
- Koulut ja koulutusjärjestelmät
- Lasten tukipalvelut
- Lainvalvonta ja oikeusjärjestelmä
- Vanhemmat ja huoltajat

BCG:n kyselytutkimuksen lähteinä on käytetty UNICEF:a, DQ Instituutiota ja Global Kids Onlinea, ja ovat lähteiden avulla laatineet mainituille viidelle osalliselle

tapoja, miten he voivat paremmin suojella lapsia kyberympäristöissä. Tässä tutkielmassa syvennytään vain osiin *teknologiayritykset ja yksityiset sektorit*, *koulut ja koulutusjärjestelmät* sekä *vanhemmat ja huoltajat*.

Teknologia yritysten ja yksityisten sektoreiden BCG:n ehdottamana tulisi kiinnittää enemmän huomiota lapsiystävällisyyteen heidän luomilla alustoilla, sovelluksissa ja laitteissa, kuten lisätä sisäänrakennettuja suojelumekanismeja, sekä asettaa selkeät alan standardit [37]. Lapsiystävällisyyden lisäksi, tulisi heidän myös kiinnittää huomiota käyttäjätietojen hallinnointiin ja jälleenmyyntiin. Covid-19 pandemian aikana useampi lasten käyttämä EdTech -sovellus oli kerännyt ja myynyt lasten henkilökohtaisia tietoja kolmansille osapuolille (aiemmin mainittu luvussa 3.2). [39] Yrityksiltä ja yksityisiltä sektoreilta toivotaan enemmän avoimuutta lapsiin kohdistuvista ongelmista, mikä tukisi näihin ongelmiin ratkaisujen löytämistä. [37]

Koulujen ja koulutuksen rooli on hyvinkin tärkeä, kun kyseessä on lasten tietoturvaosaaminen [37]. Tiffany O'Dell ja Arup Kumar Ghosh tutkimuksessaan [39] kertoivat, että osasyynä lasten heikkoon tietoturvaosaamiseen on opettajien, kouluttajien ja vanhempien oma puutteellinen tietoturvaosaaminen tai se, että tietoturvan opetus on hyvin heikkoa tai olematonta. Tutkimuksessa viitattiin useampaan tutkimukseen, joissa kävi ilmi, että hyvin pieni osa kouluttajista, opettajista ja rehtoreista kertoivat osaavansa suojautua verkossa. Ja vielä pienempi osa sanoi, että heidän opiskelijoillaan oli loistavat taidot verkkosuojautumiseen. Tutkimuksissa ilmeni alueellisia eroja, mutta pääpiirteittäin kaikenikäisten opiskelijoiden osaaminen pysyi alle 50 %:n.

Vanhemmat ja huoltajat ovat erityisen merkittävässä roolissa, kun kyse on lastensuojelusta verkossa. Vanhemmilla on mahdollisuus monitoroida lastensa toimintaa ja käyttäytymistä verkossa erilaisilla työkaluilla [41]. Martin Stoev ja Dipti K. Sarmah tutkimuksessaan *Online Protection for Children Using a Developed Parental Monitoring Tool* [41] arvioivat erilaisten seurantatyökalujen ominaisuuksia, niiden

saavutettavuutta eli hintaa, sekä ominaisuuksien parantamistärkeyttä. Tutkimuksessa on listaukset erilaisista sekä maksullisista että maksuttomista seurantatyökaluista. Useimmissa työkaluissa yhteiset ominaisuudet olivat sovellusten sekä sisältöjen monitorointi, sovellusten ja näyttöajan rajoittaminen tai säätely, verkossa sisällön filtteröinti, sekä lapsen puhelimen paikantaminen. Joissakin sovelluksissa, mutta ei kaikissa, oli myös mahdollisuus monitoroida tekstiviestejä ja chat-alustoja, sekä seurata sosiaalista mediaa epäilyttävien viestien varalta. Martin Stoev ja Dipti K. Sarmah tutkimuksessaan löysivät kuitenkin yleisistä monitorointi työkaluista puutteita. [41]

Monitoroinnin lisäksi on huomioitava vanhempien ja huoltajien tietoturvan tietämättömyys, sekä yleinen verkkokäyttäytyminen ja -asennoituminen. Internet ja erilaiset sosiaalisen median alustat ovat mahdollistaneet kaikille helpon tavan jakaa kuvia ja videoita omasta elämästä. Mutta moni vanhempi pahaa tarkoittamattaan jakaa sisältöä lapsistaan, johon tämä lapsi ei ole pystynyt antamaan suostumustaan. Kuvien julkaiseminen sosiaalisen median alustoille voi altistaa lapsen erilaisille vaaroille. Yksi näistä voi olla kuvien jakaminen tuntemattomille ihmisille, vaikka kuvan oli tarkoitus näkyä vain läheisimmille ystäville. Toinen on lapsen sijainnin paljastaminen. Kuvien julkaisun yhteydessä on tärkeää huomioida kuvan tausta ja jättää kaikki tuntomerkit pois kuvista, joilla voisi mahdollisesti paikantaa lapsen sijainnin. Kolmas lapsiin kohdistuva vaara on digitaalinen kaappaus. Tämä tarkoittaa ilmiötä, jossa joku julkaisee kuvan toisen alaikäisestä lapsesta omanaan, käyttää sitä mainostarkoituksiin tai jakaa sen pedofiilien kuvajakosivustolla [42]. [43]

4 Tekoäly ja sen hyödyntäminen poliisitoiminnassa

4.1 Tekoäly - Mikä on tekoäly?

Tutkielman viimeisissä kappaleissa keskitytään tekoälyn määrittelyyn ja sen menetelmiin. Lisäksi käydään läpi käsiteltyjen menetelmien tehokkuutta ja käytettävyyttä, sekä vastataan tutkielman tutkimuskysymyksiin.

Tekoäly (engl. Artificial Intelligence, AI) on tietokone tai tietokoneohjattu järjestelmä, joka pyrkii jäljittelemään ihmisen toimintaa ja ylittämään ihmisen älykkyyden tietyissä tehtävissä. Tekoäly on tehokas ongelmanratkaisussa ja pystyy analysoimaan sekä käsittelemään suuria määriä dataa. [44] Nykyinen tekoäly on yhä kapeaa tekoälyä, joka tarkoittaa tekoälyä, joka suoriutuu ihmistä paremmin tietyissä ennalta määritellyissä tehtävissä. Esimerkkejä kapeasta tekoälystä ovat erilaiset chatbotit, itseajavat autot ja kuva-analyysi. [45] Koneoppiminen on tekoälyn osaalue, jossa tekoäly oppii uutta analysoimalla ja löytämällä malleja, ja käyttämällä näitä hyödyksi jäljittelemään ihmisen toimintaa. Koneoppiminen vaatii suuren määrän dataa, jotta tämän tarkkuus paranee. [46][47] On kuitenkin tärkeä huomioida, että tekoäly tekee virheitä. Koska tekoälyn oppiminen tapahtuu prosessoimalla isoa määrää dataa, on vaikeaa hallita, mitä tämä data oikeasti sisältää, ja mitä tekoäly siitä oppii. Tämä voi koitua ongelmalliseksi, sillä tekoäly voi oppia materiaalista

sopimatonta sisältöä ymmärtämättä tämän olevan haitallista. Tällainen haitallinen opittu materiaali voi sisältää esimerkiksi rasistisia ja seksistisiä stereotypioita sekä alisävyjä (engl. undertones). [48] On myös hyvä huomioida tekoälyn väärinkäytön mahdollisuus, ja että tekoäly vielä nykypäivänä on puutteellinen ja tekee vielä paljon virheitä. [49][50]

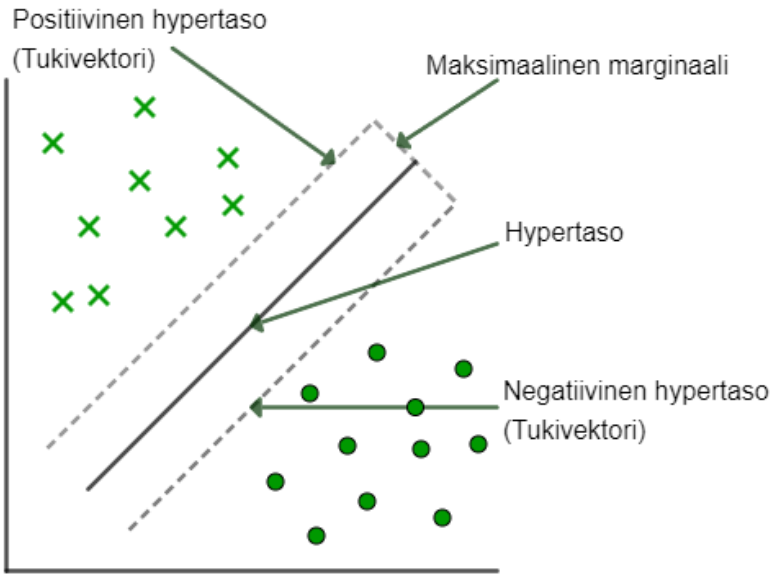
Vaikka tämän tutkielman on tarkoitus syventyä tekoälyn mahdollisuuksiin suojella lapsia, tulee myös huomioida, että tekoälyä on jo käytetty lapsia vaarantavilla tavoilla. Yksi tällainen esimerkki on tekoälyn kyky luoda kuvia lasten hyväksikäytöstä. Internet Watch Foundationin (IWF) raportin mukaan [51] yhden kuukauden aikana oli 20 254 tekoälyn luomia kuvia julkaistu pimeän verkon lasten seksuaalisen hyväksikäytön foorumeille. IWF totesi myös, että nykyisen tekoälyn luomat kuvat ovat jo sen verran realistisia, että niitä voidaan pitää tosina. [51]

4.2 Tekoälyn menetelmät

Kun lähdetään pohtimaan mahdollisia keinoja hyödyntää tekoälyä lastensuojelussa, tulee arvioida erilaisia metodeja. Tässä tutkielmassa käydään läpi kaksi erilaista algoritmia: tukivektorikone (engl. Support Vector Machine, SVM) ja konvoluutioneuroverkko (engl. Convolution Neural Network, CNN). Lisäksi tutkielmassa käsitellään PrevBot nimistä konseptia algoritmista, jota poliisi voisi potentiaalisesti hyödyntää tulevaisuudessa lastensuojeluun.

Tukivektorikone

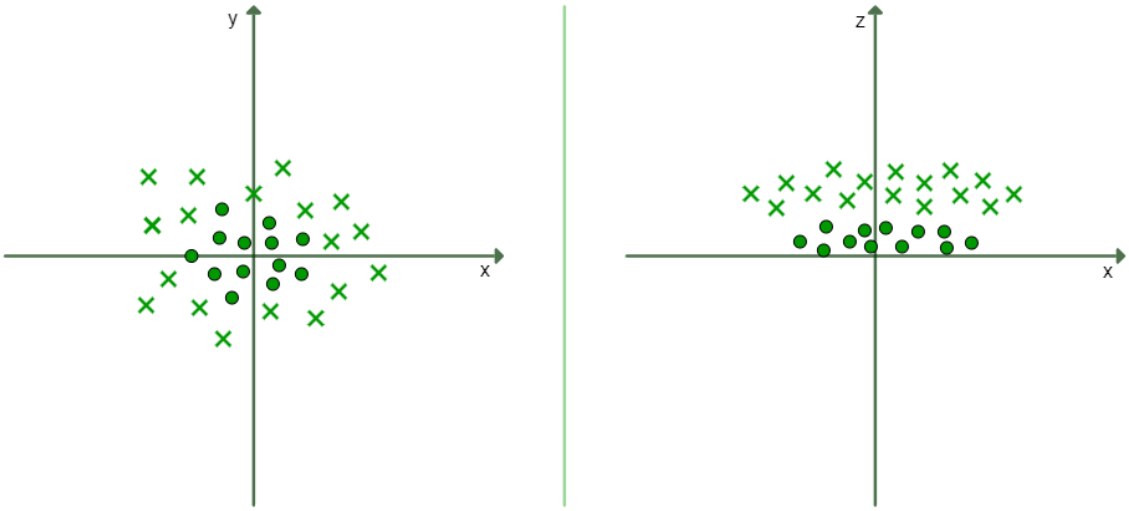
Tukivektorikone eli SVM on lajittelualgoritmi, joka pyrkii jakamaan aineiston kahteen luokkaan. SVM algoritmia käytetään enimmäkseen tekstin ja kuvien lajittelussa. SVM on ohjatun oppimisen algoritmi (engl. Supervised Machine Learning algorithm) [52], joka ei tarvitse paljoa opetusta, ennen kuin algoritmi alkaa antamaan



Kuva 4.1: Esimerkki tukivektorikoneen luokittelusta [58][52]

täsmällisiä vastauksia [53]. SVM on helposti ymmärrettävä ja käyttöinen, sekä yksi tehokkaimmista luokittelualgoritmeista [54]. SVM algoritmia ja sen muunnelmia on tutkittu paljon viimevuosien aikana monilla eri tieteenaloilla. SVM ei voi kuitenkaan sovi isolle datamäärälle, on laskennallisesti kallis eikä pysty jakamaan aineistoa isompaan kuin kahteen luokkaan. [54, 55, 56, 57]

SVM algoritmi toimii siten, että lajittelemattomat datapisteet (kuvassa 4.1 merkittynä risteinä ja ympyröinä) sijoitetaan X- ja Y-akselle. Datapisteiden välistä löytyy sekä positiivinen että negatiivinen hypertaso ja yksi maksimaalisen etäisyyden löytänyt hypertaso. **Hypertaso** on kahteen luokkaan jakava lineaarinen jakaja, joka on 2-uloitteisessa viiva ja 3-uloitteisessa taso. SVM algoritmin päätavoite on löytää hypertasolle mahdollisimman kaukainen etäisyys tukiverkoista, jotta algoritmi voi jakaa datapisteet mahdollisimman tasaisesti kahteen luokkaan. **Tukiverkot** ovat hypertasoa lähimmäisenä olevat datapisteet. **Marginaalilla** ilmaistaan kahden objektin etäisyyttä. [52] Datasta riippuen, on olemassa kaksi erilaista SVM algoritmia: Lineaarinen-SVM ja epälineaarinen-SVM. Lineaarinen-SVM:ssä (kuvassa 4.1) datapisteet voidaan jakaa luokkiin yksittäisellä suoralla, kun taas epälineaarinen-SVM



Kuva 4.2: Non-lineaarinen (vasemmalla) ja lineaarinen (oikealla) tukivektorikoneen esitelmä [58]

sijoittuu n-ulotteiseen avaruuteen (kuva 4.2). Kernel tekniikalla voidaan muokata epälineaarinen-SVM lineaariseksi kääntämällä akselit 90-astetta x-akselin ympäri. X- ja y-akseleille sijoittuva avaruus muutetaan siten z- ja x-akselille sijoittuvaksi avaruudeksi (kuva 4.2). [59]

Kun algoritmi alkaa käymään läpi annettua tietokokonaisuutta (engl. dataset), alkaa algoritmi jakamaan datapisteitä luokkiin algoritmille annetun koulutusdatan perusteella. Algoritmi pyrkii siis löytämään optimaalisen sijainnin hypertasolle, jotta hypertaso voi jakaa tietokokonaisuuden tasaisesti kahteen osaan. Hypertason sijainnin löytäminen onnistuu monella eri tavalla, mutta tässä tapauksessa algoritmissa hyödynnetään marginaaleja ja tukivektoreita. Algoritmi etsii monta erilaista mahdollista hypertason sijaintia, joista valitaan yksi sijainti, joka on mahdollisimman kaukana datapisteistä, sijoittaen itsensä datapisteiden väliin luomalla jakajan. Marginaalien avulla mitataan hypertason etäisyys lähimmäisistä tukivektoreista. [60][59] Marginaalin pituudella on merkitys siihen, kuinka tarkkoja tuloksia algoritmi antaa. Mitä isompi maksimaalinen marginaali on (kuva 4.1), sitä helpompaa algoritmin on luokitella tietokokonaisuutta. [52]

Hypertaso lasketaan 2-uloitteisella tasolla suoran laskukaavalla:

$$y = a * x + b$$

$$a * x + b - y = 0$$

Olkoon $z = (x, y)$ ja $w_0 = (a, -1)$. Tällöin hypertaso on vektorimuodossa:

$$w_0 * z + b = 0,$$

missä z ilmaisee vektoria, b ilmaisee hypertason sijaintia eli etäisyyttä origosta ja w_0 ilmaisee tukiverkkojen painoa, joka voidaan laskea tukivektoreiden z_i ja perseptronin ulostulon painon α_i tulon lineaarisummana: [61, 59, 58]

$$w_0 = \sum_{support\ vectors} \alpha_i * z_i$$

Koska algoritmi on ohjaamalla oppiva algoritmi, tulee algoritmi opettaa siis ennen sen käyttöä. Algoritmille annetaan jo valmiiksi sarja luokkiin jaettuja opetuspisteitä

$$(y_1, x_1), ..., (y_1, x_1), \qquad y_i \in -1, 1$$

Hypertason lisäksi algoritmi luo sekä positiivisen että negatiivisen hypertason, jotka ilmaisevat näitä kahta luokkaa $y = 1$ ja $y = -1$. Datapaketin jokainen piste jaetaan luokkiin sen mukaan, missä yksittäinen datapiste sijaitsee (kuva 4.1) [56] [59]

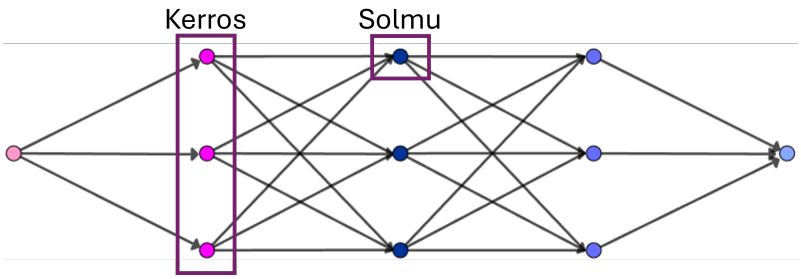
$$w * x_i + b >= 1 \quad kun\ y_i = 1,$$

$$w * x_i + b >= -1 \quad kun\ y_i = -1.$$

Konvoluutioneuroverkko

Konvoluutioneuroverkko eli CNN on koneoppimisen osa-alue, jota hyödynnetään visuaalisten tietokokonaisuuksien kanssa, kuten kuvista ja videoista datamallien (engl. datapattern) löytämisessä. CNN pohjautuu tekokeinoneuroverkkoihin (engl. Artificial Neural network, ANN), joka jakautuu erilaisiin erikoistuneisiin neuroverkkoihin, joita voidaan käyttää erilaisiin tarkoituksiin, kuten tekstin, kuvien ja äänen analysointiin. CNN on siis näistä neuroverkkojen ryhmästä erikoistunut visuaalisiin tietokokonaisuuksiin. [62][63] CNN on myös yksi tunnetuimmista ja käytetyimmistä tekoälyn menetelmistä, jota on sovellettu laajasti useilla eri aloilla [64].

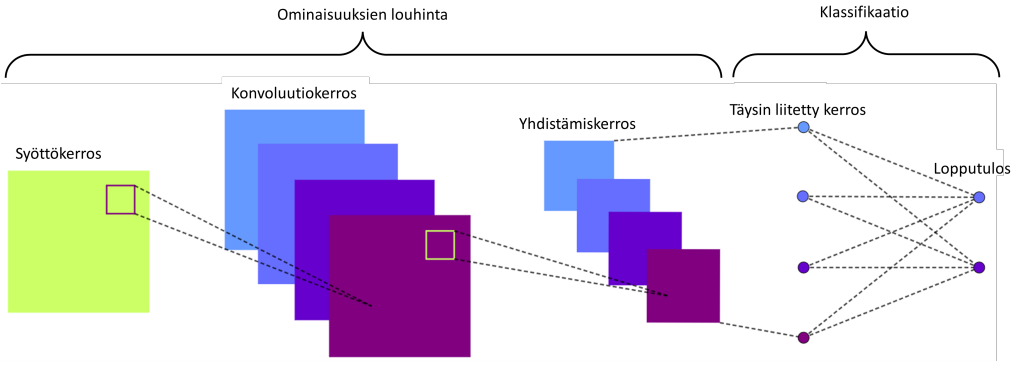
Jotta on helpompi ymmärtää CNN toimintaa, on hyvä ymmärtää sen pohjautuvan neuroverkon (NN) toimintaa. NN pyrkii mallintamaan ihmisten aivoja, ja niiden toimintaa. Yksittäiset solmut muodostavat kerroksia, ja nämä kerrokset verkkokokonaisuuden. Tämä rakenne mallintaa yksittäisiä aivojen neuroneja, jotka muodostavat yhteyksiä eri aivoalueiden välillä: [65]



Kuva 4.3: Usean kerroksen neuroverkko, missä jokaisella kerroksella useampi solmu [65]

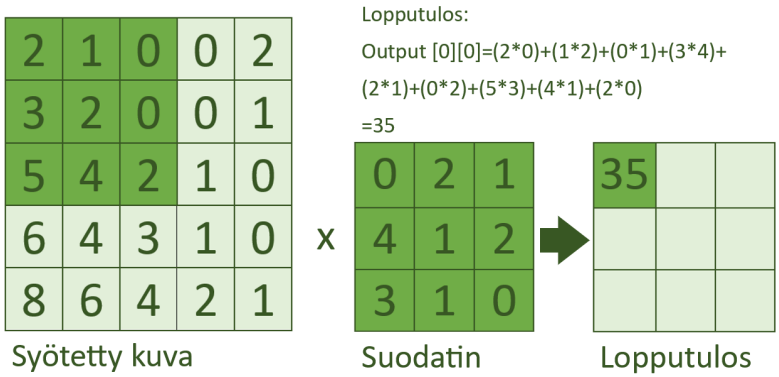
NN sisältää kolme erilaista kerrosta, syötekerros (engl. Input Layer), piilokerros (engl. Hidden Layer) ja ulostulokerros (engl. Output Layer). **Syötekerros** nimensä mukaisesti vastaanottaa neuroverkkorakenteelle annetun tietokokonaisuuden. **Piilokerros** taas käy läpi tietokokonaisuuden ennalta määritetyllä tavalla. Ja **ulostulokerros** palauttaa algoritmin saadun lopputuloksen. [62, 63, 65] Konvoluutioneuroverkko on ANN:n laajennettu versio. CNN sisältää seuraavat pää kerrokset: syö-

tekerroksen, konvoluutio kerroksen (engl. convolutional layer), yhdistämiskerroksen (engl. pooling layer) ja täysin liitetyn kerroksen (engl. fully-connected layer). [62] [63] CNN rakenteessa voi olla useampi konvoluutio ja yhdistämiskerros, mutta neuroverkkorakenne tulee päättyä täysin liitettyyn kerrokseen. [63] Esim:



Kuva 4.4: CNN kerrokset [66]

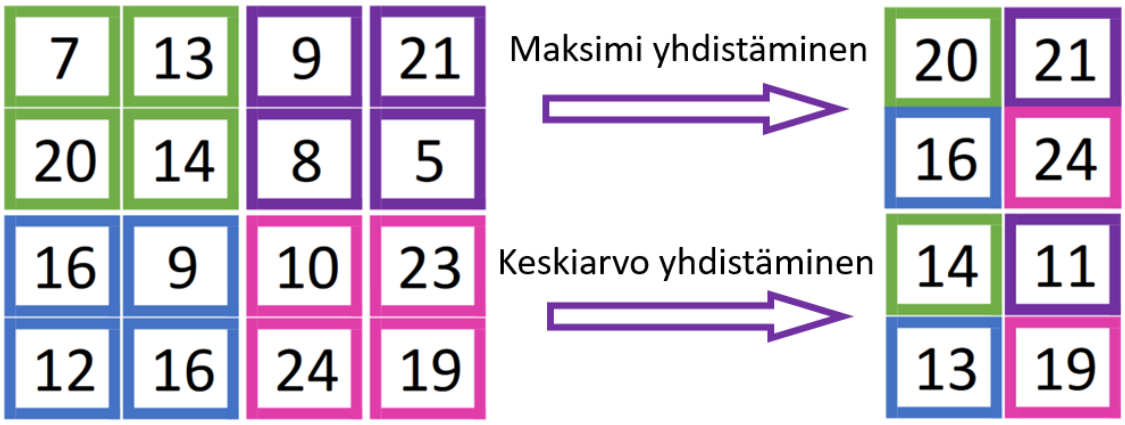
CNN toimii siten, että syötekerrokseen syötetään haluttu tietokokonaisuus. Tässä tapauksessa käsitellään kuvan käsittelyä. Syötetyn kuvan koko on 32 x 32 x 3. Syötekerroksen jälkeen konvoluutiokerros aloittaa kuvan läpikäynnin erilaisia suodattimia käyttäen. Suodattimet ovat pienempiä matriiseja usein kokoa 2x2, 3x3 tai 5x5. Jos kuvan läpikäynnissä käytetään esimerkiksi 12 suodatinta, niin saadaan tulostilavuudeksi 32 x 32 x 12, sillä suodattimien määrä vaikuttaa lopputuloksen syvyyteen [63]. Suodattimet käyvät siis läpi syötetyn kuvan pikseli kerrallaan, ja laskevat oman painon ja kuvapikselin tulon: [62, 65, 63]



Kuva 4.5: CNN konvoluutio suodattimella [63]

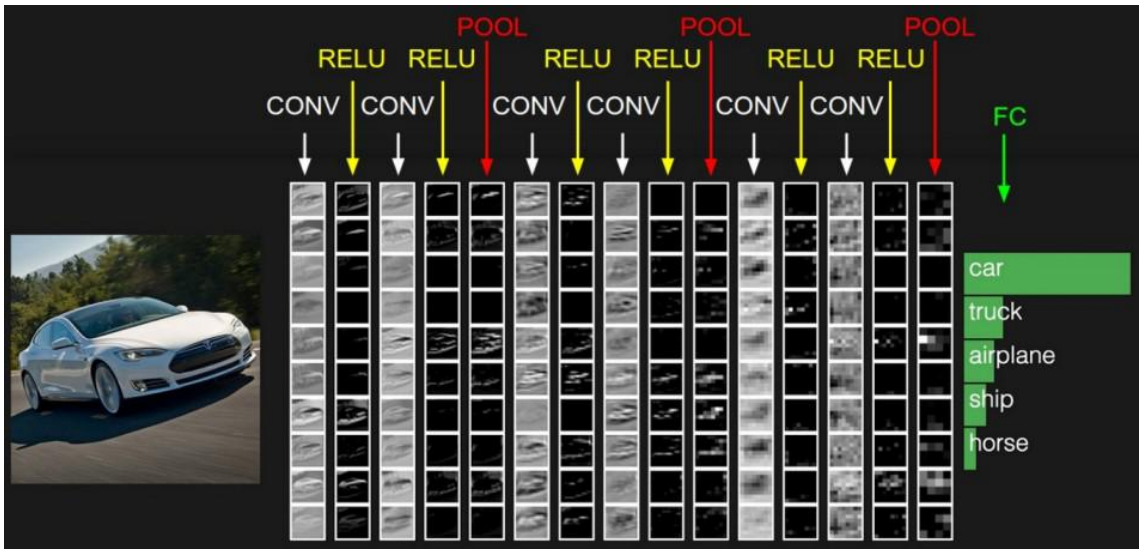
Suodattimen ja kuvan tulon siirtyminen seuraavalle kerrokselle toimii aktivaatiofunktiolla. [67]. Aktivaatiofunktio siis ilmaisee sen, onko kyseinen pikseli kuvassa aktivoituneena vai ei [68]. Yleisenä aktivaatiofunktiona käytetään Rectified Linear Activation Unit (ReLU) funktiota [62]. Funktio ReLU: $max(0, x)$ toimii siten, että positiiviset tulot syötetään seuraavalle kerrokselle suoraan, mutta negatiiviset muutetaan nolliksi. [69] Kerroksen lopputulosta kutsutaan piirrekartaksi (engl. feature maps). [62] Piirrekartat korostavat tärkeitä piirteitä syötetystä kuvasta pikselin arvojen, kuten värien, perusteella. CNN alkaa hahmottamaan ensimmäisenä kuvan reunoja, kuva siirretään seuraavalle kerrokselle, ja alkaa havaita mm. kulmia ja väriryhmittymiä. Kyseinen hahmottelu jatkuu, kunnes kaikki suodattimet ovat käyneet kuvan läpi. [65]

Seuraava kerros on yhdistämiskerros. Yhdistämiskerroksen tärkein tehtävä on pienentää kerrosten kokoa, jotta laskenta nopeutuu ja muistia kuluu vähemmän. Pienentäminen voidaan tehdä kahdella yleisimmällä kerroksella: maksimi yhdistämisellä (engl. max pooling) tai keskiarvo yhdistämisellä (engl. average pooling). [62][63] Maksimi yhdistämisessä filtteri valitsee pikselin korkeimman arvon, ja syöttää ulos kyseisen maksimiarvon seuraavalle kerrokselle [70]. Keskiarvo yhdistämisessä taas filtteri laskee pikselin arvojen keskiarvon, ja syöttää ulos lasketun arvon. [63]



Kuva 4.6: Maksimi ja keskiarvo yhdistäminen [62][71]

Ennen kuvan luokittelua, neuroverkko tasoittaa kerroksilta saadut piirrekartat yksiulotteisiksi vektoreiksi. Täysin liitetty kerros suorittaa luokittelutehtävän edellisten kerrosten ja suodattimien avulla poimittujen ominaisuuksien avulla. [63] [62] Ulostulokerros laskee logististen funktioiden avulla arvion kunkin luokan todennäköisyyspistemääräksi [62]:



Kuva 4.7: Syötetty kuva (vasemmalla), CNN kerrokset (keskellä) ja arvio lopputuloksesta prosentteina (oikealla) (kuvalähde: [62][72])

4.3 Tekoäly kyberrikostutkinnassa

Kun kyse on lastensuojelusta, olisi tärkeää havaita mahdollinen epäilyttävä toiminta mahdollisimman nopeasti [73]. Tämä kuitenkin koituu hyvin haasteelliseksi, sillä poliisin täytyisi käydä läpi suuria määriä dataa tehokkaasti ja nopeasti, jotta epäillyt saataisiin kiinni mahdollisimman pian. Resurssipula kuitenkin vaikeuttaa tätä tehtävää [74]. Lapsiin kohdistuva hyväksikäyttöön liittyvä digitaalinen rikostutkinta on todettu olevan hyvin raskas, pitkäkestoinen ja työläs. Sen lisäksi tutkimukset voivat olla luonteeltaan hyvin intensiivisesti toistuvia, mikä voi lisätä tutkijoiden uupumusta, ja sitä kautta virheiden tekoa. Työtaakka voi olla hyvin stressaava työntekijälle,

ja aiheuttaa psyykkistä stressiä. Lisäksi tutkintaan voi vaikuttaa tutkijoiden tahaton tai tahallinen puolueellisuus. Vähentämään tutkijoiden taakka, ja vähentämään virheiden tekoa, on ehdotettu tekoälyn hyödyntämistä rikostutkinnassa. [75] Tekoälyä voisi tukea kyberrikostutkijoita tutkinnassa [29, 73, 76, 74] ja ehkäistä lapsiin kohdistuvaa vaaraa internetissä [77, 78].

Taulukossa 4.1 on listattuna muutama tutkimus, joissa on pohdittu erilaisia menetelmiä sille, kuinka tekoälyä voitaisiin hyödyntää lastensuojelussa. Tutkielman rajauksen kannalta on otettu huomioon vain aiemmin käsittelyt SVM ja CNN menetelmät, ja arvioitu niiden suorituskykyä ja potentiaalia lastensuojelussa. Näiden tutkimusten lisäksi otetaan huomioon muutama tutkimus, jossa on pohdittu kriittisesti tekoälyn tämänhetkistä kykyä työskentelemään lastensuojelussa [29].

Anderson ja hänen tiiminsä [75] ovat tutkimuksessaan arvioineet, kuinka hyvin erilaiset tekstiklassifikaattorit eli tekstin luokittelijat pystyvät erottamaan seksuaalisesti houkuttelevan (engl. online grooming) tekstin ja epähoukuttelevan tekstin toisistaan. Tutkimuksessa käytettiin muunnelmia SVM menetelmästä kuten lineaarista ja epälineaarista coordinate descent fuzzy twin support vector machine (CDFTSVM), joista LCDFTSVM on lineaarinen ja RCDFTSVM epälineaarinen. CDFTSVM on SVM muunnelma, jolla pyritään edistämään jaottelua poistamalla ylimääräisiä häiriötekijöitä sumeajäsenyysfunktion (engl. fuzzy membership function) avulla ja vähentämällä laskennallista monimuotoisuutta koordinaattien laskustrategialla (engl. coordinate descent strategy). Tutkimuksessa käytettiin useampaa tekoälyn menetelmää, ja lisätäkseen luokitteluun tehoa, otettiin tutkimuksessa huomioon myös psykometrisiä ja kategorisoivia tietotekniikoita. Tutkimuksen lopputulos osoitti, että tekoäly pystyy arvioimaan, onko teksti seksuaalisesti houkutteleva vai ei. Tutkimuksessa kuitenkin myös korostettiin, että lisätutkimuksille on tarvetta. Esimerkiksi tekoälyn erilaisia menetelmiä tulisi hyödyntää enemmän, tekoälyn opetusmateriaali tulisi olla laajempi ja kattavampi, sekä on myös tärkeää ottaa huomion

internetin nopeasti muuttuvan kielen.

Tekoälyn menetelmiä on myös haluttu yhdistää psykologiseen perustaan. Jevremovic ja hänen kollegansa [77] viittasivat tutkimuksessaan Escalanten tiimin [79] tutkimukseen, jossa todettiin tekstin luokittelijoiden suorituskyvyn parantuvan, kun luokittelijat pohjautuvat psykologisen 3. vaiheen hypoteesiin. 3. vaiheen hypoteesi on hypoteesi siitä, kuinka lasten hyväksikäyttäjät käyttäytyvät lähestyessään lasta [73]. Morris tiiminsä kanssa [78] opettivat SVM algoritminsa arvioimaan sanojen arvoa. Mitä enemmän lauseesta löytyi hyväksikäyttöön liittyvää sanastoa, eli mitä suurempi sanojen arvojen summa, sitä suuremmalla todennäköisyydellä teksti tunnistetaan luonteeltaan hyväksikäyttäväksi.

Jevremovic ja hänen tiiminsä [77] tutkimuksessa arvioitiin erilaisten tekoälymenetelmien hyödyntämistä eri osa-alueissa, kuten kuva-analyysissa, äänen tunnistuksessa ja tekstianalyysissä. Tutkimuksessa kehiteltiin CASPER-järjestelmä, joka pyrkii tunnistamaan kaikki mahdolliset haittatoiminnat ja ilmoittamaan siitä lapsen huoltajille. CASPER-järjestelmä pystyisi analysoimaan kaiken näytöllä ja kaiuttimien kautta lapsen kohtaaman materiaalin.

Ratnark Gandhi [53] tutkimuksessaan tarkasteli myös, miten tekstiluokittelijoita voitaisiin hyödyntää kyberseksuaalisen häirinnän ja hyväksikäytön ehkäisyssä, jotka kohdistuvat teini-ikäisiin. Tutkimuksessa käsitellään tekstin luokittelijaa SVM algoritmia. Tutkimuksessa arvioidaan SVM algoritmin tehokkuutta verrattuna tekstiluokittelijaan Multinomial Naive Bayes (MNB). Gandhin tutkimuksen mukaan SVM menestyi paremmin kuin MNB algoritmi. Leen tiimin [74] tutkimuskoosteessa arvioidaan myös SVMn käyttöä tiedostonimien erottelussa. SVM algoritmia voi käyttää myös integroituna muihin tekoälyn menetelmiin sen tehokkaan luokittelutaitonsa takia [76].

Vitorino ja hänen tiiminsä [76] tutkimuksessa arvioitiin CNN menetelmän kykyä erottamaan lapsiin kohdistuvan pornografisen materiaalin (LHM) ja tavallisesta

pornografisesta materiaalista. Tämä koitui vaikeaksi toteuttaa, sillä CNN opettamista varten tarvittaisiin paljon opetusmateriaalia. Tutkimuksessa tämä ongelma kuitenkin sivuutettiin siten, että ensin CNN opetettiin tunnistamaan seksuaalisia kuvia käyttämällä helpommin saatavaa materiaalia, kuten tavallista pornografiaa. Tämän jälkeen CNN erikoistettiin tunnistamaan LHM materiaalia käyttämällä opetusmateriaalina LHM materiaalia. Tutkimuksessa arvioitiin myös CNN menetelmän käyttöä videoanalysoinnissa, jossa tämä toteutettiin pilkkomalla video kuviksi, ja pilkonnan jälkeen näitä kuvia analysoimalla. Tutkimuksessa todettiin videotunnistuksen vaikeudet, jotka olivat videoiden laatu ja koko. Mitä huonompi laatu, sekä mitä isompi videotiedosto, sitä vaikeammaksi tunnistus meni. Tutkimuksen laatijat kuitenkin uskoivat heidän ratkaisunsa olevan potentiaalinen tapa auttaa rikostutkijoita suurien datamäärien prosessoinnissa. Jevremovic ja hänen tiiminsä [77] tutkimuksessa tunnistettiin myös vaikeuksia LHM materiaalin tunnistuksessa. Ongelman sivuuttamiseen tutkijat opettivat CNN menetelmän käyttämättä LHM materiaalia. CNN opetettiin tunnistamaan pornografista materiaalia, ja sitten arvioimaan materiaalissa olevien osallistujien iän.

Ebrahimin tiimin [73] tutkimuksessa tutkittiin CNN kykyä toimia tekstin luokittelijana, kun konvoluutiokerroksia on yksi. Tutkimuksessa yhden konvoluutiokerroksen omaava CNN päihittää SVM ja NN menetelmät. Tutkimuksessa kokeiltiin myös 2-konvoluutiokerroksista CNN, mutta tulokset eivät olleet yhtä hyvät. Useampi kerrosmäärä lisää CNN syvyyttä ja monimuotoistaa menetelmää, mikä tekstin louhinnassa vain vaikeuttaa menetelmän toimivuutta. Monimuotoisempi ja syvempi rakenne soveltuu paremmin kuva-analyysissä. CNN kerrottiin olevan hyvä vaihtoehto tekstin louhintaan, sillä se sopii paremmin pitkien dokumenttien läpikäyntiin. CNN koettiin suoriutuvan paremmin, kun CNN oppii sananmerkityksen itse, eikä käytä hyödykseen valmiiksi koulutettuja sanavektoreita.

Kun puhutaan tekoälyn potentiaalista kyberrikostoiminnassa, on hyvä ottaa huo-

Taulukko 4.1: Tutkielmaan käytettyjen tutkimusten käytetyt algoritmit ja niiden käyttökohteet

Tutkimus	Julkaisuvuosi	Tekoälyn menetelmä	Mihin tarkoitukseen?
Morris ja muut [78]	2012	SVM	Tekstinkäsittelyyn (klassifikaatioon); Seksuaalisesti hyväksikäyttävän tekstin tunnistus
Ebrahimi ja muut [73]	2016	CNN (vain yksi konvoluutio-kerros)	Tekstinkäsittelyyn (klassifikaatioon); Lapsia hyväksikäyttävän keskustelun tunnistus
Vitorino ja muut [76]	2018	CNN	Kuvien analysointiin; Lapsia hyväksikäyttävän pornografisen materiaalin tunnistus
			Videoiden analysointiin; Lapsia hyväksikäyttävän pornografisen videomateriaalin tunnistus
		SVM	Implementoidaan CNN menetelmässä; hyödynnetään klassifikoinnissa
Anderson ja muut [75]	2019	SVMn muunnelmia LCDFTSVM ja RCDFTSVM	Tekstikäsittelyyn (klassifikaatioon); Seksuaalisesti houkuttelevien (engl. online grooming) tekstien tunnistus
Gandhi [53]	2020	SVM	Tekstikäsittelyyn (klassifikaatioon); Seksuaalisen härinnän tunnistamiseen
Jevremovic ja muut [77]	2021	CASPER	Kuvien analysointiin; Pornografisen materiaalin
			Äänen tunnistus ja analysointi
			Tekstikäsittelyyn (klassifikaatioon); Seksuaalisesti houkuttelevien (engl. online grooming) tekstien tunnistus, myös tukee useampaa kieltä

mioon nykyinen tilanne. Stephan Raajimakers otti kantaa tekoälyn nykytilanteeseen kirjoituksessaan [29]. Raajimakers toteaa, että nykypäivänä tekoäly on vielä puutteellinen ja tekee paljon virheitä. Raajimakers toteaa CNN pohjautuvan vielä epäintuitiivisiin ja kognitiivisesti epäuskottaviin operaatioihin. Toinen huomio tekoälyyn kokonaisuutena on, että tekoäly on vielä vaikeasti selitettävissä ihmisille, mikä vaikeuttaa rikostutkinnassa tekoälyn luotettavuutta. Sen lisäksi, että täytyy rikostutkijoiden pystyä tulkitsemaan tekoälyn syöttämiä tuloksia oikein, tulee se myös pystyä selittämään selkeästi oikeuskelpoisena todistuksena. Raajimakers huomioi myös tekoälyn taipumusta ennakkoluuloon (tekoälyn ennakkoluulosta on puhuttu enemmän kappaleessa 4.1). Varsinkin, kun kyse on rikostutkinnasta, tulee tekoälyn olla ennakkoluuloton. Raajimakers otti myös huomioon, että kaikilla tekoälymalleilla on ennusteita tehdessään synnynnäisiä virheitä ja oletuksia. Luokittelijoita voidaan myös tahallisesti väärinkäyttää, pakottamalla tekoälymallin ennakkoluuloisempaan suuntaan. Koska SVM tekee induktiivisia ennusteita, on sen väärinkäyttö helppoa. Viimeisenä huomiona Raajimakers ilmaisee, että myös tekoälyn integroiminen poliisitoimintaan sisällyttää paljon kysymyksiä. Lisäksi tulee tekoälyyn liittyvää lainsäädäntöä pohtia ennen kuin tekoälyä voidaan käyttää hyödyksi rikostutkinnassa. [29]

PrevBot

PrevBot on Sunden tiimin [80] kehittämä konsepti keskustelupalstabotista (engl. chatbot), joka käyttää erilaisia tekoälyn menetelmiä tunnistamaan lapsiin kohdistunutta hyväksikäyttöä. PrevBot *"crime preventive robot"* eli rikoksia ehkäisevä robotti, on suunniteltu poliisin käytettäväksi estämään kestustelupalstoilla tapahtuvaa lapsiin kohdistunutta hyväksikäyttöä. Tällainen teknologia ei ole kuitenkaan kovin uutta. Sunde ja hänen tiiminsä tutkimuksessa [80] viitataan Sweetie 2.0 bottiin[1]

[1]Bart Schermer ja muut *Legal Aspects of Sweetie 2.0* (Springer) `https://link.springer.com/chapter/10.1007/978-94-6265-288-0_1`

[2], jonka tarkoituksena oli estää webbikameralla tapahtuvaa sekstiturismia, ja jonka avulla tunnistettiin noin 1000 rikoksentekijää 71 valtiosta, ja joka päättyi useampaan pidätykseen ja tuomioon. Sunden tiimin PrevBot konseptin päätehtävä on tunnistaa mahdollinen seksuaalinen keskustelu aikuisen ja lapsen väillä. Jos botti tunnistaa keskustelun epäilyttäväksi, arvioi tämä myös, onko keskustelun aikuinen jo ennestään ollut osallisena seksuaalirikoksessa, onko aikuisen ja lapsen välinen keskustelu seksuaalinen ja onko aikuisen sukupuoli eri kuin käyttäjäprofiilissa merkitty sukupuoli. PrevBot on yhdistettynä tietokantaan, johon on tallennettuna lapsiin kohdistuvasta hyväksikäytöstä tuomittujen chat-keskusteluja. Botti voi osallistua useampaan keskusteluun ja tunnistaa epäilyttäviä henkilöitä sekä seuraamalla keskustelua että osallistumalla siihen aktiivisesti. Koska PrevBot on vielä suunnitteilla, ei varsinaisista tekoälymenetelmistä tutkimuksessa ole kerrottu. Tutkimuksessa tulee kuitenkin ilmi, millaisia päätöksiä botin tekeminen vaatisi. Ensimmäisenä tarvitaan hyvin valmistellut ja siistit tietokokonaisuudet tekoälyn opettamiseksi, ja toisena on ominaisuuksien lisääminen, jotka muuntavat raaka-aineiston keskustelupalstoilta botille sopivaksi aineistoksi. Viimeisenä on tekoälymallin luominen, jota opetetaan tietokokonaisuuden avulla. Huomioina Sunde ja hänen kollegansa pohtivat, koska tekoälymalli opetetaan englanninkielisellä aineistolla, osaa tämä vain analysoida englanninkielisiä keskustelualustoja. Mutta jos halutaan botin hyödynnettäväksi myös muissa maissa, tulee tekoälymalli opettaa maakohtaisesti eri kielisillä tietokokonaisuuksilla. [80]

4.4 Pohdinta

Tutkielmalla oli tarkoitus saavuttaa ymmärrystä ja tietoa, sekä lapsiin kohdistuvista vaaroista internetissä, mutta myös tekoälyn hyödyistä lastensuojelussa. Tarkoituksena on avartaa käsitystä internetturvallisuudesta, ja kuinka suojella lapsia

[2] `https://terredeshommes.org`

internetissä. Jotta voidaan turvata lapsille ja nuorille turvallinen internetympäristö, tulee monien tahojen tukea lapsia internetin käytössä. Lapselle tulee opettaa, miten internetissä käyttäydytään, millaisia vaaroja internetissä on, ja millaisia asioita ei kannata julkaista. Tämän vastuun kantaa vanhemmat, joiden tehtävä on perehtyä internetin vaaroihin, ja siten tukea lapsiaan turvalliseen internetkokemukseen.

Osa vastuu kuuluu myös yrityksille, jotka mahdollistavat lapsille ja nuorille keskustelu- ja julkaisualustoja. Keskiviikkona 31.1.2024 Metan, Snapchatin, TikTokin, X:n (entinen Twitter) ja Discordin toimitusjohtajat osallistuivat Yhdysvaltain senaatin oikeuskomitean kuulusteluun, johon vain Metan Mark Zuckerberg ja TikTokin Shou Zi Chew suostuivat vapaaehtoisesti [81]. Kuulustelun aloittivat vanhemmat osoittaen huoltaan ja ärtymystään toimitusjohtajia kohtaan. Nämä vanhemmat olivat lasten vanhempia, jotka olivat menettäneet tai lähes menettäneet lapsensa näiden sosiaalisen median alustojen takia. Vanhemmat jakoivat kokemuksiaan, kuinka heidän lapsensa ovat altistuneet haitalliselle materiaalille, kuten syömishäiriöön ja itsemurhaan kannustavaan materiaaliin, sekä seksuaaliselle hyväksikäytölle. Vanhemmat syyttivät toimitusjohtajia siitä, että he eivät ole pyrkineet tarpeeksi luomaan alustoistaan turvallisia nuorille. [82] Kuulustelussa korostettiin myös seksuaalisen hyväksikäytön seurauksista. Tästä kuulustelun kohteena oli Meta, joka on haastettu oikeuteen kymmenissä osavaltioissa, sillä yritystä on syytetty tekevän Instagramista ja Facebookista lapsille tahallisesti koukuttavia [82]. New Mexicon yleinen syyttäjä haastoi Metan oikeuteen ja syyttää Metaa siitä, että se mahdollisti aikuisia löytämään ja keskustelemaan lasten kanssa ja houkuttelemaan (engl. online grooming) lapsia seksuaaliseen hyväksikäytöön. [83]

Tutkielmassa käytyjen aineistojen ja tutkimuksien pohjalta voidaan päätellä, että kyseessä on vakava ja erittäin tärkeä aihe. Lapset voivat altistua internetissä monille erilaisille vaaroille, joka tulee pyrkiä minimoida eri osapuolilta. Vastuu lasten fyysisestä ja mielenterveydellisestä hyvinvoinnista on vanhemmilla. Vanhemmat

ja huoltajat ovat se osapuoli, joka voi tehokkaasti rajoittaa lapsen internetin käyttöä. Tällä hetkellä vanhempia voi tukea lastensa suojelussa erilaiset työkalut, kuten Googlen Family Link[3], joka mahdollistaa puhelimen käyttöaikarajoituksen, estää mahdollisten haitallisten sovellusten käytön ja yleisesti suojaa lapsia haitalliselta sisällöltä. Tulevaisuudessa on realistista nähdä myös tekoälypohjaisten sovellusten tukevan vanhempia mm. keskustelualustojen analysoinnissa, johon esimerkkinä annettu Googlen Family Link ei vielä pysty.

Tekoäly mahdollistaisi vanhemmille kokopäiväisen tarkkailun, mitä vanhemmat eivät itse pysty suorittamaan. Vaikkakin yritykset eivät itse ole vastuussa lasten hyvinvoinnista, ovat he vastuussa lasten suojelusta alustallaan, varsinkin jos alusta on suunnattu lapsille tai nuorille. Isoilla alustoilla, kuten Meta ja TikTok, työllistetään noin 40 000 moderaattoria, joiden tehtävä on tarkkailla ja arvioida sisällön haitallisuutta, ja puuttua haitalliseen toimintaan, kun taas Snapchat, X ja Discord työllistävät alle 2 300 moderaattoria [81]. Jotta alustan tarkkailusta saadaan tehokkaampaa ja nopeampaa, voisi tekoäly tässä tehtävässä tukea moderaattoreita. Tekoäly ei vielä pysty tehdä omia päätöksiään vaikeissa tilanteissa, joten tässä tapauksessa moderaattorin arviointikykyä tarvitaan. Tekoäly voisi kuitenkin tukea moderaattoreiden työtä analysoimalla isoja määriä dataa, ja löytää mahdollista haitallista materiaalia. Huomioon on myös otettava tekoälyn alttius ennakkoluuloille, josta tulee päästä eroon, ennen kuin tekoälyä voidaan alkaa hyödyntämään isoilla alustoilla.

Tutkielman teema on hyvin tärkeä ja ajankohtainen aihe. Lapsiin kohdistuva hyväksikäyttö on digitalisaation myötä siirtynyt verkkoon, joka mahdollistaa pedofiileille helpon tavan olla yhteyksissä lapsiin [84]. Tekoälyn implementoiminen lastensuojeluun internetissä voisi mahdollistaa kattavamman ja nopeamman suojan lapsille erilaisilla alustoilla. Tämän takia olisi tärkeää tutkia enemmän, kuinka voidaan lapsia suojella tehokkaammin internetissä. Jatkotutkimuksina ehdotetaan tut-

[3] `https://families.google/familylink/`

kimaan enemmän tekoälyä rikostutkinnassa, ja kuinka tekoälystä voitaisiin kehittää, jotta siitä tulisi luotettavampi rikostutkinnassa.

5 Yhteenveto

Digitalisaation myötä yhä useampi lapsi käyttää jokapäiväisessä elämässä internettiä yleistyvien älypuhelimien kautta. Lapset ovat vuosi vuodelta enemmän yhteydessä internettiin, jonka seurauksena, myös lapsiin kohdistuvat rikokset yleistyvät. Tämän vuoksi olisi tärkeää huomioida lasten turvallisuus erilaisilla lapsille ja nuorille kohdistetuilla alustoilla, sekä opettaa lapsia, kuinka pysyä turvassa internetissä. Myös vanhempia tulisi opettaa, kuinka he voivat suojella lapsiaan internetin vaaroilta.

Tutkielman tarkoituksena oli perehtyä, millaisia vaaroja lapset voivat kohdata internetissä, kuinka voidaan ehkäistä lapsiin kohdistuvaa hyväksikäyttöä erilaisilla tekoälyn menetelmillä. Tutkielman aihe on erittäin tärkeä, sillä lasten internetin käytön kasvaessa, olisi tärkeää parantaa myös lastensuojelun menetelmiä, ja pyrkiä tehokkaampaan ja monipuolisempaan tapaan suojella lapsia. Tekonäly voi toimia eräänlaisena apuvälineenä isoihin datamääriin liittyvien ongelmien ratkaisemisessa sekä tukea vanhempia lastensa internetkäyttäytymisen seurannassa. Tutkielma pyrki myös vastaamaan kahteen tutkimuskysymykseen pohjaamalla vastaukset aiheen tutkimuksista ja kirjallisuuskatsauksista (Taulukko 4.1). Tutkimuskysymyksiin vastataan seuraavasti:

TK1 Tekoäly on potentiaalinen apuväline lastensuojelussa, koska se pystyy nopeasti havaitsemaan lapsiin kohdistuvaa epäilyttävää toimintaa ja reagoimaan siihen. Tekoälyä voisivat vanhemmat hyödyntää lastensa keskustelujen kokopäiväisessä seuraamisessa, ja mahdollisten epäilyttävien keskustelujen havaitse-

misessa. Tekoälyllä voitaisiin myös tukea kyberrikostutkijoita mm. datan läpikäynnissä, arviointien tekemisessä, ja epäilyttävien ihmisten identifioinnissa ja paikantamisessa. Kyberrikostutkinnan kannalta on tärkeää ottaa huomioon, että nykyinen tekoäly tekee vielä päätöksiä ennakko-olettamuksia käyttäen. Nykyinen tekoäly ei ole vielä kykenevä suorittamaan haluttuja tehtäviä tarkasti ja luotettavasti. Lisäksi tekoälyn hyödyntäminen rikostutkinnassa vaatisi lainsäädännöllisiä muutoksia.

TK2 Tekoälyn menetelmillä, kuten tukivektorikoneella (SVM), voidaan arvioida lasten keskustelujen sopivuutta, sekä havaitsemaan mahdollisia seksuaalisia sävyjä sisältäviä keskusteluja aikuisen ja lapsen välillä. Tukivektorikoneesta on tehty viimevuosien aikana monia tutkimuksia eri aloilla, ja tukivektorikoneen on huomattu olevan hyvin tehokas tekstiluokittelija. Tukivektorikoneen on myös havainnoitu suoriutuvan hyvin seksuaalisesti epäilyttävän tekstin löytämisessä lapsen ja aikuisen välillä. Varsinkin erilaiset muunnelmat tukivektorikoneesta ovat tuottaneet paljon lupaavia tuloksia. Tukivektorikone on osoittautunut potentiaaliseksi aputyökaluksi epäilyttävän tekstin havainnoinnissa.

Kuva-analysointiin perehtyneet tekoälyn menetelmät, kuten konvoluutioneuroverkot (CNN), voisivat havaita sekä seksuaalista materiaalia lapsen laitteelta, mutta myös lapsiin kohdistuvaa seksuaalisesti hyväksikäyttävää materiaalia internetissä. Konvoluutioneuroverkkojen on myös huomattu suoriutuvan hyvin tekstiluokittelussa, kun kyseessä on yhden konvoluutiokerroksen omaava konvoluutioneuroverkko. Konvoluutioneuroverkot ovat tuottaneet lupaavaa tulosta seksuaalisten kuvien tunnistuksessa sekä videotunnistuksessa. Kuitenkin CNN tutkimuksissa yleisenä haasteena on ollut CNN menetelmän opettaminen. Kyseessä on siis menetelmä, joka on osoittautunut potentiaaliseksi, mutta vaatii vielä jatkotutkimusta. Tekoäly toimii vielä vain työkaluna rikostutkijoille epäilyttävän materiaalin havaitsemisessa. Se ei vielä kykene teke-

mään päätöksiä, vaan päätöksenteko on edelleen ihmisten vastuulla.

Tutkimuskysymyksien vastauksissa käy ilmi, että nykyinen tekoäly ei ole vielä tarpeeksi kyvykäs antamaan luotettavia tuloksia. Sen lisäksi nykyinen tekoäly tekee vielä arvioita ennakkoluulojen perusteella, mikä tulee korjata ennen kuin tekoälyä voidaan käyttää kyberrikostutkinnassa. Lisäksi sen käyttö rikostutkinnassa edellyttäisi lainsäädännöllisiä muutoksia. Tekoälyä tulisi vielä tutkia ja kehittää enemmän, jotta tästä tulisi luotettavampi, tehokkaampi ja monipuolisempi työkalu. Tekoäly kuitenkin osoittautuu tutkimusten perusteella potentiaaliseksi menetelmäksi toimia lasten suojelussa. Vanhemmat voisivat hyödyntää sitä lastensa keskustelujen seuraamiseen ja epäilyttävien tilanteiden havaitsemiseen.

Lähdeluettelo

[1] Our World in Data, *Number of people using the Internet*. url: https://ourworldindata.org/grapher/number-of-internet-users, (accessed: 16.10.2023).

[2] Our World in Data, *Share of US adults who use the Internet, by age*. url: https://ourworldindata.org/grapher/share-us-adults-use-internet-age, (accessed: 16.10.2023).

[3] MLL, *Mediakasvatus varhaiskasvatuksessa*. url: https://www.mll.fi/ammattilaisille/varhaiskasvattajille/mediakasvatus-varhaiskasvatuksessa/, (accessed: 16.10.2023).

[4] C. Tabi ja muut. "Contemporary Issues in Child Protection: Police Use of Artificial Intelligence for Online Child Protection in the UK". (), url: https://www.researchgate.net/publication/366793302_Contemporary_Issues_in_Child_Protection_Police_Use_of_Artificial_Intelligence_for_Online_Child_Protection_in_the_UK. (accessed: 5.11.2023).

[5] S. Fischer, *Kids' daily screen time surges during coronavirus*. url: https://www.axios.com/2020/03/31/kids-screen-time-coronavirus, (accessed: 16.10.2023).

[6] Digitaalinen Helsinki, *Mitä digitalisaatio tarkoittaa?* Url: https://digi.hel.fi/esittely/mika-digi/s, (accessed: 19.10.2023).

[7] J. J. J. Kasvi, TIEKE, *Digi digi digi.* url: https://tieke.fi/digi-digi-digi/, (accessed: 19.10.2023).

[8] EK, *Vahva kyberturvallisuus mahdollistaa kestävän digitalisaation ja vihreän kasvun.* url: https://ek.fi/wp-content/uploads/2022/11/OnePager-Kyberturvallisuus.pdf, (accessed: 30.10.2023).

[9] K. Rousku, Valtioneuvosto, *Tieto- ja kyberturvallisuus luovat pohjaa digitalisoinnillen.* url: https://valtioneuvosto.fi/-/10623/tieto-ja-kyberturvallisuus-luovat-pohjaa-digitalisoinnille, (accessed: 30.10.2023).

[10] Tekoäly.info, *Tekoälyn historia.* url: https://xn--tekoly-eua.info/tekoaly_historia/, (accessed: 12.11.2023).

[11] M. Roser, Our World in Data, *The brief history of artificial intelligence: The world has changed fast – what might be next?* Url: https://ourworldindata.org/brief-history-of-ai, (accessed: 12.11.2023).

[12] McKinsey & Company, *The state of AI in 2023: Generative AI's breakout year.* url: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-AIs-breakout-year, (accessed: 12.11.2023).

[13] C. McMahon, *The Role of Artificial Intelligence in Digital Transformation.* url: https://www.ptc.com/en/blogs/corporate/artificial-intelligence-digital-transformation, (accessed: 12.11.2023).

[14] J. Holmström, *From AI to digital transformation: The AI readiness framework.* url: https://www.sciencedirect.com/science/article/pii/S0007681321000744#sec4, (accessed: 12.11.2023).

[15] Internet Society, *How many children and young people have internet access at home?* Url: https://www.internetsociety.org/wp-content/uploads/2017/11/bp-childrenandtheinternet-20129017-en.pdf, (accessed: 30.10.2023).

[16] unicef, *How many children and young people have internet access at home?* Url: https://data.unicef.org/resources/children-and-young-people-internet-access-at-home-during-covid19/, (accessed: 30.10.2023).

[17] OECD iLibrary, *Chapter 1. Childhood in the digital age.* url: https://www.oecd-ilibrary.org/sites/2d4352c2-en/index.html?itemId=/content/component/2d4352c2-en, (accessed: 5.11.2023).

[18] OECD iLibrary, *Chapter 2. Children and digital technologies: Trends and outcomes.* url: https://www.oecd-ilibrary.org/sites/71b7058a-en/index.html?itemId=/content/component/71b7058a-en, (accessed: 5.11.2023).

[19] M. Prensky, *Digital Natives, Digital Immigrants Part 1.* url: https://www.emerald.com/insight/content/doi/10.1108/10748120110424816/full/html, (accessed: 23.1.2024).

[20] E. J. Helsper, R. Eynon, BERA, *Digital natives: Where is the evidence?* Url: https://bera-journals.onlinelibrary.wiley.com/doi/10.1080/01411920902989227, (accessed: 5.11.2023).

[21] M. Zhou, K. K. L. Lam, *Metacognitive scaffolding for online information search in K-12 and higher education settings: a systematic review.* url: https://link.springer.com/article/10.1007/s11423-019-09646-7#Abs1, (accessed: 5.11.2023).

[22] L. Hietajärvi, *Diginatiiveja ei ole.* url: https://acadsci.fi/sofi/wp-content/uploads/Diginatiiveja_ei_ole_Ilmiokartta_Sofi_2021_Hietajarvi.pdf, (accessed: 5.11.2023).

[23] E. Staksrud ja muut, *What do we know about children's use of online technologies?: a report on data availability and research gaps in Europe [2nd edition].* url: http://eprints.lse.ac.uk/24367/1/What%20do%20we%20know%

20about%20children%E2%80%99s%20use%20of%20online%20technologies%28Isero%29.pdf, (accessed: 5.11.2023).

[24] Europol, *COVID-19 sparks upward trend in cybercrime*. url: https://www.europol.europa.eu/media-press/newsroom/news/covid-19-sparks-upward-trend-in-cybercrime, (accessed: 20.11.2023).

[25] International Telecommunication Union, *Keeping children safe in the digital environment: The importance of protection and empowerment*. url: https://www.itu.int/en/ITU-D/Cybersecurity/Documents/COP/20-00802_COP-Policy_Brief.pdf, (accessed: 20.11.2023).

[26] Europol, *Cybercrime presents a major challenge for law enforcement*. url: https://www.europol.europa.eu/media-press/newsroom/news/cybercrime-presents-major-challenge-for-law-enforcement, (accessed: 20.11.2023).

[27] Interpol, *Cybercrime*. url: https://www.interpol.int/en/Crimes/Cybercrime, (accessed: 20.11.2023).

[28] Europian Parlament, *Artificial Intelligence and Law Enforcement*. url: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf, (accessed: 20.11.2023).

[29] S. Raaijmakers, *Artificial Intelligence for Law Enforcement: Challenges and Opportunities*. url: https://ieeexplore.ieee.org/abstract/document/8821442, (accessed: 20.11.2023).

[30] Microsoft, *The State of Cybercrime*. url: https://www.microsoft.com/fi-fi/security/security-insider/microsoft-digital-defense-report-2023-state-of-cybercrime, (accessed: 20.11.2023).

[31] C. J. Moloney, Ph.D. ja muut, *Assessing Law Enforcement's Cybercrime Capacity and Capability*. url: https://leb.fbi.gov/articles/featured-

articles/assessing-law-enforcements-cybercrime-capacity-and-capability-, (accessed: 20.11.2023).

[32] Kaspersky, *What is Cyber Security?* Url: https://www.kaspersky.com/resource-center/definitions/what-is-cyber-security, (accessed: 20.11.2023).

[33] TeckTarget Network, *What is cyberterrorism?* Url: https://www.techtarget.com/searchsecurity/definition/cyberterrorism, (accessed: 20.11.2023).

[34] LexisNexis® Risk Solutions, *LexisNexis® Risk Solutions Cybercrime Report.* url: https://risk.lexisnexis.co.uk/insights-resources/research/cybercrime-report, (accessed: 30.10.2023).

[35] A. Gaskell, Cybernews, *Who is most vulnerable to cybercrime: new report reveals surprising insights.* url: https://cybernews.com/security/who-is-most-vulnerable-to-cybercrime-new-report-reveals-surprising-insights/, (accessed: 30.10.2023).

[36] E&T, *Younger people more likely to fall victim to cyber crime, survey finds.* url: https://eandt.theiet.org/content/articles/2021/11/younger-people-more-likely-to-fall-victim-to-cybercrime-survey-finds/, (accessed: 30.10.2023).

[37] D. Panhans ja muut, *Why Children Are Unsafe in Cyberspace.* url: https://www.bcg.com/publications/2022/why-children-are-unsafe-in-cyberspace, (accessed: 20.11.2023).

[38] Human Rights Watch, *"How Dare They Peep into My Private Life?"* Url: https://www.hrw.org/report/2022/05/25/how-dare-they-peep-my-private-life/childrens-rights-violations-governments, (accessed: 26.1.2024).

[39] T. O'Dell, A. K. Ghosh, *Online Threats vs. Mitigation Efforts: Keeping Children Safe in the Era of Online Learning*. url: https://ieeexplore.ieee.org/abstract/document/10115142, (accessed: 22.11.2023).

[40] N. Ahmad ja muut, *Cyber Security Situational Awareness among Parents*. url: https://ieeexplore.ieee.org/abstract/document/8626830, (accessed: 15.12.2023).

[41] M. Stoev, D. K. Sarmah, *Online Protection for Children Using a Developed Parental Monitoring Tool*. url: https://www.researchgate.net/publication/372761248_Online_Protection_for_Children_Using_a_Developed_Parental_Monitoring_Tool, (accessed: 5.12.2023).

[42] F. Molloy, *'Sharenting' alert: the risks of sharing pics of your kids online*. url: https://lighthouse.mq.edu.au/article/june/Sharenting-alert-the-risks-of-sharing-pics-of-your-kids-online, (accessed: 5.12.2023).

[43] S. Sokol, *The Problems and Danger of Posting Kids on Social Media*. url: https://www.familyeducation.com/kids/safety/online/the-problems-and-danger-of-posting-kids-on-social-media, (accessed: 5.12.2023).

[44] B. J. Copeland, *artificial intelligence*. url: https://www.britannica.com/technology/artificial-intelligence#ref219078, (accessed: 9.12.2023).

[45] M. Labbe, I. Wigmore, *narrow AI (weak AI)*. url: https://www.techtarget.com/searchenterpriseai/definition/narrow-AI-weak-AI, (accessed: 9.12.2023).

[46] IBM, *What is machine learning?* Url: https://www.ibm.com/topics/machine-learning, (accessed: 9.12.2023).

[47] Columbia Engineering, *Artificial Intelligence (AI) vs. Machine Learning*. url: https://ai.engineering.columbia.edu/ai-vs-machine-learning/, (accessed: 9.12.2023).

[48] J. Barrera, A. Leung, *AI has a racism problem, but fixing it is complicated, say experts.* url: https://www.cbc.ca/news/science/artificial-intelligence-racism-bias-1.6027150, (accessed: 9.12.2023).

[49] G. Lazaro, *Understanding Gender and Racial Bias in AI.* url: https://www.sir.advancedleadership.harvard.edu/articles/understanding-gender-and-racial-bias-in-ai, (accessed: 9.12.2023).

[50] C. L. Dancy, P. K. Saucier, *AI and Blackness: Torward Moving Beyond Bias and Representation.* url: https://ieeexplore.ieee.org/document/9606203, (accessed: 9.12.2023).

[51] Internet Watch Foundation, *How AI is being abused to create child sexual abuse imagery.* url: https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf, (accessed: 9.12.2023).

[52] B. Alam, *SVM Python – Easy Implementation of SVM algorithm.* url: https://hands-on.cloud/svm-python-tutorial/, (accessed: 15.12.2023).

[53] R. Gandhi, *Termination of Cyber-Sexual Harassment and Abuse with Teenagers using Artificial Intelligence.* url: https://ndpublisher.in/admin/issues/EQv11n3c.pdf, (accessed: 15.12.2023).

[54] Jair Cervantes ja muut, *A comprehensive survey on support vector machine classification: Applications, challenges and trends.* url: https://www.sciencedirect.com/science/article/pii/S0925231220307153#ab005, (accessed: 17.4.2024).

[55] GeeksforGeeks, *Support vector machine in Machine Learning.* url: https://www.geeksforgeeks.org/support-vector-machine-in-machine-learning/, (accessed: 15.12.2023).

[56] Visually Explained, *Support Vector Machine (SVM) in 2 minutes*. url: https://www.youtube.com/watch?v=_YPScrckx28&ab_channel=VisuallyExplained, (accessed: 15.12.2023).

[57] A. Zhang ja muut, *Dive into Deep Learning*. url: https://d2l.ai/d2l-en.pdf, (accessed: 15.12.2023).

[58] C. Cortes, V. Vapnik, *Support-vector networks*. url: https://link.springer.com/article/10.1007/BF00994018, (accessed: 15.12.2023).

[59] MLMath.io, *Math behind SVM (Support Vector Machine)*. url: https://ankitnitjsr13.medium.com/math-behind-support-vector-machine-svm-5e7376d0ee4d, (accessed: 15.12.2023).

[60] N. Piepenbreier, *Support Vector Machines (SVM) in Python with Sklearn*. url: https://datagy.io/python-support-vector-machines/, (accessed: 15.12.2023).

[61] D. Nelson, *Mitä tukivektorikoneet ovat?* Url: https://www.unite.ai/fi/mit%C3%A4-ovat-tukivektorikoneet/, (accessed: 15.12.2023).

[62] GeeksforGeeks, *Introduction to Convolution Neural Network*. url: https://www.geeksforgeeks.org/introduction-convolution-neural-network/, (accessed: 31.12.2023).

[63] IBM, *What are convolutional neural networks?* Url: https://www.ibm.com/topics/convolutional-neural-networks, (accessed: 31.12.2023).

[64] Laith Alzubaidi ja muut, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*. url: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8, (accessed: 17.4.2024).

[65] M. McGregor, *What Is a Convolutional Neural Network? A Beginner's Tutorial for Machine Learning and Deep Learning*. url: https://www.freecodecamp.org/news/convolutional-neural-network-tutorial-for-beginners/, (accessed: 31.12.2023).

[66] V. H. Phung, E. J. Rhee, *A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets*. url: https://www.researchgate.net/publication/336805909_A_High-Accuracy_Model_Average_Ensemble_of_Convolutional_Neural_Networks_for_Classification_of_Cloud_Image_Patches_on_Small_Datasets, (accessed: 24.1.2024).

[67] P. Baheti, *Activation Functions in Neural Networks [12 Types & Use Cases] Learning and Deep Learning*. url: https://www.v7labs.com/blog/neural-networks-activation-functions, (accessed: 31.12.2023).

[68] GeeksforGeeks, *Activation functions in Neural Networks*. url: https://www.geeksforgeeks.org/activation-functions-neural-networks/, (accessed: 31.12.2023).

[69] OpenAI, *Rectified Linear Units*. url: https://deepai.org/machine-learning-glossary-and-terms/rectified-linear-units, (accessed: 31.12.2023).

[70] P. Skalski, *Rectified Linear Units*. url: https://towardsdatascience.com/gentle-dive-into-math-behind-convolutional-neural-networks-79a07dd44cf9, (accessed: 31.12.2023).

[71] R. D. Rakshit ja muut, *Cross-resolution face identification using deep-convolutional neural network*. url: https://www.researchgate.net/publication/349921480_Cross-resolution_face_identification_using_deep-convolutional_neural_network, (accessed: 25.1.2024).

[72] Stanford, *CS231n: Deep Learning for Computer Vision*. url: http://cs231n.stanford.edu, (accessed: 25.1.2024).

[73] M. Ebrahimi ja muut, *Detecting predatory conversations in social media by deep Convolutional Neural Networks*. url: https://www.sciencedirect.com/science/article/pii/S1742287616300731, (accessed: 13.1.2024).

[74] H. Lee ja muut, *Detecting child sexual abuse material: A comprehensive survey*. url: https://www.sciencedirect.com/science/article/pii/S2666281720301554#bib8, (accessed: 13.1.2024).

[75] P. Anderson ja muut, *An Intelligent Online Grooming Detection System Using AI Technologies*. url: https://ieeexplore.ieee.org/abstract/document/8858973, (accessed: 13.1.2024).

[76] P. Vitorino ja muut, *Leveraging deep neural networks to fight child pornography in the age of social media*. url: https://www.sciencedirect.com/science/article/pii/S1047320317302377, (accessed: 13.1.2024).

[77] A. Jevremovic ja muut, *Keeping Children Safe Online With Limited Resources: Analyzing What is Seen and Heard*. url: https://ieeexplore.ieee.org/abstract/document/9541357, (accessed: 13.1.2024).

[78] C. Morris ja muut, *Identifying Sexual Predators by SVM Classification with Lexical and Behavioral Features*. url: http://www.cs.utoronto.ca/pub/gh/Morris+Hirst-PAN-2012.pdf, (accessed: 17.1.2024).

[79] H. J. Escalante ja muut, *Sexual predator detection is chats with chained classifiers*. url: https://aclanthology.org/W13-1607.pdf, (accessed: 17.1.2024).

[80] N. Sunde ja I. M. Sunde, *Conceptualizingan AI-basedPolice Robot forPreventing Online ChildSexual ExploitationandAbuse: Part I – The Theoretical and Technical Foundations for PrevBOT*. url: https://www.idunn.no/doi/epdf/10.18261/issn.2703-7045-2021-02-01, (accessed: 17.1.2024).

[81] BBC, *Meta boss Mark Zuckerberg apologises to familis in fiery US Senate hearing*. url: https://www.bbc.com/news/technology-68161632, (accessed: 3.2.2024).

[82] B. Ortutay ja H. Hadero, *Meta, TikTok and other social media CEOs testify in heated Senate hearing on child exploitation*. url: https://www.washingtonpost.com/business/2024/01/31/meta-tiktok-snap-discord-zuckerberg-testify-senate/e15fcfd0-bff5-11ee-a4c6-8f5c350e9316_story.html, (accessed: 3.2.2024).

[83] K. Paul, *Zuckerberg tells parents of social media victims at Senate hearing: 'I'm sorry for everything you've been through'*. url: https://www.theguardian.com/us-news/2024/jan/31/tiktok-meta-x-congress-hearing-child-sexual-exploitation, (accessed: 3.2.2024).

[84] FBI, *Child Predators*. url: https://www.fbi.gov/news/stories/child-predators, (accessed: 21.1.2024).