



Generatiivisen tekoälyn hyödyntäminen parisuhdeväkivallan uhrien auttamisessa

Mia Sjövall

Haaga-Helia ammattikorkeakoulu

Tradenomi, tietojenkäsittely

AMK opinnäytetyö

2025

Tiivistelmä

Tekijä(t) Sjövall Mia
Tutkinto Tradenomi
Raportin/Opinnäytetyön nimi Generatiivisen tekoälyn hyödyntäminen parisuhdeväkivallan uhrien auttamisessa
Sivu- ja liitesivumäärä 33 + 0
<p>Generatiivinen tekoäly on viime vuosina kehittynyt nopeasti ja sen soveltaminen eri aloilla, kuten sosiaalialalla kasvaa jatkuvasti. Tässä opinnäytetyössä tarkastellaan, että miten generatiivista tekoälyä voidaan hyödyntää parisuhdeväkivallan uhrien auttamisessa. Parisuhdeväkivalta on laaja ja vaikeasti tunnistettava ongelma, johon liittyy usein häpeää, pelkoa ja isoa kynnystä hakea apua. Työn taustalla on tarve selvittää, että millä tavoin generatiivinen tekoäly voisi auttaa parisuhdeväkivallan uhreja. Työssä kartoitetaan myös tekoälyn käytön eettisiä haasteita ja mahdollisia riskejä.</p> <p>Opinnäytetyön tavoitteena on tuottaa kattava kuva generatiivisen tekoälyn mahdollisuuksista ja rajoituksista parisuhdeväkivallan uhrien apuna. Työ on rajattu käsittelemään ainoastaan generatiivista tekoälyä kuten kielimalleja, eikä esimerkiksi ennakoivaa analytiikkaa tai valvontateknologiaa tarkastella tässä työssä. Lisäksi työ ei ota kantaa varsinaisten tekoälysovellusten tekniseen toteutukseen, vaan keskittyy niiden käyttömahdollisuuksiin sosiaalisen tuen näkökulmasta. Kohderyhmänä työlle ovat väkivallan uhrit sekä ammattilaiset, jotka toimivat väkivallan uhrien parissa.</p> <p>Tutkimus toteutettiin kuvailevana kirjallisuuskatsauksena, jossa aineistona käytettiin kuutta valikoitua lähdettä. Lähteet valittiin erityisesti niiden ajankohtaisuuden, luotettavuuden ja aiheen relevanttiuden perusteella. Näiden valittujen lähteiden avulla vastattiin kolmeen tutkimuskysymykseen, jotka koskevat generatiivisen tekoälyn hyödyntämistapoja, riskejä ja eettisiä näkökulmia sekä generatiivisen tekoälyn tarjoamia mahdollisuuksia. Käsiteltävä aineisto analysoitiin systemaattisesti ja jokaisen lähteen sisältöä peilattiin valittuihin tutkimuskysymyksiin. Työ toteutettiin vuoden 2025 kevään aikana.</p> <p>Työn johtopäätöksenä voidaan todeta, että generatiivinen tekoäly tarjoaa kiinnostavia mahdollisuuksia parisuhdeväkivallan uhrien tukemiseksi, mutta sen käyttö on vasta alkuvaiheessa. Tekoäly voi tukea uhria tiedonsaannissa ja ohjata uhrin oikeiden palveluiden äärelle, mutta on kuitenkin pidettävä mielessä teknologian rajoitukset. Rajoituksiin sisältyy esimerkiksi empatian puute, virheellisen tiedon riski ja tietosuojakysymykset.</p> <p>Generatiivista tekoälyä voidaan käyttää parisuhdeväkivallan uhrien auttamisessa, mutta se ei voi täysin korvata ihmiskontaktia. Sen sijaan se voi toimia osana laajempaa palveluverkostoa ja toimia täydentävänä välineenä tilanteissa, joissa perinteinen apu ei ole saavutettavissa. Tekoälyn toimivuus uhrien auttamisessa edellyttää tarkkoja eettisiä reunaehdotuksia, ihmiskeskeistä suunnittelua ja viranomaisten aktiivista roolia valvonnassa. Jatkotutkimusta tarvitaan erityisesti siitä, että miten tekoälyyn perustuvat palvelut voidaan integroida osaksi olemassa olevaa tukiverkostoa niin, että ne täydentävät eivätkä korvaa ihmisten tukea.</p>
Asiasanat Tekoäly, generatiivinen tekoäly, parisuhdeväkivalta, lähisuhdeväkivalta

Sisällys

1	Johdanto.....	1
2	Generatiivinen tekoäly.....	2
2.1	Generatiivisen tekoälyn tyypit.....	2
2.2	Suuri kielimalli	3
2.3	GPT-4.....	4
2.4	GPT-4.5.....	5
2.5	Tekoäly keskusteluapuna	6
2.6	Eettiset haasteet ja riskit.....	7
3	Parisuhdeväkivalta	9
3.1	Parisuhdeväkivalta Suomessa	10
3.2	Parisuhdeväkivallan ehkäisy ja auttavat tahot	11
4	Kirjallisuuskatsaus tutkimusmetodina	14
4.1	Kuvaileva kirjallisuuskatsaus	14
4.2	Tutkimuskysymykset ja tiedonhaku.....	15
4.3	ScienceDirect tarkennettu haku.....	16
4.4	Google Search tarkennettu haku.....	19
5	Tulokset.....	22
5.1	Tekoäly keskusteluapuna	22
5.2	Eettinen näkökulma	25
5.3	Tekoäly apuna avun haussa.....	26
6	Generatiivinen tekoäly parisuhdeväkivallan uhrien tukena	30
7	Pohdinta	33
	Lähteet	35

1 Johdanto

Tekoäly on nopeasti kehittyvä teknologinen ilmiö, joka muuttaa monia yhteiskunnan osa-alueita mukaan lukien sosiaali- ja terveystaloudellisia palveluita. Erityisesti generatiivisen tekoälyn mahdollisuudet tarjota tukea ja ohjausta haastavissa tilanteissa ovat herättäneet kasvavaa kiinnostusta.

Generatiivinen tekoäly voi toimia esimerkiksi keskustelurobottina, joka kykenee tuottamaan vuorovaikutteista sisältöä ja tukemaan käyttäjää erilaisissa tarpeissa. Tämän kehityksen myötä herää kysymys siitä, että voidaanko tällaista teknologiaa hyödyntää myös erityisen haavoittuvissa tilanteissa, kuten parisuhdeväkivallan uhrien tukemisessa.

Ajatus aiheen tutkimiseen lähti omasta havainnostani vuoden 2024 aikana, jolloin suomalaisessa uutisoinnissa nousi esiin useita tapauksia, joissa parisuhdeväkivalta johti vakaviin seurauksiin tai jopa kuolemantapauksiin, kuten esimerkiksi Tampereen Hämeenkadun ampuminen loppuvuodesta 2024 (Mantsinen 18.12.2024). Näiden uutisten herättämä huoli sai pohtimaan, että voisiko teknologian kehitys tuoda uusia keinoja ennaltaehkäistä väkivaltaa tai tukea ajoissa apua tarvitsevia henkilöitä.

Työn tavoitteena on selvittää, että millä tavoin generatiivista tekoälyä voidaan hyödyntää parisuhdeväkivallan uhrien tukemisessa, mitä eettisiä kysymyksiä ja riskejä siihen liittyy, sekä miten tekoäly voisi madaltaa uhrien kynnystä hakea apua. Työ liittyy ajankohtaiseen keskusteluun tekoälyn käytön eettisyydestä ja soveltuvuudesta ihmisten hyvinvointia tukevissa palveluissa.

Opinnäytetyö toteutetaan kirjallisuuskatsauksena. Opinnäytetyön avulla kartoitetaan aiempaa tutkimusta ja asiantuntijanäkemyksiä aiheesta. Tiedonhaku toteutettiin tarkennetulla haulla ScienceDirect- tietokannassa sekä Google Search- palvelussa. Menetelmän avulla on pyritty kokoamaan ajankohtaista ja relevanttia tietoa generatiivisen tekoälyn soveltamisesta parisuhdeväkivallan uhrien tukemiseen.

Työn alkuosassa käsitellään generatiivisen tekoälyn peruskäsitteitä, tyyppisiä ja suuria kielimalleja kuten GPT-4 ja GPT-4.5. Lisäksi tarkastellaan tekoälyn roolia keskusteluapuna sekä siihen liittyviä eettisiä haasteita ja riskejä. Tämän jälkeen siirrytään tarkastelemaan parisuhdeväkivallan ilmiötä, sen yleisyyttä Suomessa sekä olemassa olevia tukimuotoja ja palveluita.

Työssä tarkastelu on rajattu koskemaan erityisesti generatiivista tekoälyä eikä esimerkiksi muita tekoälyn sovellusmuotoja käsitellä syvällisesti. Lisäksi huomio kiinnittyy uhrien tukemiseen ja avun saatavuuden parantamiseen. Keskeisiä käsitteitä tässä opinnäytetyössä ovat generatiivinen tekoäly, suuri kielimalli (LLM), keskustelurobotti sekä lähisuhde- ja parisuhdeväkivalta.

2 Generatiivinen tekoäly

Tekoäly on tietokoneiden ja ohjelmistojen kykyä suorittaa tehtäviä, jotka normaalisti vaatisivat ihmisen älykkyyttä. Tämä voi sisältää kykyjä kuten ongelmanratkaisu, päätöksenteko sekä kuvantunnistus. Tekoäly jakautuu useisiin osa-alueisiin, kuten koneoppimiseen, syväoppimiseen ja luonnollisen kielen käsittelyyn. (Monahan s.a.) Rajaan tässä opinnäytetyössä tietoperustan generatiiviseen tekoälyyn.

Generatiivinen tekoäly (Generative AI) tarkoittaa tekoällyn osa-aluetta, jossa luodaan uutta sisältöä, kuten esimerkiksi tekstiä, kuvia tai ääniä (HAMK 2025). Perinteinen tekoäly tarkoittaa tekoälyjärjestelmiä, jotka suorittavat määriteltyjä tehtäviä annettujen sääntöjen perusteella. Perinteisen tekoällyn tekoälyjärjestelmät ovat siis pääasiassa sääntöpohjaisia järjestelmiä, jotka eivät opi annetuista tiedoista, eivätkä osaa tuottaa uutta sisältöä annettujen tietojen perusteella. (SAP s.a.)

Generatiivinen tekoäly hyödyntää syväoppimismalleja, eli tekoäly kehittää itseään paremmaksi jatkuvasti esimerkiksi saadun palautteen kautta. Tiivistettynä Generatiivinen tekoäly ei vain opi ymmärtämään olemassa olevaa dataa, vaan luo uutta sisältöä sen perusteella, että mitä se on oppinut. (HAMK 2025.)

Generatiivista tekoälyä voi käyttää erittäin laajasti erilaisiin tarpeisiin. Muun muassa generatiivista tekoälyä voi hyödyntää taiteen ja viihteen saralla esimerkiksi luomalla käsikirjoituksia elokuvaan tai generatiivinen tekoäly voi toimia arkkitehtien apuna luomalla pohjapiirustuksia, jotka perustuvat asiaankuuluvaan koulutustietoon. Generatiivisesta tekoälystä on hyötyä myös lääketieteessä, jossa generatiivista tekoälyä on käytetty uusien lääkkeiden suunnitteluun. Viestinnässä generatiivista tekoälyä hyödynnetään luonnollisten, ihmisen kaltaisten vastauksien tuottamisessa. Tällä tavoin tekoäly pystyy ylläpitämään luonnollisempia keskusteluja kirjoittajan kanssa. (SAP s.a.)

2.1 Generatiivisen tekoällyn tyypit

Syväoppiminen on koneoppimisen osa-alue. Syvät oppimismallit käyttävät monimutkaisia arkkitehtuureja, ja nämä tunnetaan keinotekoisina neuroverkkoina. Nämä mallit voivat tuottaa erittäin monimutkaisia ja abstrakteja piirteitä datasta, kuten kuvia, tekstiä ja ääntä. Generatiivinen tekoäly toimii yleensä joko generatiivisten kontradiktorisilla verkoilla (GAN), vaihtelevilla autoenkoodereilla (VAE), tai transformer-pohjaisilla malleilla. (SAP s.a.)

Generatiiviset kontradiktoriset verkot (GAN) koostuvat kahdesta pääosasta: generaattoreista sekä syrjivästä tekijästä. Generaattori on vastuussa uusien tietoinstanssien luomisesta, kun taas syrjivä

tekijä arvioi kuinka realistisia luodut instanssit ovat. Käytännössä generaattori pyrkii luomaan niin aidon näköisiä tietoja, että niitä ei pystyisi erottamaan todellisesta datasta ja syrjivän tekijän tehtävä on parantaa kykyään tunnistaa väärennettyjä tietoja. Tämän myötä generaattori oppii tuottamaan erittäin realistista dataa. (SAP s.a.)

Vaihtelevat autoenkooderit (VAE) ovat toisenlainen generatiivisten mallien tyyppi, joka perustuu tilastolliseen päättelyyn. VAE- mallit toimivat siten, että ne koodaavat syötteen piilevään avaruuteen tiivistetyksi esitykseksi, joka on yksinkertaistettu versio alkuperäisestä datasta. Tämän jälkeen tiivistetty esitys dekodataan uudeksi dataksi. Prosessiin sisältyy satunnaisuutta, joka mahdollistaa VAE- mallien luoda monenlaisia mutta silti alkuperäisen datan kaltaisia instansseja. (SAP s.a.)

Transformer-pohjaiset mallit, kuten GPT-3 ja GPT-4 ovat olleet keskeisiä malleja tekstin tuotannossa. Mallit hyödyntävät arkkitehtuuria, joka mahdollistaa koko syötetyn tekstin kontekstin ymmärtämisen. Tämän kyvyn ansiosta ne voivat luoda yhtenäisiä ja sujuvia tekstikokonaisuuksia. (SAP s.a.)

Transformer-pohjaiset mallit käyttävät itsehuomiointimekanismia, jonka avulla malli voi kiinnittää huomiota eri osiin syötteestä samalla kertaa ja arvioida, kuinka tärkeitä ne ovat suhteessa toisiinsa, parantaen siten kontekstin ymmärtämistä ja tiedon käsittelyä. Toinen mekanismi, jota transformer-mallit käyttävät on monipäinen huomiointi, jonka avulla esimerkiksi GPT-4 analysoi tekstin eri osa alueita samanaikaisesti tehostaen tekstin sisäistämistä. (AIMultiple 2025.)

2.2 Suuri kielimalli

Suuri kielimalli (Large Language model, LLM) on tekoälyn osa-alue, joka on suunniteltu käsittelemään, ymmärtämään ja tuottamaan ihmiskieltä. LLM:t hyödyntävät syväoppimista, jossa neuroverkkojen rakenteita hyödynnetään suurien tekstiaineistojen analysointiin. Mallit oppivat tekstin kaavoja ja merkityksiä ja niiden toiminta perustuu siihen, että ne ennustavat käyttäjän syötteeseen sopivia sanoja ja lauseita asiayhteyden perusteella. (SAP 2024.)

LLM:t rakentuvat erityisesti transformer- arkkitehtuurin ja itsehuomiointimekanismien (self-attention) varaan. Näiden ansiosta ne osaavat arvioida syötetyn tiedon eri osien tärkeyttä ja muodostaa johdonmukaisia vastauksia ennustaen paitsi seuraavan sanan, mutta myös kokonaisia kappaleita ja tekstiosioita. Käytännössä LLM:n tehtävänä on päätellä, että mikä sana, lauserakenne tai sisältö sopii parhaiten tiettyyn asiayhteyteen. (SAP 2024.)

LLM:t ovat keskeisessä roolissa generatiivisessa tekoälyssä ja niiden sovelluskohteet ulottuvat laajasti eri aloille. Niitä käytetään esimerkiksi tiedonhakuun, tekstin kääntämiseen, tiivistämiseen,

kysymyksiin vastaamiseen ja uuden sisällön, kuten tekstien, kuvien, musiikin ja ohjelmistokoodin tuottamiseen. LLM- mallien kyky yhdistää suuria tietomääriä, analysoida dataa ja havaita trendejä mahdollistaa niiden käytön monilla asiantuntija-aloilla, kuten esimerkiksi lääketieteessä, rahoituksessa, petoksentorjunnassa ja jopa maataloudessa sekä vähittäiskaupassa. (SAP 2024.)

Vaikka LLM:t osaavat tuottaa vaikuttavan luonnollista kieltä, niiden ymmärrys ei perustu tietoisuuteen vaan opittuihin tilastollisiin yhteyksiin. Mallit ennustavat todennäköisimpiä jatkumojä valtavan tietomäärän perusteella ja niiden tuottama sisältö voi joskus vaikuttaa aidolta ymmärrykseltä ilman, että taustalla olisi todellista ymmärrystä asiasta. (SAP 2024.)

2.3 GPT-4

GPT-4-mallin yksi erottuvista piirteistä on sen kyky ymmärtää ja tuottaa monipuolisia sekä asiayhteyteen sopivia vastauksia. Edeltäjiinsä verrattuna GPT-4 pystyy käsittelemään monimutkaisempia, pidempikestoisia keskustelupohjaisia datapankkeja, kykenee pysymään kontekstissa pidemmän aikaa, tehden siitä loistavan mallin mm. keskustelupohjaisille työkaluille, pitkien tekstien tiivistämiseen ja monimutkaiseen ongelmanratkaisuun. GPT-4:sen yksi tärkeä ominaisuus on myös tekstimassojen analysoinnin lisäksi kuvien prosessointi. GPT-4 kykenee tuottamaan kuvailevaa tekstiä kuvista, joka on hyödyllistä esimerkiksi sairaanhoitoalalla hoitotarveanalyysissä tai luovilla aloilla palautetta varten. (AIMultiple 2025.)

GPT-4:sen suurimpia eroja edeltäjänsä on se, että GPT-4 pystyy käsitellä yli 25 000 sanaa kerralla, tehden siitä mainion työkalun suurien tekstimassojen analysointiin ja pitkien keskustelujen ylläpitämiseen. Edeltäjällensä GPT-3.5:sella sanaraja oli n. 3000. (AIMultiple 2025.)

Kehitystä on tapahtunut lisäksi NLU:ssa sekä ongelmanratkaisukyvyssä. Natural Language Understanding eli NLU kuvaa koneiden kykyä ymmärtää ja analysoida luonnollista kieltä. Tämä prosessi sisältää kielen rakenteiden tunnistamisen, sanojen merkitysten selvittämisen sekä kontekstin huomioon ottamisen, mikä mahdollistaa viestinnän syvällisemmän ymmärtämisen. NLU ei siis rajoitu pelkästään sanojen ja lauseiden tunnistamiseen, vaan se kykenee myös erottamaan hienovaraiset vivahteet ja piilotetut merkitykset, jotka ovat tärkeitä monimutkaisissa viestintätilanteissa. (AIMultiple 2025.)

GPT-4 hyödyntää näitä kehittyneitä NLU-menetelmiä analysoidakseen käyttäjien antamia syötteitä ja tuottaakseen niihin loogisia sekä luonnollisia vastauksia. Järjestelmä osaa käsitellä monimutkaisia kielellisiä rakenteita ja huomioida laajemman keskustelukontekstin, mikä tekee sen vastauksista sekä tarkkoja että mielekkäitä. Näin ollen NLU toimii perustana sille, että GPT-4 pystyy kommunikoimaan sujuvasti ja mukautumaan erilaisiin vuorovaikutustilanteisiin. (AIMultiple 2025.)

GPT-4:n arkkitehtuuri perustuu monikerroksiseen hermoverkkoon, jossa on satoja tai jopa tuhansia itsehuomiomekanismeja (self-attention modules). Näissä kerroksissa malli kykenee tunnistamaan ja mallintamaan sanojen ja lauseiden välisiä monimutkaisia suhteita. Mitä enemmän kerroksia järjestelmässä on, sitä syvällisemmin se pystyy käsittelemään syötteen kontekstia ja tekemään kehittyneempiä ennusteita. Jokainen kerros hyödyntää edellisen kerroksen tuottamaa ymmärrystä ja tarkentaa sitä edelleen, mikä johtaa lopulta entistä tarkempiin ja monipuolisempiin vastauksiin. Tämä kerroksittainen lähestymistapa mahdollistaa sen, että GPT-4 pystyy asteittain kehittämään kielellistä ymmärrystään, mikä on oleellista monimutkaisten kielellisten rakenteiden käsittelyssä ja luonnollisen kielen prosessoinnissa. (AIMultiple 2025.)

2.4 GPT-4.5

OpenAI on julkaissut GPT-4.5-version tutkielmaversion 27.2.2025. Kyseessä on tähän asti suurin ja kehittynein chat-malli, jossa on hyödynnetty sekä esikoulutuksen että jälkikoulutuksen tehokkaampaa skaalausta. Skaalatun ohjaamattoman oppimisen ansiosta malli tunnistaa aiempaa paremmin yhteyksiä ja kaavoja sekä tuottaa luovia oivalluksia ilman varsinaista päättelyä. (OpenAI 2025.)

Alustavissa testeissä GPT-4.5:n kanssa vuorovaikutus tuntuu luonnollisemmalta kuin aiemmilla malleilla. Mallilla on aiempaa laajempi tietopohja, parempi kyky tunnistaa käyttäjän tarkoituksia sekä korkeampi emotionaalinen älykkyys (EQ). Tämän ansiosta se soveltuu erityisen hyvin muun muassa kirjoittamisen ja ohjelmoinnin tukemiseen sekä käytännön ongelmien ratkaisemiseen. Lisäksi mallin odotetaan hallusinoivan aikaisempaa vähemmän, mikä tekee siitä luotettavamman verrattuna aiempiin versioihin. (OpenAI 2025.)

GPT-4.5 julkaistaan tutkimustarkoituksiin, jotta mallin vahvuudet ja rajoitteet tulisivat paremmin esiin ja jotta sen soveltamismahdollisuudet voitaisiin kartoittaa kattavammin. Malli on laajan mittakaavan tekoälymalli, jonka koulutuksessa on hyödynnetty Microsoft Azuren tekoälysupertietokoneita. Sen kehittämisessä on yhdistetty lisääntynyttä laskentatehoa ja suurempaa datamäärää sekä tehty innovaatioita mallin arkkitehtuurissa ja optimointimenetelmissä. Tämä on mahdollistanut ohjaamattoman oppimisen skaalauksen entistä korkeammalle tasolle, mikä näkyy erityisesti mallin kyvyssä analysoida ja tulkita monimutkaisia tietokokonaisuuksia. (OpenAI 2025.)

Tuloksena syntyneellä GPT-4.5-mallilla on aiempaa laajempi tietopohja ja syvempi maailman ymmärrys. Näiden ominaisuuksien ansiosta malli on luotettavampi monilla eri aihealueilla ja soveltuu entistä paremmin vaativiin kielellisiin tehtäviin, kuten käännöksiin, analytiikkaan ja asiantuntijatason tekstien tuottamiseen. (OpenAI 2025.)

GPT-4.5:n julkaisun hyötyjä verrattuna kilpailijoiden malleihin tai OpenAI:n aiempiin tekoälyversioihin ei ole heti havaittavissa. OpenAI:n tutkijat ovat kuitenkin korostaneet mallin parannuksia erityisesti sen intuitiossa ja kyvyssä ymmärtää tunteita. Vaikuttaa siltä, että sen kehityksessä on painotettu ihmismäisiä piirteitä aiempaa enemmän, mikä erottaa sen muista markkinoilla olevista malleista. (OpenAI 2025.)

Vaikka OpenAI on kasvattanut yritysasiakkaidensa määrää, GPT-4.5 ei keskity ensisijaisesti käytännön sovelluksiin vaan muistuttaa enemmän akateemista tutkimushanketta, jossa pyritään kohti yleistä tekoälyä (AGI). Tämä suuntaus voi kuitenkin olla ristiriidassa lyhyen aikavälin liiketoimintatavoitteiden kanssa, sillä malli on kallis ja sen tuomat parannukset huomataan lähinnä kokeneempien käyttäjien toimesta. (OpenAI 2025.)

GPT-4.5 on aluksi saatavilla vain maksullisille käyttäjille, ensin ChatGPT Pro -tilaajille, ja myöhemmin laajennetaan muihin maksullisiin tasoihin. Ilmaiskäyttäjille sen saatavuus on epävarmaa laskentatehon vaatimusten vuoksi. Käyttökokemuksissa se on osoittautunut paremmaksi esimerkiksi otsikoinnissa, jossa se tuotti selkeämpiä ja vaikuttavampia ehdotuksia verrattuna aiempiin versioihin. Tämä viittaa siihen, että mallin kehitys keskittyy entistä enemmän hienovaraisiin kielellisiin ja kognitiivisiin parannuksiin, joiden avulla pyritään tekemään vuorovaikutuksesta entistä luonnollisempaa ja tehokkaampaa. (Wired 2025.)

2.5 Tekoäly keskusteluapuna

Tekoälyn hyödyntäminen keskusteluapuna perustuu käyttäjän tarpeiden ymmärtämiseen ja vuorovaikutuksen suunnitteluun. Prosessi alkaa yleisimpien kysymysten (FAQs) tunnistamisesta, jotka muodostavat tekoälyn vastausten perustan. Näiden kysymysten avulla voidaan kartoittaa käyttäjien tavallisimmat tiedontarpeet ja vähentää asiakaspalveluun kohdistuvaa kuormitusta. FAQ-listan avulla määritellään tekoälyn tavoitteet eli käyttäjän ilmaisemat aikomukset, kuten kirjautumisprosessin ohjeistaminen. Käyttäjä voi esittää saman kysymyksen eri tavoilla, joten tekoälyn on kyettävä tunnistamaan synonyymiset ilmaisut ja erilaiset kieliopilliset rakenteet. (IBM 2021.)

Tekoälymallin kehityksessä huomioidaan myös substantiivit ja avainsanat, jotka liittyvät tiettyihin käyttäjien pyyntöihin. Esimerkiksi pankkiasioinnissa näitä voivat olla tilinumero, salasana ja käyttäjätunnus. Näiden elementtien avulla tekoäly pystyy rakentamaan loogisen keskustelupolun ja tarjoamaan käyttäjän tarpeisiin vastaavia ratkaisuja. (IBM 2021.)

Keskustelevan tekoälyn sovelluskohteet ovat laajat. Sen yleisin käyttömuoto on asiakaspalvelun automatisointi, jossa chatbotit korvaavat ihmistyöntekijöitä vastaamalla kysymyksiin ja suosittelemalla tuotteita. Lisäksi tekoäly voi edistää saavutettavuutta esimerkiksi tekstistä puheeksi

diktointiominaisuuksien ja kielikäännösten avulla. Myös henkilöstöhallinnossa, terveydenhuollossa ja IoT-laitteissa keskustelevaa tekoälyä hyödynnetään prosessien tehostamiseen. (IBM 2021.)

Tekoälyn merkittävimmät hyödyt liittyvät kustannustehokkuuteen, asiakaskokemuksen parantamiseen ja skaalautuvuuteen. Chatbotit voivat tarjota asiakaspalvelua ympäri vuorokauden ilman lisäkustannuksia. Lisäksi ne voivat käsitellä suuria tietomääriä ja tarjota käyttäjäkohtaisia suosituksia, mikä voi lisätä myyntiä ja asiakasuskollisuutta. (IBM 2021.)

Teknologian haasteisiin kuuluu muun muassa kielellisen syötteen tulkinta, sillä murteet, slangisanat ja tunnesävyt voivat vaikeuttaa tekoälyn kykyä ymmärtää käyttäjän tarkoitusta. Lisäksi yksityisyys ja tietoturva ovat keskeisiä huolenaiheita, sillä tekoälyjärjestelmien on käsiteltävä ja suojattava käyttäjätietoja asianmukaisesti. Myös käyttäjien asenteet voivat vaikuttaa teknologian käyttöönottoon, sillä epäluottamus koneelliseen viestintään voi vähentää järjestelmän tehokkuutta. (IBM 2021.)

Kokonaisuudessaan keskusteleva tekoäly tarjoaa yrityksille merkittäviä mahdollisuuksia parantaa tehokkuutta ja asiakaspalvelua, mutta sen kehittämisessä ja käyttöönotossa on otettava huomioon sekä teknologiset että sosiaaliset haasteet. (IBM 2021.)

2.6 Eettiset haasteet ja riskit

Generatiivisen tekoälyn käyttöön liittyy useita eettisiä sekä liiketoiminnallisia riskejä, jotka eroavat merkittävästi muista tekoälyteknologioista. Generatiivisen tekoälyn järjestelmät voivat luoda vahingollista sisältöä joko tahattomasti tai tarkoituksellisesti, mikä edellyttää jatkuvaa valvontaa sen käytössä. Erityisesti automaattisesti tuotetuissa viesteissä voi esiintyä epäasiallista tai harhaanjohtavaa sisältöä. (TechTarget 2025.)

Koska tekoälymallit koulutetaan valtavilla datamäärillä, joiden alkuperä ei ole aina tiedossa, yritykset altistuvat huomattaville tekijänoikeus- ja mainehaitoille. Myös henkilötietojen suojaaminen on haastavaa, koska kuluttajien voi olla vaikeaa selvittää ja poistaa omia tietojaan kielimalleista. Yritysten tulee varmistaa, ettei henkilötietoja jää kielimalleihin. (TechTarget 2025.)

Generatiivisen tekoälyn käyttö lisää myös riskiä luottamuksellisen tiedon vuotamisesta, mikä voi vahingoittaa pysyvästi asiakkaiden ja työntekijöiden luottamusta sekä johtaa oikeudellisiin seurauksiin. Selkeiden käytäntöjen ja vastuun jakamisen merkitys korostuu, kun pyritään suojaamaan arkaluontoisia tietoja. (TechTarget 2025.)

Lisäksi generatiivinen tekoäly voi voimistaa olemassa olevia vinoumia, mikä vaatii organisaatioilta monimuotoista arviointia sekä mallien tarkkaa seuranta. Työvoiman näkökulmasta Generatiivisen

tekoälyn nopea kehitys vaikuttaa työntekijöiden rooleihin ja moraaliin, mikä edellyttää työntekijöiden jatkuvaa osaamisen kehittämistä sekä uudenlaisia työtehtäviä kuten tekoälyn ohjaamista ja valvontaa. (TechTarget 2025.)

Myös datan alkuperän epävarmuus vaarantaa sisältöjen tarkkuutta ja luotettavuutta. Mallien läpinäkymättömyys vaikeuttaa niiden antamien tulosten tulkintaa, jolloin inhimillinen arviointi mallin tuottamien vastausten osalta pysyy välttämättömänä. Lisäksi generatiivisten mallien kyky tuottaa uskottavalta näyttävää mutta virheellistä sisältöä eli hallusinaatioita altistaa yrityksiä ja käyttäjiä erilaisille riskeille. (TechTarget 2025.)

Tekoälyn kehittämisen suuri energiantarve lisää myös ympäristövaikutuksia, kuten hiilijalanjälkeä. Generatiivisella tekoälyllä on mahdollisia poliittisia vaikutuksia, kuten polarisaation kärjistyminen algoritmien painottaessa jakavaa sisältöä rakentavan sijaan. Tämä haastaa yhteiskuntia määrittämään, kuinka tekoälyä voidaan hyödyntää vastuullisesti yhteisen hyvän edistämiseksi. (TechTarget 2025.)

Uudemmat generatiivisen tekoälyn mallit kuitenkin pyrkivät ennaltaehkäisemään tekoälyn käyttöä ns. harmaalla alueella. GPT-4:sen yhtenä tärkeimpänä ominaisuutena pidetään sen kykyä tunnistaa epäsoveliaat pyynnöt, kuten vaikka ohjeita pommin rakentamiseen tai murha-aseen piilottamiseen. Edeltäjiänsä verrattuna, GPT-4 pystyy n. 82 % todennäköisyydellä karsimaan kysymyksiä liittyen epäsoveliaisiin ja kiellettyihin aiheisiin. (AIMultiple 2025.)

3 Parisuhdeväkivalta

Parisuhdeväkivalta voidaan määritellä käyttäytymiskaavaksi parisuhteessa, jonka tavoitteena on saada tai ylläpitää valtaa ja kontrollia kumppaniin. Parisuhdeväkivalta voi olla fyysistä, seksuaalista, emotionaalista, taloudellista tai psykologista toimintaa tai sen uhkaamista. Parisuhdeväkivaltaan sisältyy kaikki sellainen käyttäytyminen, joka pelottaa, uhkaa, terrorisoi, manipuloi, vahingoittaa, nöyryyttää, syyllistää tai loukkaa toista osapuolta. Yksittäiset väkivaltatapaukset ovat harvoin erillisiä, vaan ne yleensä lisääntyvät sekä tiheydeltään, että vakavuudeltaan ajan mittaan. Parisuhdeväkivaltaa voi kokea kuka tahansa riippumatta ulkoisista tekijöistä tai taustasta ja sitä voi tapahtua monenlaisissa ihmissuhteissa. Ääritapauksissa parisuhdeväkivalta voi johtaa vakavaan fyysiseen vammaan tai kuolemaan. (United Nations s.a.)

Parisuhdeväkivalta voi olla vaikeasti tunnistettavissa silloin, jos parisuhdeväkivallan uhri ei itsekään tunnista olevansa parisuhdeväkivallan uhri. Parisuhdeväkivallan uhri saattaa pohtia, että onko hän vain ylireagoimassa tilanteeseen sen sijaan, että tunnistaisi tilanteet väkivaltaiseksi. (Naisten Linja 2024.)

Parisuhdeväkivallan ensimmäisiin merkkeihin voi kuulua esimerkiksi liiallinen huomionsoittaminen ja tavoite edetä suhteessa nopeasti. Tähän voi liittyä myös jatkuva yhteydenotto toiseen osapuoleen ja vahvat tunnereaktiot, jos yhteydenottoihin ei vastata riittävän nopeasti. Tämän lisäksi väkivaltainen kumppani usein pyrkii kontrolloimaan toisen osapuolen viestintää sekä ystävyyssuhteita. (Naisten Linja 2024.)

Parisuhdeväkivaltaan voi liittyä toisen osapuolen rajojen kunnioittamatta jättäminen sekä jatkuvat yritykset neuvotella uudelleen jo asetettuja rajoja. Parisuhdeväkivallan uhria saatetaan painostaa toimimaan tavalla, joka ei tunnu hänestä hyvältä. Manipuloiva käytös voi ilmetä esimerkiksi niin, että uhria syyllistetään siitä, ettei hän välitä riittävästi, ellei suostu kumppanin vaatimuksiin. Uhrin omia tunteita voidaan vähätellä ja hänelle aiheutetaan herkästi syyllisyyden tunnetta tilanteista, joihin hän ei ole syyllinen. Lisäksi kumppani voi toistuvasti pyrkiä kritisoimaan ja nöyryyttämään uhria julkisesti ja pyrkii eristämään sinua läheisistäsi. (Naisten Linja 2024.)

Väkivaltaisen parisuhteen tunnusmerkkejä ovat myös toistuvat riitatilanteet ilman selkeää syytä. Kumppani saattaa myös käyttäytyä aggressiivisesti tai uhkaavasti, rikkoa esineitä tai uhata fyysisellä väkivallalla. Riitatilanteissa väkivallan uhri voi usein tuntea, ettei uskalla ilmaista omaa näkökulmaansa, sillä pelkää kumppanin reaktiota. Fyysinen väkivalta voi ensin alkaa lievempinä tekoina, kuten kiinni tarttumisena tai tönimisenä, mutta väkivalta voi kehittyä myös esimerkiksi lyömiseen, potkimiseen, hiuksista repimiseen tai kuristamiseen. Tekijä voi myös pyrkiä kääntämään tilanteet niin, että väkivallan uhri tuntee olevansa vastuussa tekijän käytöksestä,

jolloin uhri saattaa alkaa kyseenalaistamaan omia kokemuksiaan ja tunteitaan. (Naisten Linja 2024.)

Väkivallalla on taipumus vaikuttaa syvästi kielteisellä tavalla uhrien identiteetin tunteeseen ja heidän käsitykseensä omasta paikastaan maailmassa. Vaikka väkivallan muodoissa esiintyy vaihtelua, on väkivalta useimmiten prosessi pikemminkin kuin yksittäinen tapahtuma. Tämä ei tarkoita, että väkivallan kehittyminen veisi välttämättä vuosia, vaikka näin usein onkin. (Babushkina & de Boer 2024.)

3.1 Parisuhdeväkivalta Suomessa

Vuonna 2023 Suomen viranomaisten tietoon tuli 12 300 pari- ja lähisuhdeväkivallan uhria, mikä on 5,3 % enemmän kuin vuonna 2022. Näistä uhreista vajaat 9 300 oli aikuisia, ja heidän määränsä kasvoi edellisvuodesta 3,2 %. Alaikäisiä uhreja oli noin 3 100, mikä tarkoittaa 12,2 % kasvua verrattuna vuoteen 2022. Aviopuolison tekemissä parisuhdeväkivaltatilanteissa uhrien määrä kasvoi 10 % vuoteen 2022 verrattuna. Pari- ja lähisuhdeväkivaltatapauksissa vuonna 2023 syylliseksi epäillyistä 75 % oli miehiä. Mikäli uhri oli alaikäinen, syylliseksi epäillyistä miehiä oli 65 %. Aikuisista uhreista 74 % on ollut naisia. Vuonna 2023 viranomaisten tietoon tulleista yli 18-vuotiaista uhreista 6 800 on ollut naisia, kun taas miesuhreja on ollut vajaat 2 500. Naisuhrien määrä on kasvanut 1,3 % edellisvuoteen verrattuna ja miesuhrien määrä on kasvanut 9,1 %. Täysi-ikäisiin uhreihin kohdistuvasta pari- ja lähisuhdeväkivallasta 49 % oli nykyisen avio- tai avopuolison tekemää väkivaltaa. Tilastossa näkyvät ainoastaan ne tapaukset, jotka ovat tulleet viranomaisten tietoon, mutta suuri osa perhe- ja lähisuhdeväkivallasta jää ilmoittamatta. (Tilastokeskus 2024.)

Väkivalta ja häirintä ovat sukupuolittuneita ilmiöitä, jotka kohdistuvat erityisesti naisiin. Naiset kohtaavat väkivaltaa usein kotona, työpaikoilla ja muissa yksityisissä tiloissa, sekä yhä useammin myös verkossa. Väkivaltaa esiintyy kaikissa yhteiskuntaluokissa. Naisiin kohdistuvan väkivallan tekijänä on useimmiten nykyinen tai entinen kumppani, kun taas miehet kokevat väkivaltaa useammin julkisissa tiloissa ja tekijänä on toinen mies. Toistuva ja uusiutuva parisuhdeväkivalta kohdistuu erityisesti naisiin. Ilmiön taustalla on sukupuolten välinen epätasa-arvo ja käsitykset naisten alemmasta asemasta, johon vaikuttavat sekä kulttuuriset että fyysiset erot. Suomessa on 2010-luvulla surmattu keskimäärin noin 15 naista vuosittain nykyisen tai entisen kumppanin toimesta. (Naisten Linja 2024.)

Naisiin kohdistuvan väkivallan ja perheväkivallan ehkäisemiseksi Suomi liittyi Istanbulin sopimukseen vuonna 2015. Sopimus velvoittaa ehkäisemään väkivaltaa, suojelemaan uhreja ja tarjoamaan heille tukipalveluita, sekä asettamaan tekijät vastuuseen teoistaan. Tilastojen ja

asiantuntijoiden mukaan ei ole kuitenkaan viitteitä siihen, että naisiin kohdistuva väkivalta olisi vähentymässä. Vuonna 2022 Ensi- ja turvakotien liitto auttoi yli 11 000 lähisuhdeväkivallan uhria, kun vuonna 2019 vastaava määrä oli alle 8000. Vuonna 2022 turvakotien asiakkaista yli puolet oli aikuisia, joista valtaosa (91 %) oli naisia. Asiakasmäärät kasvoivat neljä prosenttia edellisvuoteen verrattuna. (Turun Ensi- ja Turvakoti Ry 2023.)

Fyysinen parisuhdeväkivalta naisia kohtaan aiheuttaa Suomessa arviolta 150 miljoonan euron lisäkustannukset, mutta todelliset kulut sosiaali-, terveys- ja oikeuspalveluissa ovat tätäkin suuremmat. Vaikka Istanbulin sopimus on ollut voimassa jo useita vuosia, Suomi on edelleen Euroopan toiseksi turvattomin maa naisille. Yli puolet naisista (57 %) on kokenut fyysistä tai seksuaalista väkivaltaa tai sen uhkailua. Henkistä väkivaltaa parisuhteessa on kokenut joka toinen suomalainen nainen ja lähes 30 % naisista on joutunut vainon kohteeksi elämänsä aikana. (Turun Ensi- ja Turvakoti Ry 2023.)

3.2 Parisuhdeväkivallan ehkäisy ja auttavat tahot

Kansainväliset ihmisoikeussopimukset ohjaavat kansallista työtä lähisuhde- ja perheväkivallan ehkäisemiseksi. Tärkeimpiä sopimuksia on Istanbulin sopimus, joka on Euroopan neuvoston yleissopimus naisiin kohdistuvan ja perheväkivallan ehkäisemisestä ja torjumisesta. Suomi allekirjoitti sopimuksen vuonna 2011 ja se astui voimaan vuonna 2015. Sopimuksen toimeenpanoa ohjaavat suunnitelmat on laadittu sosiaali- ja terveysministeriön yhteydessä toimivan NAPE-toimikunnan toimesta. Toimeenpanossa on myös huomioitu Euroopan neuvoston GREVIO-asiantuntijaryhmän Suomelle antamat suositukset vuodelta 2019. Lisäksi kesäkuussa 2024 astui voimaan EU:n direktiivi naisiin kohdistuvan väkivallan ehkäisemisestä ja torjumisesta. (Helsingin kaupunki 2024.)

Lainsäädäntö Suomessa ohjaa myös lähisuhdeväkivallan ehkäisyä ja torjuntaa. Nykyinen sääntely on kuitenkin hajanaista ja siksi on ehdotettu erillisen lähisuhdeväkivallan ehkäisyä koskevan lain säätämistä. Perustuslaki turvaa jokaiselle oikeuden vapauteen, koskemattomuuteen ja yhdenvertaisuuteen. Lisäksi sosiaalihuoltolaki, terveydenhuoltolaki, lastensuojelulaki, turvakotilaki ja rikoslaki määrittelevät väkivaltaan liittyviä viranomaisvelvoitteita. Hyvinvointialueiden toimintaa ohjaa sosiaali- ja terveydenhuollon järjestämislaki. (Helsingin kaupunki 2024.)

Keskeisiä kansallisia ohjelmia lähisuhdeväkivallan torjumiseksi ovat muun muassa Istanbulin sopimuksen toimeenpanosuunnitelma 2022–2025, Naisiin kohdistuvan väkivallan torjuntaohjelma 2020–2023 sekä useat muut ohjelmat, kuten Väkivallaton lapsuus- toimenpideohjelma ja turvallisuussuunnittelun kansalliset linjaukset. Eri ohjelmissa on pyritty välttämään

päällekkäisyyksiä ja niiden tavoitteet ja toimenpiteet painottuvat eri näkökulmiin. (Helsingin kaupunki 2024.)

Lähisuhdeväkivallan ehkäisyyn voidaan vaikuttaa vahvistamalla lasten ja nuorten turvataitoja sekä edistämällä sukupuolten tasa-arvoa ja yhdenvertaisuutta. Helsingin kaupunki on sitoutunut tunne- ja vuorovaikutustaitojen opetuksen vahvistamiseen ja laajentamiseen kouluissa. Helsingin kaupungin perusopetuksen strategiassa vuosille 2021–2025 tavoitteena on, että kaikissa kaupungin kouluissa opetetaan näitä taitoja tutkittuihin ohjelmiin perustuen. Koulut saavat itse valita, että mitä ohjelmaa ne käyttävät ja opetusta toteutetaan kaikissa Helsingin 102 peruskoulussa. (Helsingin kaupunki 2024.)

Sosiaali- ja terveysministeriö vastaa Suomessa lähisuhde- ja perheväkivallan vastaisen työn suunnittelusta ja ohjauksesta. Ministeriön tavoitteena on väkivallan ennaltaehkäisy sekä sellaisen palvelujärjestelmän kehittäminen, joka pystyy tarjoamaan apua ja tukea kaikille väkivallan osapuolille, riippumatta heidän asuinpaikastaan. Hyvinvointialueet vastaavat sosiaali- ja terveyspalveluiden tarjoamisesta väkivallan uhreille ja valtion vastuulla on turvakotipalveluiden sekä Nollalinjan auttavan puhelimen järjestäminen. (Sosiaali- ja terveysministeriö 2023.)

Turvakodit tarjoavat suojaa ja tukea kaikille, jotka ovat kokeneet lähisuhdeväkivaltaa tai sen uhkaa. Turvakodissa asiakkaille tarjotaan turvallista asumista ympäri vuorokauden, kriisiapua ja psykososiaalista tukea. Palvelut ovat asiakkaille maksuttomia ja turvakotiin voi hakeutua oma-aloitteisesti. (Sosiaali- ja terveysministeriö 2023.)

Nollalinja (puhelinnumero 080 005 005) tarjoaa tukea lähisuhdeväkivaltaa tai naisiin kohdistuvaa väkivaltaa kokeneille, heidän läheisilleen sekä väkivaltatyötä tekeville ammattilaisille. Toiminnan järjestämisestä vastaa Terveyden ja hyvinvoinnin laitos (THL) ja palvelun toteuttaa Settlementti Tampere ry. Oikeusministeriö rahoittaa Nollalinjan toimintaa. (Sosiaali- ja terveysministeriö 2023.)

Seri-tukikeskukset tarjoavat apua yli 16-vuotiaille seksuaaliväkivallan uhreille. Tukikeskukseen voi hakeutua yksin, läheisten kanssa tai viranomaisen ohjaamana. Seri-keskuksista saa muun muassa oikeuslääketieteellisiä tutkimuksia, traumatukea, psykologista neuvontaa ja terapiaa. (Sosiaali- ja terveysministeriö 2023.)

Olennaiset teknologiset resurssit ja palvelut keskittyvät pääasiassa matkapuhelimiin, jotka toimivat ensimmäisenä yhteydenottokanavana ja tarjoavat pääsyn laajaan valikoimaan palveluita. Epäviralliset itseapupalvelut ja pääsy vertaistukiryhmiin voivat myös tapahtua puhelimen kautta, mutta on kuitenkin muistettava, että jokaisella parisuhdeväkivallan uhrilla ei ole omaa puhelinta. Teknologian saatavuus mahdollistaa uhrien kouluttamisen, jotta he voivat tunnistaa väkivaltaisen

suhteen ja hakea tukea, saada oikeudellisia palveluita ja neuvoja sekä mielenterveyspalveluita useimmiten teknologiaan perustuvan tuen muodossa. (Novitzky, Janssen & Kokkeler 2023.)

Näiden palveluiden tarjoaminen verkossa anonyymin pääsyn kautta on hyödyllistä, sillä kasvokkain tapahtuvat kohtaamisen saattavat usein aiheuttaa nöyryytystä sekä syyllisyyden ja häpeän tunteita parisuhdeväkivallan uhreille. Teknologinen tuki voi myös auttaa vähentämään parisuhdeväkivallan uhrien tuntemaa eristyneisyyttä ja toivottomuutta tarjoamalla pääsyn esimerkiksi tukiryhmiin.

Australiassa ja Yhdysvalloissa puhelinyhtiöt ovat tarjonneet ilmaista puhelinyhteyttä naisille, jotka kokevat parisuhdeväkivaltaa. Kuitenkin parisuhdeväkivallan tekijät voivat käyttää taktiikoita kumppanin eristämiseen kontrolloimalla sitä, että mitä uhri tekee, mihin hänellä on pääsy tai kenen kanssa uhri on yhteyksissä. Tämä rajoittaa entisestään vuorovaikutusta ja parisuhdeväkivallan tekijä usein selittää näitä toimia mustasukkaisuudella. (Novitzky ym. 2023.)

4 Kirjallisuuskatsaus tutkimusmetodina

Kirjallisuuskatsaus on järjestelmällinen ja kriittinen tapa kerätä ja analysoida kirjallisuutta selkeästi määritellyn tutkimusongelman pohjalta. Hakuprosessi suunnitellaan etukäteen huolellisesti ja siten, että prosessi on toistettavissa. Kirjallisuuskatsauksen avulla voidaan rakentaa ja arvioida teoriaa, luoda kokonaiskuvaa tutkimusongelmasta sekä tunnistaa kehittämiskohteita.

Ammattikorkeakouluissa kirjallisuuskatsauksen tavoitteena on usein tuottaa kokonaiskuva aiheesta esimerkiksi hankkeita tai toimeksiantajia varten. (Kreodi 2022.)

Kirjallisuuskatsaus on tutkimus tutkimuksesta, eli sen avulla kootaan aiempien tutkimuksien tuloksia, jotka toimivat perustana uuden tiedon tuottamiselle. Kirjallisuuskatsaus pohjautuu korkealaatuisista alkuperäistutkimuksista tehtyihin johtopäätöksiin. Kirjallisuuskatsausta ei pidä sekoittaa yleiskatsaukseen, pelkkään lähdeluetteloon tai kirja-arvosteluun, sillä niistä puuttuu katsaukselle olennainen kriittinen arviointi. (Salminen 2023).

4.1 Kuvaileva kirjallisuuskatsaus

Tässä opinnäytetyössä käytetään kuvailevaa kirjallisuuskatsausta. Kuvaileva kirjallisuuskatsaus on kirjallisuuskatsausten perustyyppi, jossa tutkimusaihetta tarkastellaan laaja-alaisesti ilman tiukkoja menetelmällisiä rajoituksia. Sen avulla voidaan hyödyntää laajoja aineistoja ja siitä on olemassa useita alatyyppejä, kuten narratiivinen, integroiva ja scoping-katsaus. (Kreodi 2022.)

Narratiivinen katsaus tarjoaa yleiskuvan tutkimusaiheesta tai sen kehityksestä. Se on menetelmällisesti kevyin muoto ja sen tavoitteena on helppolukuinen ja kertova kokonaisuus. Tiedonhaun ja aineiston käsittelyn vaiheet eivät yleensä ole tarkasti kuvattuja narratiivisessa katsauksessa. (Kreodi 2022.)

Integroiva katsaus yhdistää sekä kuvailevan että systemaattisen katsauksen piirteitä. Siihen sisältyy kriittinen aineiston arviointi ja synteessin muodostaminen eri menetelmiä hyödyntäen. Joidenkin tutkijoiden mukaan se voidaan nähdä systemaattisena katsauksena. (Kreodi 2022.)

Scoping-katsaus kartoittaa laajasti olemassa olevaa kirjallisuutta ja tarjoaa yleiskuvan aihealueesta. Se sopii erityisesti laajojen aihekokonaisuuksien tutkimiseen ja tutkimuskysymysten jäsentämiseen esimerkiksi hanketyössä. (Kreodi 2022.)

Tässä opinnäytetyössä käytetään kuvailevaa kirjallisuuskatsausta ilman tarkkaa rajautumista sen yksittäisiin alalajeihin. Koska tutkimuskysymyksiin vastataan joustavasti ilman sitoutumista tietyn alalajin reunaehtoihin, viitataan yleisellä tasolla kuvailevaan kirjallisuuskatsaukseen.

Menetelmällinen eteneminen ja lähteiden käsittely kuvataan kuitenkin yksityiskohtaisesti työn edetessä.

4.2 Tutkimuskysymykset ja tiedonhaku

Opinnäytetyössä pyritään selvittämään, että voisiko generatiivista tekoälyä hyödyntää parisuhdeväkivallan ehkäisyssä. Asian selvittämiseksi opinnäytetyössä on käytetty hyödyksi seuraavia alakysymyksiä:

- Mikä tekoälyn muoto on paras keskusteluapuna?
- Mitä eettisiä näkökulmia tulisi ottaa huomioon hyödyntäessä generatiivista tekoälyä parisuhdeväkivallan uhrien auttamisessa?
- Miten generatiivisen tekoälyn avulla voitaisiin madaltaa uhrien kynnystä hakea apua väkivaltatilanteissa?

Tässä opinnäytetyössä tiedonhaku toteutettiin hyödyntämällä yhtä tieteellistä tietokantaa sekä yhtä hakukonetta. Tietokannaksi valittiin ScienceDirect, joka sisältää laajan kokoelman tutkimusartikkeleita eri tieteenaloilta, kuten yhteiskunta- ja käyttäytymistieteistä, teknologiasta sekä terveystieteistä (ScienceDirect s.a.). ScienceDirect valikoitui käyttöön, koska sen tarjoamat aihealueet ovat läheisesti yhteydessä tämän opinnäytetyön aiheeseen. Tietokanta tarjosi hyvän lähtökohdan erityisesti vertaisarvioitujen tutkimusartikkelien löytämiseksi, mikä lisää haun luotettavuutta. Tässä työssä ScienceDirect tietokantaa hyödynnettiin nimenomaan tieteellisten tutkimuslähteiden etsimiseen.

Tietoa haettiin lisäksi Google Search hakukoneen avulla. Google on automatisoitu hakukone, joka käyttää indeksointirobotteja selaamaan ja järjestämään verkossa olevaa sisältöä. Robotit keräävät ja arvioivat tietoa esimerkiksi verkkosivujen sisällön, avainsanojen ja päivitysajankohdan perusteella. Koska verkkosisältö muuttuu jatkuvasti, myös hakutulokset voivat vaihdella ajankohdan mukaan, vaikka hakusanat pysyisivät samoina. (Google s.a.)

Google Search hakukonetta hyödynnettiin, koska tutkimusaiheen tueksi haluttiin löytää myös muuta luotettavaa aineistoa kuin pelkästään tieteellisiä artikkeleita. Esimerkiksi raportit, nettisivut ja kirjat tarjoavat hyödyllistä täydentävää tietoa. Tämän vuoksi tiedonhakua laajennettiin hakukoneen avulla, jotta mukaan saatiin myös muiden luotettavien tahojen, kuten viranomaisten tai asiantuntijaorganisaatioiden tuottamaa aineistoa.

Aineiston valinnassa painotettiin luotettavuutta. Mukaan otettiin vain sellaisia lähteitä, jotka ovat peräisin tunnetuilta asiantuntijoilta, organisaatioilta ja tutkimuslaitoksilta. Taulukot 1.–4. kuvaavat lukijalle prosessia, joka on käyty lähteitä hankkiessa kirjallisuuskatsausta varten.

4.3 ScienceDirect tarkennettu haku

Artikkelihaku toteutettiin käyttämällä ScienceDirect-tietokannan hakutoimintoa, jossa on mahdollista tarkentaa hakua esimerkiksi artikkelin julkaisuajan, tyyppin sekä sisällön mukaan. Haussa hyödynnettiin Boolean-operaattoreita (AND, OR) sekä tarkkoja hakusanoja, joiden avulla pyrittiin löytämään opinnäytetyön tutkimuskysymyksiin vastaavia tieteellisiä tutkimusartikkeleita. Hakuehtojen muodostamisessa kiinnitettiin huomiota siihen, että ne olisivat riittävän tarkkoja rajataksaan tuloksia, mutta samalla tarpeeksi joustavia, jotta aiheeseen liittyviä artikkeleita löytyisi riittävästi.

Artikkeleiden esivalinta tehtiin lukemalla niiden otsikot ja tiivistelmät (abstraktit), ja arvioimalla alustavasti, että vastaavatko ne tutkimuksen aihealuetta ja kysymyksiä. Joissakin hauissa tuloksia tuli useita satoja, jolloin hakuja rajattiin edelleen esimerkiksi julkaisuajankohdan, artikkelityypin tai hakusanojen osuvuuden perusteella. Mukaan otettiin vain tutkimusartikkeleita, jotka olivat sisällöltään relevantteja joko suoraan tai epäsuorasti. Koska osa artikkeleista sisälsi vain yksittäisiä kohtia, jotka olivat tutkimuksen kannalta hyödyllisiä, oli tärkeää tarkastella kutakin artikkelia huolellisesti, jotta mikään mahdollisesti merkityksellinen aineisto ei jäisi huomiotta.

Myöhemmissä hauissa oli tärkeää huomioida, että osa hakutuloksista oli jo tullut vastaan aiemmissä hauissa. Tämä vaikutti siihen, että joissain hauissa valikoitui lopulta vähemmän artikkeleja seuraavaan tarkasteluun. Lisäksi osa artikkeleista oli käyttörajoitettuja, eli niitä ei saanut auki. Niiden osuus oli kuitenkin melko pieni, eikä sillä ollut isoa vaikutusta kokonaisuuteen. Taulukossa 1 on esitetty käytetyt hakusanat ja -ehdot, jokaisen haun tulostusmäärä sekä se että kuinka monta artikkelia valikoitui alustavasti tarkempaan tarkasteluun. Tarkempaan tarkasteluun valikoitui yhteensä 11 lähdettä, joista lopullisiksi lähteiksi valittiin kolme lähdettä. Nämä artikkelit ovat kuvattu taulukossa 2.

Taulukko 1. ScienceDirectin artikkelihaut

Hakuehdot	Tulokset ja artikkelivalinta
<p>"Conversational AI" (Title) OR "Chatbot" (Title) AND "Artificial Intelligence" (All text) AND "Support" (All text)</p> <p>Julkaisuajankohta 2024–2025</p>	86 tulosta (joista valittiin 7)
<p>"Ethics" (Title) AND ("AI" OR "Artificial Intelligence" OR "Generative AI") (All text) AND ("Intimate Partner Violence" OR "Domestic Violence" OR "Gender-based Violence") (All text)</p> <p>Julkaisuajankohta 2023–2025</p>	33 tulosta (joista valittiin 3)
<p>("Help-seeking" OR "Disclosure") (Title) AND ("Generative AI" OR "Artificial Intelligence") (All text) AND ("Intimate Partner Violence" OR "Domestic Violence") (All text)</p> <p>Julkaisuajankohta 2020–2025</p>	29 tulosta (joista valittiin 1)

Taulukko 2. ScienceDirectin valitut artikkelit

Lähteen numero	Lähde	Lähteen tekstiviite
1.	Rahsepar Meadi, M., Sillekens, T., Metselaar, S., van Balkom, A., Bernstein, J. & Batelaan, N. 2025. Exploring the ethical challenges of conversational AI in mental health care: Scoping review. JMIR Ment Health 2025;12:e60432	(Rahsepar Meadi ym. 2025)
2.	Holmes, G., Tang, B., Gupta, S., Venkatesh, S., Christensen, H. & Whitton, A. 2025. Applications of large language models in the field of suicide prevention: Scoping review. Journal of Medical Internet Research, 27 e63126	(Holmes ym. 2025)
3.	Novitzky, P., Janssen, J. & Kokkeler, B. 2023. A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools. Heliyon 9 e17140.	(Novitzky ym. 2023)

4.4 Google Search tarkennettu haku

Google Search -hakukonetta käytettiin kolmen erilaisen haun suorittamiseen ennalta määritetyillä hakukriteereillä, joita pystyy rajaamaan halutuilla sanoilla tai virkkeillä. Hakutuloksia pystyy myös rajaamaan muillakin kriteereillä, kuten kielen, sijainnin, viimeisimmän päivityksen, tietyn verkkosivuston, käyttöoikeuksien tai hakusanojen sijainnin perusteella (Google s.a.).

Kuvaan taulukossa 3 käyttämiäni hakuetoja sekä tulospääriä, joita määrittämäni hakuvaihtoehdot tuottivat. Ensimmäisessä haussa tuloksia tuli reilusti yli 50 000, joten tarkasteluun valitsin ensimmäiset 50. Ensimmäisistä 50 tuloksesta valitsin myöhemmin tarkasteluun kahdeksan nettisivua, otsikon ja lähteen uskottavuuden pohjalta.

Sekä toinen, että kolmas hakumäärittely oli selkeästi rajatumpi koska tuloksia tuli toisella 49 sekä kolmannella ainoastaan 14. Näistä tuloksista tarkasteluun päätyivät kaikki, ja tarkempaan tarkasteluun toisesta kolme, sekä kolmannesta kuusi. Näistä tarkempaan tarkasteluun päätyneistä verkkosivuista valittiin lopullisiksi lähteiksi kolme verkkosivua tai artikkelia, jotka löytyivät Google Search hakukonetta käyttämällä.

Kaikista kolmesta hakuetoista tarkempaan tarkasteluun päättyi siis 14 eri verkkosivua. Ensimmäisestä kahdeksan, toisesta kolme ja kolmannesta myös kolme. Näistä 14 verkkosivusta lopullisiksi lähteiksi valitsin kolme lähdeä. Nämä lähteet ovat kuvattuina taulukossa 4.

Tulosten vähäisyydestä voin päätellä sen, että aiheesta on vielä hyvin vähän kirjallisuutta. Hakuetoja oli muitakin, mutta niillä ei löytynyt ollenkaan verkkosivuja, jotka olisivat lähelläkään aihepiiriäni. Tekoälystä yleisesti löytyy valtavasti kirjallisuutta, mutta valtavasta nettisivumassasta oli tärkeää löytää juuri aiheeseeni liittyvää materiaalia.

Taulukko 3. Google Search haut

Hakuehdot	Tulokset ja ensimmäinen valinta
Sisältää kaikki sanat (missä vain sivustolla): generative AI, intimate partner violence, prevention, ethics, risk assessment aikaväli 2020 tammikuu – 2025 huhtikuu	54 200 tulosta joista 50 ensimmäistä tarkasteluun. Valittiin 8.
Sisältää kaikki sanat (otsikossa): therapy, conversational AI	49 tulosta, kaikki tarkasteluun. Valittiin 3.
Sisältää kaikki sanat (missä vain sivustolla): keskusteleva tekoäly, chatbot, kielimalli, mielenterveys, tuki	14 tulosta, kaikki tarkasteluun. Valittiin 3.

Taulukko 4. Google Search lopulliset lähteet

Lähteen numero	Lähde	Lähteen tekstiviite
4.	Masbounji, C. & Spencer, S. 2024. Artificial intelligence in gender-based violence in emergency programming: Perils and potentials (Version 5.0).	(Masbounji & Spencer 2024)
5.	Babushkina, D. & de Boer, B. 2024. Disrupted self, therapy and the limits of conversational AI.	(Babushkina & de Boer 2024)

6.	Amnesty International Suomen osasto 2024. Tekoäly ja ihmisoikeustyö – toisiaan tukevat vai yhteensovittamattomat?	(Amnesty International 2024)
----	---	------------------------------

5 Tulokset

Science Direct sekä Google Search -hakujen jälkeen lopullisia lähteitä tulosanalyysiin päätyi yhteensä kuusi kappaletta. Tässä luvussa käydään läpi näitä kuutta valittua lähdetä ja miten näiden lähteiden avulla pystytään vastaamaan opinnäytetyön kolmeen alatutkimuskysymykseen. Kysymykset käydään yksitellen läpi tässä luvussa omina alalukuinaan.

Moni tarkastelluista lähteistä liittyi suoraan enemmän mielenterveyden tukemiseen ja itsemurhien ehkäisyyn kuin varsinaisesti parisuhdeväkivaltatilanteisiin. kuitenkin näissä tutkimuksissa käsiteltiin epäsuorasti samoja ilmiöitä, kuten tuen ja kriisiavun tarjoamista tekoälypohjaisilla ratkaisilla vaikeissa elämäntilanteissa, mikä teki niiden hyödyntämisestä mahdollista myös tämän työn näkökulmasta. Joissain lähteissä taas keskityttiin puhtaasti siihen, että millainen tekoälyratkaisu toimisi parhaana keskusteluapuna ilman, että viitattiin suoraan tiettyihin väkivallan tai kriisitilanteisiin.

Lisäksi monet lähteet tarkastelivat laajasti tekoälyn käyttöä sosiaali- ja terveydenhuollon tukipalveluissa ilman, että rajaus olisi ollut erityisesti generatiiviseen tekoälyyn. Kuitenkin kun tämän opinnäytetyön tavoitteena oli nimenomaan keskittyä generatiivisen tekoälyn mahdollisuuksiin ja rajoitteisiin, tämä rajaus asetti omat haasteensa kirjallisuuden löytämiselle. Generatiivista tekoälyä käsittelevä tutkimus on vielä verrattain uutta ja hajanaista, mikä vaikutti lähdeaineiston painottumiseen ja valintoihin työn aikana.

Tämän luvun lopussa olevassa taulukossa (taulukko 5) esitetään, että mitkä lähteet vastaavat kuhunkin tutkimuskysymykseen. Lähteet on numeroitu samalla tavalla kuin niiden valintavaiheessa (1-6). Tämän jälkeen kunkin lähteen tarjoamat vastaukset ja näkökulmat käsitellään alaluvuissa tutkimuskysymysten mukaan ryhmiteltyinä.

Kuudennessa luvussa ”Generatiivinen tekoäly parisuhdeväkivallan uhrien tukena” käsitellään lähteiden tarjoamia vastauksia ja näkökulmia kokonaiskuvana.

5.1 Tekoäly keskusteluapuna

Tässä kappaleessa käsitellään lähteiden sisältöä ensimmäisen tutkimuskysymyksen näkökulmasta. Valikoidut lähteet käsittelevät tekoälyn muotoja keskusteluapuna hieman eri lähtökohdista.

Keskustelevat tekoälyagentit (conversational agents, CAs) voivat toimia keskusteluapuna psyykkisesti kuormittuneille yksilöille, erityisesti terapian kontekstissa. Keskeinen väite on kuitenkin

se, että vaikka nämä tekoälyagentit voivat jäljittää keskustelua, ne eivät kykene aitoon dialogiin, joka on välttämätöntä psyykkisen toipumisen kannalta.

Artikkelissa käydään läpi kahta eritasoista keskustelun muotoa. ”Conversing” eli pinnallinen, funktionaalinen vuorovaikutus (esimerkiksi pienet jutustelut) ja ”Dialogue” eli syvälinen ja inhimillinen vuorovaikutus, jossa rakennetaan merkityksiä ja uudelleen rakennetaan minäkuva. Yleisimmät tekoälysovellukset kykenevät lähinnä pinnalliseen keskusteluun, mutta eivät syvälinen, koska niillä ei ole kognitiivista joustavuutta, tunnekokemusta tai jaettua inhimillistä ymmärrystä, jotka ovat terapeutin dialogin keskiössä. (Babushkina & de Boer 2024.)

Huolenaiheet keskustelevien tekoälysovellusten käytössä huomattiin myös koskevan niiden taipumusta antaa haitallisia neuvoja, virheellistä tietoa tai ”hallusinaatioita”. Nämä hallusinaatiot voivat ilmetä väärinä faktoina, sekä jopa tilanteina, joita ei ole olemassa. Haitallinen tai virheellinen sisältö voi johtaa vakaviin seurauksiin. (Rahsepar Meadi ym. 2025.)

Suuria kielimalleja (LLM) on hyödynnetty erityisesti itsemurhariskin tunnistamiseen ja ennustamiseen. Uudemmat tutkimukset kuitenkin osoittavat, että niitä voidaan käyttää myös laajemmin, esimerkiksi tarjoamaan tietoa itsemurhien ehkäisystä, havaitsemaan pahenevaa riskiä tai auttamaan hoitosuunnitelmien laadinnassa. Näissä tehtävissä LLM:t ovat osoittautuneet lupaaviksi, koska ne voivat tarjota nopeasti täsmällistä tietoa, rakentaa tavoitteellisia hoitosuunnitelmia ja tuottaa koulutusmateriaalia ammattilaisille. (Holmes ym. 2025.)

Haasteita on kuitenkin yhä, sillä mallit eivät kykene luotettavasti tunnistamaan tilanteita, joissa ihminen tulisi ohjata hoitoon. Herää myös kysymyksiä siitä, mikä on vastuunjako, jos hoitosuunnitelma on tekoälyn tuottama. (Holmes ym. 2025.)

LLM:t voivat tukea mielenterveystyötä esimerkiksi tarjoamalla ympärivuorokautista ja ilmaista apua. Ne voivat auttaa myös silloin, kun asiakas ei uskalla hakea apua ihmiseltä. Vaikka näissä malleissa on paljon potentiaalia, niiden käyttö mielenterveystyössä vaatii huolellista suunnittelua ja ihmisten asiantuntemusta tueksi. Tekoälyn ja ihmisten yhteistyö esimerkiksi niin sanottujen ”human-in-the-loop” -ratkaisujen muodossa on tärkeä tulevaisuuden suunta. (Holmes ym. 2025.)

Tekoälypohjaisten järjestelmien lisääntynyt käyttö tilanteissa, joissa inhimillinen empatia on keskeistä, saattaa hämärtää rajaa terapeutin työskenelyssä. Se vaikuttaa myös asiakkaan ja ammattilaisen väliseen suhteeseen. Tilanteet koskevat erityisesti niitä, joissa LLM:t ovat koulutettuja jäljittelemään ihmisen kaltaista emotionaalista älykkyyttä. (Masbouni & Spencer 2024.)

Keskustelurobotit (conversational agents, CAs) lupaavat myönteisiä vaikutuksia terveys- ja mielenterveyssektorille. Keskustelurobotit osallistuvat keskusteluihin ja tukevat niitä käyttämällä tekstiä, kuvia, ääntä tai niiden yhdistelmiä. Keskustelurobottien kyky käydä vuoropuhelua käyttäjien kanssa on kehittynyt merkittävästi suurien kielimallien kehityksen myötä. Alun perin keskustelurobotteja käytettiin lyhytaikaisten transaktioon perustuvien tehtävien tukemiseen, mutta nykyään niille annetaan yhä enemmän itsenäisiä rooleja ja niitä pidetään merkityksellisinä myös pitkäaikaisissa prosesseissa, kuten esimerkiksi psykoterapeuttisessa neuvonnassa tai sairauden hallinnassa. Yhä useammin pohditaan, että voisiko keskustelurobottia pitää esimerkiksi lääkärin tai terapeutin korvikkeena. (Babushkina & de Boer 2024.)

On kuitenkin epävarmaa, että voivatko keskustelurobotit todella edistää hoidon parempaa kohdentamista. Tämä riippuu siitä, että missä määrin keskustelurobotteja voidaan pitää merkityksellisinä vaihtoehtoina olemassa oleville tukimuodoille ilman että ne vaarantavat hoidon varsinaiset tavoitteet. (Babushkina & de Boer 2024.)

Vaikka teknologialähtöinen tutkimus, joka korostaa keskustelurobottien hyötyjä on runsasta, kriittinen arvio näiden tavoitteiden vaikuttavuudesta vaikuttaa olevan jäljessä. Esimerkiksi tutkimuksissa käyttäjät raportoivat myönteisiä kokemuksia keskusteluroboteista eri yhteyksissä, mutta pitkäaikaisista vaikutuksista ei vielä tiedetä tarpeeksi, samoin kuin taustalla vaikuttavista mekanismeista, jotka selittäisivät koetut hyödyt. Lisäksi on epäselvää, että miten keskustelurobotit sopivat yhteen terapian perusoletusten kanssa, kuten esimerkiksi asiakkaan ja terapeutin välisen kiintymyssuhteen muodostuminen. (Babushkina & de Boer 2024.)

Yksi keskustelurobottien houkuttelevista piirteistä terapeuttisessa kontekstissa on se, että ne voivat käydä vuoropuhelua. Keskustelurobotti kykenee merkitykselliseen vuoropuheluun potilaan kanssa siten, että saavutetaan oivalluksia ja syvällistä ymmärrystä. Keskustelurobotti voi käsitellä potilaan ongelmaa vaikuttamalla toivotulla tavalla hänen ajatusmalleihinsa, itsereflektioonsa, tunteisiinsa ja toimintaansa. (Babushkina & de Boer 2024.)

On kuitenkin tärkeää muistaa, että keskustelurobotin tekniikan vuoksi se ei koskaan vastaa yhdelle yksilölle. Keskustelurobotti vastaa tyypilliseen pyyntöön sellaisena kuin algoritmi on sen tunnistanut ja tulkinut. Lisäksi keskustelurobotin vastaus rakentuu siten, että se koostuu yleisluontoisista vastauksista tilanteisiin, joissa saatat olla. Vastausten yleisluontoisuuden lisäksi ne voivat olla myös rajoittuneet siihen sisältöön, jonka pohjalta robotti on koulutettu. Jos oma tilanteesi poikkeaa robotin koulutetusta aineistosta, ei keskustelurobotilla ole mahdollisuuksia käsitellä sitä. Tämän vuoksi keskustelurobotit eivät pohjimmiltaan kykene käsittelemään avoimia ongelmia, esimerkiksi

kysymyksiä: ”Mikä minussa on vialla?” tai ”Miksi olen surullinen?”. Nämä kysymykset edellyttävät henkilökohtaisen tilanteen syvällistä analysointia syyn löytämiseksi. (Babushkina & de Boer 2024.)

5.2 Eettinen näkökulma

Tässä kappaleessa käsitellään lähteiden sisältöä toisen tutkimuskysymyksen näkökulmasta. Generatiivisen tekoälyn käyttöön lähisuhdeväkivallan uhrien auttamisessa liittyy huomattava määrä riskejä sekä eettisiä näkökulmia.

Yksi keskeisimmistä eettisistä huolista liittyy yksityisyydensuojaan sekä tietoturvaan.

Tekoälyratkaisut kuten chatbotit tai automatisoidut työkalut käsittelevät usein arkaluonteisia henkilötietoja. Väärinkäytön mahdollisuus kasvaa, kun järjestelmät eivät ole täysin läpinäkyviä ja tietoturva ei ole riittävällä tasolla. (Masboungi & Spencer 2024.)

UNICEFin julkaiseman raportin mukaan tekoäly voi myös vahingoittaa asiakassuhteita tai heikentää inhimillistä vuorovaikutusta. Usein on riski sille, että asiakas nähdään ensisijaisesti datalähteenä, eikä yksilönä, jolla on omat kokemuksensa ja tarpeensa. Tämä on ristiriidassa parisuhdeväkivallan palveluiden ydinarvojen kanssa, joihin kuuluvat luottamus, empatia ja uhrin toimijuuden tukeminen. (Masboungi & Spencer 2024.)

Generatiivisen tekoälyn tuottamat sopimattomat vastaukset voivat joissain tapauksissa estää käyttäjiä hakeutumasta mielenterveyspalveluiden piiriin. On tuotu esiin huolia siitä, että ihmisiä voidaan myös helposti manipuloida tai painostamaan tekemään tekoja, joita he eivät muuten tekisi ilman generatiivisen tekoälyn ehdotusta. (Rahsepar Meadi ym. 2025.)

Tekoälyn tietoisuuden puute estää sitä olemasta toimija, joka voisi kantaa täyttä vastuuta hätätilanteissa tai riskinarvioinneissa. Tekoälyssä voi ilmetä teknisiä virheitä, sekä virheitä, jotka johtuvat inhimillisten piirteiden jäljittelystä. Riittämätöntä empaattisuutta pidetään sellaisena riskitekijänä, joka voi vaikeuttaa oikeanlaisen avun saantia. (Rahsepar Meadi ym. 2025.)

Empatia nähdään tekoälyn käytössä kaksipiippuisena asiana. Koska empatia nähdään ja koetaan yleisesti inhimillisenä ominaisuutena, esimerkiksi keskustelurobottien simuloima empatia ei vastaa aitoa empatiaa. Toisaalta myös jäljitelty empatia voi riittää tukemaan terapeuttisia oivalluksia, koska terapeutit ajoittain voivat osoittaa näennäistä empatiaa esimerkiksi keskittymisen herpaantumisen, tai myötätuntouupumuksen vuoksi. Tästä huolimatta, tekoälyn empatian puute saattaa heikentää vuorovaikutusta, aiheuttaa väärinkäsityksiä sekä vaikuttaa haitallisesti mahdollisiin hoitotuloksiin. Empatian puutteeseen on ehdotettu ratkaisuksi generatiivisen tekoälyn täydennystä ihmisen tarjoamalla mielenterveystuella. (Rahsepar Meadi ym. 2025.)

Tekoälysovellukset myös mahdollistavat haitallisen ja epäaidon materiaalin luomisen haitallisiin tarkoituksiin, kuten seksuaalissävyytteisten kuvien luomiseen ilman suostumusta (non-consensual intimate images, NCI) tai deepfake teknologiaan perustuvan pornografian tuottamiseen. Näillä materiaaleilla ja sisällöillä pyritään häpäisemään, uhkailemaan, sekä hallitsemaan ja/tai rahallisesti kiristää naisia sekä tyttöjä, joka usein johtaa fyysiseen väkivaltaan ja itsemurhiin. (Masboungi & Spencer 2024.)

Osa asiantuntijoista ovat ehdottaneet, että tekoäly lisää kognitiivista kuormitusta ihmisille, kun tekoälyä integroidaan korkean riskin päätöksentekoprosesseihin. Esimerkiksi tilanteissa, jossa ihmiset tulkitsevat jatkuvasti monimutkaisia tekoälyn tuottamia tuloksia, seuraavat järjestelmän suorituskykyä ja joutuvat tekemään nopeita sekä kriittisiä päätöksiä. Tämä lisääntynyt kognitiivinen kuormitus voi mahdollisesti johtaa esimerkiksi stressiin ja heikentyneeseen laatuun erityisesti paineen alaisissa tehtävissä. (Masboungi & Spencer 2024.)

Toiset tutkimukset viittaavat siihen, että jotkut ihmiset käyttävät enemmän aikaa algoritmien tuottamien tulosten tarkasteluun kuin mitä he käyttäisivät ihmisten tuottamien tulosten tarkasteluun. Tämä tietenkin lisää tehtävään käytettyä aikaa, mikä voi lopulta hidastaa työn tahtia sen sijaan että se parantaisi yksilön tuottavuutta. (Masboungi & Spencer 2024.)

Lisäksi tilanteissa, joissa alan ammattilaiset käyttävät tekoälyä apunaan ohjeistamaan parisuhdeväkivallan uhreja, on tärkeää pohtia muun muassa seuraavia kysymyksiä: ovatko uhrin tietoisia siitä, että tekoälyä käytetään heidän elämänsä vaikuttavien päätösten tukena sekä missä määrin heidän suostumustaan tekoälyn käyttöön on pyydetty ja saatu. (Masboungi & Spencer 2024.)

Tekoälyä hyödyntävien palvelujen kehittäjien on kiinnitettävä erityistä huomiota tietoturvaan. Datan keräämiseen liittyvät mahdolliset riskit ja eettiset kysymykset täytyy tunnistaa ja arvioida tarkasti. Käyttäjiltä täytyy pyytää lupa tietojen keräämiseen ja heille on selkeästi kerrottava, että mihin heidän tietojensa käytetään. Tietoturva on erityisen tärkeässä roolissa silloin, kun palveluita kehitetään lähisuhdeväkivallan ehkäisyyn, sillä käyttäjien tiedot voivat olla hyvin arkaluontoisia. Siksi avoimuus eli läpinäkyvyys on ehdoton edellytys tekoälyä hyödyntävien palveluiden suunnittelussa ja toteutuksessa. (Amnesty International 2024.)

5.3 Tekoäly apuna avun haussa

Tässä kappaleessa käsitellään lähteiden sisältöä kolmannen tutkimuskysymyksen näkökulmasta. Generatiivinen tekoäly ja siihen perustuvat sovellukset voisivat tukea eritoten sukupuoleen

perustuvan väkivallan uhreja erityisesti hätätilanteissa ja haastavissa ympäristöissä. Näillä teknologioilla voi olla merkittävä rooli matalan kynnyksen tuen tarjoamisessa.

Yksi keskeisimmistä sovelluksista on tekoälypohjaisten chatbottien tai virtuaalisten avustajien käyttö yksilöllisessä tuessa. Näiden avulla voidaan tarjota anonymia ja ajasta tai paikasta riippumatonta tukea väkivallan uhreille. Tällainen lähestymistapa mahdollistaa tuen piiriin hakeutumisen myös niille, joilla on korkea kynnyksensä ottaa yhteyttä perinteisiin palveluihin tai jotka elävät alueilla, joissa fyysinen palveluverkosto on heikko. (Masboungi & Spencer 2024.)

Tekoäly voi myös tunnistaa riskitilanteita ennakoivasti analysoimalla suuria määriä tietoa, mikä voi auttaa palveluntarjoajia priorisoimaan akuutit tilanteet ja reagoimaan nopeammin. Tämä voi lyhentää uhrien odotusaikaa ja nopeuttaa avun saamista. (Masboungi & Spencer 2024.)

Keskustelevaa tekoälyä kohtaan muodostuvat suhteet eivät ole samanlaisia kuin ihmisten väliset suhteet. Ne eivät ole tasavertaisia tai vastavuoroisia. Käyttäjät saattavat suosia keskustelevaa tekoälyä ihmiskontaktien sijaan, sillä se on aina positiivinen ja jatkuvasti saatavilla. On kuitenkin hyvä pitää mielessä, että jos tekoäly korvaa ihmiskontaktit, se voi lisätä yksinäisyyttä, heikentää kykyä käsitellä ristiriitoja tai estää hakeutumista ammattimaiseen hoitoon. (Rahsepar Meadi ym. 2025.)

Tekoäly voi parantaa mahdollisuuksia saada tukea ja tietoa etänä esimerkiksi chatbottien ja muiden virtuaaliavustajien avulla. Nämä voivat auttaa naisia ja tyttöjä saamaan luottamuksellista tietoa omista oikeuksistaan sekä sukupuolittuneeseen väkivaltaan liittyvistä palveluista mihin vuorokauden aikaan tahansa. Virtuaaliavustajat voivat myös tarjota perustason psykologista tukea, vaikka tällaisessa käytössä on suuria riskejä ja tutkimusnäyttö käytön vaikutuksista on vielä rajallista. Virtuaaliavustajiin liittyy myös riskejä, kuten virheellisiä tai harhaanjohtavia vastauksia, joten niiden käyttö erityisen herkissä tilanteissa vaatii tarkkaa riskien tunnistamista ja hallintaa. (Masboungi & Spencer 2024.)

Lisäksi tekoälypohjaiset puhe-, kieli- ja käännöstyökalut voisivat vähentää kielimuureja ja tukea yhteisökonsultaatioita ja paikallistason arviointeja mukaan lukien sukupuolittuneeseen väkivaltaan hätätilanteissa liittyvät tarpeiden kartoitukset ja turvallisuusarviointit. Tämä voisi edistää palveluiden saavutettavuutta ja yhdenvertaisuutta. Palvelut voisivat näin olla helpommin saatavilla myös niille, jotka eivät puhu valtakieltä tai joilla on vaikeuksia kommunikoida. Silti on hyvä huomioda, että monet nykyisistä käännöstyökaluista eivät vielä pysty käsittelemään vähemmistökieliä tai keskeisiä termejä, joita tarvitaan sukupuolittuneessa väkivaltatyössä. Lisäksi kielimallit tukevat usein vain hyvin rajattua määrää kieliä. (Masboungi & Spencer 2024.)

Keskustelevat tekoälyt voisivat sopia hyvin sellaisten terapian osa-alueiden tueksi, joissa ei tarvita ihmisten välistä syvää vuorovaikutusta tai terapeutin kykyä samaistua potilaan elämäntilanteeseen. Ne voisivat esimerkiksi auttaa opetuksessa, tukea taitojen harjoittelussa, kuten mielialan ja ajatusmallien seurannassa, tai ohjata rentoutumis- ja hengitysharjoituksia, kunhan käyttöön liittyvät riskit huomioidaan. Tällöin tekoälyn rooli muistuttaisi enemmän perinteisiä itseapuvälineitä, kuten kirjoja tai videoita. Eräässä tutkimuksessa havaittiin, että käyttäjät, jotka saivat käyttöönsä käyttäytymisterapiapohjaisen chatbotin itseapukirjan kaltaiseksi työkaluksi, kokivat masennusoireidensa lievittyneen enemmän kuin ne, jotka käyttivät tavallista sähköistä masennukseen liittyvää e-kirjaa. (Babushkina & de Boer 2024.)

AinoAid on esimerkki tekoälyä hyödyntävästä palvelusta, joka on kehitetty tukemaan lähisuhdeväkivallan vastaista työtä. AinoAid on suomalaisessa tutkimusryhmässä kehitetty tekoälypohjainen palvelu, joka on suunniteltu matalan kynnyksen avuksi lähisuhdeväkivaltaa kohdanneille. Palvelu koostuu chatbotista ja tietopankista. Chatbot toimii ikään kuin ensiapupakkauksena, joka ohjaa käyttäjää oikeiden tukipalveluiden äärelle. Tietopankki puolestaan tarjoaa hyödyllistä tietoa esimerkiksi sosiaalityöntekijöille ja muille ammattilaisille, jotka kohtaavat työssään lähisuhdeväkivaltaa. (Amnesty International 2024.)

Vaikka digitaaliset teknologiat ovat nykyään laajasti käytössä monilla elämänalueilla, niiden soveltaminen parisuhdeväkivallan ehkäisyyn ja siihen liittyviin tukipalveluihin on edelleen vähäistä. Erityisesti tietoisuuden lisäämisen, seulonnan, ennaltaehkäisyyn ja hoidon tukemiseen liittyviä digitaalisia ratkaisuja ei ole kehitetty riittävästi. Tämä koskee myös parisuhdeväkivallan jälkeisiä haasteita, kuten mielenterveyden tukipalveluita. Viime vuosina on kuitenkin alkanut kehittyä uusi tutkimusala, jossa parisuhdeväkivaltaa pyritään ehkäisemään tekoälyn, koneoppimisen ja datan louhinnan keinoin. Näissä tutkimuksissa hyödynnetään suuria tietomassoja, esimerkiksi viranomaisten rekistereitä, sosiaalisen median aineistoja ja äänitallenteita, joiden avulla pyritään kehittämään algoritmeja, jotka pystyvät ennustamaan riskiä joutua parisuhdeväkivallan uhriksi. Vaikka monet käytetyt menetelmät keskittyvät vielä retrospektiivisiin analyysiin, joissa tarkastellaan menneitä tapahtumia, koneoppimisen ja syvien neuroverkkojen avulla on mahdollista löytää monimutkaisia kaavoja ja riskitekijöitä, jotka eivät välttämättä muuten nousisi esiin. Näiden kehittyneiden mallien avulla voidaan kehittää varhaisen varoituksen järjestelmiä, jotka auttaisivat tunnistamaan riskissä olevia henkilöitä ajoissa ja kohdentamaan tukitoimenpiteitä paremmin. (Novitzky ym. 2023.)

Taulukko 5. Kirjallisuuskatsaukseen valittujen lähteiden liittyminen tutkimuskysymyksiin

Lähteet / kysymykset	Mikä tekoälyn muoto on paras keskusteluapuna?	Mitä eettisiä näkökulmia tulisi ottaa huomioon hyödyntäessä generatiivista tekoälyä parisuhdeväkivallan uhrien auttamisessa?	Miten generatiivisen tekoälyn avulla voitaisiin madaltaa uhrien kynnystä hakea apua väkivaltatilanteissa?
1.	x	x	x
2.	x		x
3.			x
4.	x	x	x
5.	x		
6.		x	x

6 Generatiivinen tekoäly parisuhdeväkivallan uhrien tukena

Tekoälyhankkeita varten tulisi kehittää ja ottaa käyttöön riskien ja hyötyjen arviointimenetelmä, jolla kartoitetaan sekä mahdollisuudet että mahdolliset haitat. Arvioinnissa ei riitä pelkkä taloudellisten kustannusten ja henkilöstöresurssien tarkastelu, vaan mukaan pitäisi ottaa myös vaikutukset esimerkiksi työpaikkojen vähenemiseen, paikallisiin työllistymismahdollisuuksiin, yhteisöjen ja naisjärjestöjen vaikutusmahdollisuuksiin, yhteisöjen yhtenäisyyteen sekä eettisiin periaatteisiin. Kaiken arvioinnin tulisi olla linjassa feminististen tekoälynäkökulmien kanssa. Lisäksi jokaisessa projektissa pitäisi tehdä erillinen eettinen tarkastelu ja suojeluriskien arviointi, koska riskit voivat vaihdella ympäristön ja ratkaisun mukaan. (Masboungi & Spencer 2024.)

Käännös- ja puhesynteesiohjelmat voivat helpottaa palveluiden saavutettavuutta ihmisille, jotka puhuvat vähemmistökieliä ja ne voivat parantaa sukupuolittuneen väkivallan ohjelmien arviointia ja kehittämistä. Silti niihin liittyy riskejä, algoritmien virheet voivat vahingoittaa yksilöitä sekä paikallista työvoimaa voidaan syrjäyttää ja henkilöstön määrää vähentää, mikä voi heikentää sitoutumista paikallisyhteisöihin. Ohjelmien kehittämisessä tulisi edetä maltillisesti, suunnitella huolella ja testata huolellisesti useaan otteeseen. Kaikki tekoälyn tuottamat sisällöt pitää tarkistaa ihmisen toimesta. (Masboungi & Spencer 2024.)

Keskustelurobotteja voidaan pitää hyödyllisinä välineinä tiedon jakamiseen, mutta terapeutisessa vuorovaikutuksessa niiden puutteet tulevat esiin. Vaikka ne osaavat jäljitellä keskustelua, niiltä kuitenkin puuttuvat ihmismäiset kyvyt, kuten sosiaalinen herkkyys, kyky luoda ihmissuhteita ja moraalinen näkökulma. Tämän vuoksi niitä ei tulisi pitää varsinaisina keskustelukumppaneina terapian kaltaisissa tilanteissa. (Babushkina & de Boer 2024.)

On tärkeää, että ihminen saa kertoa kokemuksensa ja tulla kuulluksi, jotta kokemus tunnustetaan ja vahvistetaan oikeaksi. Keskustelulla on myös tiedollinen rooli, sen avulla ihminen voi jäsentää ja ymmärtää kokemustaan paremmin ja kehittyä vuorovaikutuksessa toisten kanssa.

Keskustelurobotit pyrkivät osittain korvaamaan tätä terapeutin vuoropuhelun kumppanuutta, jossa toinen osapuoli toimii itsenäisenä ja aidosti vuorovaikutteisena tahona. Jotta robotteja voitaisiin arvioida tässä roolissa, on tärkeää ymmärtää ero tavallisen jutustelun ja syvällisen, molempia osapuolia aidosti sitovan vuoropuhelun välillä. (Babushkina & de Boer 2024.)

On vaikea osoittaa, että keskustelurobotit todella pystyisivät aitoon, syvälliseen vuoropuheluun ihmisten tavoin. Itse asiassa niiden vahvuus ei ole aidossa keskustelussa, vaan siinä, että ne osaavat jäljitellä keskustelua riittävän hyvin saavuttaakseen joitain terapian tavoitteita. Vaikka

tekoäly on rakennettu simuloimaan vuorovaikutusta, myös ihmiset toisinaan simuloivat, esittävät ja teeskentelevät, eivätkä kaikki simulaatiot ole automaattisesti huonoja. Joskus simuloiminen on hyväksyttävää ja kannustettuakin. Olennaista on kuitenkin kysyä, että voiko tällainen jäljittely todella tukea aitoa terapeutista keskustelua silloin, kun käsitellään ihmisen olemassaoloon liittyviä syviä kysymyksiä. (Babushkina & de Boer 2024.)

Koska keskustelurobotit ovat tekoälypohjaisia, niiltä puuttuu kyky aitoon inhimilliseen vuorovaikutukseen ja toisen kokemukseen samaistumiseen. Siksi ne eivät pysty tarjoamaan syvällistä ja merkityksellistä keskusteluyhteyttä, joka on olennainen osa trauman käsittelyä. Tämä on erityisen merkittävää monimutkaisissa tapauksissa, joissa ihmisen kokemus ja sen vaikutukset ovat hyvin yksilöllisiä ja moniulotteisia. Keskustelurobotit toimivat algoritmien ohjaamina ja niiden tuottavat vastaukset ovat usein pinnallisia ja toistavat opittuja kaavoja. Tämä ei riitä herättämään käyttäjässä tunnetta, että hänen kokemuksensa olisi todella tunnistettu ja hyväksytty. Ne eivät myöskään pysty tarjoamaan sellaista kognitiivista ja emotionaalista syvyyttä, jota tarvitaan trauman monimutkaisuuden ymmärtämiseen ja kokemuksen analysoimiseen. Kyse ei ole vain yksittäisten empaattisten lauseiden lausumisesta, vaan siitä, että keskustelun pitäisi tukea aidosti potilaan omaa ajattelua ja sisäistä prosessointia. Keskustelurobottien kaavamainen lähestymistapa ei riitä traumaattisten kokemusten käsittelyyn ja joissakin tapauksissa ne voivat jopa pahentaa tilannetta. (Babushkina & de Boer 2024.)

Koska tekoäly kehittyy nopeasti, voidaan odottaa, että tulevaisuudessa keskustelurobotit pystyvät entistä monipuolisempaan kielelliseen vuorovaikutukseen. Tämä johtaa siihen, että niitä ei enää nähdä vain lisätukivälineinä, kuten itseapukirjoina vaan niitä verrataan enemmän psykoterapeutteihin. Mikäli näin käy, keskustelurobotteja tulisi arvioida teknisen tarkkuuden ja tehokkuuden lisäksi myös sen perusteella, että kuinka hyvin ne edistävät terapeutisia tavoitteita ja prosesseja. (Babushkina & de Boer 2024.)

Tekoälyratkaisut tulisi kehittää yhdessä naisten ja tyttöjen kanssa ja vastuullisuus on huomioitava jo suunnitteluvaiheessa. On tärkeää arvioida paitsi itse ratkaisun ja aineistojen laatu ja kestävyys, mutta myös se ympäristö, jossa ratkaisu otetaan käyttöön, mukaan lukien lainsäädäntö ja poliittinen tilanne. Tekoälyyn liittyvät riskit eivät johdu pelkästään sen teknisestä suunnittelusta, vaan myös siitä, että miten ja missä sitä käytetään ja ketkä sitä käyttävät. (Masboungi & Spencer 2024.)

Keskustelurobottien tuominen terapiaan kuitenkin paljastaa ristiriidan kahden erilaisen lähestymistavan välillä. Yksilöllistetyssä lääketieteessä pyritään toisaalta mittaamaan terveyteen liittyviä asioita mahdollisimman tehokkaasti ja optimoidusti ja toisaalta korostetaan potilaan

henkilökohtaisia tarpeita ja hyvinvointia. Ensimmäinen näkökulma näkee terapian järjestelmänä, joka vaatii paljon resursseja ja jota voidaan tehostaa digitaalisilla ratkaisuilla, jotta siitä saadaan nopeampaa, edullisempaa ja helpompaa käyttöä. Toisaalta potilaskeskeisessä lähestymistavassa korostetaan sitä, että jokaisen potilaan tilanne on ainutlaatuinen, eikä ongelmia voida aina helposti sovittaa algoritmien käsiteltäväksi. (Babushkina & de Boer 2024.)

UNICEF on laatinut kuvitteellisen, mutta tositapahtumien innoittaman skenaarion, jossa suuri kansainvälinen järjestö on kehittänyt tekoälyyn perustuvan keskustelurobotin nimeltä mAlya. mAlya-chatbotti tarjoaa naisille ja tytöille ympärivuorokautista tietoa sukupuolittuneesta väkivallasta, uhrien oikeuksista ja paikallisesti saatavilla olevista palveluista. Chatbot jäljittelee keskustelua sosiaalityöntekijän ja uhrin eli käyttäjän välillä käyttäjien lähettämien kysymysten perusteella. Chatbot tarjoaa myös perustason emotionaalista tukea. Kuvitteellisessa skenaariossa chatbot mAlyan mainostetaan olevan täysin salattu palvelu, josta ei jää digitaalista jälkeä. (Masboungi & Spencer 2024.)

mAlya on rakennettu suuren kielimallin (LLM) varaan, joka kehitettiin ja koulutettiin Euroopassa käyttäen dataa ympäristöistä ja väestöistä, jotka eroavat huomattavasti siitä kontekstista, jossa chatbottia parhaillaan käytetään. Tämä lisää riskiä siitä, että chatbotin vastaukset eivät vastaa paikallisten uhrien todellista kokemusta ja tarpeita. Lisäksi järjestöllä ei ole resursseja chatbotin hienosäätöön, mikä johtaa siihen, että chatbotin mallin tarkkuus heikentyy ajan myötä ilman, että järjestö huomaa sitä. (Masboungi & Spencer 2024.)

Kuvitteellisessa tarinassa nuori pakolaistyttö Ana saa tukea järjestön työntekijältä ja suostuu kokeilemaan mAlyaa. Ana tuntee olonsa yksinäiseksi uudessa kaupungissa ja keskustele chatbotin kanssa usein. Chatbotti tarjoaa hänelle tietoa palveluista ja myös perustason emotionaalista tukea koulutusdatansa pohjalta. Ana kiintyy voimakkaasti chatbottiin ja kertoo tapaustyöntekijälleen, että chatbotti on auttanut häntä selviytymään yksinäisyydestä. Hän voi aina luottaa siihen, että mAlya kuuntelee ja antaa hyviä neuvoja tuomitsematta tai saamatta tuntemaan häpeää. Tämän vuoksi Ana alkaa käymään tapaustyöntekijän luona harvemmin ja muutaman viikon kuluttua työntekijä ei saa Anaan enää yhteyttä. (Masboungi & Spencer 2024.)

Yllä oleva skenaario oli luotu herättämään keskustelua siitä, että tekoälypohjaiset ratkaisut kuten chatbotit voivat tarjota merkittävää tukea väkivaltaa kokeneille henkilöille, mutta samalla niihin liittyy riskejä, erityisesti silloin kun teknologian toimintaa ei valvota riittävästi. Skenaariossa pakolaistyttö Ana myös luopuu tapaustyöntekijän näkemisestä kokonaan, sillä hän luottaa chatbottiin niin paljon, vaikka chatbotti ei välttämättä edes osaa tarjota Analle sopivaa apua ja tukea. (Masboungi & Spencer 2024.)

7 Pohdinta

Tämän opinnäytetyön aikana nousi esiin useita huomionarvioisia havaintoja liittyen generatiivisen tekoälyn hyödyntämiseen parisuhdeväkivallan uhrien tukemisessa. Lähes kaikissa tarkastelluissa lähteissä korostettiin vahvasti sitä, että tekoälyn tuottamat sisällöt tulee aina varmistaa ihmisen toimesta, etenkin silloin kun niillä on vaikutusta yksilön hyvinvointiin tai muihin merkittäviin päätöksiin. Tämä oli yksi johdonmukaisimmista viesteistä aineistossa.

Tiedonlähteistä löytyi paljon muutakin kiinnostavaa tietoa generatiiviseen tekoälyyn liittyen, joita en kuitenkaan lisännyt työhön, koska ne eivät kuitenkaan vastanneet valittuihin tutkimuskysymyksiin tai opinnäytetyön aiheeseen. Eettisestä näkökulmasta löytyi laajasti tietoa, mikä osoittaa, että kuinka keskeisenä aiheena tekoälyn etiikka tällä hetkellä nähdään, erityisesti kun kyseessä on haavoittuvassa asemassa olevien ryhmien tukeminen.

Mielenkiintoisena havaintona nousi esiin myös se, että chatbotteja kehittäessä on tärkeä ottaa huomioon paitsi tekniset ominaisuudet, mutta myös se, että missä maantieteellisessä ja lainsäädännöllisessä ympäristössä niitä käytetään. Alueelliset erot esimerkiksi tukipalveluiden saatavuudessa ja sääntelyssä voivat vaikuttaa merkittävästi siihen, että miten chatbotit voivat aidosti tukea väkivallan uhreja.

Huomasin myös, että moni valituista lähteistä löytynyt hyödyllinen tieto liittyi jossain määrin kaikkiin tutkimuskysymyksiin. Tämä mahdollisesti viittaa siihen, että tutkimuskysymyksiä olisi voinut vielä eriyttää enemmän toisistaan, jotta eri näkökulmia olisi ollut helpompi tarkastella selkeämmin erillisinä kokonaisuuksina. Lisäksi jälkikäteen arvioiden tutkimuskysymys, joka käsitteli kynnyksen madaltamista avun hakemisessa, osoittautui haastavaksi. Löydetyistä aineistoista löytyi tähän aiheeseen verrattain todella vähän suoraa tietoa.

Opinnäytetyön nimen muutin alun perin suunnitellusta työn edetessä, sillä alkuperäisen aiheen ”Generatiivisen tekoälyn hyödyntäminen parisuhdeväkivallan ehkäisyssä” osalta löytyi vain vähän olemassa olevaa tutkimusta. Sen sijaan aineistot keskittyivät enemmänkin generatiivisen tekoälyn käyttöön uhrien tukemisessa tilanteiden jälkeen. Tämä mahdollisesti heijastaa sitä, että teknologian soveltaminen ennaltaehkäisyyn on vasta kehittymässä oleva alue.

Tämän opinnäytetyön perusteella voidaan todeta, että toistaiseksi generatiivista tekoälyä ei voida vielä pitää luotettavana ja turvallisena välineenä parisuhdeväkivallan uhrien auttamisessa. Teknologia ei ole vielä niin kehittynyttä, että chatbotit pystyisivät toimimaan terapeutin tavoin, ymmärtämään vaikeita traumaattisia kokemuksia tai tukemaan uhreja oikealla sosiaalisella tavalla herkissä asioissa. Lisäksi chatbottien antamaan tietoon ja ohjeisiin ei voi täysin luottaa, joten kaikki

annetut neuvot ja tiedot tulisi varmistaa ihmisen tarkastamana ennen niiden hyödyntämistä käytännössä.

Tulevaisuuden näkökulmasta tulee olla mielenkiintoista seurata, että miten generatiivinen tekoäly kehittyi parisuhdeväkivallan ehkäisyssä ja uhrien tukemisessa. On mahdollista, että jatkossa keskustelurobotit kykenevät paremmin tunnistamaan käyttäjien tunnetiloja, tarjoamaan yksilöllisempää ja empaattisempaa tukea sekä ohjaamaan uhrin herkemmin oikeiden tukipalvelujen pariin. Kehittyneempi tekoäly voisi myös mukautua paremmin erilaisiin kulttuuri- ja lainsäädäntöympäristöihin, jolloin tuen tarjoaminen olisi sensitiivisempää ja turvallisempaa. Edellytyksenä kuitenkin on, että eettiset periaatteet, laadunvarmistus ja ihmisen rooli säilyvät keskeisinä, jotta uhrien auttaminen on jatkossakin luotettavaa ja turvallista.

Työtä tehdessä huomasin, että aihe oli paikoin haastava, erityisesti siksi, että generatiivisen tekoälyn hyödyntäminen parisuhdeväkivallan uhrien tukemisessa on vielä uusi tutkimusalue. Osa valituista lähteistä käsitteli esimerkiksi generatiivisen tekoälyn hyödyntämistä yleisesti terapiassa, mutta tämäkin aihe ja tutkimus siitä oli kuitenkin sopivaa myös tähän opinnäytetyöhön. Lisäksi, koska tekoälyyn liittyvä tutkimus etenee nopeasti, voi työn ajantasaisuus muuttua tulevaisuudessa nopeasti uusien sovellusten ja tutkimustulosten myötä.

Työskentelyn aikana olen oppinut paljon sekä tekoälyn mahdollisuuksista että siihen liittyvistä riskeistä sosiaalialalla. Lisäksi opin lisää parisuhdeväkivallan ilmiöistä ja siihen liittyvistä tukipalveluista, sekä vahvistin omia taitojani lähteiden kriittisessä analysoinnissa ja tutkimuksellisessa kirjoittamisessa. Kokonaisuutena opinnäytetyö on ollut minulle tärkeä oppimisprosessi ja uskon sen tarjoavan hyödyllistä tietoa aiheen jatkokehittämiselle.

Lähteet

AIMultiple 2025. GPT-4: 12 Features, Pricing & Accessibility in 2025. Luettavissa: <https://research.aimultiple.com/gpt4/>. Luettu: 29.1.2025

Amnesty International 2024. Tekoäly ja ihmisoikeustyö – toisiaan tukevat vai yhteensovittamattomat? Luettavissa: <https://www.amnesty.fi/tekoaly-ihmisoikeustyo/>. Luettu: 6.4.2025

Babushkina, D. & de Boer, B 2024. Disrupted self, therapy and the limits of conversational AI. Taylor & Francis Group. Luettavissa: <https://www.tandfonline.com/doi/epdf/10.1080/09515089.2024.2397004?needAccess=true>. Luettu: 18.4.2025

Google s.a. Miten Google Haku järjestää tietoja. Luettavissa: <https://www.google.com/intl/fi/search/howsearchworks/how-search-works/organizing-information/>. Luettu: 4.4.2025

HAMK 2025. Generatiivinen tekoäly. Luettavissa: <https://digipedaohjeet.hamk.fi/ohje/generatiivinen-tekoaly/>. Luettu: 29.1.2025

Helsingin Kaupunki 2024. Helsingin lähisuhdeväkivallan ehkäisyn toimintaohjelma. Luettavissa: <https://www.hel.fi/static/kanslia/hyte/helsingin-lahisuhdevakivallan-ehkaisyn-toimintaohjelma-2024%E2%80%932025.pdf>. Luettu: 22.4.2025

Holmes, G., Tang, B., Gupta, S., Venkatesh, S., Christensen, H. & Whitton, A. 2025. Applications of large language models in the field of suicide prevention: Scoping review. Journal of Medical Internet Research, 27 e63126. Luettavissa: <https://www.jmir.org/2025/1/e63126/PDF>. Luettu 16.4.2025

IBM 2021. What is Conversational AI. Luettavissa: <https://www.ibm.com/think/topics/conversational-ai>. Luettu: 12.3.2025

Kreodi 2022. Kirjallisuuskatsauksen ohjaus – perustana tutkimuskysymys ja ohjaushaastattelu. Luettavissa: <https://www.kreodi.fi/arkisto/artikkelit/kirjallisuuskatsauksen-ohjaus-perustana-tutkimuskysymys-ja-ohjaushaastattelu.html>. Luettu 18.3.2025

Naisten Linja 2024. Tunnista väkivaltaisen parisuhteen varoitus. Luettavissa: <https://naistenlinja.fi/vakivaltaisen-parisuhteen-varoituserkit/>. Luettu: 3.3.2025

Novitzky, P., Janssen, J. & Kokkeler, B. 2023. A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools. *Heliyon* 9 e17140. Luettavissa: <https://www.cell.com/action/showPdf?pii=S2405-8440%2823%2904348-7>.
Luettu: 16.4.2025

Mantsinen, J. 18.12.2024. Tampereen Hämeenkadun ampumisen motiivin epäillään liittyneen parisuhteeseen. Luettavissa: <https://www.hs.fi/suomi/art-2000010910496.html>. Luettu: 3.5.2025

Masbounji, C. & Spencer, S. 2024. Artificial intelligence in gender-based violence in emergency programming: Perils and potentials. UNICEF. Luettavissa: <https://clearinghouse.unicef.org/sites/ch/files/ch/sites-PD-ChildProtection-Knowledge%20at%20UNICEF-AI%20in%20GBV%20Emergencies-5.0.pdf>. Luettu: 12.4.2025

Monahan, J. s.a. Artificial Intelligence, Explained. Carnegie Mellon University's Heinz College. Luettavissa: <https://www.heinz.cmu.edu/media/2023/July/artificial-intelligence-explained>. Luettu: 28.1.2025

OpenAI 2025. Introducing GPT-4.5. Luettavissa: <https://openai.com/index/introducing-gpt-4-5/>.
Luettu: 4.2.2025

Rahsepar Meadi, M., Sillekens, T., Metselaar, S., van Balkom, A., Bernstein, J. & Batelaan, N. 2025. *Exploring the ethical challenges of conversational AI in mental health care: Scoping review*. *JMIR Ment Health* 2025;12:e60432. Luettavissa: <https://mental.jmir.org/2025/1/e60432/PDF>.
Luettu: 16.4.2025

Salminen, A. 2023. Mikä kirjallisuuskatsaus? Johdatus kirjallisuuskatsauksen tyypeihin ja joihinkin hallintotieteellisiin sovelluksiin. Vaasan yliopisto. Vaasa. Luettavissa: <https://osuva.uwasa.fi/bitstream/handle/10024/15470/978-952-395-081-8%20%28PDF%29.pdf?sequence=2&isAllowed=y>. 19.3.2025

ScienceDirect s.a. Explore scientific, technical, and medical research on ScienceDirect. Luettavissa: <https://www.sciencedirect.com/>. Luettu 5.4.2025

SAP 2024. What is a large language model? Luettavissa: <https://www.sap.com/finland/resources/what-is-large-language-model>. Luettu: 22.4.2025

SAP s.a. Mitä on generatiivinen tekoäly? Luettavissa:

<https://www.sap.com/finland/products/artificial-intelligence/what-is-generative-ai.html>. Luettu: 27.1.2025

Sosiaali- ja terveysministeriö 2023. Lähisuhde- ja perheväkivallan ehkäiseminen. Luettavissa:

<https://stm.fi/lahisuhde-ja-perhevakivalta>. Luettu: 19.4.2025

TechTarget 2025. Generative AI ethics: 11 biggest concerns and risks. Luettavissa:

<https://www.techtarget.com/searchenterpriseai/tip/Generative-AI-ethics-8-biggest-concerns>. Luettu: 3.2.2025

Tilastokeskus 2024. Rikos- ja pakkokeinotilasto. Luettavissa:

<https://stat.fi/julkaisu/cln32gnj49hev0cutpmcsm4ko>. Luettu: 3.3.2025

Turun Ensi- ja Turvakoti Ry 2023. Suomi on väkivaltainen maa naiselle. Luettavissa:

<https://tuentu.fi/suomi-on-vakivaltainen-maa-naisel->

<le#:~:text=Suomi%20on%20viel%C3%A4%20kuitenkin%20Euroopan,on%20kokenut%20parisuhteessa%20henkist%C3%A4%20v%C3%A4kivaltaa>. Luettu: 20.4.2025

United Nations s.a. What Is Domestic Abuse? Luettavissa:

<https://www.un.org/en/coronavirus/what-is-domestic-abuse>. Luettu: 3.3.2025

Wired 2025. With GPT-4.5, OpenAI Trips Over Its Own AGI Ambitions. Luettavissa:

<https://www.wired.com/story/gpt-4-5-openai-first-impressions/>. Luettu: 11.3.2025