Organisaation valmiudet generatiivisen chatbotin käyttöönottoon asiakaspalvelussa

Riina Kokko

Haaga-Helia ammattikorkeakoulu
Liiketoiminnan teknologiat
Master-opinnäytetyö
2024

Tiivistelmä

Tekijä(t)
Riina Kokko

Tutkinto
Tradenomi (YAMK)

Raportin/Opinnäytetyön nimi
Organisaation valmiudet generatiivisen chatbotin käyttöönottoon asiakaspalvelussa

Sivu- ja liitesivumäärä
77 + 1

Generatiivinen tekoäly on tällä hetkellä yksi lupaavimpia teknologiota, jonka hyödyntämisestä ovat kiinnostuneet niin yksilöt, organisaatiot kuin yhteiskunnatkin. Lupaavien ominaisuuksien lisäksi uudessa teknologiassa on kuitenkin myös haasteita ja riskejä. Näiden haasteiden ja riskien kartoittamiseksi toteutettiin tämä opinnäytetyö, jonka tavoitteena oli selvittää Helsingin seudun opiskelija-asuntosäätiö Hoasin valmiuksia generatiivisen chatbotin käyttöönottoon asiakaspalvelussa.

Opinnäytetyön teoreettisessa viitekehyksessä perehdyttiin nykyisten chatbottien toteuttamisessa käytettyihin teknologioihin, painopisteen ollessa generatiivisissa chatboteissa ja isoissa kielimalleissa. Tämän jälkeen siirryttiin chatbottien käyttöönoton edellytyksiin, joiden avulla kartoitettiin niitä tekijöitä, jotka edesauttavat generatiivisen chatbotin onnistunutta käyttöönottoa organisaatiossa. Teoreettisessa viitekehyksessä käsiteltiin myös chatbotteihin ja isoihin kielimalleihin liittyviä riskejä niin yksilön, organisaation kuin yhteiskunnankin tasolla.

Tutkimus toteutettiin laadullisena tutkimuksena haastattelemalla seitsemää Hoasilla eri rooleissa työskentelevää asiantuntijaa. Haastattelut toteutettiin puolistrukturoidusti, ja litteroitu aineisto analysoitiin teoriaohjauksisen sisällönanalyysin periaatteita noudattaen.

Tutkimuksen tuloksista selvisi, mitkä olivat organisaation valmiudet generatiivisen chatbotin käyttöönottoon ja miten valmiuksia voidaan kehittää. Chatbotin tavoitteiden määrittäminen ja mittareiden asettaminen niille, generatiivisen chatbotin kouluttamiseen, ylläpitoon ja toiminnan seuraamiseen panostaminen, käyttötapauksten määrittäminen ja ihmisen vuorovaikutuksen tasapainottaminen nousivat esiin tutkimuksen tuloksissa. Lisäksi uuden teknologian yhteensopivuus sekä olemassa olevien tietojärjestelmien että organisaation kulttuurin ja arvojen kanssa huomioitiin.

Tietoturva, tietosuoja, eettisyys ja vastuullisuus olivat olennaisia aiheita, joita on priorisoitava generatiivisen tekoälyn käyttöönotossa. Myös isoihin kielimalleihin liittyvät erityiset, uudet riskit on huomioitava, kun generatiivista chatbottia otetaan käyttöön organisaation toiminnassa. Opinnäytetyön tuloksissa myös chatbotteihin liittyvä lainsäädäntö yleisen tietosuoja-asetuksen ja tekoälysäädöksen muodossa nousivat esiin. Yleisen tietosuoja-asetuksen vaatimukset pysyvät samoina ja tekoälysäädös vaatii tekoälyjärjestelmän käyttöönottajaorganisaatiolta läpinäkyvyyttä tekoälyn käytön suhteen. Näiden lisäksi organisaation tulee panostaa sisäisen asiantuntijuuden kehittämiseen chatbotti-, tekoäly- ja datateknologioiden suhteen, jotta organisaatio pysyy mukana digitalisaation muutoksissa.

Asiasanat
chatbotit, generatiivinen tekoäly, luonnollisen kielen käsittely, kielimallit, organisaatiot, vastuullisuus

Sisällys

1 Johdanto 1
1.1 Toimeksiantaja 1
1.2 Opinnäytetyön tavoitteet ja rajaus 2
1.3 Opinnäytetyön rakenne 3
2 Chatbottien teknologiat 5
2.1 Luonnollisen kielen käsittely 5
2.2 Perinteiset chatbotit 7
2.3 Generatiiviset chatbotit 9
2.3.1 Isot kielimallit 10
2.3.2 Perusmallit 11
2.3.3 Kehotesuunnittelu 12
2.3.4 Hienosäätö 13
2.3.5 Retrieval-Augmented Generation 13
2.3.6 Suomen kieli isoissa kielimalleissa 15
2.4 Hybridichatbotit 15
3 Chatbottien käyttöönoton edellytykset 17
3.1 Tavoitteet ja mittareiden asettaminen 17
3.2 Ketterä kehittäminen 18
3.3 Käyttäjäkokemus 19
3.4 Toiminnan seuraaminen, ylläpito ja kouluttaminen 20
3.5 Yhteensopivuus 21
3.6 Vastuullisuus 22
3.7 Läpinäkyvyys 24
3.8 Tiedonhallinta 24
3.9 Organisaation kulttuuri ja arvot 25
3.10 Chatbottien kehittämiseen liittyvä lainsäädäntö 27
3.10.1 Yleinen tietosuoja-asetus (GDPR) 27
3.10.2 Tekoälysäädös 29
4 Chatbottien riskit ja niiden hallinta 35
4.1 Tietoturva 35
4.1.1 Jailbreaking, kehoteinjektiot ja tietojen myrkyttäminen 36
4.1.2 Keinoja tietoturvan vahvistamiseen 37
4.2 Hallusinointi 39
4.2.1 Mallin tuottaman sisällön kontrolloiminen 39
4.3 Yksityisyys ja tietosuoja 41

4.3.1 Luottamuksen vahvistaminen ... 41
4.3.2 Tahattomat väärinkäytöt ... 43
4.3.3 Keinoja yksityisyyden ja tietosuojan vahvistamiseksi ... 43
4.4 Vastuu generatiivisen tekoälyn tuottamasta sisällöstä ... 44
4.5 Yhteiskunnalliset vaikutukset ... 45
4.5.1 Taloudellisten erojen lisääntyminen ... 45
4.5.2 Automaation vaikutus työllisyyteen ... 45
4.5.3 Geopoliittiset vaikutukset ja sosioekonomisen epätasa-arvon lisääntyminen ... 46
4.5.4 Tekijänoikeudet ... 47
4.5.5 Disinformaatio ... 47
4.5.6 Ennakkoluulot ja puolueellisuus ... 48
4.5.7 Generatiivisen tekoälyn hyödyntäminen haitallisiin tarkoituksiin ... 50
5 Tutkimuksen toteutus ... 51
5.1 Tutkimus- ja kehittämishankkeen metodologia ... 51
5.1.1 Aineiston hankintamenetelmät ... 51
5.1.2 Aineiston analyysimenetelmät ... 52
5.2 Tutkimuksen tulokset ... 53
5.2.1 Chatbotin tavoitteet ... 53
5.2.2 Generatiivisesta chatbotista haetut hyödyt ... 54
5.2.3 Chatbotin käyttötapaukset ... 55
5.2.4 Asiakaspalvelussa käytettävissä oleva materiaali ... 56
5.2.5 Nykyisen chatbotin haasteet ... 57
5.2.6 Chatbotin toiminnan ja tavoitteiden toteutumisen seuraaminen ... 58
5.2.7 Chatbotin ylläpito ... 59
5.2.8 Tietoturva ja tietosuoja ... 59
5.2.9 Lainsäädäntö ... 60
5.2.10 Muita huomioita ... 60
6 Pohdinta ... 62
6.1 Johtopäätökset ... 62
6.1.1 Tavoitteiden asettaminen ... 62
6.1.2 Generatiivisen chatbotin kouluttaminen, toiminnan seuraaminen ja ylläpito ... 63
6.1.3 Käyttötapaukset ja vuorovaikutuksen tasapainottaminen ... 64
6.1.4 Yhteensopivuus ja tiedonhallinta ... 65
6.1.5 Tietoturva ja tietosuoja ... 66
6.1.6 Isojen kielimallien puutteet ja riskien hallinta ... 68
6.1.7 Eettisyys ja vastuullisuus ... 69

6.1.8 Lainsäädäntö 70
6.1.9 Organisaation arvot, kulttuuri ja käytännöt 71
6.2 Tutkimuksen luotettavuuden arviointi 72
6.3 Opinnäytetyöprosessin arviointi 73
Lähteet 75
Liitteet 78
Liite 1. Haastattelukysymykset 78

1 Johdanto

Generatiivisesta tekoälystä on viime vuosina tullut yksi lupaavimmista uusista teknologioista. OpenAI:n ChatGPT:n julkaisemisen jälkeen generatiivinen tekoäly on noussut yleiseen tietoisuuteen ja saanut runsaasti julkisuutta. Generatiivisen tekoälyn myötä myös muihin tekoälyn alueisiin ja aiheisiin liittyvät asiat ovat nousseet yleiseen keskusteluun. Generatiivinen tekoäly tarjoaa suuria mahdollisuuksia yksilöiden ja organisaatioiden lisäksi yhteiskunnille, mutta mahdollisuuksien lisäksi siihen liittyvät ongelmat ja haasteet tulee huomioida, jotta uusi teknologia voidaan ottaa käyttöön vastuullisesti.

Chatbotit ovat viime vuosikymmenen aikana nousseet yhä suositummiksi. Monet organisaatiot hyödyntävät chatbotteja toiminnassaan, tarjoten sekä sisäisiä että ulkoisia palveluja, kuten asiakaspalvelua, chatbottien avulla. Chatbottien avulla voidaan tarjota asiakkaille palvelua jokaisena vuorokaudenaikana ympärivuotisesti, palvella useita asiakkaita samanaikaisesti ja kohdistaa asiakaspalvelun resursseja tehokkaammin. Chatbottien hyödyt ovat siis kiistattomat.

Generatiivinen tekoäly, ChatGPT etunenässä, on kuitenkin luonut aivan uudenlaisia odotuksia chatboteille. Nykyiset chatbotit ovat usein hieman kömpelöitä, niiden ymmärryskyky on rajallinen, ja ne tarjoavat samoja ja toistuvia vastauksia kaikille käyttäjille. ChatGPT on näyttänyt, miten mullistava generatiivinen tekoäly voi olla chatbottien kehitykselle. Uudenlainen kontekstin ymmärtäminen, sujuvat ja ainutlaatuiset vastaukset sekä hämmästyttävän hyvät keskustelutaidot ovat osoittaneet, että nykyisissä chatboteissa on paljon parantamisen varaa.

Tämän opinnäytetyön tavoitteena onkin selvittää, mitä organisaation tulee huomioida, kun lähdetään uudistamaan organisaation chatbottia generatiivisen tekoälyn avulla. Näissä uuden tyyppisissä, generatiivisissa chatboteissa on paljon potentiaalia, mutta generatiivinen tekoäly tuo mukanaan myös uudenlaisia haasteita. Opinnäytetyössä selvitetään, miten generatiiviseen chatbottiin liittyvä potentiaali voidaan realisoida mahdollisimman tehokkaasti, samalla kuitenkin huomioiden riskit ja haasteet, joita uusi teknologia tuo mukanaan.

1.1 Toimeksiantaja

Opinnäytetyön toimeksiantajana toimii Helsingin seudun opiskelija-asuntosäätiö Hoas. Hoas on yleishyödyllinen säätiö, joka tarjoaa vuokra-asuntoja opiskelijoille pääkaupunkiseudulla. Organisaation tavoitteena on peruskoulun jälkeisissä oppilaitoksissa tutkintoaan suorittavien opiskelijoiden asuntotilanteen helpottaminen edullisia vuokra-asuntoja tarjoamalla. Tällä hetkellä Hoasin vuokra-asunnoissa asuu noin 18 000 asukasta. Hoasin toimintaa ohjaavat vastuullisuus, oikeudenmukaisuus, yksilön huomioon ottaminen, avoimuus ja turvallisuus. Nopeiden ja helppojen

digitaalisten palveluiden tarjoaminen opiskelijoille on yksi Hoasin toiminnan pohjalla olevista teemoista. (Hoas s.a.)

Hoasilla on tällä hetkellä käytössä Helmi-chatbot. Helmi osaa vastata muun muassa asunnon hakemiseen ja asumiseen liittyviin kysymyksiin, tukien Hoasin palveluneuvojia asiakaspalveluun tuleviin kysymyksiin vastaamisessa. Organisaatiossa tiedostetaan kuitenkin nykyisen chatbottiratkaisun rajoitteet ja haasteet, ja chatbotin kautta tarjottavaa asiakaspalvelua ollaan kiinnostuneita parantamaan generatiivisen tekoälyn avulla. Generatiivinen tekoäly on kuitenkin aihealueena laaja, eikä organisaation sisältä löydy resursseja perehtymään uuteen teknologiaan niin syvällisesti kuin tahtotila olisi. Hoasin tavoitteena on kuitenkin tarjota nopeita ja helppoja digitaalisia palveluita, minkä edistämiseksi generatiivinen chatbotti tarjoaa hyviä mahdollisuuksia. Hoasin arvojen mukaisesti käyttöönotossa tulee kuitenkin huomioida vastuullisuus, oikeudenmukaisuus, turvallisuus ja avoimuus. Nämä kaikki ovat oleellisia teemoja, kun puhutaan generatiivisesta tekoälystä.

Opinnäytetyön tarkoituksena on tarjota organisaatiolle tietoa siitä, mitä organisaatiossa on huomioitava, kun generatiivinen chatbotti otetaan käyttöön asiakaspalvelussa. Jotta generatiivinen chatbotti voidaan ottaa organisaatiossa käyttöön Hoasin arvojen mukaisesti, on selvitettävä, miten varmistutaan generatiivisen chatbotin vastuullisuudesta, avoimuudesta ja turvallisuudesta, ottaen huomioon myös oikeudenmukaisuus ja yksilön huomioiminen. Opinnäytetyössä perehdytäänkin kattavasti näihin aihealueisiin, tukien Hoasin valmiuksia generatiivisen tekoälyn hyödyntämiseen toiminnassaan.

1.2 Opinnäytetyön tavoitteet ja rajaus

Opinnäytetyön tavoitteena on selvittää organisaation valmiuksia generatiivisen chatbotin käyttöönottoon asiakaspalvelun kontekstissa. Valmiuksia selvitetään selkiyttämällä generatiivisen chatbotin tavoitteita ja onnistuneen käyttöönoton edellytyksiä. Lisäksi opinnäytetyössä selvitetään, miten varmistutaan generatiivisen chatbotin luotettavuudesta, tietoturvasta ja eettisyydestä sekä pohditaan lainsäädännön vaatimuksia generatiiviselle chatbotille.

Opinnäytetyön tutkimuskysymykset ovat seuraavat:

- Mitä chatbotin uudistamisella tavoitellaan?
- Miten generatiivinen chatbotti otetaan onnistuneesti käyttöön osana organisaation asiakaspalvelua?
- Miten varmistutaan generatiivisen chatbotin luotettavuudesta, tietoturvasta ja eettisyydestä?
- Mitä vaatimuksia lainsäädäntö asettaa generatiiviselle chatbotille?

Opinnäytetyössä käsitellään generatiivista tekoälyä vain chatbottien kontekstissa, eli käytännössä opinnäytetyössä keskitytään isoihin kielimalleihin. Generatiivisella tekoälyllä tarkoitetaan kaikkea

tekoälyä, jolla voidaan luoda jotain uutta, mutta esimerkiksi kuvien luominen ei ole relevanttia chatbotin näkökulmasta katsottuna. Opinnäytetyössä käsitellään vain pintapuolisesti generatiiviseen tekoälyyn, isoihin kielimalleihin ja luonnollisen kielen käsittelyyn liittyviä teknologisia näkökulmia. Painopiste on näiden teknologioiden vaikutuksessa yksilöihin, organisaatioihin ja yhteiskuntiin. Myös päinvastainen näkökulma huomioidaan, eli yksilöiden, organisaatioiden ja yhteiskuntien vaikutus teknologioihin.

Opinnäytetyössä ei käsitellä generatiivisen chatbotin varsinaista teknistä toteutusta tai kehittämistä, vaan käydään läpi teemoja ja aihealueita liittyen esimerkiksi palveluntarjoajan valintaan, jotta toimeksiantaja voi varautua siihen, mitä generatiivisen chatbotin käyttöönotossa on huomioitava. Lopullinen tekninen toteutus ja päätökset teknologisista valinnoista jäävät generatiivisen chatbotin kehittäjän päätettäväksi.

1.3 Opinnäytetyön rakenne

Opinnäytetyö koostuu teoreettisesta viitekehyksestä, empiriasta eli opinnäytetyön metodologiasta, tutkimuksen toteutuksesta ja tuloksista sekä pohdinnasta. Opinnäytetyön luvuissa 2–4 käsitellään teoreettista viitekehystä. Toisessa luvussa käsitellään chatbottien teknologioita, luonnollisen kielen käsittelyn kehittymisestä isojen kielimallien soveltamiseen. Kolmas luku käsittelee chatbottien käyttöönoton edellytyksiä, huomioiden laajasti eri teemoja, kuten tavoitteiden ja mittareiden asettaminen, käyttäjäkokemus, vastuullisuus ja tiedonhallinta. Lisäksi kolmannessa luvussa käsitellään chatbotteihin liittyvää lainsäädäntöä, eli yleistä tietosuoja-asetusta ja tekoälysäädöstä. Neljännessä luvussa käydään läpi chatbotteihin liittyviä moninaisia riskejä ja keinoja niiden hallitsemiseen. Riskejä liittyen tietoturvaan, käyttäjien yksityisyyteen ja luottamukseen, vastuun määrittelemiseen sekä yhteiskunnallisiin vaikutuksiin käsitellään neljännessä luvussa.

Opinnäytetyön luvussa 5 esitellään tutkimuksen toteutus. Ensimmäinen alaluku (5.1) käsittelee tutkimuksen metodologiaa, lähestymistapaa sekä aineiston hankinta- ja analyysimenetelmiä. Toisessa alaluvussa (5.2) esitellään tutkimuksen tulokset. Opinnäytetyön 6. luvusta löytyy pohdinta. Kuudennen luvun ensimmäinen alaluku (6.1) liittää yhteen teoreettisen viitekehyksen ja tutkimuksesta saadut tulokset eli opinnäytetyön johtopäätökset. Toinen alaluku (6.2) käsittelee tutkimuksen luotettavuutta. Kolmas alaluku (6.3) kuvaa opinnäytetyöprosessia, sen onnistumisia ja epäonnistumisia sekä jatkokehittämisideoita.

Teoreettisen viitekehyksen ja tutkimuksen tulosten välinen yhteys on kuvattu peittomatriisissa (taulukko 1).

Taulukko 1. Peittomatriisi

Tutkimuskysymys	Tietoperusta	Tulokset	Haastattelu-kysymykset
Mitä chatbotin uudistamisella tavoitellaan?	2.1, 2.2, 2.3, 2.4, 3.1	5.2.1, 5.2.2, 5.2.5	1–4
Miten generatiivinen chatbotti otetaan onnistuneesti käyttöön osana organisaation asiakaspalvelua?	3.2, 3.3, 3.5, 3.8, 3.9, 4.3, 4.4	5.2.3, 5.2.4, 5.2.6	5–9
Miten varmistutaan generatiivisen chatbotin luotettavuudesta, tietoturvasta ja eettisyydestä?	3.4, 3.6, 3.7, 4.1, 4.2	5.2.7, 5.2.8	10–14
Mitä vaatimuksia lainsäädäntö asettaa generatiiviselle chatbotille?	3.10	5.2.9	15

2 Chatbottien teknologiat

Tässä luvussa pyritään avaamaan chatbottien eri teknologioita. Luonnollisen kielen käsittely ja sen kehittyminen vuosien mittaan on oleellinen teema, kun puhutaan chatboteista. Ensimmäisessä alaluvussa perehdytäänkin luonnollisen kielen käsittelyn historiaan ja kehittymiseen. Toisessa alaluvussa käydään läpi niin sanottujen perinteisten chatbottien toimintaa ja teknologiaa niiden taustalla. Kolmas alaluku on omistettu generatiivisille chatboteille ja isoille kielimalleille. Viimeinen tämän luvun alaluku käsittelee lyhyesti hybridichatbotteja, eli chatbotteja, joissa on yhdistetty generatiivista tekoälyä ja perinteisten chatbottien arkkitehtuuria.

2.1 Luonnollisen kielen käsittely

Luonnollisen kielen käsittely (natural language processing, NLP) on olennainen osa kaikkia chatbotteja. Se on yksi vanhimmista tekoälyn tutkimuksen aloista, jonka juuret ulottuvat aina 1950-luvulle asti. Luonnollisen kielen käsittelyn tavoitteena on ihmisten luonnollisen kielen ja tietokoneiden ymmärryskyvyn välisen kuilun ylittäminen: tietokoneiden tulisi kyetä käsittelemään, analysoimaan ja tulkitsemaan luonnollista kieltä sekä vastaamaan siihen merkityksellisesti ja hyödyllisesti. (Amaratunga 2023, luku 2.) Luonnollisen kielen käsittelyn osa-alueita ovat luonnollisen kielen ymmärtäminen ja luonnollisen kielen tuottaminen (Adamopoulou & Moussiades 2020). Luonnollisen kielen käsittelyn avulla tehdään muun muassa tekstin luokittelua, nimettyjen entiteettien tunnistamista, konekääntämistä, tekstin luomista, puheen tunnistamista, tekstin tiivistämistä, kysymyksiin vastaamista ja kielen mallintamista (Amaratunga 2023, luku 2).

Ensimmäiset luonnollisen kielen käsittelyn sovellukset 1960- ja 1970-luvuilla olivat sääntöpohjaisia järjestelmiä. Sääntöpohjaiset järjestelmät perustuivat yksinkertaisiin sääntöihin ja hahmonsovitukseen (pattern matching). Esimerkiksi yksi ensimmäisestä chatboteista, Joseph Weizenbaumin 1960-luvulla kehittämä ELIZA, hyödynsi näitä tekniikoita. (Amaratunga 2023, luku 2.) ELIZA etsi avainsanoja käyttäjän syötteestä ja muodosti vastauksen ennalta määriteltyjen sääntöjen perusteella. Tämä metodologia on vielä tänäkin päivänä yleisesti käytössä chatbottien kehittämisessä. (Khan & Das 2017, luku 1.)

Seuraavien vuosikymmenten aikana luonnollisen kielen käsittelyyn sisällytettiin kielitieteen teorioita ja sääntöjä, tavoitteena kielitieteen tiedon ja muodollisten kielioppisääntöjen hyödyntäminen. Kielitieteen teorioita hyödyntämällä pyrittiin saavuttamaan syvällisempi ymmärrys kielestä muun muassa syntaktisen ja semanttisen tiedon avulla. Tämä syvällisempi ymmärrys on ratkaisevaa kehittyneemmille luonnollisen kielen käsittelyn tehtäville, kuten kysymyksiin vastaamiselle ja luonnollisen kielen ymmärtämiselle. Kielitieteen teoriaan pohjautuvat tekniikat kuitenkin vaativat paljon aikaa ja vaivaa, kun kielitieteen säännöt ja rakenteet täytyi määritellä manuaalisesti, mikä rajoitti myös

luonnollisen kielen käsittelyn järjestelmien skaalautuvuutta. Lisäksi luonnollisen kielen monimutkaisuus ja lukemattomat poikkeukset kielioppisääntöihin aiheuttivat haasteita. (Amaratunga 2023, luku 2.)

1990- ja 2000-luvuilla luonnollisen kielen käsittelyssä siirryttiin datavetoisiin ja tilastollisiin menetelmiin. Tilastolliset menetelmät perustuivat isoihin määriin kielitietoa, joiden avulla rakennettiin todennäköisyyteen pohjautuvia kielimalleja. (Amaratunga 2023, luku 2.) Niitä käytettiin muun muassa puheentunnistuksen parantamiseen, oikeinkirjoituksen korjaamiseen, optiseen merkkien tunnistamiseen ja käsialan tunnistamiseen (McTear & Ashurkina 2024, luku 4). Vaikka tilastolliset lähestymistavat olivat lupaavia, niihinkin liittyi haasteita: datan vähyys, pitkän välimatkan riippuvuuksien käsitteleminen ja monimutkaisten semanttisten riippuvuuksien havaitseminen (Amaratunga 2023, luku 2).

Nykypäivänä useimmat luonnollisen kielen käsittelyn tekniikat perustuvat koneoppimiseen (Adamopoulou & Moussiades 2020). Koneoppimiseen pohjautuvat luonnollisen kielen käsittelyn järjestelmät ovat käytössä monissa eri tehtävissä, kuten chatboteissa, virtuaalisissa avustajissa, tunneanalyysityökaluissa ja konekääntämisen palveluissa (Amaratunga 2023, luku 2). Koneoppimiseen perustuvaa luonnollisen kielen käsittelyä voidaan hyödyntää esimerkiksi käyttäjän syötteen sisällön analysoimiseen ja sopivimman vastauksen löytämiseen. Lisäksi niiden etuna on kyky oppia jatkuvasti käydyistä keskusteluista. (Adamopoulou & Moussiades 2020.)

Koneoppimiseen pohjautuva luonnollisen kielen käsittely saapui 2000- ja 2010-luvuilla. Edistykset sekä koneoppimisen algoritmeissa että tietokoneiden laskentatehossa ja saatavilla olevan datan määrässä veivät eteenpäin myös luonnollisen kielen käsittelyä. Tilastolliset menetelmät vakiintuivat, ja yhdistettyinä koneoppimisen algoritmeihin alkoivat toteuttaa luonnollisen kielen käsittelyn tehtäviä. Esimerkiksi tunneanalyysi, syväoppiminen, nimetyn entiteetin tunnistaminen (Named Entity Recognition, NER), sanojen muuttaminen vektoreiksi ja neuroverkot paransivat merkittävästi useita luonnollisen kielen käsittelyn tehtäviä. (Amaratunga 2023, luku 2.)

Koneoppimiseen pohjautuvissa chatboteissa hyödynnetään usein keinotekoisia neuroverkkoja (Amadopoulou & Moussiades 2020). Neuroverkkoja hyödynnetään kielimalleissa, jotka oppivat tilastollisia rakenteita ja suhteita sanojen välillä. Erilaiset tekniikat mahdollistavat kehittyneemmän luonnollisen kielen käsittelyn. (Amaratunga 2023, luku 2.) Sanojen upotus (word embedding) tarkoittaa sanojen muuttamista vektoreiksi, ja tähän on olemassa monia eri tekniikoita, kuten syväoppimista hyödyntävä Word2vec. Vektorimuotoiset sanat syötetään neuroverkkoon ja neuroverkko tuottaa vastauksen. (Adamopoulou & Moussiades 2020.) Takaisinkytketyt neuroverkot (Recurrent Neural Networks, RNNs) mahdollistavat edellisiin sanoihin liittyvien kontekstien muistamisen ja pitkän välimatkan riippuvuuksien havaitsemisen (Amaratunga 2023, luku 2). Pitkän ja

lyhyen aikavälin muistiverkot (Long Short Term Memory networks, LSTMs) ovat tietyn tyyppisiä takaisinkytkettyjä neuroverkkoja, joita voi myös hyödyntää pitkän välimatkan riippuvuuksien havaitsemiseen tai aiemman tiedon muistamiseen. Takaisinkytketyt neuroverkot ottavat huomioon keskustelun aikaisemman kontekstin, ja tietoa siirretään yhdeltä neuroverkon osa-alueelta toiselle. (Adamopoulou & Moussiades 2020.)

Sequence-to-Sequence (Seq2Seq) on tyypillinen generatiivinen malli, jota voidaan hyödyntää vastauksen luomisessa. Malli luo vastauksen käyttäjän syötteen perusteella. Kahta takaisinkytkettyä neuroverkkoa voidaan käyttää mallissa enkooderina ja dekooderina, mikä on mallin tavallisin ja perinteisin versio. Pidemmissä lauseissa voidaan hyödyntää pitkän ja lyhyen aikavälin muistiverkkoja. Generatiivisen mallin avulla chatbotissa ei tarvita ennalta määriteltyjä vastauksia jokaiseen mahdolliseen käyttäjän syötteeseen. (Adamopoulou & Moussiades 2020.)

Transformereihin pohjautuvat kielimallit käyttävät tarkkaavaisuusmekanismeja havaitakseen sanojen väliset riippuvuudet samanaikaisesti. Tämä mahdollistaa tehokkaamman pitkien välimatkojen riippuvuuksien käsittelyn kuin takaisinkytketyt neuroverkot, ja transformer-arkkitehtuuri onkin pohjana monille edistyksellisille kielimalleille. (Amaratunga 2023, luku 2.) Transformerit ovat käytännössä korvanneet takaisinkytketyt enkooderi-dekooderi neuroverkot (McTear & Ashurkina 2024, luku 3). Transformer-arkkitehtuuri oli vallankumouksellinen luonnollisen kielen käsittelyssä ja sen tehokkuus on mahdollistanut suurten kielimallien ilmaantumisen (Amaratunga 2023, luku 3). Muun muassa GPT-3 ja BERT hyödyntävät transformereita arkkitehtuureissaan (McTear & Ashurkina 2024, luku 3). 2010-luvun lopussa nämä transformereita hyödyntävät esikoulutetut kielimallit pääsivät huippuluokan tuloksiin luonnollisen kielen käsittelyn tehtävissä. Esikoulutetut kielimallit on koulutettu valtavalla määrällä tietoa ja hienosäädetty tiettyihin luonnollisen kielen käsittelyn tehtäviin. Tämä on edesauttanut merkittävästi kielen ymmärtämisen, tekstin luomisen ja muiden luonnollisen kielen käsittelyn tehtävien kehittymisessä. (Amaratunga 2023, luku 2.)

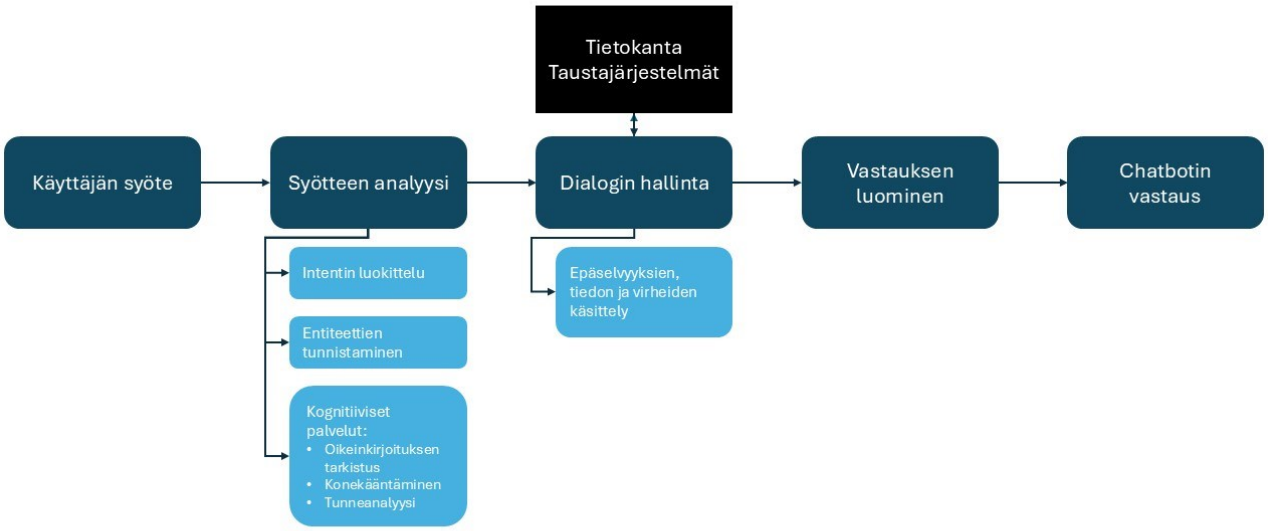
Sanojen upotus on yksi merkittävistä eroista neurokielimallien ja tilastollisten kielimallien välillä: isot kielimallit hyödyntävät sanojen vektorimuotoisia esitystapoja perussanojen sijaan. Tämä mahdollistaa hienovaraisempien erojen käsittelemisen, minkä lisäksi transformerit ja huomiomekanismi antavat neurokielimalleille paljon suuremman konteksti-ikkunan. (McTear & Ashurkina 2024, luku 4.)

2.2 Perinteiset chatbotit

Perinteiset chatbotit rakentuvat useasta komponentista, joiden avulla pyritään ymmärtämään luonnollista kieltä, hallitsemaan dialogia sekä tuottamaan vastaus luonnollisella kielellä. Näiden chatbottien oleellisia elementtejä ovat muun muassa intentit ja entiteetit. Intenteillä viitataan siihen käyttäjän tarkoitukseen tai tavoitteeseen, jonka takia hän on antanut syötteen chatbotille. Entiteetit

tarkoittavat merkityksellistä tietoa, joka poimitaan käyttäjän syötteestä myöhempää käyttöä varten. (McTear & Ashurkina 2024, luku 2.)

Adomopoulou ja Moussiades (2020) esittelevät perinteisen chatbotin arkkitehtuurin, jossa ei oteta kantaa eri teknologisiin valintoihin. Chatbottien lopullinen toteutustapa riippuu siitä, minkä tyyppistä chatbottia ollaan kehittämässä ja missä kontekstissa sitä tullaan käyttämään. Tämä arkkitehtuuri kuitenkin kattaa perinteisten chatbottien olennaisimmat osat ja niihin liittyvät asiat. Chatbottien tyypillinen arkkitehtuuri esitetään kuvassa 1.



Kuva 1. Perinteisen chatbotin arkkitehtuuri (mukaillen Adomopoulou & Moussiades 2020)

Chatbotin toiminta käynnistyy, kun se saa käyttäjältä pyynnön teksti- tai äänimuodossa. Tähän tarvitaan käyttöliittymä, esimerkiksi jokin viestisovellus, kuten Facebook, Slack tai WhatsApp. Käyttöliittymältä käyttäjän pyyntö siirtyy komponenttiin, joka analysoi käyttäjän viestin. Komponentti etsii käyttäjän viestistä intentin ja entiteetit. Intentti tarkoittaa sitä syytä tai aihetta, johon käyttäjän viesti voidaan yhdistää. Entiteetit ovat erilaisia parametrejä, jotka sisältävät viestissä olevia yksityiskohtia. Intenttien ja entiteettien tunnistaminen voidaan toteuttaa sääntöpohjaisesti tai luonnollisen kielen käsittelyn avulla. Käyttäjän syötteen analysointikomponentti voi sisältää myös joitain kognitiivisia palveluita analyysin parantamiseksi, kuten oikeinkirjoituksen tarkistusta, konekääntämistä tai tunneanalyysiä. (Adamopoulou & Moussiades 2020.) Tämä komponentti sisältää luonnollisen keskustelun ymmärtämiseen vaadittavat teknologiat. Nykypäivänä suurin osa chatboteista hyödyntää

koneoppimisen tekniikoita käyttäjän intentin luokittelemiseksi ja oleellisten entiteettien poimimiseksi. (McTear & Ashurkina 2024, luku 2.)

Tilanteessa, jossa keskustelu kestää useamman kuin yhden sananvaihdon, chatbottiin täytyy luoda keskustelupolkuja dialogin etenemisen hallitsemiseksi (McTear & Ashurkina 2024, luku 2). Dialogin hallintakomponentti vastaa keskustelun sujuvuudesta. Se hallitsee ja päivittää keskustelun kontekstia sekä tallentaa intentin ja entiteetit keskustelun kontekstiin. Dialogin hallinta tunnistaa, jos kaikkia tarvittavia tietoja ei kerätty, ja pyytää käyttäjältä lisätietoa esimerkiksi puuttuvien entiteettien täyttämiseksi. Myös jatkokysymyksien hallinta intentin tunnistamisen jälkeen kuuluu dialogin hallintakomponentin tehtäviin. Dialogin hallintakomponentti sisältää yleensä myös erilaisia moduuleja epäselvyyksien, virheiden ja tiedonkäsittelyyn. Epäselvyyksien käsittelymoduuli antaa vastauksen, kun chatbotti ei löydä intenttiä käyttäjän syötteestä tai syötettä ei ole lainkaan. Virheiden käsittelykomponentti varmistaa chatbotin toiminnan ja reagoi odottamattomiin virheisiin. Tietojenkäsittelykomponentti tallentaa keskustelun tiedot, joita chatbotti voi myöhemmin hyödyntää vastauksen muodostamisessa. (Adamopoulou & Moussiades 2020.)

Chatbotti hakee tarvitsemansa tiedot taustajärjestelmästä API-kutsun tai tietokantahaun avulla. Kun sopiva tieto on löytynyt, se välitetään takaisin dialogin hallintakomponentille vastauksen luomista varten. Vastauksen luomisessa chatbotissa voidaan hyödyntää sääntöpohjaista, hakupohjaista tai generatiivista luomistapaa. Dialogin hallintakomponentti välittää vastauksen luomisen komponentille arvot, joita vastauksessa tarvitaan. Vastauksen muodostettuaan chatbotti välittää sen käyttäjälle, minkä jälkeen se jää odottamaan uutta syötettä. (Adamopoulou & Moussiades 2020.) Perinteisissä chatboteissa vastaukset ovat tyypillisesti ennalta määriteltyjä (McTear & Ashurkina 2024, luku 2).

Perinteisissä chatboteissa on monia haasteita. Intentit ovat olleet vallitseva lähestymistapa luonnollisen kielen ymmärtämiseksi, mutta koska standardeja intenttejä ei ole olemassa, ne tulee luoda aina tarpeen mukaan. Intenttien määrä voi kasvaa jopa satoihin, jolloin ylläpitokin muuttuu hankalaksi. Tyypilliset, ennalta kirjoitetut vastaukset eivät ole kovin joustavia ja niiden lokalisointi on monimutkaista. Vaatii paljon työtä, että vastauksissa voidaan ennakoida kaikki mahdolliset eri tilanteet. Lisäksi dialogin hallinta voi käydä hyvin työlääksi, kun mahdollisia keskustelupolkuja on useita. Etenkin avoimempien keskustelujen kohdalla erilaisten päätöspuiden käyttäminen muuttuu nopeasti hallitsemattomaksi. (McTear & Ashurkina 2024, luku 2.)

2.3 Generatiiviset chatbotit

Aiemmin esitelty chatbottien perinteinen arkkitehtuuri on muuttumassa neuroverkkoihin pohjautuvien chatbottien myötä. Neuroverkkoihin pohjautuvat chatbotit eivät käsittele käyttäjän syötettä ja

vastauksen luomista eri vaiheina, vaan kaikki tapahtuu yhden vaiheen aikana. Ajatuksena on, että järjestelmä pystyy päättelemään vastauksen syötteen perusteella, ja sitä pystytään kouluttamaan automaattisesti tietoaineiston pohjalta, jolloin esimerkiksi erillistä dialogin hallintaa ei tarvita. (McTear & Ashurkina 2024, luku 3.)

Generatiiviset chatbotit ovat chatbotteja, joissa hyödynnetään isoja kielimalleja. Tässä opinnäytetyössä ”perinteisillä chatboteilla” tarkoitetaan sellaisia chatbotteja, joissa ei hyödynnetä isoja kielimalleja. On lukuisia eri tapoja ja tekniikoita toteuttaa perinteisiä chatbotteja, ja monia perinteisissäkin chatboteissa käytettyjä luonnollisen kielen käsittelyn ja koneoppimisen menetelmiä käytetään myös generatiivisissa chatboteissa. Isot kielimallit ovat kuitenkin kaikista edistyneimpiä luonnollisen kielen käsittelyn teknologioita (Engler & Dhamani 2024, luku 1). Tässä luvussa kerrotaan, mikä tekee generatiivisista chatboteista muista poikkeavia.

2.3.1 Isot kielimallit

Kielimallit pyrkivät ennustamaan sanajonon todennäköisyyttä tietyssä kielessä. Luodakseen uutta tekstiä tai arvioidakseen lauseen todennäköisyyttä, mallien on osattava tietyn kielen tilastolliset ominaisuudet ja sanojen väliset suhteet. (McTear & Ashurkina 2024, luku 4.) Kielimallit ovat oleellinen osa edistyneempiä luonnollisen kielen käsittelyn sovelluksia, kuten isoja kielimalleja. Kielimallit voidaan jakaa kahteen kategoriaan sen perusteella, mitä tehtäviä ne suorittavat: generatiivisiin ja ennustaviin kielimalleihin. Generatiiviset kielimallit luovat uutta tekstiä opitun koulutusmateriaalin pohjalta, ja niitä voi käyttää esimerkiksi tekstin ja tarinoiden sepittämiseen sekä runojen kirjoittamiseen. Ennustavat kielimallit ennustavat seuraavan sanan todennäköisyyttä, ja niitä käytetään esimerkiksi ennakoivassa tekstinsyötössä ja konekääntämisessä. (Amaratunga 2023, luku 2.) Keskustelevassa tekoälyssä, kuten chatbotissa, kielimallia hyödynnetään johdonmukaisen ja asiayhteyteen sopivan tekstin luomiseen ennustamalla edeltävien sanojen perusteella seuraava todennäköinen sana sanajonossa (McTear & Ashurkina 2024, luku 4).

Isoista kielimalleista tekee isoja niiden parametrien määrä (McTear & Ashurkina 2024, luku 4). Esimerkiksi GPT-3:n parametrimäärän arvioidaan olevan 175 miljardia parametriä. Parametrien määrä on yleisesti suhteessa siihen, kuinka paljon ja kuinka monimutkaisia asioita kielimalli voi oppia, mutta suuri määrä parametrejä vaatii myös paljon laskentatehoa. Isojen kielimallien kouluttaminen voi olla kallista ja aikaa vievää. Lisäksi niiden riskinä on ylisovittaminen, jos kielimalleja koulutetaan liian pienellä tietoaineistolla. Ylisovittamisen myötä kielimalli oppii koulutustiedon, mutta ei kykene yleistämään sitä. Tätä voidaan kuitenkin ehkäistä käyttämällä suurta tietomäärää ison kielimallin kouluttamiseen. (Amaratunga 2023, luku 4.)

Isot kielimallit koulutetaan valtavilla tietoaineistoilla, minkä ansiosta ne saavat oppinsa monipuolisesta tekstiaineistosta erilaisia kirjoitustyylejä ja kieliä. Tietoa ison kielimallin kouluttamiseen hankitaan sekä erilaisista olemassa olevista tekstiaineistoista että internetissä olevasta sisällöstä, kuten Wikipediasta. (Amaratunga 2023, luku 4.) Lisäksi kouluttamisessa hyödynnetään tekstiä muun muassa kirjoista, artikkeleista ja muusta tekstimuotoisesta tiedosta. Koulutustiedon määrä, laatu ja monimuotoisuus ovat olennaisia isoja kielimalleja koulutettaessa. (McTear & Ashurkina 2024, luku 4.) Ison tietoaineiston myötä kielimalli myös oppii yleistämään tietonsa niihin tapauksiin, joihin se ei ole aiemmin törmännyt, jolloin ylisovittamisen riski vähenee. Suuri määrä tietoa voi myös sisältää harvinaisempia skenaarioita, joita ei pienemmissä tietomäärissä ilmene, mikä auttaa mallia vastaamaan myös esimerkiksi hyvin erikoistuneisiin aihealueisiin liittyviin kysymyksiin. (Amaratunga 2023, luku 4.) Kouluttamisen aikana kielimalli oppii sanojen väliset tilastolliset suhteet, minkä perusteella se pystyy ennustamaan todennäköisintä seuraavaa sanaa (McTear & Ashurkina 2024, luku 4).

Koulutusaineiston valmistelu on olennaista, koska hankitussa aineistossa voi olla virheitä tai toistoa, tai se voi olla muuten sopimatonta kouluttamiseen. Tietoaineiston siivoaminen ja valmistelu varmistaa sen, että kielimalli oppii laadukkaasta tiedosta. Tiedon määrä lisää myös vaatimuksia laskentateholle, minkä lisäksi tiedon säilyttäminen vaatii valtavia määriä tallennustilaa. Lisäksi internetistä hankittu koulutusaineisto voi sisältää ennakkoluuloja, jotka kielimalli voi oppia. Jos koulutusaineistoa ei ole valmisteltu kunnolla, on olemassa myös riski, että malli oppii väärää tai harhaanjohtavaa tietoa. Parametrien ja tiedon valtava määrä kuitenkin mahdollistaa sen, että isoilla kielimalleilla on merkittäviä kielen ymmärtämisen ja luomisen taitoja. (Amaratunga 2023, luku 4.)

Isoja kielimalleja käytetään monessa eri sovelluksessa, kuten dialogin ja sisällön luomiseen, tiedon poimimiseen, tekstin luokitteluun, yhteenvetoon, konekääntämiseen ja koodin luomiseen. Eri kielimalleja käytetään eri tehtäviin, esimerkiksi BERT toimii hyvin tiedon poimimiseen ja tekstin luokitteluun, kun taas GPT-3 ja LLaMA sopivat hyvin keskustelevaan tekoälyyn ja sisällön luomiseen. (McTear & Ashurkina 2024, luku 4.)

2.3.2 Perusmallit

Perusmalleista puhuttaessa tarkoitetaan sellaisia isoja kielimalleja, jotka on tarkoitettu toimimaan monien erilaisten sovellusten perustana (McTear & Ashurkina 2024, luku 4). Tämä on muutos aiemmasta, jolloin kielimalleja on ollut tapana kouluttaa jotakin tiettyä tehtävää varten. Hienosäätämällä ja kustomoimalla vahvoja perusmalleja saadaan monipuolista osaamista hyödynnettyä monenlaisiin tehtäviin ja sovelluksiin. Perusmallilla voidaan tarkoittaa yleisesti esikoulutettuja isoja malleja, mutta yleensä niillä viitataan isoihin koneoppimista hyödyntäviin kielimalleihin. (Amaratunga 2023, luku 4.)

Perusmallit perustuvat siihen, että ne on esikoulutettu valtavalla määrällä monipuolista tietoa. Tämän ansiosta perusmalli oppii laajasti ja kirjavasti erilaisia muotoja, rakenteita ja tietoa. Tämä pohjakoulutus mahdollistaa sen, että ne voivat toimia pohjana monille erilaisille sovelluksille ja tehtäville. Perusmalleja voidaan hienosäätää tiettyihin toimialoihin tai tehtäviin soveltuviksi. Hienosäädön avulla esikoulutettu yleinen tieto voidaan yhdistää uuteen, tehtävä- tai alakohtaiseen tietoon. Resurssien käytön näkökulmasta on yleensä järkevämpää kouluttaa yksi iso perusmalli, jota voidaan soveltaa erilaisiin tehtäviin. (Amaratunga 2023, luku 4.)

Isot kielimallit kuuluvat perusmalleihin, koska niiden ominaisuuksia ja piirteitä voidaan hyödyntää monissa eri sovelluksissa. Hienosäädön avulla niitä voidaan soveltaa eri tehtäviin ja eri aloille. Esikoulutettu tieto voidaan siirtää moniin sovelluksiin, jolloin tehtäväkohtaisen tiedon ja koulutuksen tarpeen määrä vähenee. Kun isoja kielimalleja hyödynnetään perusmalleina, sovellusten kehittäminen nopeutuu huomattavasti. Isoihin kielimalleihin liittyvää perusmallio minaisuutta korostaa se, että niitä voidaan käyttää myös muihin kuin tekstiin liittyviin tehtäviin, kuten taiteen luomiseen tai koodin kirjoittamiseen. (Amaratunga 2023, luku 4.)

Isojen kielimallien esikouluttaminen on hyvin kallista ja vaatii paljon resursseja, minkä takia se ei suurimmalle osalle yrityksistä ja organisaatioista ole realistista tai mahdollista. Monissa käyttötapauksissa perusmallien hyödyntäminen on kuitenkin hyvä vaihtoehto, jolla vältetään uuden mallin kouluttamisen kustannukset. Perusmallit on kuitenkin koulutettu hyvin geneerisiksi, eivätkä ne siksi välttämättä toimi hyvin tilanteissa, joissa tarvitaan paljon monimutkaista tiettyyn alaan liittyvää tietoa ja terminologiaa, jota ei ollut mallin kouluttamiseen käytetyssä aineistossa mukana. (McTear & Ashurkina 2024, luku 4.) Isojen kielimallien soveltamiseksi tiettyyn tehtävään tai sovellukseen voi käyttää joko kehotesuunnittelua (prompt engineering) tai hienosäätöä (fine-tuning).

2.3.3 Kehotesuunnittelu

Kehotesuunnittelun avulla voidaan syötteen muotoilun avulla vaikuttaa ison kielimallin antamaan vastaukseen. Kehotteiden muotoilussa tulee pyrkiä tarkkuuteen ja yksityiskohtaisuuteen. Esimerkkien avulla mallille voidaan kertoa, minkälaista vastausta haetaan. Iteratiivinen lähestymistapa on suositeltavaa, ja erilaisia sana- ja lausemuotoja kannattaa kokeilla, jos vastaus ei ole riittävän hyvä. (Amaratunga 2023, Luku 4.) Yksinkertaisimmillaan kehotemuotoilu tarkoittaa sitä, että käyttäjä syöttää kehotteen mallille. Tätä kutsutaan nollakehottamiseksi. Kehotetta voidaan muotoilla erilaisilla ohjeilla, kuten tehtävän kuvauksella ja esimerkillä halutusta vastauksesta (yhden ohjauksen kehottaminen) tai kokonaisella kokoelmalla koulutusesimerkkejä mallin ohjaamiseksi (vähäisen ohjauksen kehottaminen). (McTear & Ashurkina 2024, luku 4.) Kehotemuotoilun avulla mallia voidaan hyödyntää optimaalisesti, vähentäen vastausten epäselvyyttä. Käytännössä kehotemuotoilu

mahdollistaa mallin kustomoinnin tiettyihin käyttötarkoituksiin ilman, että sitä tarvitsisi erikseen kouluttaa niitä varten. (Amaratunga 2023, luku 4.)

Myös järjestelmäkehotteilla voidaan vaikuttaa kielimallin tuottamiin vastauksiin. Järjestelmäkehotteella tarkoitetaan kattavaa ohjeistusta, joka asettaa raamit kielimallin ja käyttäjän keskustelulle. Lisäksi kehotteen asetukset tai parametrit ovat hyödyllisiä työkaluja kielimallin tuloksen parantamiseksi ja hallitsemiseksi. Asetukset sisältävät lisätietoja, jotka välitetään kielimallille kehotteen mukana. Parametrien saatavuus riippuu mallista, mutta esimerkiksi yleisellä lämpötilaparametrillä voidaan vaikuttaa siihen, kuinka satunnaista luotu teksti on. Pienempi arvo vaikuttaa siihen, kuinka vakaasti kielimalli toimii, ja se toimii hyvin tehtäviin, joissa tarvitaan tarkkoja vastauksia. Korkeamman arvon avulla kielimalli saadaan toimimaan luovemmin. Muita parametrejä on muun muassa toiston rangaistus, sanojen luomisen lopettaminen sekä maksimi- ja minimipituus. Parametrejä säätämällä voidaan vaikuttaa haluttuun lopputulokseen. (McTear & Ashurkina 2024, luku 6.)

2.3.4 Hienosäätö

Joissain tapauksissa kehotesuunnittelu ei välttämättä riitä halutun lopputuloksen saavuttamiseksi. Tällöin mallin hienosäätö tulee tarpeeseen. Hienosäädössä esikoulutettu kielimalli sovelletaan tiettyyn tehtävään tai alaan sopivaksi. Tällöin hyödynnetään mallin yleistä tietoa, ja mukautetaan se soveltumaan paremmin haluttuun tehtävään tai sovellukseen. Hienosäätö tapahtuu kouluttamalla mallia pienemmällä, kapea-alaisemmalla ja tehtäväkohtaisella tietoaineistolla. Yleensä tietoaineisto liittyy juuri haluttuun tehtävään tai sovellukseen, kuten esimerkiksi kysymyksiin vastaamiseen, sentimenttianalyysiin tai lääketieteellisen tekstin luokitteluun. (Amaratunga 2023, luku 4.)

Hienosäädöllä voidaan vaikuttaa myös mallin käyttäytymiseen, kuten vastausten kohteliaisuuteen ja ytimekkyyteen. Hienosäätö voi myös vähentää mallin taipumusta hallusinointiin sekä parantaa vastausten yhdenmukaisuutta. Lisäksi hienosäätö antaa organisaatiolle paremmat mahdollisuudet hallita mallin kouluttamista sekä parantaa sen läpinäkyvyyttä ja yksityisyyttä. (McTear & Ashurkina 2024, luku 4.) Käytännössä esikoulutettuja malleja päivitetään tehtäväkohtaisella tietoaineistolla. Hienosäätämisessä on kuitenkin myös haasteensa. Ylisovittamisen ja aiemmin opitun yleisen tiedon unohtamisen riski on olemassa. (Amaratunga 2023, luku 4.) Hienosäätäminen on myös huomattavasti kalliimpaa kuin pelkkä kehotesuunnittelu, minkä lisäksi tilanteissa, joissa koulutusaineistossa oleva tieto muuttuu usein, hienosäätö voi olla epäkäytännöllistä (McTear & Ashurkina 2024, luku 4).

2.3.5 Retrieval-Augmented Generation

Isoissa kielimalleissa on haasteita, kuten ajan tasaisen tiedon puute. Isot kielimallit koulutetaan tietoaineistolla, ja tämän tietoaineiston ajantasaisuus määrittää kielimallin vastaukset. Vastauksista

siis puuttuu esimerkiksi sellaiset uutiset ja tieteelliset läpimurrot, jotka ovat ilmenneet vasta kouluttamisen jälkeen. Isoa kielimallia ei siis suoraan voida hyödyntää tehtäviin, joissa tarvitaan ajantasaista tietoa. (Gheorghiu 2024, luku 1.)

Isot kielimallit eivät myöskään kykene erottamaan totta valheesta. Ne pystyvät luomaan vakuuttavan kuuloista misinformaatiota, minkä lisäksi vastaukset eivät välttämättä ole loogisia, tietoon pohjautuvia tai harmittomia. Myös koulutusaineistossa olevat ennakkoluulot vaikuttavat mallin toimintaan, mikä saattaa johtaa siihen, että kielimalli vahvistaa ja levittää vahingollisia ennakkoluuloja tai muuten haitallista sisältöä. (Gheorghiu 2024, luku 1.) Koska kielimallin tieto rajoittuu siihen aineistoon, mikä sillä oli käytettävissään koulutuksen aikana, malli saattaa luoda väärää tietoa antamalla vain todennäköisimmän seuraavan sanajonon, riippumatta sen todenperäisyydestä (McTear & Ashurkina 2024, luku 4).

Pitkien dokumenttien kontekstin ja sisällön muistaminen voi olla isoille kielimalleille vaikeaa. Malli saattaa unohtaa osia aiemmasta keskustelusta, mikä voi johtaa vaillinaisiin vastauksiin. Lisäksi mallien päättelykyky saattaa olla puutteellista, minkä lisäksi toiminnan läpinäkymättömyys vaikeuttaa päättelyprosessin virheiden tutkimista. (Gheorghiu 2024, luku 1.)

Retrieval-Augmented Generation (RAG) on tekniikka, joka yhdistää tiedon haun ja generatiiviset mallit. Tarkoituksena on hakea oleellista tietoa annetusta tietolähteestä ja hyödyntää tätä tietoa paremman, tarkemman vastauksen luomisessa. RAG:in avulla on mahdollista ainakin osittain paikata isoissa kielimalleissa olevia puutteita. (Gheorghiu 2024, luku 1.) Tieto haetaan ulkoisesta tietokannasta ja sisällytetään kehotteeseen. Tällä tavalla vastaus todennäköisemmin sisältää ajan tasaista ja oikeaa tietoa, vähentäen hallusinointia ja muita riskejä. Erityisen käyttökelpoinen tekniikka on sovelluksille, joissa käytetään suojattuja tietoja tai tietoa aiemmista keskusteluista käyttäjien kanssa. (McTear & Ashurkina 2024, luku 4.) RAG lisää mallin hallittavuutta ja luotettavuutta. Järjestelmää voidaan myös pyytää esittämään lähde, josta tieto on peräisin, mikä helpottaa tiedon todenperäisyyden arviointia. (Gheorghiu 2024, luku 1.)

RAG-metodiin kuuluu kaksi vaihetta: haku ja generointi. Haun aikana etsitään dokumentti, joka vastaa eniten käyttäjän kehotetta. Uusi kehote muodostetaan käyttäjän alkuperäisestä kehotteesta ja haetun dokumentin tai tiedon sisällöstä. Generointivaiheessa kielimalli luo vastauksen uuden kehotteen ja esikoulutuksen aikana opitun tiedon avulla. RAG:in avulla malli voi hyödyntää ajantasaista ja tarkempaa tietoa. (McTear & Ashurkina 2024, luku 4.) On kuitenkin hyvä muistaa, että vaikka isojen kielimallien haasteita ja rajoitteita voidaan jonkin verran paikata RAG-menetelmällä, malli saattaa silti hallusinoida tai antaa vääriä vastauksia, minkä takia vastauksien laatua on olennaista seurata (Gheorghiu 2024, luku 1).

2.3.6 Suomen kieli isoissa kielimalleissa

Suomi on pieni kieli, jota puhuu alle 0.1 % maailman väestöstä. Isojen kielimallien tutkimus onkin keskittynyt pääasiassa englannin kieleen, ja englanniksi löytyy myös paljon tekstiaineistoa isojen kielimallien kouluttamista varten. Suomen kielistä tekstiaineistoa on saatavilla paljon vähemmän, mikä luo haasteita isojen kielimallien kouluttamiselle. (Luukkonen ym. 2023, 1.) Isojen kielimallien demokratisoiminen on tärkeää, ja se vahvistaa digitaalista suvereniteettia (Turun yliopisto 2024).

Turun yliopiston TurkuNLP-tutkimusryhmä on kehittänyt innovatiivista lähestymistapaa suomen kaltaisille vähäisen koulutusaineiston kielille. Poro 34B on avoin kielimalli, jonka kouluttamisessa on yhdistetty suomen kielen resurssit englannin kieleen, jolla on saatavilla paljon koulutusaineistoa. Lähestymistavan avulla voidaan parantaa mallin suorituskykyä vähäresurssissa kielissä, vähentämättä kuitenkaan mallin toimivuutta englanniksi. (Turun yliopisto 2024.)

Poro 34B on perusmalli, jota voidaan hyödyntää eri sovelluksissa. Tällä hetkellä malli on tarkoitettu lähinnä tutkimuskäyttöön, ja se tarvitsee hienosäätöä, testausta ja lisäkoulutusta, jos sitä halutaan käyttää tuotannossa. (Turun yliopisto 2024.) TurkuNLP on yhteistyössä SiloAI:n kanssa kehittänyt myös Viking 13B:n, ison kielimallin, joka taitaa suomen kielen lisäksi myös muut pohjoismaiset kielet sekä useita ohjelmointikieliä (SiloAI 2024). Poro 34B ja Viking 13B kuuluvat malliperheisiin, joita kehitetään jatkuvasti.

2.4 Hybridichatbotit

Hybridichatboteilla tarkoitetaan tässä opinnäytetyössä sellaisia chatbotteja, joissa hyödynnetään sekä perinteisten chatbottien arkkitehtuuria, että isoja kielimalleja. Moniin perinteisten chatbottien rakentamista varten tarkoitettuihin alustoihin, kuten Dialogflow CX, IBM Watson ja Microsoft Bot Framework, on viime aikoina integroitu generatiivista tekoälyä. Tällä hetkellä jää vielä nähtäväksi, tuleeko generatiivinen tekoäly korvaamaan täysin perinteiset chatbotin rakentamiseen suunnatut alustat. (McTear & Ashurkina 2024, luku 7.)

Tällaisten niin kutsuttujen hybridialustojen avulla perinteisiin chatbotteihin voidaan yhdistää generatiivisen tekoälyn ominaisuuksia. Generatiivista tekoälyä voidaan liittää esimerkiksi vastauksen luomisen komponenttiin, mikä mahdollistaa ainutlaatuiset ja personoidut vastaukset. Lisäksi useat alustat tarjoavat mahdollisuuden eri kielimallien käyttöön, mikä mahdollistaa useamman eri kielimallin hyödyntämisen saman chatbotin sisällä. (McTear & Ashurkina 2024, luku 7.)

Esimerkiksi Dialogflow CX tarjoaa mahdollisuuksia hyödyntää isoja kielimalleja sisällön jäsentämiseen ja ymmärtämiseen, chatbotin vastausten luomiseen ja keskustelupolun kontrolloimiseen. Generatiivisen tekoälyn avulla on mahdollista vähentää chatbottien suunnitteluun käytettyä aikaa sekä

parantaa chatbotin laatua. (Google Cloud 2024.) IBM Watson taas tarjoaa mahdollisuuden hyödyntää isoja kielimalleja esimerkiksi älykkääseen tietojen keräämiseen käyttäjän syötteestä ja muodostamaan keskustelevan vastauksen käyttäjän kyselyyn (IBM Cloud 2024a, IBM Cloud 2024b).

Alustavaa tutkimusta on tehty myös isojen kielimallien hyödyntämisestä tiettyjen perinteisten chatbottien komponenttien parantamiseksi edellä mainittujen alustojen ulkopuolella. Isoja kielimalleja voidaan hyödyntää esimerkiksi intenttien ja entiteettien tunnistamisessa (Villa ym. 2024). Tutkimusta hybridichatboteista, joissa hyödynnetään sekä isoja kielimalleja että perinteisten chatbottien arkkitehtuuria, löytyy toistaiseksi kuitenkin verrattain vähän.

3 Chatbottien käyttöönoton edellytykset

Chatbotin kehittämiseen liittyy lukuisia erilaisia haasteita, ja näiden haasteiden määrä lisääntyy, kun käytetään generatiivista tekoälyä. Haasteisiin voidaan kuitenkin vastata erilaisin toimenpitein ja huomioitavin asioin, kun chatbottia lähdetään suunnittelemaan, toteuttamaan ja lopulta käyttöönottamaan. Jotta organisaatio voi turvallisin mielin hyödyntää chatbotteja ja tekoälyä, on hyvä tutustua olemassa oleviin tutkittuihin parhaisiin lähestymistapoihin ja käytäntöihin.

Seuraavissa luvuissa käydään läpi eri teemoja, joihin on oleellista kiinnittää huomiota, kun organisaatiossa harkitaan generatiivisen chatbotin käyttöönottoa. Chatbotin kehittymistä lähestytään useista eri näkökulmista ja eri asioita painottaen. Teemat antavat kattavan kehyksen sille, mitä asioita organisaation tulee ottaa huomioon generatiivisen chatbotin elinkaaren aikana, jotta lopputulos olisi vastuullinen ja hyödyllinen sekä organisaatiolle että asiakkaalle.

3.1 Tavoitteet ja mittareiden asettaminen

Selkeän liiketoimintatavoitteen asettaminen on chatbotin onnistuneen kehityksen ja käyttöönoton kannalta tärkeässä asemassa. Skuridin ja Wynn (2024) suosittelevat, että projektissa keskitytään vain yhden liiketoimintatavoitteen saavuttamiseen. Useampaan tavoitteeseen keskittyminen samanaikaisesti voi johtaa siihen, että keskittyminen hajautuu eikä resursseja optimoida, mikä puolestaan vaikuttaa valmiin chatbotin laatuun tai pahimmillaan hankkeen epäonnistumiseen (Skuridin & Wynn 2024). Vastuullisen tekoälyhankinnan näkökulmasta liiketoiminnan käyttötapaus, päämäärä ja kriteerit on syytä määritellä huolella. Myös nykyisten prosessien ja niissä havaittujen aukkojen dokumentoiminen on suositeltavaa, jotta organisaatiolla on selkeä kuva siitä, miten tekoälyratkaisulla voidaan parantaa nykyisiä prosesseja. (World Economic Forum 2023, 10.) Tavoitteiden kommunikoiminen selkeästi sekä tiimille, palveluntuottajille että johdolle on oleellista (Skuridin & Wynn 2024).

Tavoitteen määrittämisen yhteydessä organisaation kannattaa myös pohtia vaihtoehtoja chatbotille. Organisaation on varmistuttava siitä, että chatbotti on paras tapa saavuttaa halutut liiketoimintatavoitteet, minkä lisäksi se tuottaa hyötyjä myös asiakkaalle (Skuridin & Wynn 2024). Tekoälyn hyödyntämisen tulee olla paras mahdollinen ratkaisu käsillä olevaan ongelmaan, kilpailukyvyn parantamiseen ja positiivisten liiketoimintatulosten saavuttamiseen. Organisaation tulee selvittää, että samoja tuloksia ei voida saavuttaa sellaisella teknologialla, joka ei hyödynnä tekoälyä, sekä ymmärtää, miten tekoäly auttaa saavuttamaan liiketoimintatavoitteet. (World Economic Forum 2023, 9–13.)

Mittareiden asettaminen tavoitteiden seuraamista varten on yhtä lailla oleellista. Tekoälyhankinnoissa tulee keskittyä ratkaisuihin, jotka tuottavat organisaatiolle arvoa, ja arvon kehittymistä tulee seurata keskeisten suorituskykyindikaattoreiden avulla. Suorituskykyindikaattorien valinta ja määrittely on kriittistä tavoitteiden saavuttamisen seuraamiseksi. (World Economic Forum 2023, 16.) Skuridinin ja Wynnin (2024) mukaan chatbotin tuottavuus on kriittinen menestystekijä, ja sen käytön helppoutta, nopeutta ja mukavuutta tulee seurata onnistuneen käyttöönoton takaamiseksi. Chatbotin käyttäytymis- ja onnistumismittareita on seurattava, jotta pystytään varmistumaan siitä, että chatbotti saavuttaa tavoitteensa (Skuridin & Wynn 2024). On hyvä huomioida, että tekoälyratkaisujen arvon tuottaminen tapahtuu yleensä pidemmällä aikavälillä kuin sellaisilla ratkaisuilla, joissa ei hyödynnetä tekoälyä. Jatkuvasta kouluttamisesta koituvat kustannukset sekä uuteen teknologiaan liittyvät epävarmuudet ja riskit on hyvä ottaa huomioon mittareita suunniteltaessa. (World Economic Forum 2023, 16.)

Palveluntarjoajan valinnassa on tärkeää kiinnittää huomiota siihen, että palveluntarjoaja vastaa organisaation liiketoiminnan vaatimuksia (Skuridin & Wynn 2024). Palveluntarjoajan tulee ymmärtää organisaation liiketoimintatavoitteet ja kyetä selittämään, miten heidän ratkaisunsa auttaa saavuttamaan ne. On myös hyvä selvittää, minkä tasoisia takuita palveluntarjoaja voi antaa prosessin ja liiketoiminnan tuloksista. (World Economic Forum 2023, 12–13.) Palveluntarjoajan tulee olla sitoutunut liiketoimintatavoitteiden saavuttamiseen. Sitoutuminen on avainasemassa onnistuneessa kumppanuudessa. (World Economic Forum 2023, 17.)

Organisaation on hyvä tutkia eri palveluntarjoajien ja palveluiden tarjontaa alalla ja alueella (Skuridin & Wynn 2024). Organisaation tulee katselmoida saatavilla olevia tekoälyratkaisuja ja niiden tarjoajia. Tähän liittyy myös tiedon monimutkaisuuden ja ratkaisun teknisten tietojen ymmärtäminen sekä lukuisten sellaisten palveluntarjoajien tunnistaminen, jotka voisivat auttaa liiketoimintatavoitteiden saavuttamisessa. (World Economic Forum 2023, 10.) Palveluntarjoajien projektiportfolioihin on hyvä perehtyä, sekä mahdollisuuksien mukaan pyytää suosituksia palveluntarjoajan asiakkailta (Skuridin & Wynn 2024). Palveluntarjoajalta tulee myös vaatia ratkaisun kyvykkyyksien avaamista, ja mahdollisten referenssien esittämistä. Valitun ratkaisun tulisi myös tarjota näkemystä liiketoimintatavoitteiden saavuttamiseen ja käyttäjätyytyväisyyden mittaamiseen. (World Economic Forum 2023, 12–13.)

3.2 Ketterä kehittäminen

Ketterät kehittämismenetelmät auttavat organisaatiota menestymään epävarmassa ympäristössä. Muutokseen ja epävarmuuteen reagoiminen on helpompaa, kun ketterät menetelmät ovat käytössä. Tärkeää on kokeileminen, palautteen kerääminen ja muutosten tekeminen palautteen perusteella. (Agile Alliance s.a.)

Ketteriä kehitysmenetelmiä kannattaa noudattaa chatbotin suunnittelun, toteutuksen ja ylläpidon aikana. Chatbotin kehittämistä tulisi lähestyä iteratiivisesti ja keskittyen laadukkaisiin, mutta rajattuihin toiminnallisuuksiin. Chatbotin toimintaa tulee testata jokaisen iteraation tai uuden toiminnallisuuden jälkeen ja mukauttaa suunnitelmaa saadun palautteen perusteella. Chatbottia kannattaa ensin testata sisäisesti tiimissä ja organisaation liiketoiminta-asiantuntijoiden kanssa. Tämän jälkeen chatbotin toimintaa on hyvä testata pienellä oikeiden asiakkaiden pilottiryhmällä, oikeaa tietoaineistoa hyödyntäen. (Skuridin & Wynn 2024.)

Ketterän kehittämisen avulla voidaan vähentää uuden teknologian epävarmuuksia ja riskejä. Lisäksi ratkaisun toteuttaminen vaiheittain auttaa välttämään isoja etukäteiskustannuksia. Esimerkiksi proof-of-concept (POC) ennen täyttä implementaatiota voi olla varteenotettava vaihtoehto. (World Economic Forum 2023, 17.)

3.3 Käyttäjäkokemus

Käyttäjäkokemuksen arvioiminen on oleellinen osa onnistuneen chatbotin kehittämistä ja käyttöönottoa. Skuridinin ja Wynnin (2024) mukaan chatbotin on oltava paras tapa ratkaista käyttäjän ongelmat ja sen tulee toimia paremmin kuin organisaation muut järjestelmät. Chatbotin käytön helppous, nopeus ja mukavuus ovat tärkeitä, ja käyttäjäkokemusta tulee tutkia ja testata. Käyttäjäkokemukseen liittyvät vaatimukset on kirjattava chatbotin määrittelyyn. Intuitiivisuus ja yksinkertaisuus luovat hyvät lähtökohdat laadukkaalle käyttäjäkokemukselle. (Skuridin & Wynn 2024.)

Chatbotin suunnittelussa on otettava huomioon kulttuuriset normit. Käyttäjien demografia, tuki chatbotin käytölle ja chatbotin viestintätyyli vaikuttavat chatbotin käyttäjäkokemukseen, ja nämä tulee huomioida chatbotin suunnittelussa ja kehityksessä. Huomioimalla kulttuurinen konteksti, jossa chatbottia tullaan käyttämään, voidaan edesauttaa suotuisia olosuhteita chatbotin käyttöönotolle. (Urbani, Ferreira & Lam 2024.) Generatiivinen tekoäly mahdollistaa esimerkiksi personoituja kokemuksia, ja sitä kannattaa hyödyntää teknologian tekemisessä helpommin lähestyttäväksi kaikille käyttäjille (Sidaoui, Mahr & Odekerken-Schröder 2024).

Chatbotin persoonallisuuden suunnitteluun kannattaa myös panostaa. On tärkeää, että chatbotin persoonallisuus vastaa organisaation viestintätyyliä (Skuridin & Wynn 2024). Lisäksi kannattaa kiinnittää huomiota siihen, että viestintätyyli on kustomoitu vastaamaan organisaation asiakaskunnan tarpeita (Sidaoui ym. 2024).

Älykkäät chatbotit voivat aiheuttaa käyttäjissä myös eristyneisyyden ja turhautuneisuuden tunteita. Ferraro, Demsar, Sands, Restrepo ja Campbell (2024) toteavat, että yhteenkuuluvuuden tunne on oleellinen osa ihmisenä oloa. Älykkäät chatbotit voivat lisätä yhteenkuuluvuutta organisaation ja asiakkaan välillä tarjoamalla esimerkiksi personoidumpia palveluita, mutta ihmisen kohtaamisen

poistaminen yhtälöstä voi myös saada asiakkaat kokemaan olevansa enemmän eristyksissä. Lisäksi kehittyneimmätkään tekoälymallit eivät kykene empatiaan kuten ihminen. Ihminen ymmärtää sävyjä, vivahteita ja kontekseja tavalla, joka auttaa tarjoamaan personoituja ja empaattisia vastauksia. Chatbotin vastaukset voivat olla geneerisiä, robottimaisia ja vähemmän empaattisia. Chatbotti eivät välttämättä ymmärrä luonnollisen kielen vivahteita tai siihen sisältyviä tunteita kuten ihminen. Asiakas saattaa turhautua, etenkin jos kyseessä on monimutkainen, hyvin henkilökohtainen tai emotionaalisesti latautunut ongelma, ja kokea olonsa väärinymmärretyksi. (Ferraro ym. 2024.)

Välttääkseen asiakkaiden turhautumista ja eristyneisyyden tunnetta, organisaation on huomioitava chatbotin puutteet. On tärkeää, että organisaatio ei kokonaan korvaa asiakaspalvelua chatboteilla, vaan käyttää niitä tukemaan asiakaspalvelussa työskentelevien ihmisten työtä. Chatbottia kannattaa hyödyntää esimerkiksi ohjaamalla helpot kysymykset chatbotille ja monimutkaiset, emotionaalisesti latautuneet kysymykset ihmiselle. Lisäksi turhautumisen välttämiseksi chatbotti tulisi ottaa käyttöön vain arkipäiväisissä tai rutiininomaisissa tilanteissa, joissa väärinymmärretyksi tulemisen riski ei johda merkittävään turhautumiseen. (Ferraro ym. 2024.)

Chatbotin kyvykkyyttä vastata empaattisesti ja monimutkaisiin kysymyksiin voidaan lisätä myös esimerkiksi riittävän ison, empaattisuutta ja monipuolisuutta korostavan tietoaineiston käyttämisellä chatbotin kouluttamisessa ja edistyneillä luonnollisen kielen käsittelyn tekniikoilla, kuten tunneanalyysillä. Chatbotti voidaan kouluttaa vastaamaan asiakkaan kysymykseen empaattisemmin tai tunnistamaan sellaiset tilanteet, joissa asiakas on turhautunut, ja siirtämään keskustelu ihmiselle. (Ferraro ym. 2024.) Erityisesti aloilla, joilla empatia ja ymmärtäminen ovat kriittisiä, kuten terveydenhoito ja mielenterveys, tulisi välttää ylitukeutumista chatbotteihin. Organisaation on löydettävä tasapaino automatisoidun ja inhimillisen palvelun välillä säilyttääkseen henkilökohtaisen otteen. (Sidaoui ym. 2024.)

3.4 Toiminnan seuraaminen, ylläpito ja kouluttaminen

Chatbotin toiminnan seuraaminen käyttöönoton jälkeen on olennaista, jotta voidaan varmistua siitä, että chatbotti toimii odotetulla tavalla. Etenkin generatiivisen tekoälyn myötä toiminnan seuraaminen on olennaista, jotta voidaan varmistua chatbotin antamien vastausten puolueettomuudesta ja virheettömyydestä.

Skuridin ja Wynn (2024) suosittelevat luomaan työpöydän (dashboard), jolta chatbotin käyttäytymisen ja onnistumisen mittareita voidaan seurata käyttöönoton jälkeen. Urbani ja muut (2024) korostavat teknologian helppokäyttöisyyttä ja hyödyllisyyttä, mikä vaikuttaa teknologian hyväksymiseen. Helposti seurattavan työpöydän avulla myös ei-teknisesti orientoituneet henkilöt voivat helposti seurata chatbotin toimintaa ja sille asetettujen tavoitteiden toteutumista.

Chatbotin käyttöä ja suorituskykyä tulee seurata jatkuvasti. Chatbotin käymiä keskusteluja täytyy analysoida, ja niistä tulee tunnistaa sellaiset pyynnöt, joita chatbotti ei ole käsitellyt oikein. (Skuridin & Wynn 2024.) Puolueellisuuden tunnistamiseksi ja vähentämiseksi vahvat seuranta- ja valvontajärjestelmät ovat oleellisia. Eettisiä käytäntöjä ylläpidon aikana on tuettava johdonmukaisesti. (Sidaoui ym. 2024.)

Määrällisen tiedon lisäksi on hyvä kerätä myös laadullista tietoa. Palautteen kerääminen on tärkeä osa generatiivisen chatbotin toiminnan seuraamista. Tavat ja työkalut, joilla käyttäjien palautetta kerätään, tulee suunnitella (Skuridin & Wynn 2024). Palautemekanismit ovat olennaisia sekä sisäisissä että ulkoisissa tekoälyä hyödyntävissä chatboteissa (Urbani ym. 2024). Palautetta kannattaa kerätä sekä asiakkailta että työntekijöiltä (Sidaoui ym. 2024), ja sitä tulee analysoida ja hyödyntää uusien toiminnallisuuksien suunnittelussa (Skuridin & Wynn 2024). Tekoälyjärjestelmää tulee pyrkiä jatkuvasti parantamaan, ja palautteiden avulla saadaan selville, täyttyykö käyttäjien tarpeet ja eettiset standardit (Sidaoui ym. 2024).

Organisaation on varattava resursseja chatbotin toiminnan seuraamiseen, ylläpitoon ja kouluttamiseen. Lisäksi chatbotin toiminnallisuutta tulee jatkuvasti pyrkiä parantamaan ottamalla käyttöön uusimpia tekoälyteknologioita ja päivittämällä kielimalleja. (Skuridin & Wynn 2024.)

Palveluntarjoajan valinnassa kannattaa kiinnittää huomiota siihen, että palveluntarjoaja kykenee tuottamaan yksityiskohtaisia raportteja chatbotin toiminnasta ja suorituskyvystä (Skuridin & Wynn 2024). Chatbottiratkaisua tulee voida päivittää vaatimusten muuttuessa. Palveluntarjoajan on oleellista kyetä tukemaan organisaatiota, jos tekoälymalli tuottaa odottamattomia tuloksia. (World Economic Forum 2023, 14.) Organisaation on hyvä varmistua myös siitä, että chatbottia voidaan kehittää organisaation sisällä ja palveluntarjoajan vaihto on tarvittaessa mahdollista (Skuridin & Wynn 2024).

3.5 Yhteensopivuus

Valitun generatiivista tekoälyä hyödyntävän chatbottiratkaisun on oleellista olla yhteensopiva sekä organisaation olemassa olevien tietojärjestelmien että organisaation prosessien, käytäntöjen ja arvojen kanssa. Skuridin ja Wynn (2024) toteavat yhteensopivuuden olevan kriittinen menestystekijä: valitun alustan ja teknologian tulee sopia organisaation tarpeisiin, tavoitteisiin ja toimintatapoihin. Lisäksi on tärkeää huomioida ja investoida järjestelmän integraatio- ja tiedonsiirtotoimenpiteisiin, koska chatbotti tyypillisesti integroidaan organisaation muihin tietojärjestelmiin. Yhteensopivuus organisaation sisäisten tietojärjestelmien kanssa on syytä suunnitella strategisesti. (Skuridin & Wynn 2024.)

Palveluntarjoajan rajapintojen yhteensopivuutta organisaation olemassa oleviin tietojärjestelmiin kannattaa analysoida (Skuridin & Wynn 2024). Valitun tekoälyratkaisun tulee kyetä skaalautumaan, jotta kysynnän lisääntyessä tai vähetessä palvelua voidaan muokata. Tekoälyratkaisu saattaa vaatia kustomointia sopiakseen organisaation vaatimuksiin, ja on hyvä selvittää, kuinka paljon kustomointia ratkaisu vaatii. (World Economic Forum 2023, 12–13.)

Yhteensopivuus on kuitenkin miellettävä kokonaisvaltaisemmin kuin pelkän teknologisen tai tietojärjestelmien yhteensopivuuden kautta. Tekoälypohjaisten chatbottien integroiminen osaksi liiketoimintaa vaatii, että ne sopivat osaksi organisaation olemassa olevia prosesseja ja käytäntöjä. Yhteensopivuuden avulla voidaan varmistaa, että chatbotti tukee nykyisiä työnkulkuja ja sopii yhteen olemassa olevien prosessien kanssa. Yhteensopivuus on olennaisessa roolissa siinä, miten hyvin teknologian käyttöönotto onnistuu ja liiketoimintatavoitteet saavutetaan. (Urbani ym. 2024.) Lisäksi tekoälyn sopivuutta organisaation toimintaan kannattaa arvioida peilaamalla strategista yhteensopivuutta organisaation nykyisiin ja oletettaviin tulevaisuuden tarpeisiin. (World Economic Forum 2023, 13.)

3.6 Vastuullisuus

Vastuulliseen tekoälyjärjestelmään kuuluu eettisten standardien, kuten oikeudenmukaisuuden, avoimuuden, osallisuuden ja vastuullisuuden ylläpitäminen. Lisäksi tulee ottaa huomioon ympäristön kestävyys ja sosiaalinen vastuu. (World Economic Forum 2023, 21.)

Sidaoui ja muut (2024) lähestyvät chatbotteja yritysten digitaalisen vastuun (Corporate Digital Responsibility, CDR) kautta. Yritysten digitaalisen vastuun konseptin avulla voidaan vähentää chatbottiin liittyviä riskejä ja edistää positiivisia vaikutuksia. Yritysten digitaalinen vastuu aihealueena on kuitenkin vielä kehittyvä, ja konseptin tarkka määrittely on haastavaa. Yritysten digitaalinen vastuu voidaan kuitenkin määritellä käytänteinä, menettelytapoina ja hallintorakenteina, jotka liittyvät digitaaliseen muutokseen. Yritysten digitaalisen vastuun keskeisinä käsitteinä ovat vastuulliset digitaaliset käytännöt, kestävä kasvu ja kehitys, sekä luottamuksen lisääminen digitaalisissa ekosysteemeissä. Keskeistä on sen huomioiminen, kuinka digitalisaatio muokkaa yhteiskuntaa ja ympäristöä, ja kuinka se vaikuttaa yksilöihin, yhteisöihin ja valtioihin. (van der Merwe & Al Achkar 2022.) Yritysten digitaalisen vastuun voi kuvata myös periaatteina, jotka määrittävät yrityksen eettisen, reilun, ja suojelevan datan ja teknologian käytön, kun kohteena on asiakas yrityksen digitaalisessa ekosysteemissä (Wirtz, Kunz, Hartley & Tarbit 2022).

Organisaation kulttuuri on olennaisessa roolissa vastuullisen tekoälyratkaisun kehittämisessä. Eettisten tekoäly- ja tiedonhallintakäytäntöjen ja vastuullisuuden tulee olla arvossa. Eettisiä sääntöjä ja periaatteita on aktiivisesti määriteltävä, ja varmistettava, että ne ovat yhdenmukaisia organisaation

tavoitteiden kanssa. Eettisten näkökohtien huomioimiseen on panostettava koko chatbotin elinkaaren ajan. Resursseja tulee allokoida eettisten näkökulmien pohtimiseen, ja johdon tulee tukea kokonaisvaltaista eettistä suunnittelua vastuullisen toteutuksen takaamiseksi. (Sidaoui ym. 2024.)

Palveluntarjoajan valinta on oleellisessa asemassa, kun halutaan ottaa käyttöön vastuullinen tekoälyjärjestelmä. Esimerkiksi puolueellisuus ja epäeettisyys voivat huonontaa tekoälyjärjestelmän laatua, ja palveluntarjoajan on tarjottava keinoja mahdollisten ennakkoluulojen ja eettisten ongelmien poistamiseen tai vähentämiseen. Tekoälyyn liittyvien eettisten riskien vähentämiseksi palveluntarjoajan tulee tarjota ratkaisuja riskien vähentämiseksi organisaation kontekstissa. (World Economic Forum 2023, 21.)

Myös palveluntarjoajan käyttämät koulutusmenetelmät tekoälymallin kouluttamiseksi on oltava oikeudenmukaisuutta, tulkittavuutta, yksityisyyttä ja turvallisuutta lisääviä. Koulutusaineiston laatu on oleellista, ja palveluntarjoajan käyttämän koulutusaineiston alkuperään, edustavuuteen ja kokonaisuuteen kannattaa perehtyä. Palveluntarjoajan luotettavuutta lisää, jos he käyttävät erilaisia työkaluja puolueellisuuden arvioimiseen, monitoroivat myös välillisiä muuttujia, tai julkaisevat raportteja tai teknisiä arvioita mallin puolueettomuudesta. Palveluntarjoajan tulee tiedostaa myös tekoälyyn liittyvät uudet riskit, kuten disinformaatio ja liiallinen luottamus tekoälyjärjestelmiin. (World Economic Forum 2023, 22.)

Palveluntarjoajan suhtautumista eettisyyteen kannattaa tarkastella myös organisaatiotasolla: eettinen digitaalisten teknologioiden käyttö ja eettinen tekoäly tulee olla palveluntarjoajalle prioriteetteja. Myös organisaation yleinen politiikka monimuotoisuuden, tasa-arvon ja osallisuuden edistämisen suhteen kannattaa ottaa huomioon, kun halutaan valita vastuullinen palveluntarjoaja. (World Economic Forum 2023, 23.)

Ympäristön kestävyys on otettava huomioon, kun suunnitellaan vastuullista tekoälyjärjestelmää. Tällä hetkellä tekoäly ei ole ympäristöystävällinen vaihtoehto, ja vastuullinen palveluntarjoaja pyrkii vähentämään teknologian haitallisia ympäristövaikutuksia. (World Economic Forum 2023, 21.) Palveluntarjoajan tulee kyetä vastaamaan esimerkiksi kysymyksiin siitä, kuinka paljon energiaa palveluntarjoaja käyttää tekoälyratkaisujen kouluttamiseen kuukaudessa. Palveluntarjoajan käyttämän laitteiston tulisi olla optimoitu energiankulutuksen vähentämiseen ja käyttää uusiutuvia energialähteitä. On hyvä selvittää, miten palveluntarjoaja suhtautuu aiheuttamiinsa ympäristövaikutuksiin, ja pyrkiikö se vähentämään niitä. (World Economic Forum 2023, 23.)

Lainsäädännön, etenkin EU:n tekoälysäädöksen, noudattaminen edesauttaa eettistä ja vastuullista tekoälyn käyttöönottoa. Organisaation tulee pysyä ajan tasalla viimeisimmistä kehityksistä tekoälyyn liittyvässä lainsäädännössä ja varmistaa, että chatbotti noudattaa tekoälysäädöstä. (Sidaoui

ym. 2024.) Myös palveluntarjoajalta tulee varmistaa, että tekoälymalli on lainsäädännön mukainen ja vastaa tekoälysäädöksen vaatimuksiin. Myös muiden kansainvälisten organisaatioiden luomien tekoälyn hallinnan standardien noudattaminen edistää palveluntarjoajan vastuullisuutta. (World Economic Forum 2023, 27.)

3.7 Läpinäkyvyys

Tekoälyn vastuullisessa käyttöönotossa läpinäkyvyys on oleellisessa roolissa. Tekoälyn avulla asiakkaista voidaan kerätä paljon tietoa, ja tiedon avulla tarjota personoituja ja henkilökohtaisia palveluja. Saavutetun hyödyn vastakohtana on kuitenkin huoli asiakkaiden yksityisyydensuojasta. (Ferraro ym. 2024.) Erityisesti arkaluonteista tietoa käsittelevien chatbottien luotettavuus ja turvallisuus on ensisijaisen tärkeää (Urbani ym. 2024).

Asiakkaiden yksityisyydestä onkin erittäin tärkeää pitää hyvää huolta. Organisaation tulee olla mahdollisimman läpinäkyvä sen suhteen, mitä tietoja asiakkaista kerätään ja miten kerättyä tietoa käytetään. (Ferraro ym. 2024.) Keskittymällä läpinäkyvyyteen tietojen keräämisessä ja varmistamalla, että asiakkaiden huolet tulevat kuulluiksi ja vastatuiksi, voidaan luoda luottamusta tekoälyä hyödyntävää chatbottia kohtaan (Urbani ym. 2024). Tekoälyyn pohjautuvan vuorovaikutuksen luonne tulee kertoa asiakkaille selkeästi (Sidaoui ym. 2024). Asiakkailla tulee olla mahdollisuus kieltäytyä tietojen keräämisestä ja poistaa kerätyt tiedot koska tahansa (Urbani ym. 2024). Tiukat tietoturvastandardit asiakkaiden tietojen suojaamiseksi lisäävät luottamusta. Organisaation on varauduttava ongelmatilanteisiin, joita chatbotin käytöstä voi seurata, ja asetettava vastuumekanismeja niitä varten. (Sidaoui ym. 2024.)

Myös palveluntarjoajan valinnassa on syytä kiinnittää huomiota läpinäkyvyyteen. Palveluntarjoajan on kyettävä selittämään organisaatiolle tekoälyratkaisun toiminnan periaatteet, sen kyvykkyydet ja rajoitteet. Myös koulutusmenetelmien suhteen tulee odottaa läpinäkyvyyttä, ja jatkuvan oppimisen järjestelmien suhteen on hyvä kiinnittää huomiota siihen, tapahtuuko jatkuva oppiminen automaattisesti vai onko prosessissa myös ihminen mukana. (World Economic Forum 2023, 13.)

3.8 Tiedonhallinta

Vankka tiedonhallintastrategia on perustavanlaatuinen edellytys onnistuneelle tekoälyjärjestelmän käyttöönotolle. Organisaation kannattaa pohtia nykyisiä tiedonhallintamenetelmiään ja miettiä, miten uuden tekoälyjärjestelmän käyttöönotto tulee vaikuttamaan niihin. Tiedonhallintastrategia tulee määritellä ja siitä tulee viestiä selkeästi, minkä lisäksi työntekijöitä on koulutettava laadukkaan tiedonhallinnan periaatteista. Tiedonhallinnan eettisyyteen on kiinnitettävä huomiota, ja vastuut siihen liittyen on hyvä määrittää. (World Economic Forum 2023, 19.)

Suuri osa tekoälyjärjestelmien kustannuksista liittyy tiedon valmisteluun tekoälyjärjestelmää varten. Organisaation on valmisteltava käytettävä tietoaineisto, ja valmisteluun liittyvät kustannukset vaihtelevat paljon. On varmistettava, että tietoaineisto on olemassa, olennaista ja hyödyllistä tekoälyratkaisun näkökulmasta. Tiedon laatu on oleellista onnistuneen tekoälyratkaisun toteuttamisessa, ja organisaation on varmistuttava siitä, että tietoaineisto on täsmällistä, kokonaista, johdonmukaista ja ajan tasalla. (World Economic Forum 2023, 17–19.)

Lisäksi on hyvä pohtia, riittääkö sisäinen tietoaineisto vai tarvitaanko myös ulkoista tai synteettistä tietoaineistoa liiketoimintatavoitteiden saavuttamiseksi. Ulkoisia tietoaineistoja käytettäessä on varmistuttava tietoaineiston laadusta ja luotettavuudesta sekä kuka siitä vastaa. Organisaatio voi myös kerätä lisää tietoaineistoa tekoälyjärjestelmän optimaalisen toimivuuden takaamiseksi. (World Economic Forum, 19.)

Organisaation on varmistettava, että tietoja hallitaan ja käsitellään paikallisen lainsäädännön vaatimusten mukaisesti. Palveluntarjoajalla on oltava riittävät tiedon laadun varmistamisen prosessit tiedon säilyttämistä ja käsittelyä varten. On selvitettävä, kuka omistaa ja on vastuussa tietoaineistosta ja siitä johdetuista malleista. Organisaation kannattaa selvittää palveluntarjoajalta, onko asiakkaista kerättyä tietoa mahdollista poistaa tekoälymallista jälkikäteen. Palveluntarjoajalla on oltava menetelmiä tiedon yksityisyyden ja turvallisuuden takaamiseksi sekä proaktiivinen ote kyberhyökkäysten tunnistamiseen ja niiden estämiseen. Palveluntarjoajan tulee pyrkiä minimoimaan hyökkäysten vaikutus ja hallinnoida haavoittuvuuksia. Vastuut tietomurron sattuessa on oltava selvillä. (World Economic Forum 2023, 19; 25–26.)

3.9 Organisaation kulttuuri ja arvot

Organisaation kulttuuri ja arvot vaikuttavat paljon siihen, miten hyvin organisaatio on valmis ottamaan generatiivisen chatbotin käyttöön. Urbani ja muut (2024) esittävät laajennetun teknologian hyväksymismallin, jonka avulla voidaan edesauttaa organisaatiota ottamaan tehokkaasti uusi teknologia käyttöön. Alkuperäinen teknologian hyväksymismalli (technology acceptance model, TAM) sisältää teknologian helppokäyttöisyyden ja hyödyllisyyden sen todelliseen käyttöön vaikuttavina tekijöinä (Davis 1989, 320; 332–334). Modernien teknologioiden myötä nousee esiin sellaisia haasteita, joihin alkuperäinen malli ei ota kantaa. Modernin teknologian, kuten generatiivisen tekoälyn, hyväksymiseen ja todelliseen käyttöön organisaatiossa vaikuttaa helppokäyttöisyyden ja hyödyllisyyden lisäksi myös subjektiiviset normit, yhteensopivuus, suotuisat olosuhteet sekä luottamus. (Urbani ym. 2024.) Onnistuneeseen generatiivisen chatbotin käyttöönottoon vaikuttaa siis myös organisaation kulttuuri, arvot ja toimintatavat.

Organisaation tulee tukea työntekijöitään uuden teknologian käyttöönotossa. Subjektiiviset normit, kuten kollegoiden ja esihenkilöiden mielipiteet ja kannustus, vaikuttavat teknologian hyväksymiseen. Organisaation kannattaa ylläpitää kulttuuria, jossa ollaan avoimia uusille teknologioille, esimerkiksi rohkaisemalla työntekijöitä käyttämään uutta teknologiaa omissa työtehtävissään. Tietoisuuden lisääminen ja työntekijöille koituvien hyötyjen korostaminen voivat auttaa sosiaalisen paineen luomisessa teknologian käyttöönoton edistämiseksi. (Urbani ym. 2024.) Generatiivisen chatbotin toiminnallisuuksia kannattaa mainostaa sekä sisäisesti että ulkoisesti (Skuridin & Wynn 2024). Menestystarinoiden jakaminen organisaation sisällä voi lisätä teknologian hyväksymistä. Uuden teknologian hyväksymisellä on vahva sosiaalinen kiinnekohta, jota organisaatiossa kannattaa vahvistaa. (Urbani ym. 2024.)

Organisaation kannattaa myös panostaa sellaisiin tekijöihin, jotka vaikuttavat työntekijöiden kykyyn hyväksyä teknologia ja ottaa se käyttöön tehokkaasti. Muun muassa tekninen tuki ja käyttäjien kouluttaminen lisäävät suotuisia olosuhteita teknologian käyttöönotolle. Suotuisten olosuhteiden avulla voidaan helpottaa työntekijöiden oppimisprosessia ja ylläpitää positiivista suhtautumista teknologiaan. (Urbani ym. 2024.)

Generatiivisella chatbotilla haetaan tyypillisesti organisaatiossa kustannussäästöjä prosessien ja tehtävien automatisoinnilla. On kuitenkin hyvä huomata, että kustannussäästöjen lisäksi tekoälyn hyödyntäminen voi johtaa korkeaan hintaan yhteiskunnan tasolla. Ferraro ja muut (2024) esittävät tähän liittyen paradoksin ”edullisempi kustannus mutta korkeampi hinta”. Organisaation tekoälyn hyödyntämisen myötä kokonaisia työtehtäviä voi poistua automatisoinnin seurauksena, mikä voi johtaa työttömyyden lisääntymiseen. Työttömyys voi johtaa ekonomiseen epävakauteen ja epätasa-arvoisuuden lisääntymiseen. Organisaation tasolla näihin seuraamuksiin voidaan kuitenkin varautua. Organisaatio voi kouluttaa asiakaspalvelijoita esimerkiksi uusiin rooleihin tai kokonaan uusiin työnkuviin tekoälyyn liittyen. Asiakaspalvelijoita voidaan myös siirtää monimutkaisempiin, ongelmanratkaisua ja luovuutta vaativiin tehtäviin sekä syvällisempään ja merkityksellisempään kanssakäymiseen asiakkaiden kanssa. (Urbani ym. 2024.)

Tekoälyhankintoja tulisikin lähestyä kokonaisvaltaisesti, organisaation eri tiimien ja sidosryhmien yhteistyötä korostaen. Hankintoja tehdessä tulee osallistaa liiketoiminta- ja loppukäyttäjät, IT- ja tietoturva-asiantuntijat, tekoäly- ja data-asiantuntijat sekä hankinta-asiantuntijat. Organisaatiossa on oltava konsensus siitä, että tekoäly on paras mahdollinen ratkaisu käsillä olevaan haasteeseen. (World Economic Forum 2023, 9.) Poikkitoiminnallinen yhteistyö ja vahva hallinto varmistavat, että tekoälyjärjestelmästä ei koidu organisaatiolle suuria operationaalisia, taloudellisia, oikeudellisia tai maineellisia riskejä. Vahvalla tekoälyn hallinnalla organisaatiossa varmistutaan siitä, että tekoälyä käytetään vastuullisesti ja hyödyllisesti. (World Economic Forum 2023, 25.)

Chatbotin onnistunutta käyttöönottoa edesauttaa, jos organisaatiolla on sisäistä asiantuntemusta chatbotteihin ja tekoälyyn liittyen. Skuridin ja Wynn (2024) määrittävät chatbotti-asiantuntijan mukana olon tiimissä kriittiseksi menestystekijäksi. Asiantuntija voi olla joko organisaation sisältä tai sen ulkopuolelta, mutta asiantuntijuuden olemassaolo tiimissä on oleellista. He suosittelevat myös tuotetiimin muodostamista. Tuotetiimissä tulisi olla tuoteomistaja, analyyttinen dialogisuunnittelija, datatieteilijä, UX-suunnittelija ja tekninen johtaja. Lisäksi kokenut projektipäällikkö edesauttaa onnistunutta projektin läpivientiä. Myös ylläpitotiimin muodostaminen on tärkeää, sekä sen vastuiden määrittäminen ja chatbotin pääkäyttäjien kouluttaminen. (Skuridin & Wynn 2024.) Työntekijöiden jatkuva kouluttaminen ja asiantuntijuuden kehittäminen on muutenkin olennaista, kun työskennellään tekoälyn kanssa. Organisaation tulee varmistaa, että henkilökunnalla on riittävät tiedot ja taidot tekoälyn kehityksestä ja eettisistä käytännöistä, jotta he ovat kyvykkäitä työskentelemään chatbottien rinnalla ja ymmärtävät niiden kyvyt ja rajoitteet. (Sidaoui ym. 2024.)

Organisaation kannattaa panostaa huolelliseen palveluntarjoajan valintaan. Palveluntarjoaja vaikuttaa paljon siihen, minkälainen lopputuloksesta tulee. Edellisissä luvuissa käytiin läpi jo monia aihealueita, joiden yhteydessä palveluntarjoajan rooli on merkittävä. Näiden lisäksi on hyvä huomioida, että tekoäly on vielä verrattain uusi teknologia.

Palveluntarjoajalta on, uuden teknologian ollessa kyseessä, hyvä löytyä niin sanottua ajatusjohtajuutta. Palveluntarjoajan visio teknologian tulevaisuuden kehityksestä vaikuttaa siihen, että organisaation chatbottiratkaisua voidaan tulevaisuudessa kehittää hyödyllisesti. Ajatusjohtajat investoivat tutkimukseen ja kehitykseen sekä pysyvät kartalla markkinoiden trendeistä. Myös raportit alan parhaista käytännöistä tekoälyn tulosten optimoimiseksi ja verkostoitumistapahtumat asiakkaiden tietoisuuden lisäämiseksi viittaavat ajatusjohtajuuteen. Henkilökunnan kouluttamiseen investoiminen on myös oleellista, kun puhutaan jatkuvasti kehittyvästä uudesta teknologiasta. (World Economic Forum 2023, 14.)

3.10 Chatbottien kehittämiseen liittyvä lainsäädäntö

Chatbottia kehitettäessä organisaation on oltava selvillä ajantasaisesta lainsäädännöstä. Chatbotti tyypillisesti käsittelee paljon tietoja, joten yleisen tietosuoja-asetuksen vaatimukset on hyvä huomioida jo suunnitteluvaiheessa. Lisäksi EU:n tekoälysäädös vaikuttaa chatbottien kehittämiseen, kun ratkaisussa hyödynnetään tekoälyä. Suurimmassa osassa moderneista chatboteista hyödynnetään tekoälyä jossakin muodossa, joten tekoälysäädöksen vaatimukset tulee ottaa huomioon.

3.10.1 Yleinen tietosuoja-asetus (GDPR)

Yleinen tietosuoja-asetus asettaa organisaatioille vaatimuksia henkilötietojen keräämiseen, säilytykseen ja hallinnointiin liittyen. Vaatimukset koskevat sekä eurooppalaisia organisaatioita, jotka

käsittelevät henkilötietoja EU:n alueella, että EU:n ulkopuolisia organisaatioita, joiden toiminnassa käsitellään EU:n alueella asuvien ihmisten henkilötietoja. Henkilötiedot tarkoittavat kaikkia niitä tietoja, joiden perusteella henkilö voidaan tunnistaa. Tämä koskee sekä suoraa tunnistamista että välillistä tunnistamista, esimerkiksi yhdistämällä tietoja tavalla, joka mahdollistaa tunnistamisen. Henkilötietoja ovat esimerkiksi nimi, kotiosoite, sähköpostiosoite, henkilötunnus, IP-osoite, paikannustiedot ja potilastiedot. Yleinen tietosuoja-asetus rajaa organisaatiolta sellaisten tietojen käsittelyn, jotka liittyvät erityisiin tietoryhmiin. Erityisiä tietoryhmiä ovat muun muassa rotu tai etninen alkuperä, sukupuolinen suuntautuminen, uskonnollinen vakaumus ja poliittiset mielipiteet. (Your Europe 2022.)

Yleinen tietosuoja-asetus määrittelee henkilötietojen käsittelylle kaksi eri profiilia. Rekisterinpitäjä on organisaatio, joka päättää henkilötietojen käsittelytarkoituksesta ja -tavasta. Tietojenkäsittelijä on organisaatio, joka säilyttää ja käsittelee henkilötietoja rekisterinpitäjän puolesta. Tietosuoja-asetuksen mukaisesti rekisterinpitäjä voi käyttää henkilötietojen käsittelyyn ainoastaan sellaista tietojenkäsittelijää, joka noudattaa tietoturvasäännöksiä ja antaa riittävät takeet näiden noudattamisesta. (Your Europe 2022.)

Henkilötietoja voi siirtää myös EU:n ulkopuolelle, jos yleisen tietosuoja-asetuksen tarjoama suoja siirtyy henkilötietojen kanssa. Henkilötietojen suojan voidaan katsoa pysyvän samana, jos asianmukaiset suojatoimenpiteet toteutuvat, EU toteaa maan suojan olevan riittävä tai tietojen siirto perustuu erityisiin poikkeuksiin, esimerkiksi henkilön suostumukseen henkilötietojen siirtämiseen EU:n ulkopuolelle. (Your Europe 2022.)

Lähtökohtaisesti henkilötietojen käsittely on organisaatiolle sallittua, jos henkilö on antanut suostumuksensa. Suostumuksen antamisen yhteydessä tulee varmistua siitä, että henkilö ymmärtää, mihin on suostumassa. Pyyntö tulee esittää selkeällä ja yksinkertaisella kielellä, ja henkilön tulee antaa suostumuksensa vapaaehtoisesti ja tietoisesti. Suostumuksen lisäksi organisaatio voi käsitellä henkilötietoja, jos organisaatio tarvitsee henkilötietoja sopimusvelvoitteen tai laillisen velvoitteen täyttämiseksi, henkilön elintärkeiden etujen suojelemiseksi, yleisen edun mukaisen tehtävän suorittamiseksi tai organisaatio toimii oikeutetun edun puitteissa. (Your Europe 2022.)

Organisaation tulee olla avoin henkilötietojen käsittelystä. Henkilöille on kerrottava, kuka tietoja käsittelee ja miksi niitä käsitellään. Yleisen tietosuoja-asetuksen mukaan joissain tapauksissa tulee kertoa myös organisaation oikeutettu etu, kun sitä pidetään perusteena henkilötietojen käsittelylle, henkilön tietosuojaoikeudet ja automaattiseen päätöksentekoon liittyvät tiedot. (Your Europe 2022.)

Henkilön tietosuojaoikeuksiin liittyy oikeus päästää tietoihin, siirtää tiedot järjestelmästä toiseen, korjata ja poistaa niitä ja vastustaa tietojen käsittelyä. Organisaation on taattava henkilöille pääsy

omiin henkilötietoihin, minkä lisäksi henkilö voi pyytää organisaatiota lähettämään ne toiselle organisaatiolle. Henkilöllä on myös oikeus oikaista tai täydentää virheelliseksi tai puutteellisiksi kokemansa henkilötiedot. Henkilö voi vastustaa henkilötietojen käsittelyä tiettyyn käyttötarkoitukseen, ja ellei organisaatiolla ole oikeutettua etua henkilötietojen käsittelyyn, käsittely tulee lopettaa. Henkilöllä on oikeus myös pyytää rekisterinpitäjää poistamaan kerätyt tiedot, ja yrityksen tulee poistaa tiedot, ellei ole välttämätöntä säilyttää niitä esimerkiksi sananvapauden ja tiedonvälityksen vapauden kunnioittamiseksi, lakisääteisen velvoitteen takia, yleisen edun vuoksi tai oikeusvaateen takia. (Your Europe 2022.)

Yleinen tietosuoja-asetus suojaa henkilöitä myös automaattiselta päätöksenteolta ja profiloinnilta. Jos henkilö on kuitenkin antanut suostumuksensa automaattiseen päätöksentekoon, automaattista käsittelyä voidaan hyödyntää. Tästä on rajattu ulkopuolelle tapaukset, joissa automaattinen päätös pohjautuu lakiin. Tällöinkin organisaation tulee kertoa henkilölle automaattisesta päätöksenteosta, sekä taattava oikeus tarkistaa ja mahdollisesti riitauttaa automaattisesti tehty päätös. (Your Europe 2022.)

Yleinen tietosuoja-asetus sisältää myös sisäänrakennetun ja oletusarvoisen tietosuojan. Tällä tarkoitetaan, että organisaation on tietosuojaperiaatteiden toteuttamiseksi ja henkilön oikeuksien suojaamiseksi toteutettava tarvittavat tekniset ja organisatoriset toimenpiteet käsittelytoimintojen suunnittelun alkuvaiheesta asti. Sisäänrakennetulla tietosuojalla tarkoitetaan esimerkiksi tietojen pseudonymisointia ja salaamista korkean yksityisyydensuojan varmistamiseksi, ja oletusarvoisella tietosuojalla varmistetaan, etteivät henkilötiedot ole oletusarvoisesti rajoittamattoman henkilömäärän saatavilla. (Euroopan komissio s.a.)

3.10.2 Tekoälysäädös

Euroopan Unionin uusi tekoälysäädös pyrkii edistämään turvallisen, ihmisoikeuksia kunnioittavan sekä terveyttä, turvallisuutta ja ympäristöä suojelevan tekoälyn käyttöönottoa EU:n alueella. Tekoälysäädös asettaa sääntöjä tekoälyn myymiselle, käyttämiselle ja valvomiselle sekä kieltää kokonaan tietyt tekoälyn sovellukset. Erityisesti tekoälysäädös asettaa velvoitteita suuririskisille tekoälyjärjestelmille ja niiden kehittäjille. Tekoälysäädös vaatii myös tekoälyjärjestelmien läpinäkyvyyttä. Lisäksi asetuksessa otetaan kantaa yleiskäyttöisten tekoälymallien myymiseen sekä toimenpiteisiin erityisesti pienten yritysten innovaation tukemiseksi. (Euroopan parlamentin ja neuvoston asetus (EU) 2024/1689 tekoälyä koskevista yhdenmukaistetuista säännöistä ja asetusten (EY) N:o 300/2008, (EU) N:o 167/2013, (EU) N:o 168/2013, (EU) 2018/858, (EU) 2018/1139 ja (EU) 2019/2144 sekä direktiivien 2014/90/EU, (EU) 2016/797 ja (EU) 2020/1828 muuttamisesta (tekoälysäädös).)

Tekoälysäädöksen mukaisesti jäsenvaltioiden on nimettävä ilmoittamisesta vastaava viranomainen, joka valvoo vaatimustenmukaisuuden ja valvontaan liittyvien menettelyjen toteuttamisesta. Menettelyt tulee kehittää yhteistyössä muiden jäsenvaltioiden kanssa. Jäsenmaiden tulee määritellä säännöt tekoälysäädöksen rikkomuksille. Säädöksen noudattamatta jättäminen voi johtaa jopa 35 miljoonan euron tai 7 % vuotuisen liikevaihdon suuruiseen sakkoon. Säädös tulee voimaan asteittain 2.2.2025 ja 2.8.2027 välillä. (Tekoälysäädös.)

Tekoälysäädös koskettaa kaikkia toimijoita, jotka tuottavat, käyttävät tai levittävät tekoälyjärjestelmiä EU:n alueella tai tuovat tekoälyjärjestelmiä EU:n alueelle. Tekoälysäädöksestä on rajattu pois tekoälyjärjestelmät, joita käytetään sotilaallisiin, puolustuksellisiin tai kansallista turvallisuutta koskeviin tarkoituksiin sekä järjestelmät, joita ulkomaan julkiset viranomaiset tai kansainväliset organisaatiot käyttävät lain valvomiseen ja oikeudelliseen yhteistyöhön, kunhan ne suojaavat yksilön oikeuksia. Tekoälysäädös ei myöskään koske sellaisia tekoälyjärjestelmiä, joita käytetään tieteelliseen tutkimukseen tai joita ei vielä ole markkinoilla. Tekoälysäädös ei vaikuta olemassa ja voimassa olevaan EU:n lainsäädäntöön liittyen tietosuojaan, yksityisyyteen ja luottamuksellisuuteen. Tekoälysäädöksessä ei myöskään oteta kantaa yksilöiden tekoälyjärjestelmien käyttöön henkilökohtaisissa, ammattiin liittymättömissä toimissa. Myös ilmaiset ja avoimen lähdekoodin lisenssit on rajattu tekoälyasetuksen ulkopuolelle, paitsi jos kyseessä on suuririskinen järjestelmä. (Tekoälysäädös.)

Tekoälysäädöksessä määritellään erilaisia rooleja tekoälyyn liittyen. Tarjoajalla viitataan toimijaan, joka kehittää tekoälyjärjestelmää ja tuo sen markkinoille. Käyttöönottajalla tarkoitetaan toimijaa, joka käyttää tekoälyjärjestelmää. (Tekoälysäädös.)

Tekoälysäädöksessä määritetään, että sekä tarjoajien että käyttöönottajien on varmistettava työntekijöiden tai muiden tekoälyjärjestelmien käyttäjien tekoälylukutaito. Tähän sisältyy heidän teknisten tietojensa, kokemuksensa ja koulutuksensa huomioiminen, ja sen huomioiminen, missä yhteydessä tekoälyjärjestelmää tullaan käyttämään sekä kuka sitä tulee käyttämään. (Tekoälysäädös.)

Organisaatioiden tulee tiedottaa käyttäjilleen, kun he käyttävät tekoälyjärjestelmää, paitsi jos se on itsestään selvää tai tekoälyä käytetään laillisiin tarkoituksiin kuten rikosten havaitsemiseen. Synteettistä sisältöä (esim. syväväärennökset) luovien järjestelmien tulee merkitä niiden tuottamaan sisältöön, että ne ovat keinotekoisesti luotuja. (Tekoälysäädös.)

Tekoälysäädös kieltää täysin tietyt tekoälyn käyttökohteet. Tähän kuuluvat sellaiset järjestelmät, jotka manipuloivat ihmisten päätöksiä tai käyttävät hyväkseen heidän haavoittuvuuksiaan. Myös järjestelmät, jotka arvioivat tai luokittelevat ihmisiä sosiaalisen käyttäytymisen tai persoonallisuuden piirteiden perusteella ja järjestelmät, jotka arvioivat henkilön rikoksen tekemisen riskiä ovat

kiellettyjä. Tekoälysäädöksessä kielletään myös järjestelmät, jotka hyödyntävät kasvojentunnistusta internetistä tai valvontakameroista, tunteiden päättelemistä työpaikoilla tai oppilaitoksissa tai luokittelevat ihmisiä biometrisen tiedon perusteella. Joitakin poikkeuksia tällaisillekin järjestelmille sallitaan, esimerkiksi lainvalvonnallisessa käyttötarkoituksessa kadonneiden henkilöiden etsimiseen ja terroristisiten hyökkäysten estämiseen. (Tekoälysäädös.)

Suuririskiseksi tekoälyjärjestelmäksi luokitellaan sellaiset järjestelmät, joissa tekoälyä hyödynnetään tuotteen turvallisuuskomponenttina tai EU:n lainsäädäntö kattaa tuotteen itsessään, ja joiden tulee läpikäydä kolmannen osapuolen arviointi ennen niiden myyntiä tai käyttöä olemassa olevan lainsäädännön puitteissa. Olemassa oleva lainsäädäntö kattaa muun muassa moottoriajoneuvot, lelujen turvallisuuden, huviveneet ja vesiskootterit, hissit ja hissien turvakomponentit sekä lääkinnälliset laitteet. (Tekoälysäädös.)

Näiden lisäksi suuririskisiksi tekoälyjärjestelmiksi luokitellaan seuraavat käyttökohteet (tekoälysäädös.):

- Biometriset tunnisteet, jos käyttö on sallittua lainsäädännön puitteissa
    - Biometriset etätunnistusjärjestelmät
    - Biometrinen luokittelu arkaluonteisten tai suojattujen ominaisuuksien mukaisesti kyseisten ominaispiirteiden päättelyn perusteella
    - Tunteiden tunnistaminen
- Kriittinen infrastruktuuri (esim. tieliikenteen, vesi-, lämmitys- tai sähköhuollon hallinta ja toiminta)
- Yleissivistävä ja ammatillinen koulutus
    - Luonnollisten henkilöiden pääsy tai hyväksyminen
    - Oppimistulosten arviointi, kun tuloksia käytetään oppimisprosessin ohjaamiseksi
    - Koulutustason arviointiin, johon henkilö voi päästä
    - Opiskelijoiden tarkkailu ja kielletyn käyttäytymisen havaitseminen kokeiden aikana
- Työllistäminen, henkilöstöhallinto ja itsenäisen ammatinharjoittamisen mahdollistaminen
    - Rekrytointi ja valinta, erityisesti kohdennettujen työpaikkojen esittäminen, työhakemusten analysointi ja suodattaminen sekä hakijoiden arviointi
    - Työsuhteen ehtojen, uralla etenemisen ja työhön liittyvien sopimussuhteiden päättämistä koskevien päätösten tekeminen, tehtävien jakaminen käytöksen, persoonallisuuspiirteiden tai ominaisuuksien perusteella, ja henkilöiden suorituksen ja käyttäytymisen seuraaminen ja arviointi
- Välttämättömien yksityisten palvelujen ja välttämättömien julkisten palvelujen ja etuuksien saatavuus ja käyttö

  - Välttämättömien julkisen avun etuuksien ja palveluiden oikeutuksen arviointi, etujen ja palvelujen myöntäminen, vähentäminen, peruuttaminen ja takaisin periminen
  - Luottokelpoisuuden arviointi ja luottopisteytyksen määrittäminen, pois lukien talouspetosten havaitseminen
  - Riskinarviointi ja hinnoittelu sairaus- ja henkivakuutusten tapauksessa
  - Hätäpuheluiden arviointi ja luokittelu pelastus- ja ensihoitopalvelujen lähettämiseksi ja tärkeysjärjestyksen määrittämiseksi, kiireellistä sairaanhoitoa tarvitsevien potilaiden luokittelu
- Lainvalvonta, jos käyttö on sallittua lainsäädännön puitteissa
  - Henkilön rikoksen uhriksi joutumisen riski
  - Valheenpaljastimet ja vastaavat välineet
  - Todistusaineiston luotettavuuden arviointi rikostutkinnassa tai rikosoikeudellisessa menettelyssä
  - Rikokseen tai rikoksen uusimiseen syyllistymisen riskin arviointi, henkilöiden tai ryhmien persoonallisuuspiirteiden ja -ominaisuuksien tai aiemman rikollisen käyttäytymisen arviointi
  - Profilointi rikosten paljastamisen, tutkimisen tai rikoksiin liittyvien syytetoimien yhteydessä
- Muuttoliikkeen hallinta, turvapaikka-asiat ja rajavalvonta, jos käyttö on sallittua lainsäädännön puitteissa
  - Valheenpaljastimet ja vastaavat välineet
  - Jäsenvaltion alueelle tulevan henkilön aiheuttaman riskin arviointi, mukaan lukien turvallisuusriski, sääntöjenvastaisen maahanmuuton riski ja terveysriski
  - Turvapaikka-, viisumi- ja oleskelulupahakemusten ja niiden liittyvien valitusten käsittelyssä kelpoisuuden tarkastaminen, myös todistusaineiston luotettavuuden arviointi
  - Muuttoliikkeen hallinnan, turvapaikka-asioiden tai rajavalvonnan yhteydessä henkilön havaitseminen, tunnistaminen tai henkilöllisyyden määrittäminen, pois lukien matkustusasiakirjojen tarkastaminen
- Oikeudenhoito ja demokraattiset prosessit
  - Tosiseikkojen tai lainsäädännön tutkiminen ja tulkinta sekä lainsäädännön soveltaminen konkreettisiin tosiseikkoihin tai vastaava käyttö vaihtoehtoisessa riidanratkaisussa
  - Vaalin tai kansanäänestyksen tulokseen tai äänestyskäyttäytymiseen vaikuttaminen, pois lukien esimerkiksi välineet, joita käytetään poliittisten kampanjoiden järjestämiseen, optimointiin tai jäsentämiseen

Näitä tekoälyjärjestelmiä ei kuitenkaan pidetä suuririskisinä, jos niistä ei aiheudu merkittävää riskiä ihmisten terveydelle, turvallisuudelle tai perusoikeuksille. Suuririskisyyttä vähentää myös se, jos

järjestelmä suorittaa suppeaa menettelyllistä tehtävää, järjestelmä parantaa ihmisen aiemmin suorittaman tehtävän tulosta, järjestelmä havaitsee päätöksentekotapoja tai -poikkeamia eikä sen tarkoituksena ole korvata ihmisen tekemää arviota tai vaikuttaa siihen, tai järjestelmä on suunniteltu suorittamaan valmistelutehtävä, joka koskee mainittujen käyttötapausten kannalta merkityksellistä arviointia. Suuririskisiä järjestelmiä on kuitenkin kaikki tekoälyjärjestelmät, joissa suoritetaan luonnollisten henkilöiden profilointia. EU:n komissio tarjoaa ohjeita ja esimerkkejä suuririskisistä tekoälyjärjestelmistä, ja saattaa lisätä tai poistaa ehtoja suuririskisyyden luokittelulle, jos siihen havaitaan tarvetta. (Tekoälysäädös.)

Suuririskisten tekoälyjärjestelmien on vastattava tiettyjä standardeja, ottaen huomioon niiden käyttötarkoituksen ja tekoälyteknologian nykyisen tilan. Jos tuote sisältää tekoälyjärjestelmän, tarjoajien tulee varmistaa, että kaikki asiaankuuluvat EU:n lainsäädännöt täyttyvät. Riskienhallinnan tulee olla jatkuva prosessi koko tekoälyn elinkaaren ajan. Mahdollisia riskejä terveydelle, turvallisuudelle ja perusoikeuksille tulee analysoida ja arvioida. Suuririskiset tekoälyjärjestelmät tulee kehittää käyttäen korkealaatuista tietoaineistoa järjestelmän kouluttamiseen, validoimiseen ja testaamiseen. Tietoaineiston laadussa tulee huomioida tiedonkeruumenetelmät, tiedon valmistelu, mahdolliset ennakkoluulot ja puutteet tietoaineistossa. (Tekoälysäädös.)

Tekoälysäädös asettaa myös vaatimuksia suuririskisten tekoälyjärjestelmien tarjoajille muun muassa teknisen dokumentaation olemassaolosta, automaattisesta järjestelmän toiminnan seuraamisesta, avoimuudesta ja riittävien tietojen antamisesta käyttöönottajille, luonnollisten henkilöiden valvonnasta sekä järjestelmien tarkkuudesta, toimintavarmuudesta ja turvallisuudesta. Yritysten, jotka tarjoavat suuririskisiä tekoälyjärjestelmiä, tulee esittää yhteystietonsa tuotteen tai sen pakkauksen yhteydessä. Tuotteeseen tulee myös laittaa merkintä siitä, että se vastaa EU:n standardeja. Järjestelmän tulee myös olla saavutettava EU:n direktiivien mukaisesti. (Tekoälysäädös.)

Myös suuririskisten tekoälyjärjestelmien käyttöönottajilla on velvollisuuksia. Järjestelmää tulee käyttää ohjeiden mukaisesti, varmistaa valvonta ihmisen toimesta, varmistaa tietoaineiston tarkoituksenmukaisuus ja seurata järjestelmän toimintaa. Riskejä havaittaessa järjestelmän tarjoajaa sekä asiaankuuluvia viranomaisia on tiedotettava välittömästi. Käyttöönottajien on pidettävä tekoälyjärjestelmän luomia lokeja vähintään kuusi kuukautta. Työntekijöitä on tiedotettava suuririskisen tekoälyjärjestelmän käyttöönotosta. Jos järjestelmää ei ole rekisteröity EU:n tietokantaan, sitä ei tulisi käyttää. Käyttöönottajien tulee myös varmistaa tietosuojan toteutuminen ja tehdä yhteistyötä asiaankuuluvien viranomaisten kanssa. (Tekoälysäädös.)

Yleiskäyttöisiin tekoälymalleihin katsotaan kuuluvan systeeminen riski, koska niillä on merkittävä suorituskyky. Yritysten, jotka luovat yleiskäyttöisiä tekoälymalleja, tulee pitää kirjaa tekoälyn kehityksestä ja testaamisesta. EU:n ulkopuolisten yritysten, jotka myyvät yleiskäyttöisiä tekoälymalleja

EU:ssa, tulee määrätä edustaja EU:n sisältä. Edustajan vastuulla on varmistaa, että tekoälymalli vastaa EU:n lainsäädäntöön, säilyttää kopiota teknisestä dokumentaatiosta kymmenen vuoden ajan, ja tarjota tarvittavat tiedot viranomaisille pyydettäessä. Suuririskisten tekoälyjärjestelmien tarjoajien tulee seurata tekoälyjärjestelmien toimintaa koko niiden elinkaaren ajan, keräten ja analysoiden tietoja ja varmistaen jatkuvan lainmukaisuuden. (Tekoälysäädös.)

4 Chatbottien riskit ja niiden hallinta

Kuten kaikkeen teknologiaan, myös chatbotteihin liittyy riskejä, jotka on hyvä tiedostaa, kun teknologia otetaan käyttöön. Generatiivisen tekoälyn myötä on syntynyt myös uusia uhkia, jotka organisaation on hyvä tiedostaa voidakseen tehdä tietoisen päätöksen generatiivisen tekoälyn hyötyjen ja riskien tasapainottamisesta. Etenkin generatiiviseen tekoälyyn liittyen uhkat kuitenkin kehittyvät jatkuvasti, koska teknologia itsessään kehittyy tällä hetkellä vauhdilla. Näin uuden teknologian kohdalla on hyvä tiedostaa, että kaikki riskit eivät välttämättä ole vielä tiedossa ja uusia riskejä voi ilmetä tulevaisuudessa.

Seuraavissa luvuissa käsitellään chatbotteihin ja generatiiviseen tekoälyyn liittyviä riskejä monesta eri näkökulmasta. Tietoturvaan ja henkilösuojaan liittyvien riskien lisäksi käsitellään isojen kielimallien erityisiä ominaisuuksia, jotka muodostavat uusia haasteita. Lisäksi generatiivisen tekoälyn yhteiskunnallisia vaikutuksia avataan niin yksilön, organisaation kuin yhteiskunnankin näkökulmasta.

4.1 Tietoturva

Chatbottien riskinä on tietoturvahyökkäysten, tietovuotojen ja muiden kyberturvallisuuteen liittyvien ongelmien realisoituminen. Kyberuhat muuttuvat ja kehittyvät jatkuvasti, mikä luo haasteita chatbotin kehittämiseen ja ylläpitoon. (Yang, Chen, Por & Ku 2023.) Generatiivisen tekoälyn kehittyminen on hyödyttänyt niin kyberturvallisuuden puolustajia kuin sitä vastaan hyökkääviäkin. Tekoäly ja koneoppiminen on lisännyt kyberhyökkäysten tehokkuutta ja tehnyt niistä voimakkaampia. Hyökkääjät käyttävät generatiivista tekoälyä vakuuttavaan sosiaaliseen manipulointiin ja tietojen kalasteluun sekä erilaisten hyökkäyskuormien, haitta- ja lunnasohjelmien luomiseen. (Gupta, Akiri, Aryal, Parker & Praharaj 2023.) Lisäksi yksityisyyteen ja tietojen vuotamiseen liittyvät ongelmat ovat yleensä yhteydessä tahattomaan arkaluontoisen tiedon jakamiseen, tietovuotoihin mallin vastausten kautta, mallin kopioimiseen ja tietojen myrkyttämiseen. Kyberhyökkäysten tarkoituksena on paljastaa tekoälymallien haavoittuvuuksia ja manipuloida mallin toimintaa syötettä muokkaamalla. (Sebastian 2023, 3–6.)

Monet isojen kielimallien kehittäjät ovat asettaneet sääntöjä sille, minkälaista sisältöä he eivät halua mallien tuottavan ja näkevät paljon vaivaa sen eteen, että mallit noudattavat näitä sääntöjä. Kaupalliset kielimallien kehittäjät eivät esimerkiksi halua, että malli tuottaa vihapuhetta tai syrjintää, koska se antaisi kuluttajille huonon vaikutelman. Isojen kielimallien turvallisuuden parantaminen liittyy pitkälti mallin tuottaman sisällön hallintaan, ja tämä vaatii teknisiä väliintuloja. Vaikka suurin osa organisaatioista asettaa sääntöjä isojen kielimallien käytölle, käyttäjät saattavat yrittää tuottaa sisältöä, joka ei ole sallittua, tai vahingossa aikaansaada sellaista. (Engler & Dhamani 2024, luku

3.) Asetettuja rajoitteita voidaan kuitenkin kiertää erilaisin tekniikoin, kuten jailbreakingin avulla tai käänteisellä psykologialla. (Gupta ym. 2023.)

Kyberuhkiin varautumisessa suuressa roolissa on kehittäjien tietoisuus uusimmista tietoturvauhkista ja tarvittavista turvallisuustoimenpiteistä chatbotin suojaamiseksi. Haasteeksi muodostuu kuitenkin myös chatbottien käyttäjäystävällisyys, ja tarpeellisten tietoturvatoimenpiteiden vaikutusta käyttäjäkokemukseen on harkittava. Chatbottien tulee olla käyttäjäystävällisiä ja helppokäyttöisiä, joten tietoturvatoimenpiteiden vaikutusta näihin asioihin tulee arvioida, ja päättää sopiva tasapaino molempien näkökohtien välillä. (Yang ym. 2023.) Riskin vähentämiseen kuuluu tiedon turvaaminen vahvojen turvallisuusmenetelmien avulla, kuten salauksella ja pääsynhallinnalla. Hyökkäyksiin vastaamisessa generatiivisen tekoälyn koulutukseen käytettävän tietoaineiston laadukkuus on olennaista. Tietojen kunnollisella turvaamisella ja yksityisyyden varmistamisella voidaan vähentää hyökkäysten riskejä. (Sebastian 2023, 3.)

4.1.1 Jailbreaking, kehoteinjektiot ja tietojen myrkyttäminen

Tiettyjä kehotesyötteitä käyttämällä voidaan kiertää isojen kielimallien kehittäjien laatimia rajoitteita. Tätä toimintaa kutsutaan nimellä jailbreaking. Rajoitteiden kiertäminen mahdollistaa haitallisen sisällön luomisen, disinformaation levittämisen ja muita haitallisia käyttökohteita. (Gupta ym. 2023.) Yritykset yrittävät tunnistaa ja estää tällaisen toiminnan, ja vahvistaa turvatoimia tämän tyyppisen toiminnan estämiseksi. Yrityksillä on kuitenkin myös teknisiä haasteita siinä, miten löytää tasapaino sen välillä, että chatbotti ei voi vastata mihinkään tai sitä voidaan käyttää väärin. (Engler & Dhamani 2024, luku 5.)

Kehoteinjektioilla (prompt injection) pyritään kehotteen muokkaamisen avulla saamaan kielimalli toimimaan vastoin rajoitteitaan tai paljastamaan arkaluontoista tietoa. (Gupta ym. 2023.) Kehoteinjektioilla voidaan yrittää syöttää chatbotille haitallista tietoa tai ohjeita, joilla pyritään vaikuttamaan mallin toimintaan. Myös epäsuorat kehoteinjektiohyökkäykset ovat mahdollisia. Epäsuorissa kehoteinjektioissa hyödynnetään kolmannen osapuolen tietolähdettä, kuten websivua, josta haitalliset ohjeet haetaan. Nämä kehoteinjektiot syötetään tietoaineistoon, jota todennäköisesti tullaan käyttämään ja täten epäsuorasti vaikutetaan mallin toimintaan. (Engler & Dhamani 2024, luku 5.) Riskeihin sisältyy misinformaation ja disinformaation leviäminen, puolueellisen ja ennakkoluuloisen sisällön luominen, yksityisyydensuojan rikkominen sekä muiden, kielimalliin liittyvien, sovellusten hyväksikäyttö. (Gupta ym. 2023.)

Tietojen myrkyttäminen tapahtuu niin, että hyökkääjä syöttää vahingollista tietoa mallin koulutusaineistoon tarkoituksenaan vaikuttaa mallin tuleviin ennusteisiin tai käyttäytymisiin. Riski on suuri etenkin järjestelmissä, jotka jatkuvasti oppivat vuorovaikutuksesta. (Sebastian 2023, 4–6.)

Koulutusaineisto vaarantuu vahingollisesta tietoaineistosta. Tietojen myrkyttämistä voidaan hyödyntää esimerkiksi älykkäämpien haittaohjelmien rakentamiseksi tai tietojen kalastelun suodattimien vahingoittamiseksi. Jo pieni määrä haitallista tietoa tietoaineistossa voi vaikuttaa mallin toimintaan. (Engler & Dhamani 2024, luku 5.)

4.1.2 Keinoja tietoturvan vahvistamiseen

Chatbottien kehittämiselle ei ole olemassa standardoituja turvallisuustoimenpiteitä. Kehittäjien tulee lähtökohtaisesti itse päättää, mitä ja koska turvallisuustoimia tarvitaan, ja miten niiden toimivuudesta varmistutaan. Esimerkiksi päästä päähän (end-to-end) -salaus on lupaava tapa suojata käyttäjien tietoja, mutta toteutus voi kustannusten vuoksi olla pienille ja keskisuurille organisaatioille haastavaa. Päästä päähän -salauksen avulla tietojen siirtäminen käyttäjän ja chatbotin välillä on turvallista, eikä niihin pääse käsiksi ulkopuolelta. (Yang ym. 2023.) Päästä päähän -salauksen avulla vain kommunikoivat osapuolet voivat lukea viestit. Keskustelu on salattu ja vain viestin vastaanottaja voi purkaa salauksen. Jos salausta ei käytetä, kolmansien osapuolien on mahdollista päästä tietoihin käsiksi. Jos kyseessä on chatbotti, joka ei käytä henkilötietoja tai muita käyttäjään liittyviä tietoja toiminnassaan, myös HTTPS-protokolla voi riittää. Päästä päähän -salauksella kuitenkin varmistutaan turvallisesta tiedonsiirrosta osapuolten välillä. Koska chatbotteja integroidaan enenevissä määrin myös erilaisiin viestintäsovelluksiin, kuten Facebook Messenger, WhatsApp ja Telegram, tulee organisaation varmistua myös siitä, että nämä alustat tukevat päästä päähän -salausta. (Hasal ym. 2021.)

Myös lohkoketjuteknologiat ovat lupaavia chatbottien kohdalla, koska tietoa voidaan tallentaa ja jakaa turvallisesti, hajautetusti ja peukaloinnilta välttyen. Pääsynhallinta ja todennus ovat oleellisia osia chatbotin turvallisuutta. Chatbotin tulee vaatia käyttäjää todentamaan henkilöllisyytensä ennen pääsyä arkaluontoiseen tietoon. Pääsynhallinnalla taataan se, että käyttäjällä on pääsy vain niihin tietoihin ja palveluihin, joihin hänellä on oikeus. (Yang ym. 2023.) Todentaminen ja pääsynhallinta eivät välttämättä ole tarpeellisia, jos järjestelmän ei ole oleellista tietää henkilön identiteettiä tai tarvitse pääsyä henkilön tietoihin. Kuitenkin aina, jos chatbotti hyödyntää käyttäjän tietoja, todennus ja pääsynhallinta ovat välttämättömiä henkilön tunnistamiseksi. Kaksi- tai monivaiheinen tunnistautuminen lisää turvallisuutta. (Hasal ym. 2021.)

Itsestään tuhoutuvat viestit voivat myös olla varteenotettava ja käytännöllinen vaihtoehto, kun siirretään arkaluontoista tietoa. Tällöin viestit tuhotaan automaattisesti tietyn ajan jälkeen. Etenkin arkaluontoista tietoa käsittelevillä aloilla, kuten terveydenhoitoon liittyen, itsestään tuhoutuvat viestit lisäävät turvallisuutta. On myös GDPR:n vaatimus, että tietoja ei säilytetä kauemmin kuin niiden käsittelyn kannalta on tarpeellista. Ongelmaksi saattaa kuitenkin muodostua se, että chatbottien kanssa käytyjä keskusteluja käytetään chatbottien kouluttamiseen, jolloin tietoja myös

tallennetaan. Keskusteluhistoriaa käytetään hyödyksi chatbotin laadun parantamisessa. Koneoppimisen menetelmät tarvitsevat tietoaineistoa, joilla niitä voidaan kouluttaa. Lähtökohtaisesti suurempi koulutusaineisto korreloi chatbotin keskustelutaitojen ja -laadun kanssa, minkä takia monet organisaatiot tallentavat keskusteluhistorian. Organisaatioiden tulee kuitenkin pseudonymisoida ja salata henkilötiedot tästä aineistosta. Keskusteluhistorian voi tallentaa, mutta siitä ei tule käydä ilmi yhteyttä tiettyyn käyttäjään tai käyttäjän tietoihin. Henkilötietojen tunnistamiseen keskusteluhistoriasta on olemassa erilaisia menetelmiä, kuten hahmonsovitus ja entiteettien tunnistaminen. (Hasal ym. 2021.)

Tietojen anonymisoinnilla ja aggregoinnilla voidaan vaikeuttaa pääsyä henkilötietoihin. Anonymisointi tarkoittaa, että henkilön tunnistamiseen liittyvät tiedot korvataan yhdellä tai useammalla keinotekoisella tunnisteella tai pseudonyymillä. Aggregointi tarkoittaa tietojen yhdistelemistä tavalla, jonka lopullinen muoto ei sisällä henkilötietoja. Erotetun yksityisyyden tekniikat tarjoavat matemaattisen määritelmän yksityisyydestä, ja mahdollistavat tietojen jakamisen tietoaineistosta ryhmien rakenteen perusteella ilman, että paljastetaan tietoja yksilöstä. (Sebastian 2023, 6–10.)

Anonymisointi- ja salaustekniikoilla on tärkeä rooli tietojen suojaamisessa ja yksityisyyden varmistamisessa. Tietoaineisto, jota chatbotin koulutuksessa käytetään, tulee olla suojassa luvattomalta pääsyltä ja väärinkäytöltä. Anonymisoinnin avulla on vaikeaa tai mahdotonta yhdistää saatu tieto takaisin siihen yksilöön, jota tieto koskee. Anonymisointia käytetään usein käyttäjän tietojen suojaamiseen tekoälymallin koulutuksen aikana. Anonymisointitekniikoihin kuuluu tietojen peittäminen, pseudonymisointi, yleistäminen ja erotettu yksityisyys. (Sebastian 2023, 6–10.)

Isojen kielimallien vahvuutta voidaan lisätä esimerkiksi kouluttamalla mallia esimerkeillä mahdollisista hyökkäyksistä. Kestävyyttä testaamalla voidaan varmistaa, että malli pystyy käsittelemään epätavalliset tai odottamattomat syötteet epäonnistumatta. Myös kyselyiden määrän rajaaminen ja automaattisten kyselyiden estäminen voi auttaa suojaamaan järjestelmää väärinkäytöiltä ja automaattisilta hyökkäyksiltä. (Sebastian 2023, 6–10.)

Myös tekniset hallintamekanismit ja monitorointityökalut ovat oleellisia turvallisuusuhkien havaitsemiseksi ja niihin reagoimiseksi. Chatbottien ja niiden lokien seuraaminen ja analysoiminen on tärkeää epätavallisen toiminnan ja tietomurtojen havaitsemiseksi. Kokonaisvaltaisia turvallisuusarvioita on hyvä toteuttaa koko chatbotin elinkaaren ajan. Chatbottia tulee päivittää ja haavoittuvaisuuksia paikata säännöllisesti, jotta voidaan varmistua chatbotin turvallisuudesta. (Yang ym. 2023.)

Chatbottien kehitystiimien tietoturvaosaaminen ja asiantuntemus on myös usein vaihtelevaa, ja on hyvä varmistua siitä, että kaikilla tiimin jäsenillä on tarvittava osaaminen ja koulutus tietoturvan suhteen. (Yang ym. 2023.) Ihmisen käyttäytyminen on kriittistä tietoturvan kannalta. Koulutuksessa

tulee panostaa siihen, miten internetissä pysytään turvassa ja jakaa tietoa edistyksellisistä kyberhyökkäysten muodoista. Tietoturvakoulutuksen on siirryttävä säännöistä ja käytännöistä siihen, miten voidaan hyödyntää tilanteeseen liittyvää tietoa generatiiviseen tekoälyyn liittyvien monien uhkien huomaamiseen. Yritysten tulee kouluttaa työntekijöitään kyberuhkien moninaisuudesta, ja siitä, miten generatiivista tekoälyä voidaan hyödyntää niissä. Yksilöiden kouluttaminen auttaa tiedostamaan uhkia ja pysymään turvassa. (Engler & Dhamani 2024, luku 5.) Myös käyttäjiä tulee kouluttaa riskeistä ja kannustaa tekoälyn vastuulliseen käyttöön sekä kehottaa ilmoittamaan epäilyttävästä toiminnasta ja sisällöstä. (Gupta ym. 2023.)

Generatiivisen tekoälyn tietoturvauhkiin ei ole olemassa nopeita tai valmiita ratkaisuja. Tekoälyä voidaan hyödyntää myös tietoturvan parantamiseen ja uhkien havaitsemiseen. (Engler & Dhamani 2024, luku 5.)

4.2 Hallusinointi

Isot kielimallit voivat hallusinoida eli tuottaa sellaista tietoa, joka ei ole totta tai todellista. Hallusinointiin voi olla monia eri syitä, kuten koulutusaineiston yleistäminen, puutteellinen totuuden määrittely, ennakkoluulot ja virheet koulutusaineistossa, ylisovittaminen ja ulkoa oppiminen. Myös kehotteiden muodostaminen vaikuttaa merkittävästi saatuun tulokseen, ja epäselvät tai johdattelevat syötteet lisäävät hallusinoinnin riskiä. Isoissa kielimalleissa ei myöskään ole ulkoisia faktantarkistusmekanismeja, joten kaikki vastaukset pohjautuvat niiden aiemmin oppimaan tietoon, jolloin esimerkiksi reaaliaikaisia tietoja ei ole mahdollista hyödyntää. (Amaratunga 2023, luku 6.) Toisinaan tekoäly voi luoda vastauksia, jotka vaikuttavat viittaavan tiettyyn tietoon tai paljastavan arkaluonteista tietoa. Yleensä kyseessä on kuitenkin hallusinaatio, jossa malli tarjoaa arkaluonteisilta vaikuttavia tietoja vastauksissaan. Vastaukset on kuitenkin luotu koulutuksen aikana opittujen rakenteiden perusteella, eivätkä kuvasta pääsyä mihinkään tietolähteeseen tai luottamuksellisiin tietokantoihin. (Sebastian 2023, 4–6.)

4.2.1 Mallin tuottaman sisällön kontrolloiminen

On muutamia strategioita, joilla voidaan yrittää estää isoa kielimallia tuottamasta sisältöä, jota sen ei pitäisi tuottaa. Yksi yksinkertainen vaihtoehto on toteuttaa jonkinlainen mekanismi tunnistamaan, kun malli tuottaa toksista sisältöä, ja antaa sen sijaan geneerinen vastaus, kuten ”on ohjeideni vastaista tuottaa tällaista sisältöä”. Tässä on kuitenkin riskinä se, että chatbotti antaa liian helposti geneerisen vastauksen, mikä voi turhauttaa käyttäjiä, jos syöte ei ollut millään tavalla haitallinen. (Engler & Dhamani 2024, luku 3.)

Yksi vaihtoehto on käydä koulutusaineistoa läpi, ja poistaa sieltä sellainen aineisto, jota mallin ei haluta oppivan. Malli voidaan myös kouluttaa havaitsemaan ja luokittelemaan turvalliset ja ei-

turvalliset aineistot, ja kouluttaa se vastaamaan kuten turvallisiksi luokitelluissa vaihtoehdoissa. Näin voidaan vähentää mallin luoman haitallisen sisällön määrää. Jos koulutusaineistosta poistetaan aineisto, jota mallin ei haluta oppivan, muodostuu kuitenkin riski siitä, että malli ei enää tunnista haitallista sisältöä yhtä hyvin. (Engler & Dhamani 2024, luku 3.)

Vahvistusoppiminen ihmisen palautteen kautta (Reinforcement learning from human feedback, RLHF) on menetelmä, jolla mallia voisi kouluttaa oppimaan pois epätoivotusta käyttäytymisestä. Ihmiset arvioisivat mallin vastauksia merkitsemällä ne hyväksyttäviksi tai ongelmallisiksi tai tarkentamalla toivottua vastausta. Tämä tekniikka on todettu paremmin skaalautuvaksi ja adaptoituvaksi menetelmäksi kuin aiemmat. RLHF:llä on kuitenkin sekä rahallinen että emotionaalinen hinta. Työ on usein ulkoistettu alihankkijoille, jotka altistuvat jatkuvasti mahdolliselle traumaattiselle sisällölle. Tänä päivänä monet isot kielimallit nojaavat vahvistusoppimiseen ihmisen palautteen kautta. (Engler & Dhamani 2024, luku 3.)

Vahvistusoppiminen tekoälyn palautteen kautta on uusi tekniikka isojen kielimallien turvallisuuden parantamiseksi, jossa oleellista on ihmisen poistaminen kuviosta. Yksi malli koulutetaan valittujen periaatteiden mukaisesti ja malli arvioi muiden mallien vastauksia. Ensimmäinen malli tunnistaa vastaukset, jotka ovat periaatteiden vastaisia ja ohjaa sitten toista mallia palautteen avulla. Tällainen tekniikka mahdollistaa sen, että ihmisten ei tarvitsisi suorittaa emotionaalisesti raskasta sisällön läpikäymistä. (Engler & Dhamani 2024, luku 3.)

Isojen kielimallien hallusinointiin voidaan vaikuttaa muutamin eri tavoin. Mallit voidaan kouluttaa siteeraamaan lähteitään, mikä helpottaisi ihmisten arviointia vastauksen oikeellisuudesta. Isoihin kielimalleihin voidaan sisällyttää myös haku, jonka avulla malli voi hakea tietoa määritellyistä lähteistä esiopitun tiedon sijaan. Jos kielimallilta kysyy kysymyksen, johon se ei tiedä vastausta, hallusinaation sijaan se voi hakea tietoa sopivalla kyselyllä ja tiivistää löydetyt tiedot vastaukseen. Tämä vaatii mallilta sitä, että se on tietoinen siitä, koska se ei tiedä vastausta. (Engler & Dhamani 2024, luku 3.)

Haun avulla voidaan myös korvata sitä puutetta, että mallit on koulutettu menneisyyden tiedoilla: malli opetetaan tunnistamaan tilanteet, joissa tieto, jota siltä pyydetään, on uudempaa, kuin tieto, jolla sitä on koulutettu, ja se tekee haun uudempaan tietoon. Käyttäjät voivat myös testata hallusinaation todennäköisyyttä esimerkiksi kysymällä saman kysymyksen useampaan kertaan: jos kyseessä on hallusinaatio, vastaus on todennäköisesti joka kerralla eri. Tämä tekniikka ei kuitenkaan välttämättä toimi esimerkiksi ChatGPT:n kaltaisessa sovelluksessa, jossa malli ottaa huomioon aiemman keskusteluhistorian, minkä perusteella se voi toistaa virheensä. (Engler & Dhamani 2024, luku 3.)

Hallusinoinnin mahdollisuutta voidaan pienentää myös kehotteiden muodostamisella, esimerkiksi lisäämällä kehotteeseen, että on sallittua olla epävarma ja vastata ”en tiedä” sen sijaan, että ei puhuisi totta. Kehotemuotoilulla voidaan vaikuttaa vahvasti mallin luomaan sisältöön. Monilla saatavilla olevilla kielimalleilla on myös lämpötilaparametri, joka kontrolloi sitä, kuinka paljon mallin tulisi noudattaa koulutusdataa tai luoda ”luovempia” vastauksia. Tietoon perustuvien kysymyksien kohdalla lämpötilan tulisi olla nollassa, jolloin malli palauttaa aina suurimman todennäköisyyden vastauksen. Jos lämpötila on korkea, sitä voidaan paremmin hyödyntää luovempiin tehtäviin. Parametriä kontrolloimalla voidaan vaikuttaa hallusinoinnin määrään, vaikka sitä edelleen tapahtuu myös silloin, kun lämpötilaparametri on asetettu nollaan. Tekniikoita kehitetään jatkuvasti, kun generatiivisten mallien käyttö lisääntyy. Monet nykyiset strategiat keskittyvät haittojen vähentämiseen, mutta eivät ratkaise ongelmia kokonaan. (Engler & Dhamani 2024, luku 5.)

Hallusinoinnin välttämiseksi voidaan myös hyödyntää tekniikoita kuten hienosäätäminen tai ulkoisten tarkistusjärjestelmien rakentaminen tiedon oikeellisuuden varmistamiseksi. Myös ihmisen valvonta on tärkeää hallusinoinnin huomaamiseksi ja sen vähentämiseksi. (Amaratunga 2023, Luku 6.)

4.3 Yksityisyys ja tietosuoja

Isojen kielimallien riippuvuus suurista tietomääristä, joka saattaa sisältää arkaluontoista tietoa, herättää huolta yksityisyydestä. On olemassa riski, että vahingossa kerätään ja paljastetaan arkaluonteista tietoa käyttäjistä, erityisesti chatbottien kohdalla, jossa henkilökohtaista tai luottamuksellista tietoa usein jaetaan. (Sebastian 2023, 1.)

On tärkeää kehittää tehokkaita strategioita ja teknologioita, joilla voidaan varmistaa käyttäjien tietojen turvallisuus, samalla ylläpitäen hyötyä, jota isoista kielimalleista saadaan. Yksityisyyden ja tietojen suojaaminen tekoälyjärjestelmässä sisältää eettisiä, oikeudellisia ja käyttäjän luottamukseen liittyviä huomioita. Yksityisyys ja tietojen suojaaminen on kriittistä eettisen ja vastuullisen tekoälyn kehittämiseksi ja käyttöönottamiseksi. Yksityisyys ja tietosuoja vaikuttavat myös käyttäjien luottamukseen tekoälyjärjestelmiä kohtaan, mikä vaikuttaa kokonaisuudessaan teknologian hyväksymiseen ja sen menestymiseen. (Sebastian 2023, 2.)

4.3.1 Luottamuksen vahvistaminen

Käyttäjän luottamus chatbottiin ja käyttäjän yksityisyyden turvaaminen ovat olennaisia onnistuneen chatbotin käyttöönoton takaamiseksi. Käyttäjät voivat kokea pulaa luottamuksesta, jos eivät tiedä, miten heidän tietojaan käytetään tai kenellä on niihin pääsy. Chatbottien tulee olla läpinäkyviä käyttäjän tietojen keräämisen, käsittelyn ja säilyttämisen suhteen. Organisaation tulisi julkaista selkeät ja tiiviit yksityisyyskäytännöt ja käyttöehdot, ja varmistua siitä, että ne ovat helposti käyttäjän

saatavilla. Näihin tulee sisällyttää myös tieto siitä, mitä tietoa kerätään, kuinka sitä käytetään ja kenen kanssa tiedot jaetaan. Käyttäjältä tulee myös selkeästi pyytää lupa tietojen keräämistä ja käsittelyä varten, ja käyttäjällä tulee olla mahdollisuus kieltäytyä tästä. (Yang ym. 2023.) Organisaatioiden tulee priorisoida luottamusta ja läpinäkyvyyttä chatbottien kehittämisessä ja käyttöönotossa. Käyttäjät tarvitsevat vakuutuksia sille, että heidän henkilötietojaan käsitellään vastuullisesti ja turvallisesti. (Sebastian 2023, 3.)

Eettiset periaatteet vaativat, että henkilökohtaista ja arkaluontoista tietoa käyttäjistä suojataan ja kunnioitetaan. Isot kielimallit, jotka käsittelevät valtavaa määrää tietoa, voivat vahingossa oppia ja tuottaa arkaluontoista tietoa. Henkilötietojen paljastuminen voi johtaa huomattaviin seuraamuksiin, kuten identiteettivarkauksiin, rahalliseen menetykseen ja mainehaittaan. (Sebastian 2023, 2–3.)

Generatiivinen tekoäly mahdollistaa sosiaalisen manipuloinnin monella eri tavalla. Tähän liittyy niin disinformaation ja valeuutisten levittäminen kuin ihmisten ostopäätöksiin vaikuttaminenkin. Ihmisten emotionaalinen, kognitiivinen tai käyttäytymisen manipulointi on kuitenkin eettisesti kyseenalaista, ja usein lisäksi laitonta. Äärimmäisen personoiduilla mainossisällöillä voidaan manipuloida ihmisiä ja vaikuttaa heidän ostoskäyttäytymiseensä. Tämä voi johtaa yksilön tunteeseen siitä, että heidän yksityisyyttään on loukattu, jos he eivät ole olleet tietoisia siitä, miten heihin liittyvää tietoa on käytetty. Organisaatioiden on tärkeää olla läpinäkyviä toiminnoissaan, jotta ihmisten autonomiaa ei loukata. (Wach ym. 2023, 15–17.)

Monet chatbotit jakavat tietoja kolmansille osapuolille, kuten viestintäsovelluksille tai pilviympäristöille (Hasal ym. 2021). Organisaatioiden tulee olla tietoisia siitä, mitä tietoja mahdolliset kolmannet osapuolet keräävät käyttäjistä, ja arvioida, miten tämä vaikuttaa kokonaisuuteen. Käyttäjille tulee myös kertoa, kun tietoja annetaan kolmansille osapuolille.

Tietosuojaan liittyy paljon olemassa olevia lakeja, kuten yleinen tietosuoja-asetus EU:ssa. Nämä lait asettavat vaatimuksia tiedon keräämiselle, käsittelylle ja säilyttämiselle. Yleinen tietosuoja-asetus takaa lisäksi käyttäjälle oikeuden tietoon pääsyyn, sen korjaamiseen ja poistoon. Myös generatiivisen tekoälyn, joka käsittelee EU:ssa olevien käyttäjien tietoja, tulee noudattaa näitä lakeja. Noudattamatta jättäminen voi johtaa merkittäviin sakkoihin ja oikeudellisiin seuraamuksiin. (Sebastian 2023, 3.)

Yleisen tietosuoja-asetuksen mukaan käyttäjillä on myös oikeus kerättyjen tietojen poistamiseen. Tekoälypohjaisten chatbottien suhteen tämä voi olla monimutkaista. Menneitä keskusteluja käytetään chatbottien kouluttamiseen, ja jo luodusta mallista tietoja on mahdoton poistaa. Yleensä tämä ei kuitenkaan ole ongelma, jos käyttäjän henkilötiedot on asianmukaisesti suojattu ennen mallin kouluttamista. (Hasal ym. 2021.)

4.3.2 Tahattomat väärinkäytöt

Käyttäjien yksityisyydelle muodostaa riskin tieto, jota käyttäjät itse antavat mallille syötteissään. Tahattomasti voidaan jakaa arkaluonteista tai henkilökohtaista tietoa. Tätä tietoa voidaan käyttää parantamaan ja kouluttamaan mallia, ja se voidaan sisällyttää vastauksena muiden ihmisten syötteeseen. Isot kielimallit ovat hyviä vuotamaan arkaluonteista tietoja, kun kysytään oikeita kysymyksiä. (Engler & Dhamani 2024, luku 3.) Tahattomat väärinkäytöt organisaation ja työntekijöiden toimesta ovat myös mahdollisia. Työntekijät voivat vahingossa syöttää luottamuksellista tietoa kielimallille, jolloin siitä tulee osa kielimallin tietoaineistoa. Tavallinen käyttäjä voisi mahdollisesti saada pääsyn tähän tietoon pelkästään kysymällä siitä. (Gupta ym. 2023.) Arkaluontoisen tai luottamuksellisen tiedon paljastuminen on yksi isoimmista kaupallisista riskeistä useimmille yrityksille. (Engler & Dhamani 2024, luku 3.) Tämä on herättänyt kysymyksiä siitä, miten isot kielimallit käsittelevät käyttäjien henkilökohtaisia tietoja. (Gupta ym. 2023.)

Yritykset ovat tietoisia isojen kielimallien puutteista, ja mahdollisesta tietovuodon riskistä, mutta ne keräävät ja tallentavat käyttäjien keskusteluita ja muuta tietoa käyttäjistä, kuten IP-osoitteen, laitetietoja ja niin edelleen. Monet yritykset salaavat tai poistavat henkilötietoja ennen kuin tietoja käytetään chatbotin kouluttamiseen, mutta se ei poista kaikkia riskejä. (Engler & Dhamani 2024, luku 3.)

4.3.3 Keinoja yksityisyyden ja tietosuojan vahvistamiseksi

Henkilötietoihin ja yksityisyyden suojaan liittyviä riskejä voidaan vähentää huomioimalla ja priorisoimalla eettiset huolenaiheet generatiivista tekoälyä kehitettäessä, ja panostamalla käyttäjien yksityisyyttä ja turvallisuutta suojaaviin järjestelmiin. Tähän sisältyy toimenpiteet vahvan tietosuojan kehittämiseksi, läpinäkyvyys henkilökohtaisen tiedon keräämisestä ja käyttämisestä sekä säännölliset turvallisuusohjeiden ja -käytäntöjen arvioinnit ja päivitykset. (Wach ym. 2023, 15–16.) Tiedon luottamuksellisuuden ylläpitämiseksi yksityisyyden suojaaminen ja tietojen turvallisuus on oleellista. Ilman kunnollisia suojatoimia tietoja voidaan peukaloida tai manipuloida, mikä voi johtaa epäluotettaviin tuloksiin tekoälyjärjestelmiltä. (Sebastian 2023, 3.)

Tekoälyn vastuullista suunnittelua varten tarvitaan vahvaa hallintoa monella eri tasolla, mukaan lukien yksittäiset kehittäjät, instituutiot, alat, sektorit ja kansainväliset organisaatiot (Wach ym. 2023, 15–16). Lakien mukaan toimiminen ja kolmansien osapuolien arvioinnit esimerkiksi yleisen tietosuoja-asetuksen mukaan toimimisesta voivat edesauttaa käyttäjien luottamusta. Käyttäjille tulee myös tarjota tietoa siitä, mitä riskejä chatbotteihin liittyy ja kuinka käyttäjät itse voivat suojata omia tietojaan. (Yang ym. 2023.)

4.4 Vastuu generatiivisen tekoälyn tuottamasta sisällöstä

Generatiivisen tekoälyn malleille on tyypillistä, että ne toimivat itsenäisesti ilman ihmisen valvontaa. Tämän seurauksena on vaikea hahmottaa, kenellä on vastuu mallin tuottamasta sisällöstä. Jos malli esimerkiksi antaa vääriä ohjeita, joista koituu rahallisia tai muita menetyksiä, voi olla vaikea määrittää, kenen vastuulla virhe on. Puutteellisen lainsäädännön seurauksena vastuun määrittäminen on vaikeaa, ja siitä voi koitua niin taloudellisia kuin maineeseenkin liittyviä haittoja organisaatioille ja yksilöille. (Wach ym. 2023, 10–11.)

Isoista kielimalleista saatujen tietojen todenperäisyyttä tai tarkkuutta ei voi varmistaa, ja siksi käyttäjien odotuksia tulee hallita (Engler & Dhamani 2024, luku 5). Generatiivinen tekoäly kykenee luomaan vakuuttavaa sisältöä, minkä takia käyttäjillä on tapana luottaa generatiivisen tekoälyn vastauksiin. Tämä vaikuttaa ihmisten päätöksentekoon, ja saattaa vaikuttaa jopa moraalisiin valintoihin. Liiallinen luottamus tekoälyn tuottamaan sisältöön ilman kriittistä suhtautumista voi johtaa huonojen päätösten tekemiseen. Generatiiviseen tekoälyyn liittyvä antropomorfismi herättää kysymyksiä siitä, kuinka turvallista siihen on luottaa ja kenellä on vastuu siitä, kun generatiivinen tekoäly aiheuttaa harmia ohjeillaan. Organisaatioiden ja kehittäjien tulee ottaa vastuu siitä, että tekoälyä hyödyntävät työkalut ovat turvallisia käyttää sekä valvoa niiden toimintaa. (Wach ym. 2023, 10–11.)

Generatiivisen tekoälyn luomat vastaukset voivat olla laadultaan huonoja, epärelevantteja ja jopa loukkaavia. Pitkällä aikavälillä tämä voi johtaa huonoihin liiketoimintapäätöksiin ja operationaaliseen tehottomuuteen. Laadun parantamiseksi on oleellista käyttää monipuolista ja laadukasta, ennakkoon hyväksyttyä tietoaineistoa mallien kouluttamisessa. Myös ihmisten ja käyttäjien palautteen perusteella oppiminen parantaa tulosten laatua. Laadun varmistamiseksi asiakaspalvelussa tulisi ottaa käyttöön selkeät ohjeet ja standardit generatiivisen tekoälyn hyödyntämiselle: tekoälyn luomia vastauksia tulee monitoroida ja arvioida säännöllisesti. Myös asiakaspalvelutyöntekijöiden osaaminen vaikuttaa siihen, kuinka hyvin generatiivista tekoälyä hyödynnetään, ja työntekijöiden kouluttamiseen kannattaa panostaa. (Wach ym. 2023, 11–13.)

Isoja kielimalleja on hyvä lähestyä kriittisesti, ja niiden käyttöä on harkittava tarkkaan erityisesti aloilla, joilla ihmisen asiantuntemuksen rooli on vahva, kuten esimerkiksi mielenterveyteen liittyvissä asioissa. On tärkeää varmistua siitä, että kielimallin tuottama tieto on tosiasiallisesti oikein eikä tahattomasti levitä ennakkoluuloja tai väärää tietoa. Isoja kielimalleja käyttäessä on otettava käyttöön toimenpiteitä, jotka estävät tai vähentävät mahdollisia riskejä, kuten mallien hienosäätäminen turvallisemmaksi, ihmisen arvioinnin lisääminen ja sääntöjen muodostaminen vastuullisen käytön varmistamiseksi. (Amaratunga 2023, luku 6.) Generatiivista tekoälyä voi käyttää työn tukena, mutta siihen ei voi luottaa sokeasti (Engler & Dhamani 2024, luku 5.).

4.5 Yhteiskunnalliset vaikutukset

Generatiivinen tekoäly ja isot kielimallit vaikuttavat yksilö- ja organisaatiotasojen lisäksi myös yhteiskunnallisella tasolla. Automaation lisääntyessä työttömyys voi kasvaa ja taloudelliset erot lisääntyä, millä on yhteiskunnallisia vaikutuksia. Lisäksi disinformaation ja erilaisten vaikutusoperaatioiden vaikutuksia yhteiskuntaan on hyvä arvioida. Tekoälyn koulutusaineistoon sisältyvät ennakkoluulot ovat olleet paljon keskustelussa, ja tekoälyä käytettäessä on syytä pitää mielessä, että sen tuottamat tiedot eivät ole objektiivisia tai välttämättä edes totta.

Organisaatioiden on hyvä varautua myös yhteiskunnallisiin vaikutuksiin, kuten mahdolliseen työttömyyden lisääntymiseen. Työntekijöitä on koulutettava uusista teknologioista sekä niiden käyttöön liittyvistä riskeistä ja mahdollisuuksista. Yhteiskunnan tulee tukea yksilöitä erilaisissa siirtymätilanteissa, esimerkiksi työttömyyden kohdalla, jotta muutosten vaikutuksia voidaan hillitä.

4.5.1 Taloudellisten erojen lisääntyminen

Yhteiskunnallisella tasolla generatiivinen tekoäly voi vaikuttaa taloudellisten erojen lisääntymiseen niin yksilö- kuin organisaatiotasollakin. Automatisoinnin avulla voidaan tehdä työtä, joka nykyään on ihmisten tekemää, mikä voi johtaa työttömyyteen ja taloudellisten erojen lisääntymiseen erityisesti sellaisilla aloilla, joilla tehdään paljon toistuvia tehtäviä. Tekoälyn tuomat ekonomiset hyödyt voivat jakautua hyvin epätasaisesti. (Wach ym. 2023, 10–11.)

Organisaatiotasolla tämä voi tarkoittaa sitä, että tietyt yritykset alkavat dominoida tekoälymarkkinaa, mikä johtaa vallan keskittymiseen, ja voi johtaa kilpailun ja innovaatioiden vähenemiseen. Pienten ryhmien voi olla vaikeampi päästä tekoälyteknologioiden pariin, kun esimerkiksi pienillä organisaatioilla ei välttämättä ole mahdollisuutta kehittää ja ottaa käyttöön edistyneitä tekoälymalleja. Tällöin yritykset, joilla näihin uusiin teknologioihin on varaa, saavuttavat kilpailullisen etulyöntiaseman. Tekoälypalveluiden hinnoittelu on myös olennaista: ilman lainsäädäntöä yritykset voivat asettaa korkeita hintoja tekoälypalveluille, jolloin pienillä organisaatioilla tai yksilöillä ei ole niihin varaa. Tämä voi muodostaa esteen teknologioiden käyttöönotolle, vähentäen niiden käyttöä ja niistä mahdollisesti koituvia hyötyjä. (Wach ym. 2023, 10–11.)

4.5.2 Automaation vaikutus työllisyyteen

Tekoäly voi luoda uusia työpaikkoja ja lisätä tehokkuutta, mutta se saattaa myös korvata monia nykyisiä työpaikkoja. Tekoälyn tuoma automatisointimahdollisuus koskee sekä tietotyöläisiä että manuaalisen työn tekijöitä. Seurauksena voi olla suurta työttömyyttä tietyillä aloilla ja ammateissa. Työttömyyden riski voi johtaa negatiiviseen näkemykseen tekoälystä, mikä vaikuttaa oppimisprosessiin ja työhyvinvointiin kielteisesti. (Wach ym. 2023, 13–15.) Isojen kielimallien avulla voidaan

automatisoida paljon erilaisia tehtäviä, mutta monet työt vaativat ihmisen älykkyyttä, luovuutta ja empatiaa, joten aivan kokonaan ihmisiä ei voi korvata isoilla kielimalleilla. (Amaratunga 2023, luku 6.)

Tekoälyn kehittyminen ja käyttöönotto voi myös luoda lukuisia uusia työmahdollisuuksia. Uudet työpaikat ovat etenkin tekoälyn tutkimuksen ja kehityksen alalla, mutta myös esimerkiksi koneoppimisen, ohjelmoinnin ja tietoturvan asiantuntemus tulee olemaan tärkeää. (Wach ym. 2023, 13–15.)

Lainsäätäjien ja työnantajien tulee panostaa työntekijöiden kouluttamiseen, jotta luodaan mahdollisuus pärjätä muuttuvassa ympäristössä. Kouluttaminen on myös oleellista työttömyyden välttämiseksi ja osaavan työvoiman takaamiseksi. Työntekijöitä tulee myös tukea siirtymissä, kuten työpaikan vaihtamisessa, ja varmistaa tarvittavan lisäkoulutuksen saatavuus. Muuttuvassa maailmassa kriittisen ajattelun, ongelmanratkaisun, viestintätaitojen ja yhteistyön osaaminen korostuu. Työntekijät voivat siirtyä toistuvista työtehtävistä tehtäviin, joissa tarvitaan kyseisiä taitoja. Tekoäly todennäköisesti tulee korvaamaan joitakin työtehtäviä, mutta ei koko työtä, mikä tulee ottaa huomioon strategisia päätöksiä suunniteltaessa. Organisaatioiden tulee panostaa työntekijöiden kouluttamiseen sekä teknisten taitojen ja kriittisen ajattelun, ongelmanratkaisun ja itsensä johtamisen kehittämiseen. (Wach ym. 2023, 13–15.)

4.5.3 Geopoliittiset vaikutukset ja sosioekonomisen epätasa-arvon lisääntyminen

Geopoliittiset riskit ja seuraukset on myös hyvä huomioida osana tekoälyyn liittyviä riskejä. Maat ja organisaatiot, jotka ovat etulyöntiasemassa, voivat saavuttaa merkittäviä ekonomisia, poliittisia ja strategisia etuja tekoälyteknologioita hyödyntämällä. Tämä voi johtaa vallan epätasaiseen jakautumiseen, kansainvälisen kilpailun vähenemiseen ja kyseenalaiseen kilpailuun tekoälyteknologioiden kehityksessä. Kyseenalainen kilpailu voi johtaa epäeettisiin käytäntöihin, kuten tietojen varastamiseen ja tekijänoikeuksien loukkaamiseen, mikä voi vaikuttaa geopoliittiseen tilanteeseen ja globaaliin talouteen. (Wach ym. 2023, 10–11.)

Generatiivinen tekoäly lisää myös sosioekonomista epätasa-arvoa. Tärkeät sosioekonomiset erot liittyvät usein digitaaliseen epätasa-arvoon. Digitaalinen epätasa-arvo tarkoittaa epätasaista jakoa pääsyssä teknologiaan, digitaalisissa taidoissa ja teknologian käytössä sekä teknologian käyttöön liittyvissä eduissa ja haitoissa. Esimerkiksi ChatGPT:n käyttö ei ole levinnyt laajalle, vaikka sen käyttö onkin ilmaista. Matalan tulotason maat käyttävät ChatGPT:tä vähemmän kuin korkeamman tulotason maat, mikä voi johtua infrastruktuurisista haasteita tai tiedon puutteesta. Digitaalinen epätasa-arvoisuus lisää taloudellista epätasa-arvoa ja vähentää kilpailua. Kehittyneet taloudet ovat paremmassa asemassa hyödyntää tekoälyteknologioita, kun taas kehittyvissä talouksissa on erilaisia haasteita, kuten digitaalisen infrastruktuurin puutteellisuus ja investointipääoman puute.

Kehittyvät taloudet menettävät ekonomista ja poliittista neuvotteluvaltaa, ja niistä tulee entistä riippuvaisempia alueista, jotka kontrolloivat teknologiaa. Globaalin etelän tukemiseen tuleekin panostaa kansainvälisen yhteistyön avulla, jotta kukaan ei jää ulkopuolelle tekoälyn vallankumouksesta. (Wach ym. 2023, 17–19.)

4.5.4 Tekijänoikeudet

Tekijänoikeuksista generatiivisen tekoälyyn liittyen käydään yhä kiivasta keskustelua. Generatiivisen tekoälyn kehittäjät eivät useinkaan omista tietoaineistoa, jonka pohjalta sitä on koulutettu, mikä voi johtaa kiistoihin tietojen omistajuudesta ja käyttöoikeudesta (Gupta ym. 2023). Lisäksi keskustelua käydään siihen liittyen, saako generatiivisella tekoälyllä luotua sisältöä suojata tekijänoikeuksilla (Wach ym. 2023, 16–17).

Epäselvyydet tekijänoikeuksissa ja omistajuuden määrittelemisessä voivat johtaa oikeudellisiin riitoihin ja haasteisiin tekijänoikeusrikkomuksista. Tekoälyn luoman sisällön, kuten artikkelien ja kuvien, tekijänoikeuteen liittyvistä asioista käydään vielä keskustelua. (Wach ym. 2023, 10–11.)

Plagiointi on myös ongelma, jonka riski voi lisääntyä tekoälyn käytön myötä. Tekoälysovelluksia on hyödynnetty sekä opiskelijoiden että tutkijoiden toimesta ilman, että siitä on mainittu. Myös yritykset käyttävät generatiivista tekoälyä yhä enemmän toiminnoissaan, ja on oleellista, että myös yritykset raportoivat vastuullisesti tekoälyn käytöstään. (Wach ym. 2023, 16–17.)

4.5.5 Disinformaatio

Generatiivista tekoälyä hyödyntämällä voidaan tehokkaasti luoda ja levittää disinformaatiota ja misinformaatiota. Lisäksi vahingollisten stereotypioiden ja ennakkoluulojen ylläpitäminen on riski. (Wach ym. 2023, 11–13.) Suuren käyttäjämäärän takia virheellisen tiedon leviäminen voi olla laajaa (Gupta ym. 2023).

Viimeisen vuosikymmenen aikana erilaiset vaikutusoperaatiot etenkin internetissä ja sosiaalisen median alustoilla ovat tulleet julkiseen tietoisuuteen. Vaikutusoperaatioilla pyritään vaikuttamaan kohdeyleisön mielipiteisiin. Tällaiset vaikutusoperaatiot heikentävät luottamusta instituutioihin ja tietoympäristöön, lisäävät erimielisyyttä ja jakautumista yhteiskunnassa ja johtavat usein todellisiin seuraamuksiin, kuten taloudellisiin menetyksiin ja väkivaltaan. (Engler & Dhamani 2024, luku 5.)

Väärissä käsissä generatiivisia malleja voidaan hyödyntää vihapuheen ja disinformaation lisäämiseen ja levittämiseen. Vaikka generatiivisia malleja tai tekoälyjä ei varsinaisesti tarvita vaikutusoperaatioiden tuottamiseen, se tekee siitä kuitenkin helpompaa ja tehokkaampaa. (Engler & Dhamani 2024, luku 5.) Isot kielimallit mahdollistavat disinformaation levittämisen, koska ne kykenevät

luomaan uskottavia, mutta täysin valheellisia uutisartikkeleita. Disinformaation avulla voidaan levittää väärää tietoa, manipuloida julkista mielipidettä ja aiheuttaa poliittista epävakautta. Väärien uutisten avulla voidaan vaikuttaa myös esimerkiksi osakemarkkinoihin. Myös polarisaatiota voidaan vahvistaa isojen kielimallien avulla, jos käyttäjille tarjotaan vain sisältöä, joka tukee henkilön aiempia näkemyksiä. (Amaratunga 2023, luku 6.)

Tekoälyn luomaa epätodellista sisältöä on yhä vaikeampi tunnistaa ihmisen luomasta sisällöstä. Syväväärennösten avulla erilaiset mahdollisuudet huijaamiseen ovat käytännössä loputtomat. Generatiivisella tekoälyllä on mahdollista luoda synteettisiä kuvia ja videoita ihmisistä ilman heidän suostumustaan. (Wach ym. 2023, 11–16.) Syväväärennöksissä voidaan hyödyntää isoja kielimalleja myös esimerkiksi luomalla vakuuttavaa dialogia, joka tekee niistä uskottavampia (Amaratunga 2023, luku 6). Riskinä on moninaiset negatiiviset seuraukset, mukaan lukien identiteettivarkaudet ja kiristäminen. Syväväärennöksiä voidaan hyödyntää propagandaan tai muun disinformaation levittämiseen, mikä asettaa julkiset auktoriteetit ja median luotettavuuden kyseenalaiseksi. Syväväärennöksille tuleekin asettaa selkeät ja tiukat säännöt ja ohjeet niiden käyttämiseen muun muassa poliittisissa kampanjoissa ja uutismedioissa. (Wach ym. 2023, 11–16.) Merkinnät alkuperästä ovat yksi keino, jolla riskejä voisi mahdollisesti vähentää, ainakin kertomalla kuluttajille olennaisia tietoja kuvan alkuperästä. Tekniset haasteet tällaisen toteuttamiselle ovat kuitenkin merkittävät. (Engler & Dhamani 2024, luku 5.)

Tekoälyn luoman sisällön määrä tulee todennäköisesti myös saastuttamaan tietoympäristöä ja vaikuttamaan isojen kielimallien kouluttamiseen tulevaisuudessa: mitä enemmän disinformaatiota isoilla kielimalleilla luodaan, sitä enemmän tulevaisuuden isot kielimallit tulee olemaan koulutettuja mahdollisesti vahingollisella sisällöllä. (Engler & Dhamani 2024, luku 5.)

Medialukutaito on oleellinen keino torjua vaikutusoperaatioita, etenkin yhdistettynä muihin strategioihin. Yksilön luottamusta valeuutisiin tulisi vähentää ja todellisiin uutisiin lisätä. Ihmisten tietoisuutta disinformaatiosta tulee lisätä ja medialukutaitoja kehittää. Myös lainsäädännöllä voidaan vaikuttaa disinformaation leviämiseen, jos yritykset saadaan vastuuseen alustoillaan julkaistusta, tekoälyllä luodusta, disinformaatiota sisältävästä sisällöstä. (Engler & Dhamani 2024, luku 5.) Generatiivinen tekoäly vaatii valvontaa ja vastuumekanismeja, jotta sitä ei käytetä väärin (Wach ym. 2023, 11–13).

4.5.6 Ennakkoluulot ja puolueellisuus

Isoihin kielimalleihin, kuten kaikkiin muihinkin uusiin teknologioihin, liittyy riskejä ja rajoitteita, jotka on hyvä ymmärtää, ennen kuin teknologiaa hyödynnetään. Isot kielimallit eivät kykene tiedostamaan tai ymmärtämään sisältöä, vaan luovat tekstiä koulutusaineistonsa perusteella. Ne eivät ole

tietoisia olentoja, vaan tuottavat sisältöä ilman tunteita ja tarkoitusta. Isot kielimallit eivät ole aina oikeassa, ja ne voivat tuottaa väärää, harhaanjohtavaa tai puolueellista tietoa. Niiden tieto perustuu ainoastaan aineistoon, jota on käytetty koulutuksessa, eivätkä ne pysty luomaan esimerkiksi täysin uutta tietoa tai teorioita. Isot kielimallit eivät ole puolueettomia ja objektiivisia, koska ne voivat oppia, ja oppivatkin, koulutuksessa käytetyn tietoaineiston sisältämät ennakkoluulot. (Amaratunga 2023, luku 6.) Koulutusaineistossa olevat ennakkoluulot vaikuttavat mallin käyttäytymiseen, ja pahimmillaan johtavat syrjintään ja epäreiluun kohteluun, mikä voi lisätä olemassa olevaa epätasa-arvoa ja syrjiviä käytäntöjä. Vaikutukset voivat ulottua niin yksilöön, organisaation kuin yhteiskuntaankin. Puolueellisuus voi näkyä esimerkiksi historiallisen syrjinnän toistamisena, tietyn poliittisen suuntautumisen suosimisena tai epätoivottujen käytäntöjen vahvistamisena. (Wach ym. 2023, 10–13.)

Etenkin koneoppimisen ohjelmissa, joissa on hyödynnetty tietoaineistoa vain tietyltä demografiselta ryhmältä, puolueellisuus korostuu. Myös suodattamattomat, internetistä peräisin olevat laajat tietoaineistot sisältävät yleensä puolueellista tietoa, jonka generatiivisen tekoälyn mallit oppivat. Koska aineisto on yleensä peräisin menneisyydestä, siinä ei myöskään ole kaikkein edistyksellisimpiä asioita mukana, joita eri sosiaaliset ja ekonomiset liikkeet saavat aikaan. Puolueellisuuden vähentämiseksi on tärkeää tutkia menetelmiä, joilla läpinäkyvyyttä voidaan lisätä ja puolueellisuutta vähentää. Lisäksi laadukas, monipuolinen koulutusaineisto on tärkeää generatiivisten mallien kouluttamisessa. (Wach ym. 2023, 17–19.)

Generatiivisen tekoälyn vastaus on täysin riippuvainen koulutusaineiston laadusta ja edustavuudesta. Algoritmien kehittämiseen tulisikin omaksua holistisempi ja inklusiivisempi näkökulma. (Wach ym. 2023, 11–13.) Tekoäly alana on erittäin homogeeninen ja valkoisten miesten dominoima. Puutteet sukupuolen ja entisyyden moninaisuudessa heijastuvat mallien tuotoksiin, kun kaikkia näkökulmia ei oteta huomioon. (Wach ym. 2023, 17–19.) Eri näkökulmien huomioiminen algoritmeja kehittäessä on oleellista. (Wach ym. 2023, 11–13.)

Chatbotit usein keräävät tietoa käydyistä keskusteluista ja oppivat jatkuvasti lisää kerätystä tietoaineistosta. Tämä tarkoittaa käytännössä sitä, että chatbotti voi oppia tahattomasti stereotypioita ja ennakkoluuloja, ja jatkossa toimia puolueellisesti tai syrjivästi. (Yang ym. 2023.)

Organisaatiolle on ongelmallista, jos asiakkaita kohdellaan eriarvoisesti ja lainvastaisesti. Mekanismit ennakkoluulojen tunnistamiseksi ja vähentämiseksi voivat olla puutteellisia lainsäädännön puuttuessa. (Wach ym. 2023, 10–11.) Chatbotin suunnittelu, toteutus ja käyttöönotto tulisi tehdä vastuullisesti ja eettisesti. Lisäksi erityisesti herkissä aiheissa, kuten mielenterveyteen liittyen, tulee pohtia myös ihmisten välisen vuorovaikutuksen merkitystä. Etenkin näissä tapauksissa tulee varmistaa, että chatbotit eivät kokonaan korvaa ihmisten välistä vuorovaikutusta. (Yang ym. 2023.)

4.5.7 Generatiivisen tekoälyn hyödyntäminen haitallisiin tarkoituksiin

Tietojen kalastelussa hyökkääjät esittävät olevansa luotettavia entiteettejä kerätäkseen arkaluontoista tietoa hyväuskoisilta uhreilta. Generatiivisen tekoälyn avulla tietoja voidaan kalastella entistä tehokkaammin ja vaikeammin havaittavasti, esimerkiksi luomalla sähköposteja, jotka imitoivat tiettyä henkilöä. (Gupta ym. 2023.) Tässä voidaan käyttää hyväksi esimerkiksi sosiaalisesta mediasta löytyviä profiileja tietolähteenä, joka syötetään generatiiviselle tekoälylle, auttaen sitä luomaan uskottavampia tietojen kalasteluun tähtääviä sähköposteja (Engler & Dhamani 2024, luku 5). Isojen kielimallien kyky oppia eri kirjoitustyylejä mahdollistaa myös sen, että niitä voidaan käyttää petokseen, jossa hyödynnetään esiintymistä tiettynä henkilönä. Isot kielimallit kykenevät luomaan vakuuttavaa sisältöä, joihin ihmisten on helppo uskoa. (Amaratunga 2023, Luku 6.) Myös kieliopillisesti oikein olevan englanninkielisen sisällön luominen on helpompaa generatiivisen tekoälyn avulla, mikä lisää kalasteluyritysten uskottavuutta. Generatiivinen tekoäly ei siis mahdollista mitään uutta, mutta se tekee tietojen kalastelusta tehokkaampaa ja halvempaa. (Engler & Dhamani 2024, luku 5.)

Sosiaalinen manipulointi tähtää yksilöiden manipuloimiseen. Tämä voi tarkoittaa sitä, että yksilö huijataan suorittamaan jokin tehtävä tai paljastamaan luottamuksellista tietoa. Generatiivisen tekoälyn avulla voidaan luoda hyvin vakuuttavan kuuloista ja kontekstiin liittyvää sisältöä, jonka avulla yksilöitä on helpompi huijata. (Gupta ym. 2023.) Keräämällä tietoja esimerkiksi sosiaalisen median alustoilta ja muualta internetistä, generatiivista tekoälyä voidaan käyttää luomaan tarkkoja profiileja yksilöistä. Nämä profiilit voivat sisältää myös tietoja yksilön henkilökohtaisista mielenkiinnonkohteista, uskomuksista ja ihmissuhteista. Tällainen sosiaalinen valvonta mahdollistaa esimerkiksi julkisen mielipiteen manipuloinnin kohdennetun mainonnan avulla ja saattaa johtaa tiettyjen ihmisryhmien syrjintään. Sosiaalisella valvonnalla voidaan etsiä haavoittuvia kohteita ja generatiivisen tekoälyn avulla toteuttaa kohdennettua häirintää. (Wach ym. 2023, 15–16.)

Lunnas- ja haittaohjelmien kehittäminen on ollut aikaa vievää ja taitoa vaativaa, mutta generatiivisen tekoälyn avulla tällaisten ohjelmien luominen on entistä helpompaa. Myös erilaisten virusten, hyökkäyskuormien ja polymorfisten haittaohjelmien tekemiseen voidaan hyödyntää generatiivista tekoälyä. (Gupta ym. 2023.) Erityisesti koodin luontiin kehitetyt isot kielimallit voivat auttaa haittaohjelmien kehittäjiä, kun parempaa koodia pystytään tuottamaan nopeammin kuin aiemmin. (Engler & Dhamani 2024, luku 5.)

Generatiivista tekoälyä voidaan hyödyntää myös automaattiseen hakkerointiin tai koodin analysointiin haitallisin tarkoitusperin. Kielimalleilla on laaja tietoaineisto tunnetuista haavoittuvuuksista, ja tätä tietoa voidaan hyödyntää koodin analysoimiseen ja haavoittuvuuksien etsimiseen. (Gupta ym. 2023.)

5 Tutkimuksen toteutus

Tässä luvussa käsitellään opinnäytetyön tutkimuksen metodologiaa, toteutusta ja tuloksia. Ensimmäisessä alaluvussa (5.1) esitellään tutkimuksen lähestymistapa, aineiston hankintamenetelmät ja aineiston analyysimenetelmät. Toisessa alaluvussa (5.2) esitellään tutkimuksen tulokset.

5.1 Tutkimus- ja kehittämishankkeen metodologia

Opinnäytetyön lähestymistapa on tapaustutkimus. Tapaustutkimus sopii hyvin tilanteisiin, joissa tutkimuksen tarkoituksena on tuottaa yksityiskohtaista, syvällistä tietoa tutkittavasta tapauksesta, ja joissa tutkimuksen avulla pyritään tuottamaan kehittämisehdotuksia aiheeseen liittyen. Tapaustutkimuksen avulla ei pyritä tilastolliseen yleistämiseen, vaan keskitytään kysymyksiin ”miten?” ja ”miksi?” ilmiön todellisessa toimintaympäristössä. (Ojasalo, Moilanen & Ritalahti 2015, 52–53.) Opinnäytetyössä perehdytään syvällisesti organisaation nykytilanteeseen, jotta saadaan selville ne asiat, joita organisaatiossa tulee kehittää, jotta generatiivinen chatbotti voidaan ottaa käyttöön. Tapaustutkimus lähestymistapana mahdollistaa perusteellisen syventymisen organisaation tilanteeseen useasta eri näkökulmasta, ja tämän avulla uuden tiedon tuottamisen generatiivisen chatbotin käyttöönoton tueksi.

5.1.1 Aineiston hankintamenetelmät

Opinnäytetyön tavoitteena on saada syvällistä tietoa tutkimuksen kohteesta. Tähän tarkoitukseen puolistrukturoitu haastattelu sopii hyvin, sillä menetelmä on oiva tilanteisiin, joissa halutaan syventää tai selventää tutkimuksen kohteena olevia asioita (Ojasalo ym. 2015, 106). Puolistrukturoitu haastattelu mahdollistaa sen, että haastattelun aikana kysymyksien järjestystä tai sanamuotoa voidaan muuttaa, ja tarpeen mukaan lisätä tai poistaa kysymyksiä (Ojasalo ym. 2015, 108). Opinnäytetyössä selvitetään organisaation nykytilannetta syvällisesti, jotta valmiuksia generatiivisen chatbotin käyttöönottoon voidaan selvittää. Haastattelujen avulla saadaan kerättyä syvällistä ja selventävää tietoa aiheesta useammasta eri lähteestä ja näkökulmasta organisaation sisällä. Haastatteluissa selvitetään myös organisaatiosta löytyvän, aiheeseen liittyvän dokumentaation olemassaoloa ja käyttömahdollisuuksia.

Haastattelut toteutettiin yhteensä seitsemälle henkilölle, jotka toimivat eri rooleissa Hoasin organisaatiossa. Haastateltavien anonyymiyden säilyttämiseksi haastateltavien rooleja ei avata tämän enempää, koska roolin perusteella haastateltavat voitaisiin tunnistaa raportista. Haastatteluihin pyrittiin valitsemaan mahdollisimman laajasti eri näkökulmista chatbottiin liittyviin asioihin lähestyviä henkilöitä. Kaikki haastateltavat olivat ainakin jossain määrin tietoisia organisaation nykyisestä

chatbotista, mutta syvällinen nykyisen chatbotin tuntemus ei ollut tarpeellista. Haastattelut toteutettiin yksilöhaastatteluina, ja ne kestivät keskimäärin puolesta tunnista tuntiin.

Haastatteluissa selvitettiin organisaation valmiuksia generatiivisen chatbotin käyttöönottoa varten. Puolistrukturoidun haastattelun pohjana käytettiin kysymyksiä, jotka liittyivät chatbotin tavoitteisiin, käyttötapauksiin, ylläpitoon ja ei-toiminnallisiin määrittelyihin. Tavoitteisiin liittyvät kysymykset koskivat sekä nykyiselle chatbotille asetettuja tavoitteita, että chatbotin uudistuksesta haettuja hyötyjä. Käyttötapauksiin liittyvillä kysymyksillä kartoitettiin asiakaspalvelun kysytyimpiä teemoja, ongelmatilanteita, chatbotin ja palveluneuvojan välistä suhdetta sekä materiaalia, jota palveluneuvojilla on käytössään asiakkaiden kysymyksiin vastaamiseen. Chatbotin ylläpitoa kartoitettiin kysymällä, miten nykyisen chatbotin toimintaa seurataan, kerätäänkö siitä palautetta sekä koetaanko ylläpidossa ongelmia. Ei-toiminnallista määrittelyistä selvitettiin organisaation tietoturva- ja tietosuojavaatimuksia sekä organisaation tietoa chatbotin lainsäädännöllisistä vaatimuksista.

Haastattelujen aikana kysymyksiä poistettiin tai lisättiin, yleensä riippuen haastateltavan roolista organisaatiossa, ja siitä, kuinka tietoinen kyseinen haastateltava oli kyseisestä aiheesta. Osalla haastateltavista oli enemmän tietoa esimerkiksi tavoitteisiin ja ylläpitoon liittyen, osalla taas haastattelut painottuivat asiakaspalveluun liittyviin kysymyksiin. Haastatteluja muokattiin sen perusteella, mitä haastatteluiden aikana selvisi.

Haastattelut toteutettiin Teams-sovelluksen kautta etäyhteydellä. Teams-sovelluksen avulla haastattelut pystyttiin tallentamaan ja automaattisesti litteroimaan. Automaattinen litterointi helpotti haastattelujen viemistä kirjalliseen muotoon, mutta vaati vielä runsaasti manuaalista läpikäyntiä automaation tekemien virheiden korjaamiseksi ja puutteiden paikkaamiseksi. Haastattelut litteroitiin yleiskielellä. Tässä tapauksessa sanat tai sanavalinnat eivät olleet niin oleellisia kuin esiin tulleet asiat, joten litterointia ei tehty sanatarkasti (Ojasalo ym. 2015, 110).

5.1.2 Aineiston analyysimenetelmät

Litteroinnin jälkeen haastattelujen aineisto analysoitiin. Dokumenttianalyysissä tutkittavasta ilmiöstä kirjoitettu, puhuttu tai kuvattu aineisto analysoidaan järjestelmällisesti, ja pyritään tuottamaan tutkittavasta asiasta selkeä sanallinen kuvaus (Ojasalo ym. 2015, 136). Litteroidulle aineistolle tehtiin sisällönanalyysi, joka toteutettiin teoriaohjauksisesti. Teoriaohjauksisella sisällönanalyysillä tarkoitetaan, että analyysin luokittelu pohjautuu aikaisempaan viitekehykseen (Ojasalo ym. 2015, 140). Teoriaohjaavassa analyysissä teoria toimii ikään kuin analyysin apuna. Analyysissä tunnistetaan teorian vaikutus, mutta sitä ei pyritä testaamaan. Analyysiprosessissa vaihtelevat aineistolähtöisyys ja teorian ohjaavuus, ja näitä pyritään yhdistämään toisiinsa. (Tuomi & Sarajärvi 2018, luku 4.2.)

Sisällönanalyysin vaiheet voi karkeasti jaotella aineiston pelkistämiseen, ryhmittelyyn ja abstrahointiin. Aineisto pelkistetään karsimalla epäolennaiset asiat pois, esimerkiksi tiivistämällä tai pilkkomalla osiin. Pelkistämisen jälkeen aineisto ryhmitellään, eli etsitään käsitteitä, jotka kuvaavat joko samankaltaisuuksia tai eroavaisuuksia. Käsitteet ryhmitellään ja yhdistetään alaluokiksi. Alaluokat nimetään aineiston sisällön mukaan, ja alaluokkia yhdistelemällä muodostetaan yläluokkia ja pääluokkia. Pääluokat nimetään puolestaan ilmiötä kuvaavan aiheen mukaan. Lopuksi muodostetaan yhdistävä luokka, joka yhdistyy tutkimustehtävään. (Tuomi & Sarajärvi 2018, luku 4.4.3.)

Abstrahointi tarkoittaa aineiston käsitteellistämistä, jossa tunnistetaan olennainen tieto ja muodostetaan teoreettisia käsitteitä. Käsitteitä yhdistelemällä saadaan vastaus tutkimuskysymykseen. (Tuomi & Sarajärvi 2018, luku 4.4.3.) Teoriaohjaavan sisällönanalyysin abstrahoinnissa tutkimuksen aineisto liitetään olemassa olevaan teoriaan. Alaluokat muodostetaan aineistolähtöisesti, mutta yläluokat pohjautuvat teoriaan. (Tuomi & Sarajärvi 2018, luku 4.4.5.)

Opinnäytetyössä litteroitu aineisto pelkistettiin tiivistämällä oleelliset asiat ja karsimalla pois sellaiset aiheet, jotka eivät liittyneet tutkimukseen. Pelkistämisen jälkeen aineisto ryhmiteltiin ensin aineistosta nouseviin alaluokkiin. Alaluokista muodostettiin aineistolähtöiset yläluokat, minkä jälkeen niitä alettiin yhdistää teoriasta nouseviin käsitteisiin. Opinnäytetyön teoreettisessa viitekehyksessä määriteltiin käsitteet, joiden perusteella aineistosta esiin nousevat alaluokat luokiteltiin. Sisällönanalyysin avulla tutkimusaineisto saatiin muotoiltua tiiviiksi ja selkeäksi, minkä lisäksi tutkimusaineistosta saatiin johdettua vastaukset tutkimuskysymyksiin.

5.2 Tutkimuksen tulokset

Tulosten läpikäyminen aloitetaan chatbotin tavoitteista ja generatiivisesta tekoälystä haetuista hyödyistä. Tämän jälkeen siirrytään chatbotin käyttötapauksiin ja asiakaspalvelussa käytettävissä olevaan materiaaliin. Sitten käydään läpi nykyisen chatbotin haasteet sekä toiminnan seuraaminen ja ylläpito. Lopuksi käsitellään tulokset liittyen chatbotin tietoturvaan ja tietosuojaan sekä lainsäädäntöön. Viimeisenä käydään läpi muita haastatteluissa esille nousseita olennaisia aiheita.

5.2.1 Chatbotin tavoitteet

Haastatteluissa selvitettiin, miksi organisaatiossa on alun perin otettu chatbotti käyttöön. Suurin syy chatbotin käyttöönottoon oli asiakaspalvelun työkuorman vähentäminen. Ennen chatbotin käyttöönottoa palveluneuvojat vastasivat toistuviin kysymyksiin samoista asioista, ja chatbotin avulla asiakaspalvelun resursseja on saatu paremmin hyödynnettyä, kun asiakkaat voivat omatoimisesti hoitaa asiansa organisaation digitaalisissa kanavissa. Yksinkertaiset, usein toistuvat kysymykset on siirretty chatbotin hoidettavaksi, ja palveluneuvojat keskittyvät henkilökohtaista palvelua vaativiin asiakaspalvelutehtäviin.

> ”Siellä toistui tosi paljon kysymykset, ja se vei hirveästi resurssia meiltä. Meillä pitäisi olla moninkertainen määrä ihmisiä siellä chatissa, jos meillä ei olisi bottia siellä.”

Tämän lisäksi käyttöönoton syynä mainittiin jatkuvan palvelun tarve. Asiakaskunta koostuu opiskelijoista, ja he ovat usein aktiivisia myös toimistoaikojen ulkopuolella, iltaisin ja viikonloppuisin. Chatbotin avulla asiakkaat voivat saada vastauksia yleisimpiin kysymyksiin myös silloin, kun palveluneuvojia ei ole paikalla.

> ”Meidän opiskelijat on yleensä aktiivisempia silloin, kun me ei olla auki, eikä meillä myöskään ole mitään mahdollisuutta, että meillä olisi ihminen vastaamassa chatteihin 24/7.”

Kansainvälisten opiskelijoiden näkökulmakin huomioitiin, eli esimerkiksi aikaerojen takia palvelua on hyvä olla saatavilla ympäri vuorokauden. Lisäksi haastatteluissa kävi ilmi organisaation halu olla läsnä siellä, missä asiakkaat eli opiskelijatkin ovat, ja täksi paikaksi on organisaatiossa tunnistettu chatti.

> ”Haluttiin kuitenkin, että ollaan siellä missä meidän opiskelijatkin on, ja opiskelijathan on chateissa, niin haluttiin sitten tarjota palvelua siellä, missä sitä halutaan.”

Chatbotti nähtiin tietyssä määrin myös verkkosivujen hakutoiminnon korvikkeena: moneen kysymykseen, johon chatbotti osaa vastata, löytyy vastaus myös organisaation verkkosivuilta, ja chatbotti helpottaa oikean tiedon löytymistä. Chatbotilla haluttiin palvella asiakkaiden erilaisia tapoja etsiä informaatiota: osa asiakkaista kokee helpoimmaksi hakea tietoa suoraan verkkosivuilta, osa haluaa hoitaa asiansa puhelimitse ja osa taas kokee miellyttävimmäksi ottaa yhteyttä chatbotin kautta.

> ”Ihmisillä on kauheasti eri tapoja etsiä sitä informaatiota. Toiset selaa viimeiseen asti verkkosivuja ennen kuin ne kysyy chatbotilta tai soittaa asiakaspalveluun.”

> ”Että jos siellä yrittää hakea jotain, niin se chatbotti voisi myös vähän sitä korvata, että tavallaan se on yksi hakuoptio, joka voisi olla siellä verkkosivuilla.”

> ”Tuntuu siltä, että kun sitä tietoa on nykyään niin massiivisia määriä, ja sitä on joka puolella, niin tämä botti sitten tekisi sulle sen etsintätyön siinä, että sun ei itse tarvitse tietää, missä se on, vaan että se kertoo sulle sitten.”

5.2.2 Generatiivisesta chatbotista haetut hyödyt

Haastatteluissa selvitettiin, mitä hyötyjä chatbotin uudistamisesta generatiivisen tekoälyn avulla haetaan. Yhtenä isona huomiona haastateltavat nostivat esiin chatbotin kouluttamiseen käytetyn ajan. Tällä hetkellä, kun chatbotti on ollut jo useamman vuoden käytössä, sen kouluttamiseen käytetty aika koetaan suhteettoman suureksi. Kouluttaminen myös koetaan työlääksi.

> ”Se on kuitenkin aika iso panostus, kun puhutaan palvelusta, joka on ollut käytössä jo pitkään, että sellainen määrä työtä toki on ihan asiallista silloin kun palvelu otetaan käyttöön, mutta kun niitä periaatteessa samoja asioita on jo koulutettu vuosikausia niin voisi ajatella, että sen perusylläpitoon vaadittava tuntimäärä on kohtuuton tällä hetkellä.”

> ”Sitä on liian paljon pitänyt kouluttaa mun mielestä, siihen hyötyyn nähdä mikä siitä on tullut, että melkein siinä ajassa tai sen ajan, minkä ihminen käyttää kouluttamiseen, niin olisi sitten palvellut näitä kyselijöitäkin, kärjistäen.”

Chatbotin uudistamisella tavoitellaan entistä tehokkaampaa apuria asiakaspalvelulle eli sitä, että chatbotti kykenisi hoitamaan suuremman osan asiakkaiden kysymyksistä, jolloin palveluneuvojilla olisi paremmin aikaa palvella niitä asiakkaita, jotka tarvitsevat henkilökohtaista ohjausta ja neuvontaa. Lisäksi halutaan, että chatbotti osaa vastata kysymyksiin entistä paremmin ja laadukkaammin. Chatbotin uudistamisella tavoitellaan sitä, että asiakaspalveluun päätyy vain sellaiset asiakkaat, joiden kysymyksiä tai tarpeita ei pystytä automaattisesti käsittelemään.

Haastatteluissa kävi myös ilmi, että nykyisen chatbotin teknologiaa pidetään vanhentuneena, ja tiedostetaan, että uusia, parempia teknologioita on markkinoilla saatavilla. Organisaatiossa on halu pysyä kehityksessä mukana, ja uudemman teknologian toivotaan auttavan siinä, että chatbotti ymmärtää kokonaisuuksia paremmin ja pystyy käymään luonnollisempaa ja henkilökohtaisempaa keskustelua asiakkaiden kanssa, esimerkiksi kysymällä jatkokysymyksiä. Organisaatiossa koetaan, että tapa, jolla asiakkaat käyttävät chatbottia, on uusien, generatiivisten chatbottien myötä muuttunut, ja näihin uusiin käyttötarpeisiin halutaan vastata. Lisäksi nykyisen chatbotin ylläpitoon tarvitaan lukuisia eri järjestelmiä, ja näiden järjestelmien määrää koetaan tarpeelliseksi vähentää.

> ”Ja osaisi itse kysyä lisäkysymyksiä myös, että [kun asiakas kysyy] ”voinko hakea asuntoa”, niin sitten se [chatbotti] kysyisi, että, no riippuu vähän, että opiskeletko pääkaupunkiseudulla, onko se tutkintoon johtava koulutus? Ja sitten se [asiakas] vastaisi siihen. Että se olisi oikeasti keskustelu.”

> ”Nykyään ne kysymyksetkin on pidempiä kun ehkä odotetaan enemmän sellaista ChatGPT-tyyppistä vastaajaa, että on jo totuttu siihen tekoälyn kanssa keskusteluun niin sitten ehkä odotetaankin, että se pystyisi jo kertomaan laajemmin ja tarkemmin.”

5.2.3 Chatbotin käyttötapaukset

Haastatteluissa selvitettiin chatbotin käyttötapauksia. Tällä hetkellä asiakkaat ovat asiakaspalveluun eniten yhteydessä hakemiseen, asumiseen, elämäntilanteen muutoksiin, sähköisiin palveluihin, autopaikkoihin, asunnon kuntoon, vikailmoituksiin ja avaimiin liittyvistä asioista. Etenkin hakeminen, ja siihen liittyvä iso kokonaisuus, koettiin yleiseksi syyksi olla asiakaspalveluun yhteydessä. Hakemisen kokonaisuuteen liittyvät kysymykset siitä, missä vaiheessa hakemus on, miten asuntoa voi hakea, milloin sitä voi hakea ja milloin asunnon saa. Haastatteluissa huomioitiin myös kysymysten kausiluontoinen vaihtelu: esimerkiksi kuun alussa asiakaspalveluun tulee enemmän kysymyksiä avaimiin liittyen. Lisäksi koettiin, että yhteydenotot jakautuvat selkeästi henkilökohtaisiin kysymyksiin, jotka liittyvät asiakkaan tilanteeseen ja henkilökohtaisiin tietoihin, ja niin sanottuihin yleisiin kysymyksiin, joiden vastaukset ovat käytännössä samat kaikille kysyjille.

> ”Hakemuksista. Ehdottomasti. Ja sitten se riippuu myös vähän, että mihin aikaan kuukaudesta se on, koska kuun alussa sitten taas asiakkaat on joko avaimista, asunnon siisteydestä, asunnon yleisestä kunnosta.”

> ”Ja sitten, jos miettii että ei ole hakemuksesta, niin sitten että milloin he saa sitten sen asunnon.”

Haastateltavilta kysyttiin siitä, mihin kysymyksiin chatbotti tällä hetkellä osaa vastata, ja mitkä kysymykset ohjautuvat palveluneuvojalle. Tällä hetkellä kaikki kysymykset, jotka koskevat asiakkaan henkilökohtaisia asioita, ohjautuvat aina palveluneuvojalle. Chatbotti osaa vastata vain yleisiin kysymyksiin, joihin lähtökohtaisesti löytyy tiedot myös organisaation verkkosivuilta. Jos asiakas haluaa tietoa omasta tilanteestaan, hänen tulee keskustella palveluneuvojan kanssa. Kun chatti on siirtynyt palveluneuvojalle, myös vahva tunnistautuminen on mahdollista, jolloin asiakaspalvelija tietää, kenen kanssa keskustelee, ja voi kertoa myös asiakkaan henkilökohtaisia tietoja esimerkiksi hakemuksen tilasta. Haastatteluissa huomattiin myös monen asiakkaan taipumus pyytää chatissa suoraan palveluneuvojaa paikalle, jolloin chatbotti ohitetaan käytännössä kokonaan.

> ”Siis ne on oppinut nyt sen, että ne pystyy ohittaa sen chatbotin sillä tavalla, että ne kysyy pystyykö puhumaan ihmiselle tai fyysiselle henkilölle, niin se ohittaa sen saman tien, että eihän ne oo ehtinyt käydä mitään keskusteluja.”

Tällä hetkellä chat-keskustelu siirtyy botilta palveluneuvojalle, kun joko asiakas itse pyytää palveluneuvojaa tai chatbotti ei enää osaa itse vastata, ja tarjoaa siirtoa palveluneuvojalle. Ensin chatbotti tarkistaa, onko palveluneuvojia paikalla. Jos palveluneuvoja on paikalla, keskustelu yhdistetään asiakaspalvelijalle.

5.2.4 Asiakaspalvelussa käytettävissä oleva materiaali

Haastateltavilta kysyttiin palveluneuvojilla käytössään olevasta materiaalista kysymyksiin vastaamisen tueksi. Organisaatiolla on paljon erilaista materiaalia palveluneuvojien työn tekemisen tueksi, mutta sen koettiin olevan monessa eri paikassa, ja oikean tiedon löytyminen koettiin vaikeaksi. Tietoa on saatavilla muun muassa organisaation julkisilla verkkosivuilla, sisäisesti on olemassa erilaisia valmiita viestipohjia, minkä lisäksi materiaalia löytyy verkkolevyiltä ja Teamsistä. Lisäksi työntekijöiden henkilökohtaisilla työasemilla saattaa olla materiaalia. Kollegoiden tuki mainittiin useammassa haastattelussa, ja hiljaisen tiedon olemassaolostakin ollaan tietoisia, ja tämän tiedon dokumentoimista varten on tehty myös toimenpiteitä.

> ”Meillä on Teamsissa, meillä on verkkolevyillä, ja näin päin pois, ja meidän nettisivuilla. Ja meidän verkkolevyllä on aikamoinen viidakko.”

Palveluneuvojat myös hakevat tietoa eri järjestelmistä, kuten asiakkuus- ja kommunikaatiojärjestelmistä. Tiedon koettiin olevan hajallaan eri puolilla, ja sen järkeistäminen nähtiin tärkeäksi, myös

chatbotin uudistamisen kannalta. Organisaation tavoitteena on, että kaikki materiaali saataisiin yhteen paikkaan ja helposti löydettäviksi.

> ”Se on semmoinen työn alla oleva asia, että ne saataisiin yhteen paikkaan ja helposti löydettäväksi, että nyt jos yhtäkkiä pitäisi lähteä etsimään jotain ohjetta, niin ensin pitää miettiä, että oliko se nyt tuolla meidän verkkoasemalla vai onko se Teamsissa. Ja sitten jos se on Teamsissa, niin missä tiimissä ja millä kanavalla se nyt olikaan.”

5.2.5 Nykyisen chatbotin haasteet

Ongelmalliseksi nähtiin jonotusjärjestelmän puute. Tällä hetkellä chatissä ei ole minkäänlaista jonotusjärjestelmää, eli kaikki chatbotilta siirtyvät chatit menevät suoraan palveluneuvojalle. Tämä koettiin hieman hankalaksi, jos chatin kautta tulee samanaikaisesti paljon yhteydenottoja. Jonotusjärjestelmän avulla voitaisiin kertoa asiakkaille, jos palveluneuvojat ovat kiireisiä, ja palveluneuvojat voisivat paremmin keskittyä kesken oleviin keskusteluihin.

> ”Yhdelle henkilölle voisi vaikka maksimissaan tulla ehkä 2 chattia, että kun ne voi olla joskus vähän sellaisia vaativampia, niin sitten ei pysty keskittymään siihen niin täysillä, kun sitten on se viisi chattia auki.”

Haastatteluissa kävi myös ilmi, että kun organisaatiossa on otettu WhatsApp yhtenä yhteydenottokanavana käyttöön, chatin läpi tulevien yhteydenottojen määrä on vähentynyt. Vaikka organisaatiolla on käytössä WhatsApp, chatbottia ei ole yhdistetty siihen. Haastatteluissa koettiin, että WhatsAppin chattiä voitaisiin parantaa lisäämällä edes joitakin automaattivastauksia. Organisaatiolla on käytössä myös Facebook Messenger yhteydenottokanavana, ja tämän kanavan olemassaoloa kyseenalaistettiin. koska Messengerin kautta ei juuri koskaan tule yhteydenottoja.

> ”Ja Messengeriä ei käytä kukaan. En ole ikinä nähnyt, että sieltä olisi tullut mitään viestiä.”

Tällä hetkellä chatin poikkeusaukioloaikojen säätäminen on vaikeaa, ja vaatii monta manuaalista vaihetta sekä ennen että jälkeen poikkeusaukioloajankohdan. Chattiin toivottiin ominaisuutta, josta sen voisi yksinkertaisesti laittaa väliaikaisesti pois päältä, tai sitä, että chatbotti automaattisesti tietäisi, kun palveluneuvojia ei ole paikalla. Lisäksi mahdollisuus muotoilla viesti, jonka chatbotti kertoisi asiakkaalle poikkeusaukioloaikojen syyksi, koettiin hyödylliseksi.

> ”Että sen voi laittaa sen palveluneuvojan saatavuuden jotenkin helpommin päältä ja pois päältä, ja sitten se osaisi se botti sanoa, että valitettavasti nyt ei ole ketään.”

Haastatteluissa nostettiin esille, että chatbotti ei osaa vastata esimerkiksi kiinteistökohtaisiin kysymyksiin tai sähköisten palveluiden vikatilanteisiin liittyviin kyselyihin. Haastatteluissa kävi myös ilmi, että henkilöt, joilla ei ole suomalaista henkilötunnusta tai pankkitunnuksia, eivät tällä hetkellä voi käyttää chattiä henkilökohtaisten asioiden hoitamiseen, koska he eivät voi tunnistautua vahvasti palveluun. Lisäksi huomioitiin, että tällä hetkellä chatbotti antaa paljon pitkiä, geneerisiä vastauksia

kysymyksiin, ja peräänkuulutettiin parempia keskustelutaitoja, esimerkiksi lisäkysymysten esittämistä keskustelun jatkamiseksi.

Yleisesti chatbotti nähtiin kuitenkin hyödyllisenä apuna asiakaspalvelulle, joka osaa vastata yleisimpiin yksinkertaisiin kysymyksiin. Haittaa chatbotista ei nähty olevan, pienikin apu koettiin hyödylliseksi. Parantamisen tarve tuli ilmi: enemmän hyötyä chatbotista olisi, jos se osaisi vastata paremmin yleisimpiin kysymyksiin.

> ”Ja muut, yleiset kysymykset osaa siis tosi hyvin, ja nehän on kaikki myöskin sellaisia kysymyksiä, mistä oikeasti löytyisi kyllä vastaukset meidän nettisivuilta, mutta helpompi kysyä sitten chatissa, kun lähteä katsomaan sieltä nettisivuilta.”

> ”Jos se chatti ei tule meille, niin tietenkin se on hyödyllistä, että se chatbotti pystyy vastaamaan niihin asioihin.”

5.2.6 Chatbotin toiminnan ja tavoitteiden toteutumisen seuraaminen

Nykyisen chatbotin toimintaa seurataan lähinnä käymällä keskusteluja läpi säännöllisesti. Ne keskustelut, joihin chatbotti ei ole osannut vastata, tarkistetaan, ja tarvittaessa korjataan chatbotin toimintaa. Lisäksi chatbotin toimintaa seurataan säännöllisissä palavereissa chatbotin toimittajan kanssa, missä käydään läpi analytiikkaa siitä, kuinka chatbotti on menestynyt, mitä virhetilanteita on ollut ja mitä voisi tehdä seuraavaksi. Lisäksi tarkastellaan sitä, kuinka suureen osaan kysymyksistä chatbotti on osannut vastata, ja kuinka suuren osan se on siirtänyt palveluneuvojalle. Lisäksi asiakaspalvelussa seurataan yhteydenottoja asiasanojen perusteella.

> ”Käydään läpi botin käymiä keskusteluja säännöllisesti, niistä seurataan.”

> ”Raportteja seurataan, ja että miten hän siellä toimii, tuleeko paljon vääriä vastauksia, tarvitseeko esimerkiksi sitten jotain koulutusta, tuleeko kuinka paljon siirtoja palveluneuvojalle.”

Selkeitä chatbotin tavoitteita seuraavia mittareita ei organisaatiossa ollut asetettu, vaan tavoitteiden toteutumista seurataan lähinnä chatbotin toimintaa tarkastelemalla.

> ”Kai ne on tämänhetkisessä chatbotti-järjestelmässä, että kuinka moneen se osaa vastata, kuinka moneen ei, kuinka paljon se siirtää ihmiselle.”

Chatbotista kerätään palautetta kyselylomakkeella, jota tarjotaan asiakkaalle chatbotin käyttämisen jälkeen. Tämä kysely ei kuitenkaan ole ollut kovin suosittu, eikä siis kerää paljoa vastauksia, eikä muita palautekanavia ole. Avoimia palautteita voi aina laittaa, ja asiakastyytyväisyyskyselyissä chatbottia voi yleisellä tasolla kommentoida. Organisaatiossa ei kuitenkaan kovin aktiivisesti seurata asiakkaiden kokemuksia chatbotin käytöstä.

> ”Joo, kerätään palautetta. Semmoinen kyselylomake, mikä aukeaa siihen, että voi täyttää sen käyttämisen jälkeen.”

> ”Sen chatbotin ohessa on se kysely, mutta siihen vastataan kerran viikossa tyyliin.”

5.2.7 Chatbotin ylläpito

Chatbotin ylläpito koetaan hankalaksi. Kokonaisuutta on vaikea hallita, ja uusien ominaisuuksien ja chatbotin kehittymisen myötä nähdään, että uusia ominaisuuksia on rakennettu vanhojen päälle tavalla, joka ei välttämättä ole kannattavaa. Eri tilanteita varten on rakennettu erilaisia järjestelmiä, joiden hallinta koetaan vaikeaksi. Yleisesti chatbotin kouluttaminen koetaan todella työlääksi ja hitaaksi. Lisäksi chatbotin kaksikielisyys aiheuttaa ylläpidon kannalta ongelmia, koska kaikki muutokset pitää tehdä sekä suomeksi että englanniksi. Lisäksi palvelun toiminnassa on toisinaan havaittu vikoja, mutta koska tiedostetaan, että järjestelmää ollaan parantamassa, niin katse on jo suuntautunut tulevaisuuteen.

> ”Se ei ole kauhean helppo kokonaisuus, että mitä mä voin ottaa pois, mitä voin jättää pois.”

> ”Työlästä. Ja esimerkiksi juuri tällä hetkellä ei toimi.”

Ylläpidon parantamiseksi haastatteluissa ehdotettiin isompien massojen läpikäymistä kerralla. Lisäksi chatbotti täytyy tallentaa ja julkaista erikseen, minkä takia julkaisu unohtuu helposti. Julkaisemisessa saattaa myös kestää todella kauan, jos chatbottiin on tehty paljon muutoksia. Erilliset kieliversiot ovat myös ongelmallisia, koska tällä hetkellä eri kieliversiot tarkoittavat käytännössä kahta eri chatbottia, joita molempia täytyy erikseen päivittää. Lisäksi visuaalista työkalua keskustelupolkujen kehittämiselle toivottiin, mitä nykyinen järjestelmä ei tarjoa.

> ”Ehkä joku semmoinen, että voisin käydä ison massan kerralla läpi ja korvamerkitä ne, jotka tarvitsee toimenpiteitä. Ja sitten palata niihin korvamerkittyihin.”

5.2.8 Tietoturva ja tietosuoja

Haastatteluissa selvitettiin organisaation tietoa chatbotin tietoturva- ja tietosuojavaatimuksille. Organisaatiossa tunnistetaan, että asiakkaat helposti antavat henkilötietojaan tai kertovat hyvinkin henkilökohtaisia asioita chatbotille. Asiakkaan virheen ei haluta aiheuttavan riskiä asiakkaan tai organisaation tietoturvalle tai -suojalle. Haastateltavat kokivat tärkeänä, että asiakkaan tiedot eivät päätyisi esimerkiksi isojen kielimallien koulutusaineistoon, ja että tiedot pysyisivät EU:n sisäpuolella. Tietoturvallista ja tietoja suojaavaa chatbottijärjestelmää pidettiin tärkeänä. Haastatteluissa noteerattiin myös ihmisen valvonnan tärkeys, ja se, että chatbotilla ei tule olla pääsyä kaikkein arkaluontoisempiin paikkoihin, vaikka tarkoitus olisikin, että se pystyisi jakamaan ja etsimään tietoa laajasti. Pääsyä esimerkiksi työntekijöiden sähköposteihin ei tietoturvasyistä koeta hyvänä asiana. Organisaatiossa koetaan tärkeäksi, että vaikka chatbotti pystyisi antamaan henkilökohtaista palvelua etsimällä asiakastiedoista tietoa, se ei eksyisi väärään paikkaan.

> ”Haluttaisiin huolehtia siitä, että vaikka käyttäjä menisi laittamaan sinne henkilötietonsa, niin henkilötiedot eivät valuisi minnekään koulutusaineistoihin.”

> ”Ihmisillä tuntuu olevan tapana syöttää sinne esimerkiksi henkilötunnuksia tai saattavat kertoa hyvinkin henkilökohtaisia asioita sille chatille, niin tietysti me halutaan, että ne heidän tiedot on turvassa.”

Organisaatiolla on lupa käsitellä henkilötietoja, mutta niitä käsitellään vain organisaation sisäisesti, eikä luovuteta kolmansille osapuolille. Tällä hetkellä tietoja käsitellään organisaation tietosuojaselosteen mukaisesti, ja organisaation IT-osasto järjestää uusien järjestelmien tietoturva- ja tietosuoja-auditoinnit.

5.2.9 Lainsäädäntö

Chatbotin lainsäädännöllisistä vaatimuksista haastateltavat tunnistivat yleisen tietosuoja-asetuksen ja tekoälysäädöksen merkityksen. Organisaatiossa halutaan olla tekoälysäädöksen kannalta käyttöönottajan asemassa, ja koetaan, että jos tiedot pysyvät EU:n sisällä, eikä organisaatio tee profilointeja tekoälyn avulla, niin lainsäädännölliset vaatimukset tulee ainakin suurin piirtein täytetyiksi. Haastatteluissa huomioitiin myös tulevien tiukennuksien mahdollisuus, kun tekoälyä koskeva kansallinen lainsäädäntö tarkentuu.

> ”Chatbotin kannalta se tietosuoja pysyy ennallaan. Se ei muuta sitä tietosuojalakia tai -asetusta mitenkään, että kaikki siinä pysyy ennallaan. Se on vain yksi sovellus.”

> ”Toki GDPR:n mukaan pitää tehdä, ja sitten myös tekoälysäädöksen mukaisesti, että nää nyt molemmat pitää ottaa huomioon.”

5.2.10 Muita huomioita

Haastateltavilta kysyttiin, miten he parantaisivat chatbottia, jos saisivat siihen vapaat kädet. Vastauksissa toistui muutama jo mainittu asia: koulutukseen käytettävän ajan vähentäminen ja paremmat vastaukset asiakkaille. Chatbotin haluttaisiin olevan itseohjautuvampi, ja ihmisen osuuden olevan vähäisempi. Lisäksi chatbotista haluttaisiin älykkäämpi niin, että se kykenisi hahmottamaan isompia kokonaisuuksia, ja pitämään mielessä kerralla edes kaksi asiaa. Keskustelukohtainen muisti mahdollistaisi luontevamman keskustelun. Chatbottiin haluttaisiin myös lisää ohjeita esimerkiksi hakemuksen vikatilanteista.

> ”Jos sä ensin sanot, että oon muuttamassa lemmikin kanssa, mihin mä voin hakea? Ja sitten se [chatbotti] vastaa, että Hoasin asuntoihin kaikkialle muualle paitsi soluihin ja yksiöihin yhteiskeittiöllä jne. voi muuttaa lemmikin kanssa. Ja sitten se [asiakas] kysyy, eli voinko muuttaa soluun, niin tällä hetkellä [chatbotti] vastaisi siihen, että voit muuttaa Hoas-asuntoon, jos olet täysipäiväinen opiskelija pääkaupunkiseudulla jne. Tavallaan, että sillä ei ole mitään sellaista keskustelukohtaista muistia.”

Haastattelujen lopuksi tiedusteltiin avoimia kommentteja aiheeseen liittyen. Chatbotin yhteensopivuus olemassa olevien järjestelmien kanssa nostettiin esille: rajapintayhteys asiakaspalvelujärjestelmään on oleellinen. Chatbotista myös toivotaan helppokäyttöistä, helposti käyttöön otettavaa, ja jouhevasti sisäisesti testattavaa.

”Tämä on ominaisuus, mikä myös haluttaisiin siltä tulevalta chatbotilta, että jonkin verran ratkaisuja on markkinoilla, joissa sitä yhteyttä asiakaspalvelujärjestelmään ei ole, tai he tarjoavat sitä asiakaspalvelujärjestelmää, joka heillä itsellään on. Mutta, että voisi saada rajapinnan toiseen järjestelmään, sitä ei kaikilla ole vielä tarjota.”

”Me haluttaisiin semmoinen chatti, joka olisi helppo käyttää, tai helppo ottaa käyttöön ja helppo testata ensin sisäisesti, ja sitten vasta antaa ulkoisille kohderyhmille käyttöön.”

6 Pohdinta

Opinnäytetyön tavoitteena oli selvittää organisaation valmiuksia generatiivisen chatbotin käyttöönottoon organisaation asiakaspalvelussa. Tutkimuskysymykset olivat (1) mitä chatbotin uudistamisella tavoitellaan, (2) miten generatiivinen chatbotti otetaan onnistuneesti käyttöön osana organisaation asiakaspalvelua, (3) miten varmistutaan generatiivisen chatbotin luotettavuudesta, tietoturvasta ja eettisyydestä sekä (4) mitä vaatimuksia lainsäädäntö asettaa generatiiviselle chatbotille. Johtopäätöksissä vastataan näihin tutkimuskysymyksiin. Toimeksiantajaorganisaatio voi hyödyntää opinnäytetyön tuloksia, kun se etenee generatiivisen chatbotin käyttöönoton prosessissa. Tuloksista on myös hyötyä palveluntarjoajan valinnassa. Kaikkinensa opinnäytetyö lisää organisaation valmiuksia hyödyntää generatiivista tekoälyä toiminnassaan ja tuo ilmi kaikki ne moninaiset teemat ja aiheet, joita organisaation tulee huomioida generatiivisen chatbotin käyttöönotossa.

Johtopäätösten jälkeen käsitellään tutkimuksen luotettavuutta sekä pohditaan ja arvioidaan opinnäytetyöprosessin etenemistä.

6.1 Johtopäätökset

Seuraavaksi esitellään opinnäytetyön johtopäätökset ja toimenpidesuositukset organisaatiolle. Johtopäätökset käsittelevät generatiivisen chatbotin käyttöönottoa ja kehittämistä monesta eri näkökulmasta. Koska organisaatio on tekoälyn käyttöönottaja, johtopäätöksissä korostetaan myös palveluntarjoajan valintaan liittyviä seikkoja, joilla voidaan vaikuttaa lopullisen chatbottitoteutuksen laatuun.

6.1.1 Tavoitteiden asettaminen

Chatbotin käyttöönotossa on oleellista pohtia tavoitteita, joita organisaatiolla on chatbotille. Aikaisemmassa kirjallisuudessa (esim. Skuridin & Wynn 2024) korostetaan, että chatbotin tulisi parantaa tehokkuutta tai ongelmanratkaisua, ja vähentää merkittävästi työntekijöiden toistuviin työtehtäviin käyttämää aikaa. On syytä pohtia perusteellisesti, onko generatiivinen chatbotti varmasti paras mahdollinen ratkaisu ongelman ratkaisemiseksi vai voisiko ongelman ratkaista myös perinteisemmillä teknologioilla.

Haastatteluissa nousi esiin, että organisaatiolla on ollut jo nykyisen chatbotin käyttöönottoon selkeät tavoitteet: asiakaspalvelun työkuorman vähentäminen sekä resurssien käytön optimoiminen. Myös ympärivuorokautinen asiakaspalvelu korostui, sekä halu palvella asiakkaita niissä kanavissa, joita he käyttävät. Chatbotin avulla ratkaistaan asiakaspalvelun konkreettisia haasteita, ja chatbotin uudistamisella pyritään vahvistamaan näitä hyödyllisiä vaikutuksia. Generatiivisella tekoälyllä halutaan tehostaa chatbotin toimintaa.

Aikaisemman tutkimuksen (esim. Skuridin & Wynn 2024) perusteella tavoitteiden asettamisen lisäksi on tärkeää asettaa myös mittareita tavoitteiden seuraamista varten. On suositeltavaa, että organisaatio ottaa käyttöön konkreettisia mittareita generatiivisen chatbotin tavoitteiden saavuttamisen seuraamiselle. Organisaatio voi asettaa mittareita esimerkiksi sille, kuinka suuren osuuden keskusteluista chatbotti kykenee ratkaisemaan ilman ihmisen väliintuloa, ja kuinka usein chatbotti ei osaa vastata asiakkaiden kysymyksiin. Lisäksi yhtenä mittarina organisaatio voi seurata esimerkiksi chatbotin kouluttamiseen käytettyä työaikaa. Näille mittareille on hyvä asettaa konkreettiset tavoitteet, jotta nähdään, onko generatiivisesta chatbotista todellista hyötyä organisaation tavoitteiden saavuttamisessa.

Kirjallisuudessa nousi vahvasti esiin myös palveluntarjoajan rooli organisaation tavoitteiden saavuttamisessa (esim. World Economic Forum 2023; Skuridin & Wynn 2024). Palveluntarjoajan valinnassa on syytä kiinnittää huomiota siihen, että palveluntarjoaja ymmärtää organisaation liiketoimintatavoitteet. Eri palveluntarjoajia kannattaa vertailla ja kilpailuttaa, ja valita toimittaja, joka parhaiten vastaa organisaation vaatimuksiin. Palveluntarjoajilta voi myös pyytää referenssejä tai suosituksia palveluntarjoajan asiakkailta. Palveluntarjoajan sitoutuminen organisaation liiketoimintatavoitteiden saavuttamiseen edesauttaa onnistunutta kumppanuutta. On myös hyvä pitää mielessä, että chatbottia tulee voida kehittää organisaatiossa sisäisesti tai palveluntarjoajaa vaihtaa tarvittaessa.

6.1.2 Generatiivisen chatbotin kouluttaminen, toiminnan seuraaminen ja ylläpito

Haastatteluissa nousi esiin, että chatbotin kouluttaminen nähdään tällä hetkellä aikaa vieväksi ja vaikeaksi. Generatiivisen tekoälyn avulla chatbotin kouluttaminen muuttuu, mutta lähtökohtaisesti on vaikea ennustaa, tuleeko chatbotin kouluttamiseen käytettävä aika muuttumaan merkittävästi. Etenkin alkuvaiheessa kouluttamiseen on investoitava, ja koska generatiivisen tekoälyn tuottamaa sisältöä tulee seurata, ja tekoälyä kouluttaa, on siihen jatkossakin varattava resursseja. On suositeltavaa, että koulutukseen käytettävää aikaa seurataan ja kouluttamisen vähentämiselle asetetaan konkreettisia tavoitteita, joiden saavuttamista seurataan, kun generatiivinen tekoäly otetaan käyttöön.

Haastatteluissa selvisi, että tällä hetkellä organisaatiossa seurataan chatbotin toimintaa lähinnä seuraamalla analytiikkaa. Luvut siitä, kuinka monta keskustelua chatbotti kykenee hoitamaan itsenäisesti, ja kuinka monta keskustelua siirtyy palveluneuvojalle, ovat olennaisessa asemassa. Tämä on organisaatiolle hyvä lähtökohta chatbotin toiminnan seuraamiselle, mutta generatiivisen tekoälyn myötä chatbotin toimintaan on hyvä kiinnittää enemmän huomiota.

Aikaisemmassa kirjallisuudessa korostetaan generatiivisen chatbotin toiminnan seuraamisen tärkeyttä muun muassa puolueellisuuden ja virheellisen sisällön tunnistamisen takia (esim. Skuridin &

Wynn 2024; Sidaoui ym. 2024). Generatiivisen chatbotin käymiä keskusteluja tulee seurata, ja ne pyynnöt, joita se ei osaa käsitellä, tulee tunnistaa. Chatbotin täsmällisyyttä, luotettavuutta ja puolueetonta päätöksentekoa on olennaisen tärkeää seurata, kun generatiivista tekoälyä hyödynnetään. Chatbotin toiminnan seuraaminen tulisi tehdä mahdollisimman helpoksi ja yksinkertaiseksi, esimerkiksi rakentamalla työpöytä (dashboard), jolta chatbotin toimintaa pystyy seuraamaan myös ei-teknisesti orientoituneet ihmiset. Palveluntarjoajan valinnassa kannattaa kiinnittää huomiota siihen, että palveluntarjoaja kykenee tarjoamaan yksityiskohtaista tietoa generatiivisen chatbotin toiminnasta.

Generatiivisen chatbotin käyttöä ja suorituskykyä on seurattava jatkuvasti, ja sen käymiä keskusteluja on monitoroitava. Keskusteluista on tunnistettava sellaiset pyynnöt, joita chatbotti ei ole osannut käsitellä, ja kouluttaa sitä toimimaan oikein. Ylläpitoon kuuluu myös chatbotin toiminnallisuuden jatkuva parantaminen uusimpia tekoälyteknologioita hyödyntäen ja kielimalleja päivittäen.

Määrällisen tiedon lisäksi organisaation tulee panostaa myös laadullisen tiedon keräämiseen. Aikaisemmissa tutkimuksissa on todettu, että palautteen kerääminen on olennainen osa generatiivisen chatbotin toiminnan seuraamista (esim. Skuridin & Wynn 2024; Urbani ym. 2024; Sidaoui ym. 2024). Haastatteluissa kävi ilmi, että organisaatiossa ei kovin aktiivisesti kerätä palautetta nykyisen chatbotin toiminnasta, eikä chatbotin yhteydessä oleva palautekysely kerää juurikaan vastauksia. Organisaation tuleekin pohtia, miten nykyistä palautekyselyä voidaan parantaa niin, että siitä on enemmän hyötyä. Käyttäjien palautetta kannattaa hyödyntää uusien toiminnallisuuksien suunnittelussa. Myös käyttäjäkokemusta tulee testata, jotta voidaan varmistua siitä, että chatbotti vastaa organisaation asiakaskunnan tarpeisiin.

Aikaisemmassa tutkimuksessa on havaittu myös chatbotin testaamisen tärkeys (esim. Skuridin & Wynn 2024). Chatbottia kannattaa testata iteratiivisesti jokaisen uuden toiminnallisuuden jälkeen. On suositeltavaa, että testausta suoritetaan ensin sisäisesti tiimissä ja liiketoiminta-asiantuntijoiden kanssa sekä pienellä oikeiden asiakkaiden pilottiryhmällä hyödyntäen oikeaa dataa ennen laajamittaista käyttöönottoa.

6.1.3 Käyttötapaukset ja vuorovaikutuksen tasapainottaminen

Haastatteluissa havaittiin, että organisaatio on tunnistanut paljon käyttötapauksia, joissa chatbotti voi olla hyödyksi, ja näiden lisäksi organisaatiolla on selkeä näkemys siitä, missä käyttötapauksissa tarvitaan palveluneuvojaa. Chatbotti hoitaa yleistason kysymyksiä, ja palveluneuvojat tarjoavat henkilökohtaista palvelua niille asiakkaille, jotka sitä todella tarvitsevat. Käyttötapaukset kannattaa kirjata ylös, jotta niitä on helpompi hallinnoida. Käyttötapausten määrittelyn avulla voidaan myös tehdä näkyväksi, missä vaiheessa tarvitaan ihmisen väliintuloa. Käyttötapauksissa tulee

keskittyä ensisijaisesti rajattujen, mutta laadukkaiden toiminnallisuuksien tarjoamiseen, joten liian isoista kokonaisuuksista ei kannata aloittaa.

Aikaisemmissa tutkimuksissa on huomattu, että chatbotin käyttöönotto voi joissain tilanteissa myös heikentää asiakaspalvelun laatua (esim. Ferraro ym. 2024; Sidaoui ym. 2024). On suositeltavaa, että chatbotti otetaan käyttöön vain arkipäiväisissä ja rutiininomaisissa tehtävissä, ja hankalammat kysymykset siirretään palveluneuvojien käsiteltäviksi. Ihmisten välinen vuorovaikutus on edelleen olennainen osa asiakaspalvelua. Haastatteluiden perusteella organisaatiolla on tällä hetkellä hyvä suhtautuminen ihmisen ja tekoälyn tasapainottamiseen asiakaspalvelussa. Chatbotti tukee palveluneuvojien työtä, mutta ei korvaa sitä. Tämä on olennaista, sillä liika automaatio saattaa myös heikentää asiakaskokemusta. Kun generatiivisen tekoälyn mahdollisuuksia organisaatiossa pohditaan, ja muita kehityskohteita mietitään, on hyvä pitää mielessä, että ihmisyys on olennainen osa asiakaspalvelua.

6.1.4 Yhteensopivuus ja tiedonhallinta

Aikaisemmassa kirjallisuudessa korostetaan yhteensopivuuden merkitystä chatbottia kehitettäessä (esim. World Economic Forum 2023; Skuridin & Wynn 2024). Haastattelujen perusteella organisaatiolla on haasteita nykyisen chatbotin teknologisessa yhteensopivuudessa organisaation muiden tietojärjestelmien kanssa. Palveluntarjoajaa valittaessa tuleekin kiinnittää erityistä huomiota siihen, että uusi chatbottialusta tarjoaa kattavat integraatiomahdollisuudet organisaation olemassa oleviin järjestelmiin. Yhteensopivuuden varmistamiseksi on hyvä kartoittaa nykyisen chatbotin ongelmakohdat, kuten haastatteluissa esiin nousseet haasteet poikkeusaukioloaikojen muuttamisessa sekä WhatsAppin automaattiviesteissä, jotta vastaavilta ongelmilta voidaan tulevaisuudessa välttyä.

Aikaisemmassa tutkimuksessa todetaan myös, että yhteensopivuus liittyy muuhunkin kuin pelkästään teknologiseen yhteensopivuuteen (esim. Urbani ym. 2024). Yhteensopivuus organisaation tavoitteiden ja arvojen kanssa on myös tärkeässä osassa kokonaisvaltaista yhteensopivuutta. Organisaatiolla on kuitenkin jo käytössä chatbotti, joten uudistamisen yhteydessä tulee lähinnä miettiä generatiivisen tekoälyn vaikutuksia. Organisaation on hyvä tunnistaa ja määritellä ne arvot ja periaatteet, joiden perusteella tekoälyä halutaan organisaatiossa hyödyntää.

Haastatteluissa havaittiin, että organisaatiossa oleva tietoaineisto koetaan tällä hetkellä varsin organisoimattomaksi, eikä organisaatiossa ole yleisiä käytäntöjä tietojen tallentamiselle. Kuitenkin aiemmassa kirjallisuudessa korostetaan, että generatiivinen chatbotti vaatii toimiakseen laadukasta tietoaineistoa (esim. World Economic Forum 2023). Organisaation tulee panostaa siihen, että tietojen hallinta on kunnossa. Tietojen tulisi olla täsmällistä, kokonaista, johdonmukaista ja ajan tasalla.

Tiedonhallinnan tehostamiseksi kannattaa laatia tiedonhallintastrategia, jossa määritellään myös tiedonhallintaan liittyvät vastuut ja eettiset käytännöt sen varmistamiseksi, että tieto on laadukasta, käytettävää ja luotettavaa.

Asiakaspalvelun chatbotilla ei tarvitse olla laajaa tietämystä kaikista mahdollisista aiheista, vaan kontekstisidonnainen tieto riittää. Kouluttamisessa tulee hyödyntää vain olennaisia tietolähteitä, kuten aiempia asiakaspalvelukohtaamisia, organisaation omia nettisivuja ja sisäisiä tietokantoja. Suuri osa tekoälyjärjestelmien kustannuksista liittyy tiedon valmisteluun tekoälyjärjestelmää varten. Palveluntarjoajalta on hyvä varmistaa, kuinka paljon muutosten tekeminen olemassa oleviin järjestelmiin tai infrastruktuuriin tulee maksamaan.

6.1.5 Tietoturva ja tietosuoja

Haastattelujen perusteella organisaatiolla on tietoturvaa ja eettisyyttä korostava lähtökohta generatiivisen chatbotin kehittämiselle ja käyttöönotolle. Organisaatiossa tunnistetaan asiakkaiden taipumus antaa henkilökohtaisia tietojaan chatbotille, ja näissäkin tapauksissa asiakkaan henkilötietojen suojaaminen on organisaation prioriteettina. Asiakkaiden tietojen ei tule päätyä isojen kielimallien koulutusaineistoon, ja tietojen pysyminen EU:n alueella on tärkeää. Organisaatiossa käsitellään henkilötietoja sisäisesti, eikä niitä luovuteta kolmansille osapuolille. Tietoturva- ja tietosuoja-auditointeja järjestetään organisaation uusille järjestelmille.

Aikaisempi kirjallisuus korostaa palveluntarjoajan roolia tietoturva-asioissa, sillä chatbotin teknologiset ominaisuudet määrittävät pitkälti sen, kuinka tietoturvallinen chatbotti on (esim. Yang ym. 2023; Hasal ym. 2021; Sebastian 2023). Vahvan tietoturvan varmistamiseksi organisaation on selvitettävä, miten palveluntarjoaja suhtautuu mahdollisiin tietoturvauhkiin ja pyrkii vähentämään niitä. Eri teknologiset ratkaisut, kuten päästä-päähän-salaaminen ja pääsynhallinta lisäävät järjestelmän tietoturvaa. Palveluntarjoajan on kyettävä selittämään, miten se pyrkii vähentämään isoihin kielimalleihin liittyviä tietoturvariskejä, kuten kehoteinjektioita tai tietojen myrkyttämistä. Järjestelmän turvallisuutta tulee arvioida säännöllisesti, ja järjestelmä on tärkeää pitää päivitettynä. Tietoturvan vahvistamiseksi suunnattuja keinoja esitellään tämän opinnäytetyön luvussa 4.1.2. Palveluntarjoajan tulee myös tarjota ratkaisuja sellaisten tilanteiden varalle, joissa käyttäjä antaa loukkaavia syötteitä chatbotille. On riskinä, että nämä päätyvät koulutusaineistoon, vaarantaen chatbotin luotettavuuden. Sisällöllisesti tai kielellisesti loukkaavat sisällöt on tunnistettava ja rajattava pois chatbotin koulutusaineistosta.

Organisaation sisälläkin on hyvä olla tietoa ja taitoa chatbottiin liittyvistä tietoturva-asioista. Aikaisemmissa tutkimuksissa on käynyt selväksi, että ihmisten käyttäytyminen on kriittisessä roolissa tietoturvan kannalta, ja koulutuksen avulla ihmisten aiheuttamaa riskiä voidaan vähentää (esim.

Engler & Dhamani 2024). Erityisesti chatbottien kehityksessä osallisena olevien henkilöiden tietoturvaosaaminen on hyvä olla ajan tasalla, sillä generatiiviseen tekoälyyn ja chatbotteihin liittyvät tietoturvariskit muuttuvat ja kehittyvät jatkuvasti.

Yleinen tietosuoja-asetus takaa käyttäjille oikeuksia henkilötietoihin liittyen, ja haastattelujen perusteella organisaatio toimii tällä hetkellä asetuksen mukaisesti. Organisaatio ei tietoisesti kerää henkilötietoja chatbotin avulla, joten siltä osin asian varmistaminen on kunnossa. Kuten haastatteluissakin nostettiin esiin, on kuitenkin mahdollista, että käyttäjät tahattomasti antavat henkilötietojaan chatbotille. Aikaisemmassa tutkimuksessa on todettu, että tämä muodostaa riskin siitä, että henkilötiedot päätyvät chatbotin koulutusaineistoon, josta niitä voi olla jälkikäteen vaikeaa tai jopa mahdotonta poistaa (esim. Engler & Dhamani 2024; Gupta ym. 2023).

Palveluntarjoajan valinnassa tulee kiinnittää huomiota siihen, että generatiivisen tekoälyn tuomat riskit henkilötietojen vuotamiseen esimerkiksi tällaisessa tahattoman väärinkäytön yhteydessä eivät tuota merkittävää riskiä käyttäjille tai organisaatiolle. Palveluntarjoajan tulee kyetä määrittelemään, miten varmistutaan siitä, että koulutusaineistoon ei päädy sellaisia tietoja, jotka määritellään henkilötiedoiksi. Jos käytyjä keskusteluja tallennetaan ja käytetään chatbotin kouluttamiseen, palveluntarjoajan tulee tarjota vahvat mekanismit henkilökohtaisten tietojen tunnistamiseen ja poistamiseen koulutusmateriaalista. Näiden turvallisuustoimenpiteiden lisäksi organisaation on tarpeellista viestiä käyttäjilleen, että henkilökohtaista tietoa ei kannata chatbotille antaa.

Aikaisemmassa tutkimuksessa korostuu läpinäkyvyyden vaikutus asiakkaiden luottamukseen tekoälyjärjestelmiä kohtaan (esim. Ferraro ym. 2024; Urbani ym. 2024; Sidaoui ym. 2024). Asiakkaiden näkökulmasta generatiivisen chatbotin luottamusta lisää, kun organisaatio viestii läpinäkyvästi siitä, mitä tietoja chatbotti kerää ja tallentaa, ja mihin näitä tietoja käytetään. Organisaation verkkosivuilta löytyy tällä hetkellä tietoa henkilötietojen käsittelystä ja tietojen keräämisestä käyttäjiltä sekä lukuisia tietosuojaselosteita. Evästeitä käytetään tilastointiin, toiminnallisuuksiin ja markkinointiin. Selosteessa mainitaan lukuisia kolmansia osapuolia, joille evästeiden avulla kerättyjä tietoja käyttäjistä annetaan, kuten Google, Facebook, Leadoo ja Hotjar. Tällä hetkellä chatbotin käyttäminen vaatii evästeiden hyväksymistä.

Chatbotin luotettavuuden lisäämiseksi on suositeltavaa, että asiakkaille viestitään selkeästi, mitä kaikkia tietoja heistä kerätään chatbotin käytön yhteydessä, ja keille niitä jaetaan. Lisäksi on syytä pohtia sitä, kuinka tarpeellista chatbottia on pitää evästeiden sallimisen takana. Organisaation kannattaa panostaa chatbotin läpinäkyvyyden lisäämiseen selkeyttämällä evästeisiin ja tietojen keräämiseen liittyvää viestintää, ja viestimällä niistä käyttäjille myös chatbotin käytön yhteydessä. Selkeällä viestinnällä voidaan parantaa asiakkaiden luottamusta chatbotin käyttöä kohtaan.

Lainsäädännön noudattaminen vaikuttaa myönteisesti asiakkaiden luottamukseen, ja organisaatio voisikin harkita kolmannen osapuolen suorittamien auditointien hyödyntämistä vaatimustenmukaisuuden varmistamiseksi, ja tämän kommunikoimista asiakkaille esimerkiksi sertifikaattien muodossa.

6.1.6 Isojen kielimallien puutteet ja riskien hallinta

Haastatteluissa nousi esiin, että chatbotin toimintaa halutaan parantaa lisäämällä sen älykkyyttä, keskustelutaitoja ja isompien kokonaisuuksien hahmottamista. Näihin ominaisuuksiin generatiivisen tekoälyn isot kielimallit sopivat hyvin. Hyötyjen lisäksi isot kielimallit tuovat mukanaan myös riskejä. Aikaisemmassa tutkimuksessa on havaittu isoihin kielimalleihin liittyviä riskejä, kuten hallusinointi (esim. Amaratunga 2023; Sebastian 2023).

Hallusinointi tarkoittaa tilannetta, jossa tekoäly antaa vastauksen, joka ei ole totta. Organisaation tulisi olla tietoinen riskistä, että chatbotti saattaa toisinaan neuvoa asiakkaita väärin, ja varautua tällaisiin tapauksiin. Generatiivinen tekoäly luo vakuuttavaa sisältöä, johon käyttäjän on helppo luottaa. Organisaation tuleekin pohtia, miten toimitaan, jos generatiivinen tekoäly tuottaa virheellistä sisältöä. Etenkin tapauksissa, joissa virheellisestä sisällöstä koituu esimerkiksi taloudellista haittaa asiakkaalle, on vastuumekanismit syytä määritellä. Myös palveluntarjoajan on syytä olla tietoinen näistä riskeistä ja tarjota mekanismeja, joilla riskejä voidaan vähentää.

Riskien vähentämisen mekanismit voivat kuitenkin tuottaa chatbotille haasteita. Merkittävin haaste on, että chatbotti muuttuu harvasanaisemmaksi ja sen hyödyllisyys vähenee. Myös hakutoiminnallisuuden yhdistäminen RAG-tekniikan avulla voi auttaa vastauksien laadun parantamisessa, mutta tämä tekniikka on suhteellisen vaativa, ja saattaa nostaa chatbotin kehittämisen kustannuksia. Palveluntarjoajaa valittaessa on hyvä tiedostaa, miten ratkaisussa pyritään riskien vähentämiseen, hyödyt kuitenkin säilyttäen.

Hybridimallit, joissa generatiivista tekoälyä yhdistetään niin sanottujen perinteisten chatbottien toiminnallisuuteen tietyissä tilanteissa, voi olla ratkaisu, jolla vähennetään hallusinoinnin riskiä. Generatiivista tekoälyä voidaan hyödyntää esimerkiksi intenttien ja entiteettien tunnistamisessa tai vastausten generoimisessa. Useat suuret toimijat, kuten Google ja IBM, tarjoavat tällaisia hybridimalleja, mutta niissä voi olla haasteita käytetyn ison kielimallin rajoitteiden, erityisesti kielituen suhteen. Koska organisaatio tarjoaa palveluaan suomeksi ja englanniksi, tulisi löytää ratkaisu, jossa iso kielimalli tuottaa myös sujuvaa suomen kieltä. Yhtenä ratkaisuna tähän voi olla suomen kieltä tukevien isojen kielimallien hyödyntäminen, kuten Poro tai Viking, mutta koska nykyisen tarjonnan selvittäminen oli tämän opinnäytetyön rajauksen ulkopuolella, on vaikeaa sanoa, onko markkinoilla vielä yhtäkään ratkaisua, jossa suomenkielisiä isoja kielimalleja hyödynnetään.

Haastattelujen perusteella organisaation on vielä määriteltävä, mitä chatbotin toiminnallisuuksia ja ominaisuuksia organisaatio haluaa priorisoida. Isojen kielimallien avulla voidaan lisätä chatbotin älykkyyttä ja keskustelutaitoja. Jos on tärkeää, että chatbotin vastaus perustuu aina tietoon ja on täysin virheetön, on turvallisin ratkaisu edelleen niin sanottu perinteinen chatbotti, jossa ei hyödynnetä isoja kielimalleja.

Aikaisemmassa kirjallisuudessa mainitaan generatiivisten chatbottien puutteena myös empatiakyvyn puute (esim. Ferraro ym. 2024; Sidaoui ym. 2024). Vaikka isot kielimallit ovatkin hyvin edistyneitä ja älykkäitä, ne eivät kuitenkaan kykene empatiaan kuten ihminen. Uusimmilla luonnollisen kielen käsittelyn tekniikoilla, kuten tunneanalyysillä, voidaan parantaa chatbotin kykyä tunnistaa asiakkaan mielentila. On kuitenkin varmistettava, että tietyissä empatiaa vaativissa tilanteissa keskustelu siirtyy palveluneuvojalle. Palveluntarjoajan chatbottiratkaisun mahdollisuuksia tunnistaa empatiaa vaativia aiheita keskustelusta kannattaa tiedustella etukäteen.

6.1.7 Eettisyys ja vastuullisuus

Chatbotin eettinen ja vastuullinen kehittäminen ja käyttöönotto vaatii organisaatiolta panostusta eettisten näkökulmien huomioimiseen koko kehitysprosessin ajan. Eettisten huolenaiheiden pohtimiseen on varattava resursseja, ja sitä on tuettava myös organisaation johdon taholta. Apuna eettisten ja vastuullisten toimintatapojen luomiseen voi käyttää esimerkiksi yrityksen digitaalisen vastuun konseptia, joka antaa kehyksen ja lähestymistavan digitaalisen vastuun hahmottamiseen.

Vastuullisen tekoälyn käyttöönotto vaatii sitä, että organisaatiolla on riittävät tiedot ja taidot tunnistaa ne asiat, joiden avulla palveluntarjoajan valinta vaikuttaa järjestelmän eettisyyteen ja vastuullisuuteen. Palveluntarjoajan valinnassa kannattaa priorisoida sellaisia toimijoita, jotka ovat niin sanottuja ajatusjohtajia, jotka investoivat tutkimukseen ja kehitykseen sekä ovat perillä markkinoiden trendeistä. Tämän avulla voidaan varmistaa palvelun kehittyminen tulevaisuudessa. Myös palveluntarjoajan yleinen suhtautuminen eettisyyteen ja vastuullisuuteen kannattaa huomioida, jotta voidaan varmistua, että nämä asiat otetaan huomioon chatbottia kehitettäessä.

Aiemmassa kirjallisuudessa korostetaan palveluntarjoajan roolia tekoälyjärjestelmän eettisyyden ja vastuullisuuden varmistamiseksi (esim. World Economic Forum 2023). Vastuullinen palveluntarjoaja on avoin tekoälymallistaan ja sen koulutuksesta. Palveluntarjoajan tulee selkeästi kertoa organisaatiolle, miten tekoälymalli toimii ja mistä sen koulutukseen käytettävä tietoaineisto tulee. Palveluntarjoajan on myös esitettävä ratkaisuja riskien tunnistamiseen ja vähentämiseen, esimerkiksi puolueellisuuden huomaamiseen ja siihen reagoimiseen.

Osana vastuullista toimintaa on huomioitava myös teknologian ympäristövaikutukset. Tekoäly ei tällä hetkellä ole ympäristöystävällinen vaihtoehto, ja palveluntarjoajan valinnassa on kiinnitettävä

huomiota siihen, että palveluntarjoaja on tietoinen haitallisista ympäristövaikutuksista ja pyrkii vähentämään niitä. Vastuullinen palveluntarjoaja viestii avoimesti tekoälyratkaisujen vaatimista energiamääristä ja siitä, mitä toimenpiteitä he tekevät ympäristöhaittojen vähentämiseksi.

6.1.8 Lainsäädäntö

Lainsäädännön huolellinen noudattaminen edesauttaa turvallisen ja eettisen järjestelmän käyttöönottoa. Generatiivisten chatbottien kohdalla on syytä kiinnittää huomiota yleiseen tietosuoja-asetukseen ja tekoälysäädökseen.

Yleisen tietosuoja-asetuksen vaatimukset eivät muutu tekoälysäädöksen myötä. Chatbottia kehittäessä on siis hyvä edelleen pitää mielessä samat vaatimukset, mitä organisaatiolla on aiemminkin ollut henkilötietojen käsittelyn suhteen. Haastattelujen perusteella organisaatiolla ei ole ainakaan tällä hetkellä tarkoituksena käsitellä asiakkaan henkilötietoja chatbotissa. Riski siitä, että asiakas kuitenkin syöttää henkilötietojaan chattiin, on olemassa. Organisaation tulee huolehtia siitä, että asiakasta ohjeistetaan olemaan syöttämättä henkilökohtaisia tietojaan chattiin sekä siitä, että asiakkaan chattiin syöttämiä henkilötietoja ei käytetä chatbotin kouluttamiseen. Henkilötiedot tulisi tunnistaa ja poistaa ennen kuin keskustelutietoja käytetään chatbotin kouluttamiseen.

Yleisen tietosuoja-asetuksen näkökulmasta tämän opinnäytetyön toimeksiantajaorganisaatio on rekisterinpitäjä ja tietojenkäsittelijä on organisaatio, joka käsittelee henkilötietoja rekisterinpitäjän puolesta. Tästä näkökulmasta palveluntarjoaja tulee siis olemaan tietojenkäsittelijä, kun taas organisaatio toimii rekisterinpitäjänä. Palveluntarjoajalta tulee varmistaa riittävät takeet siitä, että he noudattavat yleistä tietosuoja-asetusta. Palveluntarjoajan valinnassa on hyvä kiinnittää huomiota siihen, että tietoja ei siirretä EU:n ulkopuolelle.

Tekoälysäädös ottaa kantaa pääasiassa suuririskisiin tekoälyjärjestelmiin ja näiden kehittäjiin. Tekoälysäädöksessä määritellään tarjoajiksi toimijat, jotka kehittävät tekoälyjärjestelmiä ja käyttöönottajiksi toimijat, jotka ottavat tekoälyjärjestelmän käyttöön toiminnassaan. Määritelmän mukaisesti tämän opinnäytetyön toimeksiantajaorganisaatio tulee olemaan käyttöönottaja, koska se ei tule itse kehittämään tekoälyjärjestelmää. Tekoälysäädös kuitenkin määrittää tiettyjä vastuita myös käyttöönottajille. Organisaation tulee ensisijaisesti olla läpinäkyvä tekoälyn käytöstä. Chatbotin yhteydessä tulee käyttäjälle selkeästi viestiä, että hän keskustelee tekoälyn kanssa. Tekoälysäädös määrittää velvoitteita myös suuririskisten tekoälyjärjestelmien käyttöönottajille. Chatbotti ei kuitenkaan lukeudu suuririskiseksi tekoälyjärjestelmäksi, joten lähtökohtaisesti riittää, että organisaatio noudattaa läpinäkyvyyden velvoitetta toiminnassaan.

Tekoälysäädöksessä otetaan erikseen kantaa yleiskäyttöisiin tekoälymalleihin, joiden katsotaan muodostavan systemaattinen riski. Yleiskäyttöisten tekoälymallien kehittäjille asetetaan velvoitteita,

joiden tarkoituksena on turvata järjestelmien turvallisuus. Isojen kielimallien voidaan katsoa lukeutuvan näihin yleiskäyttöisiin tekoälymalleihin, ja tältä osalta vaikuttavan välillisesti myös generatiivista tekoälyä hyödyntävän chatbotin käyttöönottoon. Palveluntarjoajaa valittaessa on hyvä kiinnittää huomiota siihen, että chatbotissa käytettävä iso kielimalli vastaa tekoälysäädöksen vaatimuksiin ja EU:n lainsäädäntöön.

6.1.9 Organisaation arvot, kulttuuri ja käytännöt

Organisaation käytännöt vaikuttavat onnistuneeseen chatbotin käyttöönottoon. Haastattelujen perusteella organisaation käytänteet jo tällä hetkellä osin tukevat chatbotin käyttöönottoa, koska chatbotti on ollut organisaatiossa jo vuosia käytössä. Generatiivisen tekoälyn myötä on kuitenkin hyvä pohtia myös uusia näkökulmia chatbotin kehittämiseen.

Aiemmassa kirjallisuudessa korostetaan ketterien menetelmien merkitystä chatbotin suunnittelussa, kehittämisessä ja ylläpidossa (esim. Skuridin & Wynn 2024; World Economic Forum 2023). Chatbotin kehittämistä kannattaa lähestyä ketterästi, hyödyntäen MVP-lähestymistapaa, eli lähteä liikkeelle rajatuista, laadukkaista toiminnallisuuksista, jotka tuottavat konkreettisia hyötyjä sekä asiakkaalle että organisaatiolle. Myös proof-of-concept (POC) voi olla lähestymistapana varteenotettava. Organisaation kannattaa kiinnittää huomiota siihen, että palveluntarjoajan työtavat noudattavat näitä parhaiksi havaittuja käytäntöjä. Organisaation edustajan tulisi osallistua aktiivisesti chatbotin iteratiiviseen kehittämiseen palveluntarjoajan kanssa. Myös organisaatiossa on hyvä tutustua ketteriin menetelmiin, jotta esimerkiksi käyttötapauksia voidaan toteuttaa järkevässä järjestyksessä, ja kommunikaatio palveluntarjoajan kanssa on sujuvaa.

Haastatteluissa havaittiin, että organisaation vastuut chatbotin kehittämiseen liittyen ovat mahdollisesti hieman epäselvät. Aikaisemmassa tutkimuksessa on osoitettu, että organisaation sisäinen asiantuntijuus, poikkihallinnollinen yhteistyö sekä vastuiden selkeä määrittäminen ovat olennaisia tekijöitä tekoäly- ja chatbottiprojekteissa (esim. Skuridin & Wynn 2024; Sidaoui ym. 2024; World Economic Forum 2023). Chatbotin suunnittelua ja kehittämistä varten on hyvä muodostaa oma tiimi, jolla on myös selkeät vastuut. Organisaation tulisi panostaa vähintään vastuiden määrittämiseen sille, miten chatbottia kehitetään organisaation sisällä ja kuka vastaa sen ylläpidosta ja kouluttamisesta. Organisaation tulee myös varata resursseja chatbotin kehittämiseen ja ylläpitoon. Pääkäyttäjiä tulee kouluttaa ja heidän tietoisuuttaan tekoälyn kehityksestä ja siihen liittyvistä eettisistä käytännöistä tulee lisätä. Vastuiden määrittäminen tukee vastuullista ja eettistä tekoälyn käyttöä, kun vastuuhenkilöillä on riittävät tiedot ja taidot aiheesta.

Aikaisemmassa kirjallisuudessa on todettu, että generatiivisella tekoälyllä voi olla laajoja vaikutuksia myös yhteiskunnallisella tasolla (esim. Wach ym. 2023; Amaratunga 2023; Engler & Dhamani

2024). Generatiivisen tekoälyn myötä organisaation on hyvä pohtia uuden teknologian laajempia vaikutuksia. Vaikka tässä vaiheessa generatiivista tekoälyä hyödynnettäisiin vain tietyissä asiakaspalvelun käyttötapauksissa, generatiivisella tekoälyllä voi tulevaisuudessa olla laajempi merkitys organisaation toiminnalle. Generatiivisella tekoälyllä on mahdollisuus automatisoida paljon sellaista työtä, jota organisaatiossa tällä hetkellä tekevät ihmiset. Jo tässä vaiheessa on siis hyvä pohtia, missä määrin tuetaan työntekijöitä vastaavissa työelämän muutoksissa.

Etenkin asiakaspalvelussa on syytä panostaa työntekijöiden kouluttamiseen työskentelemään uuden teknologian rinnalla. Organisaation on myös syytä pohtia, mitä vaikutuksia generatiivisen tekoälyn tuomilla muutoksilla tulee olemaan asiakaspalveluun, ja miten näihin muutoksiin halutaan reagoida. Palveluneuvojia voidaan esimerkiksi siirtää vaativampiin, inhimillistä otetta vaativiin tehtäviin. Myös asiakaspalvelun laadun parantaminen voi olla yhtenä tavoitteena, kun palveluneuvojilla on enemmän aikaa ja resursseja keskittyä henkilökohtaiseen asiakaspalveluun.

Organisaation on hyvä panostaa henkilökunnan koulutukseen tekoälyn suhteen, jotta uusi teknologia voidaan ottaa käyttöön tehokkaasti ja hyödyllisesti. Organisaation sisäistä chatbotti-, tekoäly- ja data-asiantuntijuutta kannattaa kehittää, jotta organisaatio pysyy mukana digitalisaation muutoksissa.

6.2 Tutkimuksen luotettavuuden arviointi

Laadullisen tutkimuksen luotettavuuden arviointi ei ole yksiselitteistä. Eri oppaissa tarkastellaan ja painotetaan erilaisia piirteitä. Yhtenä tekijänä voidaan arvioida tutkimuksen puolueettomuutta. Tietty puolueettomuus saavutetaan, kun tutkija pyrkii ymmärtämään tutkittavia itsenään, ilman ennakko-oletuksia. Tietyssä mielessä laadullisen tutkimuksen tekijä on aina puolueellinen, koska tutkija luo tutkimusasetelman ja tekee tulkinnat. (Tuomi & Sarajärvi 2018, luku 6.1.) Tässä opinnäytetyössä puolueettomuutta edistää tutkijan suhde tutkittavana olevaan organisaatioon ja haastateltaviin, koska tutkijalla ei ole aiempaa suhdetta tutkimusprosessin ulkopuolelta kyseiseen organisaatioon.

Triangulaatio on yksi keino, jolla laadullisenkin tutkimuksen luotettavuutta voidaan mahdollisesti lisätä. Triangulaatiolla voidaan tarkoittaa tutkimusaineiston keräämisestä monelta eri tiedonantajaryhmältä, useamman tutkijan mukanaoloa tutkimusprosessissa, teoreettisten näkökulmien moninaisuutta sekä useiden eri tutkimusmenetelmien käyttöä. (Tuomi & Sarajärvi 2018, luku 6.5.) Tässä opinnäytetyössä on pyritty saamaan useita eri näkökulmia tutkimusprosessiin pääasiassa eri tiedonantajaryhmien huomioimisella tutkimusaineiston keräämisessä ja teoreettisen näkökulmien moninaisuus huomioiden. Tutkimuksen haastateltaviksi pyrittiin valitsemaan haastateltavia,

joilla olisi erilaisia lähtökohtia ja näkökulmia lähestyttävään teemaan. Lisäksi teoreettisessa viitekehyksessä pyrittiin tuomaan esiin useita eri näkökulmia.

Laadullisessa tutkimuksessa tutkimusprosessin tarkka kuvaus ja tulkintojen perustelut ovat oleellisia luotettavuuden lisäämiseksi. Tarkan kuvausten ja perustelujen avulla lukija voi tehdä johtopäätöksiä ja arvioida tutkimuksen luotettavuutta. (Ojasalo ym. 2015, 105.) Opinnäytetyössä on pyritty etenemään järjestelmällisesti, analyyttisesti ja kriittisesti. Teoreettisen viitekehyksen ja tutkimuksen tulokset on pyritty saamaan keskustelemaan keskenään, ja tämän havainnollistamisen tueksi opinnäytetyöstä löytyy peittomatriisi (taulukko 1). Opinnäytetyössä on pyritty johdonmukaisuuteen ja tarkkaan raportointiin, mutta lopullinen luotettavuuden arviointi jää lukijan pääteltäväksi.

6.3 Opinnäytetyöprosessin arviointi

Kokonaisuudessaan opinnäytetyöprosessi oli sujuva. Suomenkielisen opinnäytetyön kirjoittamisessa oli kuitenkin aihepiirin takia erinäisiä haasteita. Generatiiviseen tekoälyyn liittyvä tutkimus painottuu hyvin vahvasti englanninkieliseen maailmaan, ja pääasiassa tutkimuksia julkaistaan englanniksi. Vakiintuneita suomennoksia tekoälyyn liittyville termeille ei vielä ole, joten opinnäytetyössä jouduttiin käyttämään myös luovuutta ja omaa harkintakykyä tiettyjen termien suomentamisessa. Oman kokemukseni mukaan IT-ala Suomessa tukeutuu hyvin pitkälti englanninkielisiin termeihin toiminnassaan, esimerkiksi sanaa ”agile” käytetään käytännössä synonyyminä suomenkieliselle vastineelle ”ketterä”. Kuitenkin suomen kieltä puhuu äidinkielenään hyvin pieni määrä maailman väestöstä, ja tästäkin syystä on tärkeää, että myös näille uusille teknologisille termeille muodostettaisiin yhtenäinen suomenkielinen sanasto.

Lisäksi, koska kyseessä on hyvin uusi teknologia, generatiiviseen tekoälyyn, isoihin kielimalleihin ja luonnollisen kielen käsittelyyn liittyvää terminologiaa käytetään jopa englanninkielisissä tutkimuksissa ristiin. Opinnäytetyössä olen pyrkinyt selventämään näiden käsitteiden eroja ja luomaan yleiskuvan siitä, miten nämä käsitteet liittyvät toisiinsa, miten ne eroavat toisistaan ja missä tilanteissa ne tarkoittavat käytännössä samoja asioita. Isoista kielimalleista puhutaan usein omana käsitteenään, vaikka ne käytännössä ovat jatkumoa luonnollisen kielen käsittelyn kehittymiselle. Terminologian ja käsitteiden erojen selvittäminen oli selkeä haaste opinnäytetyön tekemisessä, ja toivottavasti olen työssä onnistunut selkeyttämään aihepiiriin liittyvää terminologiaa.

Oman haasteensa opinnäytetyöprosessille toi myös asian ajankohtaisuus. Generatiivinen tekoäly kehittyy jatkuvasti, ja on vaikea ennustaa, miltä tutkimukset näyttävät vaikkapa vuoden kuluttua. Uutta tietoa tulee koko ajan lisää, ja opinnäytetyössä haasteena oli mahdollisimman ajantasaisen tiedon hyödyntäminen. Jossain vaiheessa teorian kerääminen oli kuitenkin lopetettava, jotta opinnäytetyö valmistuisi ajallaan. Onkin siis todettava, että tässä opinnäytetyössä esitellyt teknologiat,

riskit ja haasteet, voivat olla jo ensi vuonna täysin eri näköisiä. Opinnäytetyössä pyrittiin myös kuitenkin huomioimaan sellaisia geneerisiä, uusiin teknologioihin liittyviä asioita, joista on hyötyä, vaikka itse teknologia muuttuisi reippaasti. Tällaisia ovat esimerkiksi organisaation arvot ja kulttuuri, sekä ketterän kehittämisen periaatteet.

Opinnäytetyöprosessin aikana nousi esiin myös selkeä uusi tutkimuksen aihe: nykyisen markkinatilanteen selvittäminen. Koska tässä opinnäytetyössä ei perehdytty siihen, mitä teknologioita ja ratkaisuja generatiivisten chatbottien suhteen on tällä hetkellä tarjolla, tuloksista jäi uupumaan tietoa, joka voisi olla toimeksiantajalle hyödyllistä. Tutkimuksella voitaisiin selvittää, mitä palveluntarjoajia Suomessa tai kansainvälisesti tällä hetkellä on, mitä teknologioita he tarjoavat, miten he suhtautuvat eettisyyteen ja vastuullisuuteen tekoälyn kontekstissa, ja miten hyvin heidän ratkaisuissaan tuetaan suomen kieltä. Tutkimuksella voitaisiin vahvistaa näkemystä siitä, minkälaisia ratkaisuja Suomen markkinoilla tällä hetkellä on tarjolla.

Kaikkinensa koen opinnäytetyöprosessin sangen onnistuneena. Yhteistyö toimeksiantajaorganisaation kanssa oli koko prosessin ajan hedelmällistä ja hyödyllistä. Itse opin paljon generatiivisesta tekoälystä, chatboteista ja generatiivisista chatboteista, minkä lisäksi opinnäytetyöprosessi vahvisti itsenäisen työskentelyn, itsensä johtamisen ja projektin hallinnan taitoja. Toimeksiantajalta saatu palaute opinnäytetyöstä on ollut positiivista, ja opinnäytetyön tulokset on koettu organisaatiossa hyödyllisinä. Toivon ja uskon, että opinnäytetyöstä Hoas saa tarpeellista tietoa matkalla kohti generatiivisen chatbotin käyttöönottoa.

Lähteet

Adamopoulou, E. & Moussiades, L. 2020. Chatbots: History, technology, and applications. Machine learning with applications, 2, s. 100006.

Agile Alliance s.a. What is Agile? Luettavissa: https://www.agilealliance.org/agile101/. Luettu: 15.8.2024.

Amaratunga, T. 2023. Understanding Large Language Models: Learning Their Underlying Concepts and Technologies. Apress. New York. E-kirja. Luettu: 29.7.2024.

Davis, F. D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS quarterly, 13, 3, s. 319–340.

Engler, M. & Dhamani, N. 2024. Introduction to Generative AI. Manning Publications. E-kirja. Luettu: 16.8.2024.

Euroopan komissio s.a. Mitä tarkoittaa ’sisäänrakennettu’ ja ’oletusarvoinen’ tietosuoja? Luettavissa: https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean_fi. Luettu: 7.8.2024.

Euroopan parlamentin ja neuvoston asetus (EU) 2024/1689, annettu 13 päivänä kesäkuuta 2024, tekoälyä koskevista yhdenmukaistetuista säännöistä ja asetusten (EY) N:o 300/2008, (EU) N:o 167/2013, (EU) N:o 168/2013, (EU) 2018/858, (EU) 2018/1139 ja (EU) 2019/2144 sekä direktiivien 2014/90/EU, (EU) 2016/797 ja (EU) 2020/1828 muuttamisesta (tekoälysäädös) (ETA:n kannalta merkityksellinen teksti).

Ferraro, C., Demsar, V., Sands, S., Restrepo, M. & Campbell, C. 2024. The paradoxes of generative AI-enabled customer service: A guide for managers. Business Horizons, 67, 5, s. 549-559.

Gheorghiu, A. 2024. Building Data-Driven Applications with LlamaIndex: A Practical Guide to Retrieval-Augmented Generation (RAG) to enhance LLM Applications. E-kirja. Luettu: 20.8.2024.

Google Cloud 2024. Generative features overview. Luettavissa: https://cloud.google.com/dialogflow/cx/docs/concept/generative. Luettu: 20.8.2024.

Gupta, M., Akiri, C., Aryal, K., Parker, E. & Praharaj, L. 2023. From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. IEEE Access, 11, s. 80218–80245.

Hasal, M., Nowakova, J., Saghair, K. A., Abdulla, H. Snasel, V. & Ogiela, L. 2021. Chatbots: Security, privacy, data protection, and social aspects. Concurrency and Computation: Practice and Experience, 13, 19.

Hoas s.a. Helsingin seudun opiskelija-asuntosäätiö Hoas. Luettavissa: https://hoas.fi/hoas/helsingin-seudun-opiskelija-asuntosaatio-hoas/. Luettu: 27.8.2024.

IBM Cloud 2024a. Information gathering. Luettavissa: https://cloud.ibm.com/docs/watson-assistant?topic=watson-assistant-information-gathering. Luettu: 20.8.2024.

IBM Cloud 2024b. Conversational search. Luettavissa: https://cloud.ibm.com/docs/watson-assistant?topic=watson-assistant-conversational-search. Luettu: 20.8.2024.

Khan, R. & Das, A. 2017. Build Better Chatbots: A Complete Guide to Getting Started with Chatbots. Apress. New York. E-kirja. Luettu: 28.5.2024.

Luukkonen, R., Komulainen, V., Luoma, J., Eskelinen, A., Kanerva, J., Kupari, H., Ginter, F., Laippala, V., Muenninghoff, N., Piktus, A., Wang, T., Tazi, N., Le Scao, T., Wolf, T., Suominen, O., Sairanen, S., Merioksa, M., Heinonen, J., Vahtola, A., Antao, S. & Pyysalo, S. 2023. FinGPT: Large Generative Models for a Small Language. arXiv.org. Luettavissa: https://arxiv.org/pdf/2311.05640. Luettu: 20.8.2024.

McTear, M. & Ashurkina, M. 2024. Transforming Conversational AI: Exploring the Power of Large Language Models in Interactive Conversational Agents. Apress. New York. E-kirja. Luettu: 20.8.2024.

Ojasalo, K., Moilanen, T. & Ritalahti, J. 2015. Kehittämistyön menetelmät: Uudenlaista osaamista liiketoimintaan. 3.–4. painos. Sanoma Pro Oy. Helsinki.

Sebastian, G. 2023. Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information. International Journal of Security and Privacy in Pervasive Computing, 15, 1.

Sidaoui, K., Mahr, D. & Odekerken-Schröder, G. 2024. Generative AI in Responsible Conversational Agent Integration: Guidelines for Service Managers. Organizational Dynamic, 53, 2, s. 101045.

SiloAI 2024. Viking 13B: Scaling Nordic AI models using an open source training framework for LUMI. Luettavissa: https://www.silo.ai/blog/viking-13b-scaling-nordic-ai-models-using-an-open-source-training-framework-for-lumi. Luettu: 20.8.2024.

Skuridin, A. & Wynn, M. 2024. Chatbot Design and Implementation: Towards an Operational Model for Chatbots. Information, 15, 4, s. 226.

Tuomi, J. & Sarajärvi, A. 2018 Laadullinen tutkimus ja sisällönanalyysi. Uudistettu laitos. Kustannusosakeyhtiö Tammi. Helsinki. E-kirja. Luettu: 21.8.2024.

Turun yliopisto 2024. Euroopan avoin kielimalli Poro: eurooppalaisen tekoälyn ja kielten monimuotoisuuden virstanpylväs. Luettavissa: https://www.utu.fi/fi/ajankohtaista/uutinen/euroopan-avoin-kielimalli-poro-eurooppalaisen-tekoalyn-ja-kielten. Luettu: 20.8.2024.

Urbani, R., Ferreira, C. & Lam, J. 2024. Managerial framework for evaluating AI chatbot integration: Bridging organizational readiness and technological challenges. Business Horizons, 67, 5, s. 595-606.

van der Merwe, J. & Al Achkar, Z. 2022. Data responsibility, corporate social responsibility, and corporate digital responsibility. Data & Policy, 4, e12.

Villa, L., Carneros-Prado, D., Dobrescu, C. C., Sanchez-Miguel, A., Cubero, G. & Hervas, R. 2024. Comparative Analysis of Generic and Fine-Tuned Large Language Models for Conversational Agent Systems. Robotics (Basel), 13, 5, p. 68.

Wach, K., Doanh Duong, C., Ejdys, J., Kazlauskaite, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J. & Ziemba, E. 2023. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. Entrepreneurial Business and Economics Review, 11, 2, s. 7–30.

Wirtz, J., Kuntz, W. H., Hartley, N. & Tarbit, J. 2022. Corporate Digital Responsibility in Service Firms and Their Ecosystems. Journal of Service Research, 26, 2, s.173–190.

World Economic Forum 2023. Adopting AI Responsibly: Guidelines for Procurement of AI Solutions by the Private Sector. World Economic Forum. Luettavissa: https://www.weforum.org/publications/adopting-ai-responsibly-guidelines-for-procurement-of-ai-solutions-by-the-private-sector/. Luettu: 8.8.2024.

Yang, J., Chen, Y., Por, L. Y. & Ku, C. S. 2023. A Systematic Literature Review of Information Security in Chatbots. Applied Sciences, 13, 11, s. 6355.

Your Europe 2022. Yleinen tietosuoja-asetus. Luettavissa: https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_fi.htm. Luettu: 7.8.2024.

Liitteet

Liite 1. Haastattelukysymykset

1. Miksi chatbotti on alun perin otettu käyttöön organisaatiossa?
2. Mitä ongelmaa nykyisellä chatbotilla pyritään ratkaisemaan/mitä sillä on tavoiteltu?
3. Miten chatbotin tavoitteiden saavuttamista seurataan?
4. Mitä hyötyjä uudesta chatbotista haetaan?
5. Mistä asioista asiakkaat ovat eniten yhteydessä?
6. Mihin kysymyksiin chatbotti tällä hetkellä osaa vastata ja mitkä kysymykset ohjautuvat ihmiselle?
7. Minkälainen prosessi ns. human-handoff tällä hetkellä on?
8. Mitä materiaalia/tietoa ihmisillä on käytettävissään kysymyksiin vastaamiseen? Mistä tämä tieto löytyy?
9. Miten parantaisit nykyistä chatbottia?
10. Miten nykyisen chatbotin toimintaa seurataan?
11. Kerätäänkö nykyisestä chatbotista jotain palautetta? Mitä?
12. Mitä ongelmia nykyisen chatbotin ylläpidossa on?
13. Miten chatbotin ylläpitoa voisi parantaa?
14. Mitkä ovat tietoturva- ja/tai tietosuojavaatimukset chatbotille?
15. Tiedätkö mihin lainsäädännöllisiin vaatimuksiin chatbotin tulee vastata?