

**DATAN ROOLI MARKKINOINNIN
TEKOÄLYSOVELLUSTEN
IMPLEMENTOINTIPROSESSISSA**

**Jyväskylän yliopisto
Kauppakorkeakoulu**

Pro gradu -tutkielma

2022

**Tekijä: Aleksi Kantonen
Oppiaine: Markkinointi
Ohjaaja: Mika Skippari**



JYVÄSKYLÄN YLIOPISTO

TIIVISTELMÄ

Tekijä ALEKSI KANTONEN	
Työn nimi Datan rooli markkinoinnin tekoälysovellusten implementointiprosessissa	
Oppiaine MARKKINOINTI	Työn laji Pro gradu -tutkielma
Aika (pvm.) 5.6.2022	Sivumäärä 62
Tiivistelmä	
<p>Monet teknologiset innovaatiot ovat ottaneet vahvan jalansijan markkinoinnin kentällä digitalisaation myötä. Tekoäly on yksi uusimmista ja nykypäivän trendikkäistä teknologisista ratkaisuista, joka tarjoaa markkinoinnin asiantuntijoille tehokkaita, dataan pohjautuvia ratkaisuja.</p> <p>Vaikka tekoälyä on käsitelty markkinoinnin akateemisessa kirjallisuudessa viime aikoina enenevässä määrin, on tutkimus siitä kuitenkin verrattain vähäistä. Aiemmissä tutkimuksissa on käsitelty mm. tekoälystä markkinoinnin tukena ja tekoälysovellusten hyödyntämisestä markkinoinnin operatiivisissa tehtävissä, mutta ne tarjoavat kuitenkin varsin rajoitetusti ymmärrystä markkinoinnin tekoälysovelluksissa hyödynnettävästä datasta.</p> <p>Tämän pro gradu -tutkielman tarkoituksena oli selvittää, minkälaista dataa markkinoinnin tekoälysovellusten implementointiprosesseissa hyödynnetään ja mistä sitä tyypillisesti kerätään. Tämän toteuttamiseksi tutkielman menetelmäksi valittiin laadullinen tutkimus, sillä se soveltuu hyvin vastaavanlaiselle moniulotteiselle tutkimusaiheelle, joka vaatii syvällistä ymmärrystä. Tutkimusaineisto kerättiin seitsemän puolistrukturoidun teemahaastattelun avulla, mikä analysoitiin sisällönanalyysia sekä teemoittelua käyttäen.</p> <p>Aineiston analyysin lopputulemana muodostettiin kolme pääteemaa, jotka sisälsivät erillisiä alateemoja. Tutkimustulokset mukailivat osittain aiempaa kirjallisuutta, mutta esittivät myös uusia havaintoja. Tulokset toivat esiin, kuinka markkinoinnin tekoälysovellusten implementointiprosesseissa hyödynnettävä data on tyypillisesti strukturoitua ja yrityksen sisäistä dataa. Datan tulisi olla mahdollisimman laadukasta, sillä tekoälysovellusten toiminta perustuu vahvasti niille syötettyyn dataan, mutta samalla laadun määrittäminen koetaan erittäin kontekstisidonnaiseksi. Täydellisen laadun etsiminen ei ole relevanttia, sillä tosielämän tilanteissa data on useimmiten jokseenkin vinoutunut. Tärkeintä on siis luoda jatkuvia sekä järjestelmällisiä datan käsittelyn prosesseja organisaatioon, ja tavoitella että se sopii sille määritellyyn markkinoinnin käyttötarkoitukseen parhaalla mahdollisella tavalla.</p>	
Asiasanat Data, Digimarkkinointi, Koneoppiminen, Tekoäly	

KUVIOT JA TAULUKOT

KUVIOT

KUVIO 1 Markkinoinnin tekoälysovellusten implementointiprosessin vaiheet (Alsultanny, 2011; Haghghatnia ym., 2018; Overgoor ym., 2019; Shearer, 2000)	23
KUVIO 2 Sisäiset ja ulkoiset datalähteet (Hartmann ym., 2016)	27
KUVIO 3 Tutkimuksen teemoittelun tulokset ja niiden suhteet	49

TAULUKOT

TAULUKKO 1 Digitaalisten kanavien luokittelu (Taiminen & Karjaluoto, 2015)	13
TAULUKKO 2 Tekoäly historiallisen älykkyyden kehityksen mukaan (Huang & Rust, 2018).....	17
TAULUKKO 3 Koneoppimisen lajit (Campbell ym., 2020; De Bruyn ym., 2020; Overgoor ym., 2019)	21
TAULUKKO 4 Haastattelujoukon esittely.....	33
TAULUKKO 5 Tutkimustulosten yhteenveto.....	50

SISÄLLYS

TIIVISTELMÄ	2
KUVIOT JA TAULUKOT	3
1 JOHDANTO.....	6
1.1 Tutkimuksen taustaa	6
1.2 Tutkimustehtävä ja -kysymykset	8
1.3 Tutkimuksen rakenne ja rajaus.....	9
2 MARKKINOINNIN DIGITAALINEN MUUTOS.....	10
2.1 Markkinointi ja sen digitalisoituminen	10
2.2 Digimarkkinointi	10
2.3 Digimarkkinoinnin kanavat	11
2.4 Digimarkkinoinnin mittaaminen.....	13
3 MARKKINOINNIN TEKOÄLY.....	15
3.1 Tekoäly	15
3.2 Markkinoinnin tekoälysovellukset.....	17
3.3 Koneoppiminen.....	18
3.4 Markkinoinnin tekoälysovellusten implementointiprosessi.....	21
4 DATAN HYÖDYNTÄMINEN MARKKINOINNIN TEKOÄLYSOVELLUKSISSA.....	25
4.1 Mitä dataa hyödynnetään?	25
4.2 Mistä dataa kerätään?	26
4.3 Datan laatu ja sen edellytykset	28
5 METODOLOGIA.....	30
5.1 Laadullinen tutkimusmenetelmä	30
5.2 Aineiston kerääminen	32
5.3 Sisällönanalyysi.....	33
6 TUTKIMUKSEN TULOKSET.....	36
6.1 Strukturoidun ja strukturoimattoman datan hyödyntäminen	36
6.1.1 Strukturoitu data keskiössä	36
6.1.2 Strukturoimattoman datan potentiaali	37
6.2 Datalähteet ja niiden painotus	39
6.2.1 Sisäisen datan vahva rooli.....	39
6.2.2 Ulkoisen datan rikastava rooli	41
6.3 Datan laatu ja sen merkitys	43
6.3.1 Datan laadun kontekstisidonnaisuus.....	43
6.3.2 Datan laatuun vaikuttavat tekijät	44
6.3.3 Datan laadun merkitys	46

6.4	Tulosten yhteenveto	47
7	JOHTOPÄÄTÖKSET JA ARVIONTI	51
7.1	Tulosten tarkastelu ja johtopäätökset	51
7.2	Tutkimuksen arviointi	53
7.3	Jatkotutkimusehdotukset	54
	LÄHTEET	55

1 JOHDANTO

1.1 Tutkimuksen taustaa

Yritysten on tärkeä pystyä seuraamaan toimialojen digitaalista muutosta selvittääkseen kasvavan kilpailun keskellä (Caliskan ym., 2020). Digitaalisella muutoksella viitataan tyypillisesti organisaatioiden pyrkimykseen toimia menestyksellisesti digitaalisessa toimintaympäristössä (Hughes & Vafeas, 2019). Digitaalisen median sekä tieto- ja viestintäteknologian nopea omaksuminen yhteiskunnassa on vaikuttanut merkittävästi ihmisten tapoihin kommunikoida sekä täyttää omia sosioekonomisia, tunnepohjaisia ja materiaalisia tarpeitaan (Dwivedi ym., 2020). Muun muassa sähköpostit, verkkosivut, hakukoneet ja sosiaalisen median alustat ovat nykypäivänä suuressa suosiossa ihmisten päivittäisissä aktiviteeteissa, kuten yhteydenpidossa, tuotteiden ja palveluiden ostamisessa sekä pankkipalveluissa (Kulkarni ym., 2020).

Markkinoinnin useat digitaaliset kanavaratkaisut ja työkalut ovat luoneet yrityksille uudenlaisia mahdollisuuksia toteuttaa digitaalista mainontaa ilman maantieteellisiä rajoja, mitata markkinoinnin suorituskykyä sekä kerätä vaivattomasti yrityksille arvokasta dataa (Järvinen, Töllinen, ym., 2012). Digitaalisen muutoksen ja uusien teknologioiden kehittymisen myötä syntyi myös digitaalisen markkinoinnin käsite, joka on saavuttanut vahvan aseman markkinoinnin tutkimus- ja liiketoimintakontekstissa (Key, 2017; Royle & Laing, 2014). Sen määritelmä on ajan myötä kehittynyt digitaalisten kanavien hyödyntämisestä tuotteiden ja palvelujen markkinoinnissa laajemmaksi markkinoinnin kattotermiksi, jolla kuvataan asiakashankinnan, asiakassuhteiden ylläpidon sekä brändin ja myynnin edistämisen prosesseja käyttäen digitaalisia teknologioita (Kannan & Li, 2017).

Tekoäly on yksi viime vuosien trendikkäimmistä ja suosiota saaneista teknologisista ratkaisuista, vaikka sen käsite esiteltiin ensimmäistä kertaa maailmalle jo 1950-luvun puolessa välissä (Ma & Sun, 2020; Overgoor ym., 2019). Tekoälyteknologian sekä sen akateemisen tutkimuksen laajuus on tuona aikana ulottunut monialaisesti niin markkinoinnin kuin terveydenhoidon, rahoituksen

sekä koulutuksen aloille (Campbell ym., 2020; Ma & Sun, 2020). Tekoälyn ennustetaan muuttavan mm. yritysten johdon toimikuvaa, markkinoinnin strategista roolia ja sisältöä sekä kuluttajakäyttäytymistä lähitulevaisuudessa, minkä lisäksi sen uskotaan menestyvän yrityksissä paremmin, jos sitä ei nähdä henkilöstön korvikkeena vaan ennemmin organisaation tukena (Davenport ym., 2020).

Tekoälyn rooli markkinoinnissa on vahvistunut viime vuosina huomattavasti, sillä ihmisen ja tietokoneen välinen raja päätöksenteossa on nykypäivänä häilyvämpi aiempaan verrattuna (Stone ym., 2020). Tekoälyä hyödynnetään jo esimerkiksi markkinoinnin kohdentamisessa ja analysoinnissa sekä mainossisältöjen suunnittelussa (Stone ym., 2020). Globaalisti tunnettuja esimerkkejä tekoälysovellusten hyödyntämisestä markkinoinnissa ovat mm. Netflixin ja Amazonin suosittelujärjestelmät sekä Googlen automatisoidut digitaalisen mainonnan ratkaisut (Huang & Rust, 2018; Overgoor ym., 2019). Tekoälyn odotetaan lähivuosina auttavan entistä enemmän markkinoinnin päätöksentekoprosessien vauhdittamisessa sekä markkinoinnin päätöksentekijöille tärkeän ja kriittisen datan tuottamisessa (Overgoor ym., 2019). Sen uskotaan ennen kaikkea auttavan yrityksiä strategisessa päätöksenteossa, kuten liiketoimintamallien valinnassa, hinnoitteluratkaisuissa sekä viestintä- ja toimituskanavapäätöksissä (Stone ym., 2020).

Markkinoinnin kansainvälisessä tutkimuskirjallisuudessa tekoälystä käytetään termiä *Marketing AI*, joka kääntyy tässä tutkielmassa suomeksi muotoon markkinoinnin tekoäly (Ascarza ym., 2021; Overgoor ym., 2019; Vlačić ym., 2021). Markkinoinnin tekoälyyn keskittyvä akateeminen kirjallisuus tarjoaa kattavasti tutkimuksia mm. tekoälysovellusten hyödyntämisestä markkinoinnin operatiivisissa tehtävissä, tekoälystä markkinoinnin strategisena tukena, markkinoinnin tekoälyn tutkimuskentän trendeistä sekä datan roolista markkinoinnin tekoälyn hyödyntämisessä (Brock & von Wangenheim, 2019; Campbell ym., 2020; Klaus & Zaichkowsky, 2020; Overgoor ym., 2019; Shah ym., 2020).

Tekoälysovellusten toiminta perustuu vahvasti niille syötettyyn dataan, minkä takia datan tutkiminen ja siihen syventyminen on relevanttia ja tarpeellista (Verma ym., 2021). Tuotetun datan määrä maailmassa on kasvanut suurta vauhtia, mikä on vahvistanut tekoälyratkaisujen suosiota markkinoinnin liiketoiminnassa sekä akateemisessa tutkimuksessa (Brock & von Wangenheim, 2019; Overgoor ym., 2019). Big datasta on muovautunut suosittu termi, jolla viitataan tyypillisesti suureen, jatkuvasti kasvavaan ja moninaiseen datajoukkoon, joka pitää sisällään yksittäisen kuluttajan tietoja, joita yritykset pyrkivät hyödyntämään liiketoiminnassaan (Martin & Murphy, 2017). Big datan kehityksen myötä, dataa on saatavilla enemmän kuin koskaan ennen, miksi myös tekoälyn hyödyntäminen markkinoinnissa, kuin myös muilla liiketoiminta-aloilla, on helpompaa ja tehokkaampaa (Overgoor ym., 2019). Datan valtava määrä ja helppo saatavuus ovat tuoneet samalla myös haasteita sen hyödyntämiseen esimerkiksi tietosuojakysymysten muodossa (Martin & Murphy, 2017; Vlačić ym., 2021). Kuluttajat ovat aiempaa tietoisempia ja samalla huolissaan omien henkilötietojen jakamisesta, minkä myötä asenteet yritysten datankeruuta kohtaan ovat tiukentuneet (Martin & Murphy, 2017).

Tekoälysovellukset sekä niiden menestyksenkäs implementointi markkinoinnin eri tehtäviin ovat erittäin riippuvaisia datasta, joka näin ollen asettaa reunaehdot tekoälysovellusten toimivuudelle (Overgoor ym., 2019). Datan rooli markkinoinnin tekoälysovellusten implementoinnissa on tästä johtuen varsin merkittävä (Brock & von Wangenheim, 2019; Overgoor ym., 2019). Tekoälysovellusten implementointiprosesseissa hyödynnettävää dataa ei ole kuitenkaan tutkittu markkinoinnin alalla, miksi syvällisemmän ymmärryksen luominen, teoreettista pohjaa rakentaen, on ajankohtaista ja tärkeää (Borges ym., 2021; Overgoor ym., 2019; Pappas ym., 2018). Datan roolia painotetaan vahvasti monissa markkinoinnin tekoälyyn liittyvissä tutkimuksissa, mutta aiheen käsittely jää silti vain pintapuoliseksi kuvaukseksi datan ominaisuuksista ja potentiaalista (Campbell ym., 2020; Overgoor ym., 2019). Seuraavassa luvussa esitellään tutkielman tutkimustehtävä sekä -kysymykset tämän tutkimusaukon täyttämiseksi.

1.2 Tutkimustehtävä ja -kysymykset

Tekoälyteknologiaa sekä sen hyödyntämistä on tarkasteltu markkinoinnin akateemisessa tutkimuksessa melko kattavasti, tutkimuksen keskittyen aiheisiin, kuten tekoälysovellusten hyödyntäminen markkinoinnin eri tehtävissä sekä niiden rooli markkinoinnin strategisena tukena (Brock & von Wangenheim, 2019; Campbell ym., 2020; Klaus & Zaichkowsky, 2020; Overgoor ym., 2019). Markkinoinnin tekoälysovellusten implementointia käsitellään markkinoinnin akateemisessa tutkimuksessa kuitenkin vähäisesti. Overgoor ym. (2019) havainnoivat markkinoinnin tekoälysovellusten implementointiprosessin kulkua CRISP-DM (Cross-Industry Standard Process for Data Mining) -prosessimallin avulla, joka antaa selkeän kuvauksen tekoälysovelluksen implementoinnin eri vaiheista korostaen datan merkitystä sen toteuttamisessa. Tekoälysovellusten implementointiprosessissa hyödynnettävää dataa ei kuitenkaan ole tutkittu markkinoinnin alalla lainkaan, vaikka data on tekoälysovellusten kriittinen elementti (Hair & Sarstedt, 2021; Overgoor ym., 2019).

Data luo edellytykset toimivalle ja menestykselle markkinoinnin tekoälysovellukselle, sillä tekoäly vaatii suuria määriä korkealaatuista dataa oppiakseen autonomisesti selvittämään erilaisia ennalta määriteltyjä ongelmia (Brock & von Wangenheim, 2019; Campbell ym., 2020). Data määritellään myös CRISP-DM -prosessimallissa merkittäväksi ja aikaa vaativaksi tekijäksi, johon on syytä kiinnittää huomiota (Overgoor ym., 2019; Shearer, 2000). Markkinoinnin tekoälysovellusten implementoinnissa hyödynnettävän datan tutkimuksen puute on ilmeinen, minkä takia tämä pro gradu -tutkielma keskittyy tarkastelemaan ja selvittämään, minkälaista markkinoinnin tekoälysovellusten implementoinnissa hyödynnettävä data oikeasti on.

Tutkimukselle asetettiin kaksi päätutkimuskysymystä:

TK1: Minkälaista dataa markkinoinnin tekoälysovellusten implementointiprosesseissa hyödynnetään?

TK2: Mistä lähteistä markkinoinnin tekoälysovellusten implementointiprosesseissa hyödynnettävää dataa kerätään?

1.3 Tutkimuksen rakenne ja rajaus

Tutkimuksen aineisto tullaan keräämään Suomessa toimivien markkinointitoimistojen asiantuntijoilta, jotka omaavat osaamista tekoälysovellusten hyödyntämisestä markkinoinnin alalla. Tämä osaltaan rajaa tutkimuksen laajempaa yleistämistä, sillä aineisto koostuu ainoastaan suomalaisten markkinointitoimistojen asiantuntijoiden näkemyksistä. Tutkimusaineisto kerätään teemahaastattelujen avulla, jotka litteroidaan ja analysoidaan sisällönanalyysin avulla. Aineiston sisällönanalyysin metodina käytetään teemoittelua, jonka lopputulemana tutkimustulokset esitetään teemoittain.

Tutkimuksen ensimmäisessä luvussa esitellään tutkimuksen taustaa, esitetään tutkimustehtävä- ja kysymykset sekä tutkimuksen rakenne ja rajaus. Toinen, kolmas ja neljäs luku muodostavat tutkimuksen teoriaosion, joka käsittelee markkinoinnin digitaalista muutosta, markkinoinnin tekoälyä sekä datan hyödyntämistä markkinoinnin tekoälysovelluksissa. Tätä seuraa viides luku, jossa käydään läpi tutkimuksen metodologia, joka pitää sisällään tutkimusmenetelmän, tutkimusaineiston sekä aineiston analysoinnissa käytettävän sisällönanalyysin esittelyn. Kuudennessa luvussa esitetään sisällönanalyysin lopputulemana muodostetut tutkimustulokset teemoittain. Viimeinen luku vetää yhteen tutkimuksen johtopäätökset sekä arvioinnin, minkä lisäksi tutkimukselle annetaan myös jatkotutkimusehdotukset.

2 MARKKINOINNIN DIGITAALINEN MUUTOS

2.1 Markkinointi ja sen digitalisoituminen

Digitaalisten teknologioiden ketterä kehittyminen on muuttanut niin yritysten toimintaympäristöä kuin kuluttajien tottumuksia laajalla rintamalla (Kannan & Li, 2017). Teknologisten innovaatioiden myötä myös markkinointikanavat ovat muuttaneet muotoaan (Key, 2017). Digitaalinen kehitys on luonut markkinoijille uusia mahdollisuuksia tavoittaa tärkeimpiä sidosryhmiä, kuten kuluttajia, aiempaa tehokkaammin ja kohdistetummin (Key, 2017). Kuluttajat ovat yritysten lailla olleet suurten muutosten edessä, kun digitalisaation vaikutukset ovat muuttaneet ja kehittäneet myös heidän ympäristöä ja sen myötä käyttäytymistään (Dwivedi ym., 2020). Dwivedi ym. (2020) havainnollistavat tätä ilmiötä digitaalisen kuluttajakulttuurin (*Digital Consumer Culture*) käsitteellä, joka on markkinoinnissa vielä varsin uusi ja koskematon tutkimusalue.

Lukuisten mahdollisuuksien lisäksi, digitaalinen vallankumous on tuonut mukanaan yrityksille myös haasteita avaamalla markkinat globaalille kilpailulle toimialaan katsomatta (Caliskan ym., 2020). Tästä johtuen yritysten on tärkeä kohdistaa resurssejaan hyödyntämällä uusimpia teknologisia ratkaisuja, jotta kilpailukykyä pystytään ylläpitämään parhaalla mahdollisella tavalla (Caliskan ym., 2020). Markkinoinnin alalla tämä on tarkoittanut digitaalisen markkinointiin eli digimarkkinointiin siirtymistä sekä muiden teknologisten innovaatioiden, kuten tekoälyratkaisujen integroimista (Hurwitz ym., 2015; Kannan & Li, 2017).

Dataa syntyy digitaalisessa toimintaympäristössä valtavia määriä, mitä yritykset pyrkivät hyödyntämään mm. asiakasymmärryksessä, markkinoinnin tehokkuuden mittaamisessa ja markkinointistrategioiden suunnittelussa (Kannan & Li, 2017). Datan määrä on saavuttanut pisteen, jossa ihminen ei enää kykene käsittelemään sitä tarpeeksi tehokkaasti, mihin teknologiset innovaatiot, kuten tekoäly, on tuonut tarvittavia ratkaisuja (Hurwitz ym., 2015). Nykypäivänä puhutankin usein dataohjautuvasta markkinoinnista, jolla viitataan markkinointitoimien toteuttamiseen ja suunnitteluun datalähtöisesti (Shah & Murthi, 2021).

2.2 Digimarkkinointi

Digitaalisen markkinoinnin eli digimarkkinoinnin käsitteellä havainnollistetaan digitaalisten teknologioiden hyödyntämistä markkinoinnissa (American Marketing Association, 2020; Järvinen, Tollinen, ym., 2012; Wymbs, 2011). Digimarkkinoinnilla viitataan siis tyypillisesti markkinointitoimiin, joissa hyödynnetään digitaalisia kanavia, jotka useimmiten ovat yhteydessä Internetiin (American Marketing Association, 2020; Järvinen, Tollinen, ym., 2012; Wymbs, 2011). Myös termejä, kuten online- ja Internet-markkinointi, on käytetty

luonnehtimaan digimarkkinointia, vaikka nämä ottavat huomioon ainoastaan Internetiin yhteydessä olevan teknologian (Wymbs, 2011). Digimarkkinointi on näistä kiistatta kattavin käsite, minkä takia se on ottanut vahvan jalansijan akateemisessa kirjallisuudessa (Dwivedi, 2016; Kannan & Li, 2017).

Digimarkkinoinnin ja perinteisen markkinoinnin keskeisimpänä erona ovat digitaaliset teknologiat, jotka mahdollistavat markkinoinnin monipuolisen mittavuuden sekä helpottavat ja ovat muuttaneet asiakkaiden ja yritysten välistä kommunikaatiota (Wymbs, 2011). Digimarkkinointi on laaja kattotermi, joka pitää sisällään mm. asiakashankintaan, brändin tukemiseen, asiakkaiden säilyttämiseen sekä myynnin edistämiseen liittyvät prosessit, joiden avulla pyritään luomaan arvoa asiakkaille sekä muille sidosryhmille (Kannan & Li, 2017). Digimarkkinointia voidaan määritellä myös hieman osallistavamman näkökulman kautta luonnehtien sitä digitaalisten teknologioiden mahdollistamaksi prosessiksi, joka sallii yhteistyön asiakkaiden ja kumppanien kanssa luoden, kommunikoiden, toimittaan ja ylläpitäen arvoa sidosryhmille (Kannan & Li, 2017). Digitaalisen Markkinoinnin Instituutti (DMI) on jäsentänyt digimarkkinoinnin määritelmää seuraavanlaisesti: ”digitaalisten teknologioiden käyttämistä integroidun, kohdistetun ja mitattavan kommunikoinnin luomiseksi, jonka avulla mahdollistetaan asiakkaiden hankintaa ja ylläpitoa, samalla rakentaen entistä vahvempia asiakassuhteita” (Wymbs, 2011).

Digimarkkinoinnin lukuisat eri kanavaratkaisut ovat luoneet yrityksille uusia mahdollisuuksia tavoittaa arvokkaita sidosryhmiään, kuten kuluttajia, minkä myötä markkinoinnista on tullut entistä tärkeämpi osa yrityksen liiketoimintaa (Kannan & Li, 2017). Näiden digitaalisten kanavien kautta syntyy suuria määriä dataa, jota voidaan hyödyntää markkinoinnin tehokkuuden mittaamisessa, jonka on jo pitkään koettu olevan markkinoinnin haasteena (Järvinen, Töllinen, ym., 2012). Näin ollen seuraavissa luvuissa esitellään tarkemmin digimarkkinoinnin kanavia sekä digimarkkinoinnin tehokkuuden mittaamista.

2.3 Digimarkkinoinnin kanavat

Jotta yrityksen markkinointistrategian ja -taktiikoiden suunnittelulle sekä toteutukselle voidaan luoda parhaat mahdolliset edellytykset, on digimarkkinoinnin tyypillisiin ominaisuuksiin ja ilmentymiin syytä perehtyä huolellisesti (Taiminen & Karjaluoto, 2015). Digimarkkinoinnin kanavaratkaisut ja niiden painotukset ovat usein riippuvaisia liiketoimintakontekstista sekä markkinoinnin asiantuntijoiden näkemyksistä, minkä takia kaiken kattavia yleistyksiä tai ohjeita ei niistä monesti löydy (Desai, 2019). Digimarkkinointi on laaja käsite, joka kattaa lukuisia kanavia ja tekniikoita liiketoiminnan edistämiseksi, joita ovat mm. sosiaalisen median markkinointi, sisältömarkkinointi, markkinoinnin automaatio, hakukonemarkkinointi, sähköpostimarkkinointi, PPC eli klikkiperusteinen mainonta, natiivimainonta, inbound-markkinointi sekä kumppanuusmarkkinointi (American Marketing Association, 2020; Chaffey & Smith, 2013; Desai, 2019).

Sähköpostimarkkinointi, sosiaalisen median markkinointi sekä hakukonemarkkinointia ovat suosittuja ja yleisiä digimarkkinointimetoja (American Marketing Association, 2020; Key, 2017). Sähköpostimarkkinointi on suora digitaalinen kanava, jolla ylläpidetään tiedottavaa, myynnillistä sekä asiakassuhteita ylläpitävää markkinointia (Key, 2017). Perinteisen myyntityön sijasta, sosiaalisen median markkinointi on lähempänä tarinankerrontaa (Taiminen & Karjaluo, 2015). Sosiaalinen media määrittää eri kanaloiden ja alustojen joukoksi, jonka avulla toimitetaan ja jaetaan moninaista sisältöä (Taiminen & Karjaluo, 2015). Sosiaalinen media mahdollistaa teksti-, kuva-, video- ja äänisisällön vaivattoman jakamisen käyttäjien kesken (Key, 2017). Hakukonemarkkinointia voidaan kuvata yhtenä markkinoinnin muotona, jossa pyritään saamaan näkyvyyttä hakukoneen, kuten Googlen ja Yahoosivustojen hakutulosten sivulla (Key, 2017). Se pitää sisällään hakukonemarkkinoinnin ja -optimoinnin, joista ensimmäisenä mainitulla viitataan maksettuun, monesti klikkiperusteiseen hinnoitteluun pohjautuvaan mainontaan, jonka avulla pyritään esiintymään mahdollisimman suotuisasti hakukonetuloksissa (Key, 2017; Pan ym., 2011). Hakukoneoptimoinnilla taas tarkoitetaan prosesseja, joilla pyritään sijoittumaan hakukoneen organisaation tuloksissa mahdollisimman hyvin (Taiminen & Karjaluo, 2015). Hakukonemarkkinoinnin tarkoituksena on tiedottaa haun tehneitä ihmisiä yritysten palveluista, tuotteista sekä muusta tärkeästä informaatiosta, tavoitteena sijoittua ja erottautua parhaalla mahdollisella tavalla hakutulosten sivulla (Key, 2017; Pan ym., 2011; Taiminen & Karjaluo, 2015).

Digimarkkinoinnissa hyödynnettävät mediat voidaan tyypillisesti jakaa kolmeen luokkaan: maksettuun, omistettuun ja ansaittuun mediaan. Maksetulle medialle on ominaista, että yritykset ostavat mediatilaa tarkoituksenaan tavoittaa yleisöä sekä saada käyttäjiä tai konversiota. Omistetulla medialla viitataan yrityksen omistamaan mediaan, kuten verkkosivuihin, mobiilisovelluksiin, blogeihin tai sähköpostilistoihin, jotka ovat yrityksen hallinnassa. Ansaittu media taas on perinteisesti tarkoittanut yrityksen julkista kuvaa, jota on pyritty hallitsemaan PR-toimien avulla. Nykyään ansaittu media pitää sisällään myös digitaalisen WOM:in, joka muodostuu mm. erilaisten yhteisöjen, sosiaalisen median markkinoinnin sekä siellä käytävien keskustelujen kautta. (Chaffey & Ellis-Chadwick, 2019)

Taiminen ja Karjaluo (2015) jaottelevat digitaaliset kanavat sen mukaan, kuinka vahvasti yrityksen on mahdollista hallita kanavan kommunikaatiota, sekä onko tämä kommunikaatio yksi- vai kaksisuuntaista (Taulukko 1). Yksisuuntaisia ja yritysten vahvasti hallittavia kanavia ovat mm. yrityksen verkkosivut, bannerimainonta ja sähköposti. Vahvan kontrollin kaksisuuntaisia kanavia ovat esimerkiksi yrityksen blogit sekä yhteisöt, joissa yritykset toimivat vuorovaikutuksessa muiden ihmisten kanssa. Matalan kontrollin yhdensuuntaisina kanavina pidetään mm. hakukonemarkkinointia ja -optimointia, jotka jo aiemmin esittelyn mukaisesti mahdollistavat yrityksen sijoittumisen hakukoneiden tuloksissa mahdollisimman suotuisalla tavalla. Sosiaalisen median kanavat taas ovat tyypillinen esimerkki kahdensuuntaisista ja matalan kontrollin digitaalisista

kanavista, sillä yritykset ja kuluttajat voivat olla sosiaalisessa mediassa helposti vuorovaikutuksessa. (Taiminen & Karjaluoto, 2015)

TAULUKKO 1 Digitaalisten kanavien luokittelu (Taiminen & Karjaluoto, 2015)

	Matala kontrolli	Korkea kontrolli
Yksisuuntainen kommunikaatio	Hakukonemarkkinointi	Verkkosivut, bannerimainonta & sähköposti
Kaksisuuntainen kommunikaatio	Sosiaalinen media	Blogit & yhteisöt

2.4 Digimarkkinoinnin mittaaminen

Markkinoinnin suorituskyvyn mitattavuuden haasteet ovat digitalisaation myötä helpottaneet, kun uudet teknologiat sekä siirtyminen digitaalisiin kanaviin ovat tehneet datan keräämisestä kaikille mahdollista (Järvinen, Töllinen, ym., 2012). Saatavilla olevan datan määrä sekä lukuisat markkinoinnin suorituskyvyn mittarit ovat antaneet markkinoinnin ammattilaisille aiempaa vakuuttavampia työkaluja markkinoinnin tulosten todentamiseksi, mikä on vahvistanut markkinoinnin roolia liiketoiminnassa (Spiller & Tuten, 2015).

Frösén ym. (2016) määrittelevät markkinoinnin suorituskyvyn mittaamisen johdon työkaluksi, jolla markkinoinnille asetetaan yrityksen liiketoiminnan tehokkuuteen suhteutetut mittarit, joita lopulta hyödynnetään markkinoinnin suorituskyvyn arvioimisessa. He tarkentavat yrityksen tehokkuuden liiketoimintakentällä koostuvan taloudellisesta tuloksesta, asiakkaiden asenteista sekä suhteellisesta suorituskyvystä kilpailijoihin verrattuna. Jos markkinoinnin sijaan puhutaan rajatummin vain digimarkkinoinnista ja sen mittaamisesta, voidaan edellä esitelty määrittely rajata myös koskemaan ainoastaan sen eri kanavien ja toimintojen suorituskykyä (Järvinen, 2016).

Digitaalinen analytiikka on määrällisen ja laadullisen datan keräämistä eri digitaalisen median kanavien joukosta, sen ollen keskeisessä roolissa digimarkkinoinnin suorituskyvyn mittaamisesta (Järvinen, 2016). Se on teknologioiden mahdollistama, datan ja prosessien analysoinnin väline, jossa hyödynnetään uusimpia teknologisia ratkaisuja, kuten tekoälyä, koneoppimista ja lohkoketjuja (Gupta ym., 2020). Lisäksi digitaaliseen analytiikkaan voidaan lukea myös on- ja offline -datalähteiden hyödyntäminen personoidun sitouttamisen suunnittelussa ja toteutuksessa (Gupta ym., 2020). Digitaalisen analytiikan juuret lähtevät kuitenkin web-analytiikasta, joka nähdään Internet-datan keräämisinä, mittaamisena, analysointina ja raportointina kuluttajien tai yritysten verkkokäyttäytymisen ymmärtämiseksi ja optimoimiseksi (Järvinen, 2016; Waa, 2008).

Web-analytiikka esiintyy usein digimarkkinoinnin suorituskyvyn tyypillisenä mittaus- ja analysointikäsitteenä ja se nähdään usein kattoterminä juuri liiketoimintaan liittyvän digitaalisen datan mittaamiselle, analysoinnille ja optimoinnille (Järvinen, 2016). Koska web-analytiikka viittaa vahvasti

verkkosivuston analytiikkaan ja digimarkkinointi sisältää laajasti myös muita kanavia, kuten sosiaalisen mediaa ja hakukonemarkkinointia, on digitaalisesta analytiikasta tullut vähitellen web-analytiikkaa korvaava kattotermin (Chaffey & Patron, 2012; Järvinen, 2016).

3 MARKKINOINNIN TEKOÄLY

3.1 Tekoäly

Markkinoinnin tekoälyyn keskittyvä tutkimus on viime vuosina ollut kovassa kasvussa (Shankar, 2018). Tekoälyteknologian koetaan olevan yhteiskunnan ja liiketoiminnan muutoksen eturintamassa, ja se on saavuttanut yleisessä keskustelussa suosituksen ja trendikkään leiman (Campbell ym., 2020; Shankar, 2018). Vaikka tekoäly onkin tällä hetkellä varsin ajankohtainen aihe, ulottuvat sen juuret 1950-luvulle, jolloin Alan Turing esitteli ajatuksen koneiden mahdollisuudesta ajatella (Ma & Sun, 2020; Wirth, 2018). Tekoäly eli *Artificial Intelligence (AI)*, sai käsitteenä syntynsä nopeasti tämän jälkeen vuonna 1955, kun John McCarthy esitteli kyseisen käsitteen Dartmouthin yliopiston konferenssissa ensimmäistä kertaa maailmalle (Ma & Sun, 2020; Wirth, 2018).

Tekoäly on moniulotteinen ja laaja käsite, minkä takia sen selkeä määrittely on tärkeää (Wirth, 2018). Sillä viitataan algoritmeihin, koneisiin, ohjelmiin sekä järjestelmiin, jotka osoittavat älykkyyttä (Shankar, 2018). Tekoälylle ominaista on sen kokemusperusteinen, ihmisen kaltaista älykystä käytöstä jäljittelevä oppimiskyky, joka ilmenee edellä mainittujen tekoälyn sovellusten toiminnassa (Huang & Rust, 2021; Syam & Sharma, 2018). Tekoäly koostuu laajasta teknologioiden joukosta, jolla pyritään ratkomaan ongelmia ja tekemään päätöksiä hyödyntäen sille syötettyä dataa (Campbell ym., 2020; Huang ym., 2019). Se mahdollistaa koneiden oppia ja suoriutua niille annetuista tehtävistä lähes ihmisten kaltaisesti (Campbell ym., 2020).

Tekoäly perustuu moniin sille keskeisiin teknologioihin, joita ovat mm. koneoppiminen, syväoppiminen, sääntöpohjaiset asiantuntijajärjestelmät, luonnollisen kielen käsittely, neuroverkot, robotit ja ohjelmistorobotiikka (Davenport ym., 2020). Eri teknologiat koetaan useimmiten tekoälyn alakategorioiksi, ja tekoäly itsessään mielletään enemmänkin kattotermiksi (Davenport ym., 2020). Tekoälyn eri teknologiat mahdollistavat sille syötetyn ulkoisen datan tulkitsemisen, siitä oppimisen ja datan sisältämän tiedon omaksumisen joustavasti (Kaplan & Haenlein, 2019). Koneoppiminen on yksi olennainen osa tekoälyä, sillä se mahdollistaa tietokoneiden oppimisen ilman täsmällistä ohjelmointia, mitä tyypillisesti erilaisissa tekoälyn ilmentymissä tarvitaan (Kaplan & Haenlein, 2019). Suuri osa tekoälyratkaisuista hyödyntää koneoppimista, minkä myötä koneoppiminen on vahvasti tekoälyn teknologioiden keskiössä (Huang & Rust, 2018; Overgoor ym., 2019).

Tekoälyä voidaan tarkastella eri teknologioiden lisäksi myös liiketoiminnan ilmentymien avulla, kuten esimerkiksi datasta muodostettavien näkemysten, liiketoiminnan automatisoinnin sekä työntekijöiden ja asiakkaiden sitouttamisen kautta. Davenport ja Ronanki (2018) esittävät, kuinka ohjelmistorobotiikkaa hyödyntävistä automaatioista hallinnollisissa ja taloudellisissa tehtävissä on tullut tyypillisin tekoälyn liiketoiminnallinen ilmentymä. Toiseksi yleisin

liiketoiminnallinen ilmentymä on algoritmit ja niiden hyödyntäminen datasta löytyvien merkitysten havainnoinnissa, kuten asiakkaan ostoaikomuksen ennustamisessa tai digitaalisten mainosten personoidussa kohdentamisessa. Työntekijöiden ja asiakkaiden sitouttamiseen liittyvät toimet, kuten chatbottien käyttö sivustoilla tai erilaiset tuotteiden suosittelujärjestelmät, on tekoälyn liiketoiminnallisista ilmentymistä ollut kolmanneksi yleisin. (Davenport & Ronanki, 2018)

Tekoäly voidaan tyypillisesti jakaa myös sen kehityksen vaiheiden mukaan heikkoon ja vahvaan tekoälyyn (Kaplan & Haenlein, 2019; Tariq & Abonamah, 2021; Wirth, 2018). Lisäksi tunnistetaan myös kolmas ja tekoälyn korkein taso supertekoäly, jolla viitataan tekoälyyn, joka suoriutuu ihmistä paremmin jokaisella osa-alueella (Barrett & Baum, 2017; Gill, 2016; Kaplan & Haenlein, 2019). Heikko tekoäly on tyypillisesti suunniteltu jotain tiettyä ongelmaa tai tehtävää varten, kuten Facebookin kuvantunnistus ja Sirin äänentunnistus (Kaplan & Haenlein, 2019; Wirth, 2018). Vahvalla tai yleisellä tekoälyllä (*Artificial General Intelligence, AGI*) taas nähdään olevan kyky järkeillä, suunnitella ja ratkoa itsenäisesti erilaisia tehtäviä, joihin sitä ei varsinaisesti alun perin olisi edes suunniteltu (Kaplan & Haenlein, 2019; Wirth, 2018). Tätä tekoälyn tasoa ei kuitenkaan koeta olevan vielä saavutettu, minkä takia tekoälyllä viitataan yleisimmin juuri heikkoon tekoälyyn (Kaplan & Haenlein, 2019; Wirth, 2018). Jos tekoälyyn viitataan tulevaisuuden ylivoimaisena voimana, joka voittaa ihmisen tehtävässä kuin tehtävässä, puhutaan tällöin supertekoälystä (Barrett & Baum, 2017; Gill, 2016).

Huang ja Rust (2018) tarkastelevat tekoälyä markkinoinnin näkökulmasta sen historiallisen kehityksen kautta (Taulukko 2). He jakavat tekoälyn neljään kategoriaan sen älykkyyden kehityksen vaiheisiin pohjautuen. Nämä neljä kategoriaa sen kehityksen kronologisessa järjestyksessä ovat mekaaninen, analyttinen, vaistonvarainen ja empaattinen älykkyys. Mekaaninen älykkyys pystyy suorittamaan automaattisesti rutiininomaisia tehtäviä, mutta ei esiinny kovinkaan älykkäänä. Tähän voidaan lukea mm. tarjoilijoiden ja vähittäiskaupan myyjien työtehtävät, joita tekoäly pystyy mekaanisen älykkyyden tasolla suorittamaan. Analyttisellä älykkyydellä viitataan kykyyn käsitellä dataa ja hyödyntää sitä ongelmanratkaisuun liittyvissä tehtävissä. Näitä ovat tyypillisesti esim. matemaatikkojen, insinöörien ja tilintarkastajien työtehtävät, joissa vaaditaan varsin analyttisiä taitoja. Vaistonvarainen älykkyys on kykyä ajatella luovasti ja sopeutua uusiin tilanteisiin. Tämän nähdään perustuvan kokonaisvaltaiseen ja kokemuspohjaiseen taitoon ajatella, jota tyypillisesti nähdään esimerkiksi lakimiehillä, markkinointipääälliköillä tai johdon konsulteilla. Empaattinen älykkyys on kykyä tunnistaa ja ymmärtää muiden ihmisten tunteita, reagoida niihin asianmukaisesti sekä vaikuttaa muiden tunteisiin. Vastaavanlaisia ihmissuhdetaitoja ja sosiaalisia taitoja sisältäviä tehtäviä löytyy esimerkiksi poliitikkojen ja psykiatrien työnkuvista. (Huang & Rust, 2018)

TAULUKKO 2 Tekoäly historiallisen älykkyyden kehityksen mukaan (Huang & Rust, 2018)

Tekoälyn kehitysvaihe	Kuvaus	Työtehtävä
Mekaaninen älykkyyys	Rutiininomaisten tehtävien suorittaminen automaattisesti.	Esim. tarjoilijat ja vähittäiskaupan myyjät
Analyttinen älykkyyys	Datan hyödyntäminen ja käsittely ongelmanratkaisuun liittyvissä tehtävissä.	Esim. tilintarkastajat ja matemaatikot.
Vaistonvarainen älykkyyys	Luova ajattelu ja sopeutuminen uusiin tilanteisiin, perustuen kokonaisvaltaiseen ja kokemuspohjaiseen ajatteluun.	Esim. liikkeenjohdon konsultit ja markkinointijohtajat.
Empaattinen älykkyyys	Muiden ihmisten tunteiden tunnistaminen ja ymmärrys, niihin asianmukainen reagoiminen sekä niihin vaikuttaminen.	Esim. poliitikot ja psykiatrit.

3.2 Markkinoinnin tekoälysovellukset

Tekoälyn uskotaan vaikuttavan huomattavalla tavalla markkinointiin sekä sen strategiseen suunnitteluun, myynnin prosesseihin, asiakaspalveluun sekä liiketoimintamallien valintaan (Davenport ym., 2020). Tekoälyä on kasvavin määrin jo hyödynnetty markkinoinnin eri tehtävien tehostamisessa, kuten markkinointikampanjoiden suorituskyvyn nostamisessa, asiakaskokemuksen kehittämässä ja asiakasymmärryksen keräämisessä (Thontirawong & Chinchachokchai, 2021).

Datan määrä on kasvanut globaalisti hurjaa vauhtia, johtaen siihen, etteivät ihmiset pysty käsittelemään tai analysoimaan sitä enää tarpeeksi tehokkaasti (Hurwitz ym., 2015). Tekoäly tarjoaa niin markkinoinnin kuin monelle muulle toimialalle ratkaisuja tähän ongelmaan, ja sitä hyödynnetään jo mittavissa määrin monissa eri markkinoinnin tehtävissä (Campbell ym., 2020). Tekoälyä on hyödynnetty tyypillisesti varsinkin markkinoinnin operatiivisissa tehtävissä, kuten asiakaskohdennuksessa ja -analysoinnissa sekä mainosisältöjen valitsemisessa (Stone ym., 2020). Myös asiakaspalvelun chatbotit, digimarkkinoinnin sisältöjen valinnassa auttavat suosittelujärjestelmät, kuluttajien tiettyjä ominaisuuksia tunnistavat mallit tai työkalut, jotka mallintavat mainoskampanjoiden potentiaalisia lopputulemia, ovat tyypillisiä markkinoinnin tekoälyn tyypillisiä sovelluksia (Overgoor ym., 2019).

Markkinoinnin tekoälysovelluksia ilmenee jatkuvasti enemmän ja enemmän, minkä myötä Nair ja Gupta (2020) esittelevät nykypäivän tyypillisimpiä digimarkkinoinnin tekoälysovelluksia. Näitä ovat mm. ohjelmallinen ostaminen, älykkäät sisällöt, chatbotit, äänentunnistus ja äänihaku, liidien pisteytys, mainosten kohdistaminen, dynaaminen hinnoittelu, uudelleen kohdistaminen ja

markkinoinnin automaatio, joista muutamaa avataan hieman tarkemmin seuraavassa kappaleessa.

Tekoälyn avulla personoitu älykäs sisältö on yksi hyödyllinen tapa sitouttaa potentiaalisia asiakkaita verkkosivun sisältöihin, näyttämällä heille kiinnostavia sisältöä, kuten verkkokaupan tuotteita. Esimerkiksi edellisiin verkkosivuston hakuihin tai muiden asiakkaiden preferensseihin perustuvat suositukset ovat tästä hyvä esimerkki (Nair & Gupta, 2020). Yksi tyypillinen digitaalisen mainonnan esimerkki tekoälysovelluksesta on ohjelmallinen ostaminen, joka on digitaalisen mainostilan ostamista ja myymistä automaatioteknologiaa hyödyntäen ilman varsinaisia mediamyymiä (Davenport ym., 2020). Se on suosittu ja tehokkuutta lisäävä mainonnan muoto, joka hyödyntää dataa tarjotakseen sivustojen käyttäjille heille sopivat mainokset oikeassa paikassa, oikeaan aikaan (Nair & Gupta, 2020). Dynaamisessa hinnoittelussa tehdään reaaliaikaisia hinnoittelupäätöksiä koneoppimisen algoritmeja hyödyntäen, minkä myötä asiakkaille voidaan personoida juuri heille sopivia tarjouksia (Nair & Gupta, 2020). Yksittäiset kuluttajat voivat monesti olla valmiita maksamaan tuotteista tai palveluista eri suuruisia hintoja, mihin dynaaminen hinnoittelu antaa ratkaisun tarjoamalla yritykselle tulosta maksimoivan sekä kuluttajalle miellyttävän hintatarjouksen (Li ym., 2020; Nair & Gupta, 2020).

3.3 Koneoppiminen

Koneoppiminen eli *Machine Learning (ML)* on tekoälyn yksi alakategoria, johon tekoälyyn liittyvä tutkimus on keskittynyt vahvasti 1990-luvun jälkeen (Ma & Sun, 2020; Syam & Sharma, 2018). Koneoppiminen keskittyy tietokonejärjestelmien ympärille, jotka pystyvät oppimaan ja kehittymään oman kokemuksen kautta (Thontirawong & Chinchanchokchai, 2021). Koneoppimisen avulla on mahdollista ennustaa tulevaisuuden kehityksen kulkua ja tukea päätöksentekoa prosessoimalla sille syötettyä dataa (Miklosik ym., 2019). Yksi koneoppimisen tunnetuimmista määritelmistä on Mitchellin vuonna 1997 esittämä, jossa hän tunnistaa koneoppimiseksi toiminnan, jossa tietokoneohjelma oppii ja kehittyy kokemuksesta E (Experience), jos sen suorituskyky tehtävissä T (Tasks) kasvaa suorituskyvyn mittareilla P (Performance measures) mitattuna. Kun tietokoneelle syötetään dataa, se oppii kokemuksistaan ja kehittyy sille ohjatussa tehtävässä, voidaan kyseistä toimintaa kutsua koneoppimiseksi. Koneoppiminen nähdäänkin nykypäivän yleisimpänä tekoälyn teknologisenä ilmentymänä (Huang & Rust, 2018; Overgoor ym., 2019).

Koneoppiminen on olennainen osa tekoälyä, vaikka tekoäly ei perustukaan ainoastaan siihen (Kaplan & Haenlein, 2019). Koneoppiminen on laaja tieteen ja teknologian ala, joka mahdollistaa tietokoneiden oppimisen ilman, että niitä olisi pääasiassa ohjelmoitu siihen (Shah ym., 2020). Koneoppiminen keskittyy sellaisten tietokoneohjelmien kehittämiseen, joille datan käsittely ja hakeminen on mahdollista, sillä koneiden oppiminen perustuu suurilta osin dataan ja sen käsittelyyn (Shah ym., 2020). Koneoppimisella viitataan usein tietokoneen kykyyn

oppia kokemuksen perusteella, kuten esimerkiksi muuttamalla sen prosessointia vasta saadun tiedon perusteella (Feng ym., 2021). Tietokoneen suorittaessa sille ohjattua tehtävää, koneoppimisen algoritmit tunnistavat datasta havaittavia kaavoja ja oppivat niiden avulla (Feng ym., 2021). Algoritmit yhdistelevät historiallista dataa tuoreemman datan kanssa, ja pystyvät näin oppimaan ennakoitukykyä sekä ymmärtämään erilaisia suhteita (Hair & Sarstedt, 2021).

Markkinoinnin koneoppimisen sovelluksia voidaan hyödyntää esimerkiksi tuotesuosittelussa personoidun asiakaskokemuksen luomiseksi (Thontirawong & Chinchachokchai, 2021). Koneoppimisen sovelluksia käytetään myös esimerkiksi asiakaskohdentamisessa, sisältöjen suosittelussa, mainosten hintatarjousten optimaalisen pisteen löytämisessä sekä asiakasviestinnässä (Siau & Yang, 2017). Koneoppimisen avulla datasta voidaan havaita erilaisia kaavoja ja malleja, joita voidaan käyttää esimerkiksi markkinoinnin strategisissa päätöksissä (Thontirawong & Chinchachokchai, 2021).

Datan laatu ja sen oikeaoppinen käyttö ovat huomattavan tärkeitä elementtejä niin koneoppimiseen kuin tekoälyyn pohjautuvissa ratkaisuissa (Hair & Sarstedt, 2021). Koneoppimisen algoritmeille syötettävä datan on oltava laadukasta ja validia, sillä ilman korkealaatuista dataa, algoritmit eivät vastaanota datasta toivottuja signaaleja eivätkä sen myötä kehity tai opi halutulla tavalla (Hair & Sarstedt, 2021). Tyypillisiä datan heikkoon laatuun vaikuttavia elementtejä ovat mm. puuttuvana, puolittaisena, päällekkäisenä, virheellisenä tai vanhentuneena ilmentyvänä datana (Gudivada ym., 2017).

Taulukossa 3 havainnoidaan, miten koneoppiminen jaetaan tyypillisesti kolmeen eri luokkaan: ohjattuun oppimiseen, ohjaamattomaan oppimiseen sekä vahvistusoppimiseen (Campbell ym., 2020; De Bruyn ym., 2020; Overgoor ym., 2019). Ohjattu oppiminen on koneoppimisen käytetyin oppimismetodi, jolla viitataan koneen oppimiseen esimerkkien avulla (Campbell ym., 2020; Jordan & Mitchell, 2015). Ohjatussa oppimisessa syötetty data erotellaan opetus- ja testidataan, joita hyödyntämällä se koulutetaan ratkomaan sille määritellyjä ongelmia (De Bruyn ym., 2020; Overgoor ym., 2019). Opetusdata pitää sisällään selitettävät ja selitettävät muuttujat, joista jälkimmäisen arvot ovat selvillä (De Bruyn ym., 2020; Overgoor ym., 2019). Tätä dataa hyödyntäen kone muodostaa mallin, jonka tavoitteena on ennustaa datan selitettäviä muuttujia mahdollisimman hyvin (De Bruyn ym., 2020; Overgoor ym., 2019). Oppimisprosessin päätteeksi kone on luonut mallin, jonka avulla se voi vastata määriteltyihin kysymyksiin, eikä tarvitse aiemmin syötettyä opetusdataa enää tehtävien ratkomiseen, vaan pystyy muodostamaan ratkaisuja myös uudesta datasta (De Bruyn ym., 2020; Overgoor ym., 2019). Esimerkkejä ohjatusta oppimisesta ovat mm. luokittelu ja päätöspuut, ja käytännön tasolla chatbotit, jotka oppivat asiakkaiden esittämistä yleisimmistä kysymyksistä (Campbell ym., 2020; Overgoor ym., 2019). Ohjatussa oppimisessa on hyvä huomioida datan paikkansapitävyyden tärkeys, sillä sen merkitys oppimisprosessille on ensiarvoista (Adnan & Akbar, 2019). Jos data on heikkolaatuista tai väärää, ei kone opi muodostamaan ratkaisumalleja halutulla tavalla (Adnan & Akbar, 2019).

Toisin kuin ohjatussa oppimisessa, ohjaamattomassa oppimisessa koneella on tiedossa vain selittävät muuttujat, muttei varsinaista selitettävää muuttujaa (Campbell ym., 2020; Kaplan & Haenlein, 2019; Overgoor ym., 2019). Ohjaamattomassa oppimisessä ei yritetä ennustaa ennalta asetettuja lopputulemia, vaan siinä keskitytään datan analysointiin ilman vastaavanlaisia tavoitetta (Campbell ym., 2020). Ohjaamattoman oppimisen keskiössä onkin datasta löydettävien yhteyksien sekä ryhmien muodostaminen (Paschen ym., 2019). Ryhmittelyanalyysi tai klusterianalyysi on tyypillinen esimerkki ohjaamattomasta oppimisestä, minkä lisäksi käytännön esimerkkejä löydetään mm. puheentunnistusteknologiasta, jota Applen Siri ja Amazonin Alexa käyttävät (Kaplan & Haenlein, 2019).

Vahvistusoppiminen ja ohjaamaton oppiminen ovat luonteiltaan varsin samankaltaisia (Overgoor ym., 2019). Suurin ero näiden välillä on kuitenkin vahvistusoppimiselle ominaisessa palautteenannossa (De Bruyn ym., 2020; Overgoor ym., 2019; Sutton & Barto, 2018). Vahvistusoppimisessa koneen algoritmit saavat palautetta tehtyjen toimintojen perusteella, ja oppivat ajan myötä tekemään päätöksiä ympäristössä, jossa pyrkimyksenä on maksimoida palkintojen ja minimoida rangaistusten määrä (Sutton & Barto, 2018). Vahvistusoppimisen avulla on pystytty opettamaan pelejä, kuten shakkia, ja se on lisäksi erittäin tärkeä osatekijä robottien ja autonomisten ajoneuvojen toiminnassa (De Bruyn ym., 2020). Facebookin konversio seuranta pikselin käytön olevan yksi esimerkki vahvistusoppimisestä, jossa algoritmi testaa aluksi mainoksen suoriutumista laajasti koko kohderyhmällä, minkä jälkeen se pystyy optimoimaan mainontaa saamiensa signaalien kautta (Campbell ym., 2020).

TAULUKKO 3 Koneoppimisen lajit (Campbell ym., 2020; De Bruyn ym., 2020; Overgoor ym., 2019)

Koneoppimisen laji	Kuvaus	Käytännön esimerkki markkinoinnin tehtävistä
Ohjattu oppiminen	Syötetty data erotellaan opetus- ja testidataan, jota hyödyntämällä kone koulutetaan ratkomaan sille määritellyjä ongelmia. Kone muodostaa dataa hyödyntäen mallin, jonka avulla se ennustaa selitettäviä muuttujia mahdollisimman hyvin. Validi ja laadukas data on erityisen tärkeää, sillä heikkolaatuisella tai väärällä datalla kone ei opi muodostamaan haluttuja ratkaisumalleja.	Chatbotit, jotka oppivat vastaamaan asiakkaiden yleisimmin esittämiin kysymyksiin.
Ohjaamaton oppiminen	Koneella on tiedossa vain selittävät muuttujat, muttei varsinaista selitettävää muuttujaa. Ennalta asetettuja lopputulemia ei pyritä ennustamaan, vaan huomio kiinnitetään datan analysointiin, minkä lisäksi datasta löydetävien yhteyksien sekä ryhmien muodostaminen on keskiössä.	Puheentunnistusteknologian hyödyntäminen kaupallisessa tarkoituksessa, kuten Applen Siri ja Amazonin Alexa.
Vahvistusoppiminen	Varsin samankaltaista kuin ohjaamaton oppiminen. Näiden suurin ero kuitenkin vahvistusoppimiselle ominaisessa palautteenannossa. Algoritmit saavat tässä palautetta tehtyjen toimintojen perusteella ja oppivat ajan myötä tekemään päätöksiä ympäristössä, jossa pyrkimyksenä minimoida rangaistusten ja maksimoida palkintojen määrä. Sen avulla on pystytty opettamaan pelejä, kuten shakkia, minkä lisäksi se on tärkeä osatekijä robottien ja autonomisten ajoneuvojen toiminnassa.	Facebookin konversio seuranta-apin käyttäminen, jossa algoritmi kokeilee mainoksen suoriutumista aluksi koko kohderyhmälle, minkä jälkeen mainontaa pystytään optimoimaan saatujen signaalien kautta.

3.4 Markkinoinnin tekoälysovellusten implementointiprosessi

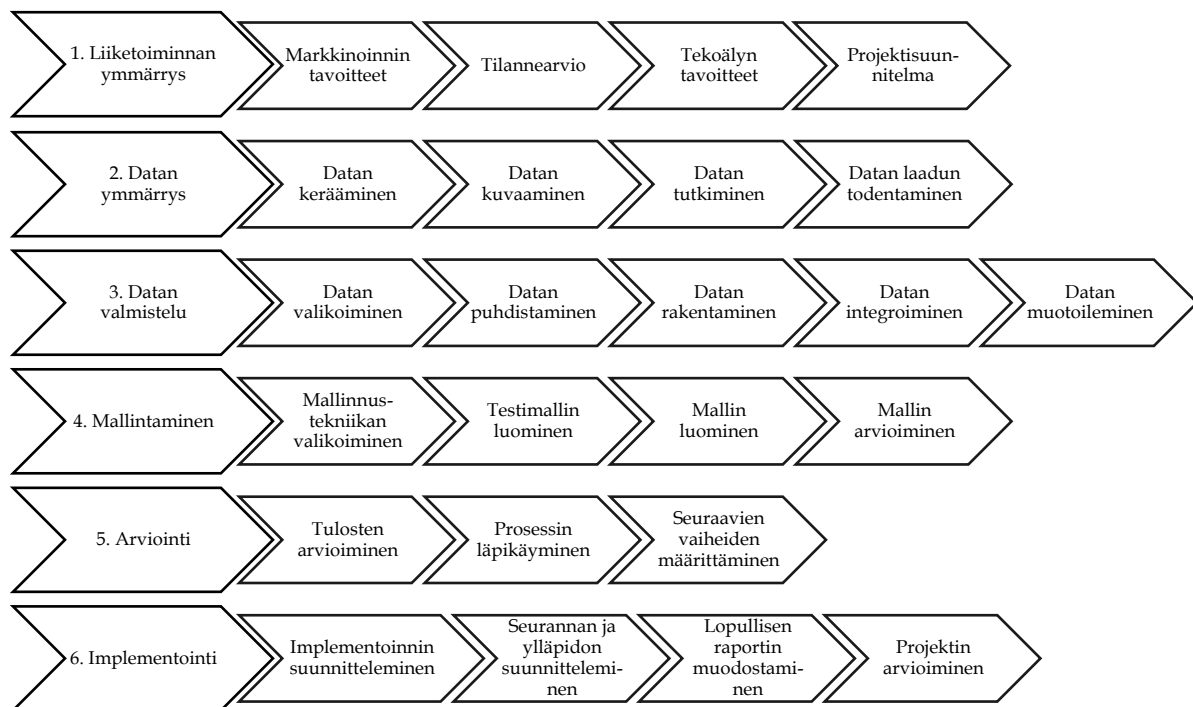
Markkinoinnin tekoälyyn keskittyvässä akateemisessa tutkimuksessa on käsitelty viime vuosina mm. tekoälysovellusten hyödyntämistä markkinoinnin operatiivisissa tehtävissä sekä markkinoinnin strategisena tukena (Brock & von Wangenheim, 2019; Klaus & Zaichkowsky, 2020; Overgoor ym., 2019; Shah ym., 2020). Aiempi tutkimus on tarjonnut hyviä esimerkkejä siitä, kuinka tekoälyteknologian avulla voidaan helpottaa ja ratkaista markkinoinnin erinäisiä tehtäviä

(Overgoor ym., 2019). Tekoälyn uskotaan saavuttavan lähitulevaisuudessa jopa sosiaalista mediaa suuremman vaikutuksen markkinoinnin toimialaan, mikä kuvastaa aiheen ajankohtaista sekä merkittävää luonnetta (Rust, 2020).

Uuden teknologian onnistunut käyttöönotto vaatii yritykseltä selkeästi määriteltyjä prosesseja, joita on helppo seurata sekä ymmärtää. Tekoälyn implementointia ja sen omaksumista on käsitelty markkinoinnin kirjallisuudessa jo aiemmin, mutta tekoälysovellusten implementointiprosessin kuvaaminen on jäänyt akateemisessa tutkimuksessa lähes olemattomaksi (Borges ym., 2021; Overgoor ym., 2019; Pappas ym., 2018). Jotta markkinoinnin tekoälyratkaisulla voidaan saavuttaa merkittäviä tuloksia, on markkinoinnin päätöksentekijöillä oltava tarvittavat valmiudet sen käyttöönottamiseksi (Overgoor ym., 2019).

Brock ja von Wangenheim (2019) havainnoivat, kuinka tekoäly otetaan tyyppillisesti käyttöön yritysten digitaalisen muutosprosessien yhteydessä, tukemaan niiden jo olemassa olevia liiketoimintoja. He esittävät myös DIGITAL-viitekehyksen onnistuneelle tekoälyn käyttöönotolle vastaavanlaisen digitaalisen muutosprosessin yhteydessä. Tutkijat havainnoivat tekoälysovellusten menestyksestä implementointia liiketoimintaan viitekehyksen seitsemän osa-alueen *Data, Intelligence, Grounded, Integral, Teaming, Agile, Leadership* kautta. Mitä vahvempi yritys on näillä osa-alueilla, sitä paremmat lähtökohdat sillä on tekoälysovellusten implementoinnin menestyksekkääseen toteuttamiseen. DIGITAL tarjoaa viitekehyksen menestyksekkään tekoälysovelluksen implementoinnin tarkasteluun, mutta ei kuvaa varsinaista implementointiprosessia kuitenkaan tarkemmin (Brock & von Wangenheim, 2019).

Overgoor ym. (2019) tarkastelevat tätä tutkimusaukkoa CRISP-DM (Cross-Industry Standard Process for Data Mining) -prosessimallia hyödyntäen, jonka avulla he havainnoivat markkinoinnin tekoälysovellusten implementointiprosessin vaiheita. CRISP-DM luotiin alun perin viitekehykseksi datanlouhinnan prosessien tarkasteluun, mutta sitä on sen myötä hyödynnetty menestyksekkäästi myös muilla liiketoiminnan osa-alueilla (Overgoor ym., 2019; Shearer, 2000). Kuvio 1 esittää CRISP-DM -prosessimallin vaiheet markkinoinnin tekoälysovellusten implementoinnissa, joita ovat liiketoiminnan ymmärrys, datan ymmärrys, datan valmistelu, mallintaminen, arviointi ja implementointi (Alsultanny, 2011; Haghghatnia ym., 2018; Overgoor ym., 2019; Shearer, 2000).



KUVIO 1 Markkinoinnin tekoälysovellusten implementointiprosessin vaiheet (Alsultanny, 2011; Haghghatnia ym., 2018; Overgoor ym., 2019; Shearer, 2000)

Markkinoinnin tekoälysovellusten implementointiprosessin ensimmäinen vaihe CRISP-DM-mallissa on liiketoiminnan ymmärrys, jossa määritellään markkinoinnin tavoitteet, jotka pyritään saavuttamaan tekoälyratkaisun avulla (Overgoor ym., 2019). Tätä seuraa kokonaisvaltaisen tilannearvion tekeminen senhetkisestä tilanteesta, minkä jälkeen on mahdollista määritellä tekoälyn ja sen implementointiprosessiin onnistumiseen liittyvät tavoitteet, joilla arvioidaan koko projektin suoriutumista (Overgoor ym., 2019). Ensimmäisen vaiheen loppuksi on syytä ottaa katsaus vielä koko prosessin läpivientiin sekä laatia projektisuunnitelma (Alsultanny, 2011; Overgoor ym., 2019; Shearer, 2000).

Mallin toinen vaihe on datan ymmärrys, joka alkaa datan keräämisellä (Haghghatnia ym., 2018; Overgoor ym., 2019; Shearer, 2000). Kerättyä dataa pyritään kuvaamaan yksityiskohtaisesti, minkä lisäksi hankitun datan ominaisuuksia, kuten datan muotoa ja määrää on syytä tutkia markkinoinnin tavoitteiden valossa (Overgoor ym., 2019; Shearer, 2000). Datan tarkemman tutkimisen jälkeen todennetaan sen laatu, esimerkiksi tarkastellen datan mahdollisia puutteita ja validiuutta (Overgoor ym., 2019; Shearer, 2000).

Datan valmistelu on viisiosainen vaihe, ja useimmiten markkinoinnin tekoälyn implementoinnissa aikaa vievin (Overgoor ym., 2019; Shearer, 2000). Se koostuu datan valikoinnista, puhdistamisesta, rakentamisesta, integroinnista sekä muotoilusta (Alsultanny, 2011; Shearer, 2000). Datan valikoinnilla viitataan päätöksentekoon juuri siitä datasta, jota halutaan sisällyttää markkinoinnin tekoälyratkaisun kehitykseen ja testaukseen. Kun datan valikoiminta on suoritettu, se on syytä käydä läpi ja puhdistaa yhteneväiseksi. Lisäksi dataa voi olla myös tarpeen

lisätä rakentamalla alkuperäisestä datasta johdettuja arvoja, jos sen ei nähdä soveltuvan mallintamiseen. Datan integrointi yhteen tiedostoon tai taulukkoon voi tulla myös tarpeeseen, jos se on hajautettu useaan eri paikkaan. Datan valmistelun viimeisessä vaiheessa voidaan dataa tarpeen tullen muotoilla asianmukaiseen muotoon. (Overgoor ym., 2019; Shearer, 2000)

Prosessin seuraava vaihe on mallintaminen, joka koostuu mallinnustekniikoiden valinnasta, testimallin ja varsinaisen mallin luomisesta sekä tämän arvioimisesta (Alsultanny, 2011; Overgoor ym., 2019; Shearer, 2000). Mallinnustekniikoiden valinnassa on tarkoitus löytää kyseessä olevaan markkinoinnin tehtävään sopiva mallinnustekniikka, jonka jälkeen mallin testaukselle luodaan testauskriteerit ja muodostetaan varsinainen testimalli (Overgoor ym., 2019; Shearer, 2000). Testimallin luomisessa datajoukko jaetaan tyypillisesti opetus- ja testidataan, joista ensimmäisenä mainittua käytetään mallin rakentamisessa ja testidataa sen laadun arvioinnissa (Overgoor ym., 2019; Shearer, 2000). Testimallin luomisen ja validoinnin myötä saadaan muodostettua varsinainen malli, joka tulee se vielä arvioida ennalta määritellyjä kriteerejä sekä asiantuntijoiden omaa näkemystä hyödyntäen (Shearer, 2000).

Ennen mallin varsinaista implementointia tulee se arvioida perusteellisesti läpi ja tarkastella sen rakennetta varmistaakseen, että se saavuttaa sille asetetut liiketoiminnalliset tavoitteet (Shearer, 2000). Arviointivaihe sisältää tulosten arvioinnin, prosessin läpikäynnin sekä projektin seuraavien vaiheiden määrittelyn (Alsultanny, 2011; Overgoor ym., 2019).

Tekoälyn implementointi on mallin viimeinen vaihe, jonka tarkoituksena on luoda liiketoiminnallista arvoa (Haghighatnia ym., 2018; Overgoor ym., 2019; Shearer, 2000). Implementointi on syytä suunnitella selkeästi läpi, jotta ymmärretään miten ja milloin implementointi tulisi toteuttaa (Overgoor ym., 2019). Ensimmäistä vaihetta seuraa tekoälyratkaisun ylläpidon ja seurannan suunnittelu, joka usein nähdäänkin tärkeässä roolissa suurissa muutosprojekteissa (Overgoor ym., 2019). Projektin päätteeksi on myös suotavaa laatia loppuraportti, jossa arvioidaan koko prosessin läpivienti (Overgoor ym., 2019; Shearer, 2000). Tässä voidaan tuoda esille mm. projektin onnistumiset ja epäonnistumiset, kuten myös mahdolliset kehitysehdotukset tulevaisuuden vastaavia projekteja varten (Shearer, 2000).

Overgoor ym. (2019) huomauttavat datan olevan markkinoinnin tekoälysovellusten keskiössä ja vaativan myös implementoinnin suunnittelussa ja läpiviennissä erityistä tarkkuutta. Se on tekoälyteknologian perusta ja myös CRISP-DM -prosessimallin tärkein elementti, joka luo edellytykset menestykselle tekoälysovellukselle (Overgoor ym., 2019). Digitaalisen datan valtava määrä on mahdollistanut tekoälysovellusten hyödyntämisen markkinoinnin eri tehtävissä, mutta luo samalla haasteita sen käsittelyyn ja hallintaan (Hair & Sarstedt, 2021). Seuraava luku esittelee markkinoinnin tekoälysovelluksissa hyödynnettävää dataa tarkemmin.

4 DATAN HYÖDYNTÄMINEN MARKKINOINNIN TEKOÄLYSOVELLUKSISSA

Data on markkinoinnin tekoälysovellusten perusta, sillä toimiakseen menestyksekkäästi, ne tarvitsevat suuria määriä laadukasta dataa (Brock & von Wangenheim, 2019). Tekoälyteknologian hyödyntäminen markkinoinnissa on yleistynyt datan määrän merkittävän kasvun myötä, jota digitalisaatio on vauhdittanut huomattavalla tavalla (Amado ym., 2018; Wedel & Kannan, 2016). Yritykset keräävät ja säilövät valtavasti dataa, mikä on johtanut datan hallintaan liittyviin haasteisiin (Amado ym., 2018). Maailmassa tuotetun datan määrä on ylittänyt ihmisten kyvyn omaksua, tulkita ja tehdä siihen perustuvia monimutkaisia päätöksiä (Hurwitz ym., 2015).

Big datalla viitataan suurten datamassojen käsittelyyn sekä varastointiin, joita ei perinteisillä tiedonhallintamenetelmillä pystyttäisi käsittelemään (Saura, 2021). Big dataa voidaan luonnehtia sen kolmen tunnusomaisen ominaisuuden avulla, joita ovat määrä, monimuotoisuus sekä nopeus (Saura, 2021). Big data muodostuu erilaisista lähteistä, kuten sosiaalisen median julkaisuista, päivittäisistä transaktioista sekä eri laitteisiin asennetuista sensoreista (Pappas ym., 2018). Tekoäly ja koneoppiminen tarjoavat big datan käsittelyyn ihmistä tehokkaamman ratkaisun, mahdollistaen suurten datamassojen vaivattoman prosessoinnin (Campbell ym., 2020).

Data tuo tukea markkinoinnin tehokkuuden mittaamiseen, ja auttaa näin ollen dataohjautuvan markkinoinnin päätöksenteossa (Wedel & Kannan, 2016). Datan rooli päätöksenteossa on varsin suuri, sillä sen avulla pystytään perustelemaan markkinoinnille tyypillisesti kriittisiä valintoja, kuten tuotteelle sopivan markkinan löytämistä tai markkinointikanavien valitsemista (Amado ym., 2018). Ripeiden teknologisten harppausten myötä, markkinoijat turvautuvat datan käsittelyssä nykyään entistä enemmän tekoälyyn, jonka hyödyntäminen jo itsessään vaatii suuria määriä olemassa olevaa dataa (Campbell ym., 2020).

4.1 Mitä dataa hyödynnetään?

Markkinoinnin tekoälysovellukset vaativat toimiakseen relevanttia dataa digitaalisessa muodossa (Brock & von Wangenheim, 2019). Tämä data voi olla rakenteeltaan joko strukturoitua tai strukturoimatonta (Kietzmann ym., 2018). Näistä ensimmäisellä viitataan standardoituihin datajoukkoihin, joista esimerkkinä ovat mm. asiakkaista kerätty verkkosivustodata, transaktiotiedot tai bannerimainoksen klikkimäärä (Balducci & Marinova, 2018; Kietzmann ym., 2018; Paschen ym., 2019). Strukturoidulle datalle ominaista on sen ennalta määritelty, yksiulotteinen ja numeerinen luonne (Balducci & Marinova, 2018; Saura, 2021). Strukturoitu data ei voi samanaikaisesti edustaa eri ilmiöitä, eikä sisältää useampaa ai-
nutlaatuista tiedon palasta (Balducci & Marinova, 2018).

Toisin kuin strukturoidulla datalla, strukturoimattomalla datalla ei ole ennalta määriteltyjä numeerisia arvoja, minkä takia sitä käsittelevän asiantuntijan tulee manuaalisesti tai automatisoituja menetelmiä, kuten ohjaamatonta koneoppimista hyödyntäen, määrittää datalle tietyt arvot (Balducci & Marinova, 2018). Strukturoimaton datayksikkö sisältää useita puolia ja edustaa näin erilaisia ilmiöitä samanaikaisesti (Balducci & Marinova, 2018). Esimerkiksi ääni, video ja kuva ovat strukturoimatonta dataa, jotka on konvertoitava strukturoituun muotoon, jotta niitä on mahdollista hyödyntää (Hair & Sarstedt, 2021; Paschen ym., 2019). Strukturoimattoman datan määrä on kasvanut erittäin nopeasti, sen osuuden yltäen jo 80–90 % yritysten hallinnoiman datan kokonaismäärästä (Hair & Sarstedt, 2021; Kietzmann ym., 2018).

Strukturoimaton data tarjoaa markkinoijille useita mahdollisuuksia sekä haasteita, sen sisältämän monitahoisen informaation sekä suuren volyymin johdosta (Balducci & Marinova, 2018; Kietzmann ym., 2018; Paschen ym., 2019). Tekoälyä ja koneoppimista hyödyntävien teknologioiden nopea kasvu on tuonut ratkaisuja strukturoimattoman datan käsittelyyn ja hyödyntämiseen, mikä on tehnyt strukturoimattomasta datasta aiempaa näkyvämmän aiheen niin markkinoinnin akateemisessa tutkimuksessa kuin liiketoiminnassa (Balducci & Marinova, 2018). Strukturoimaton data on ollut suurin ajuri datan kokonaismäärän nopeaan ja valtavaan kasvuun (Ma & Sun, 2020).

Tästä huolimatta strukturoimaton data ja sen lukemattomien eri muotojen muuntaminen strukturoituun muotoon sisältää vielä monia haasteita markkinoinnin tekoälysovellusten kanssa työskenteleville (Hair & Sarstedt, 2021). Moniulotteisen rakenteen sekä valtavan volyymin takia, strukturoimattoman datan koetaan omaavaan potentiaalia markkinoinnin tekoälysovellusten implementoinnissa, minkä takia siitä tulisi tarkkailla mielenkiinnolla (Balducci & Marinova, 2018; Hair & Sarstedt, 2021).

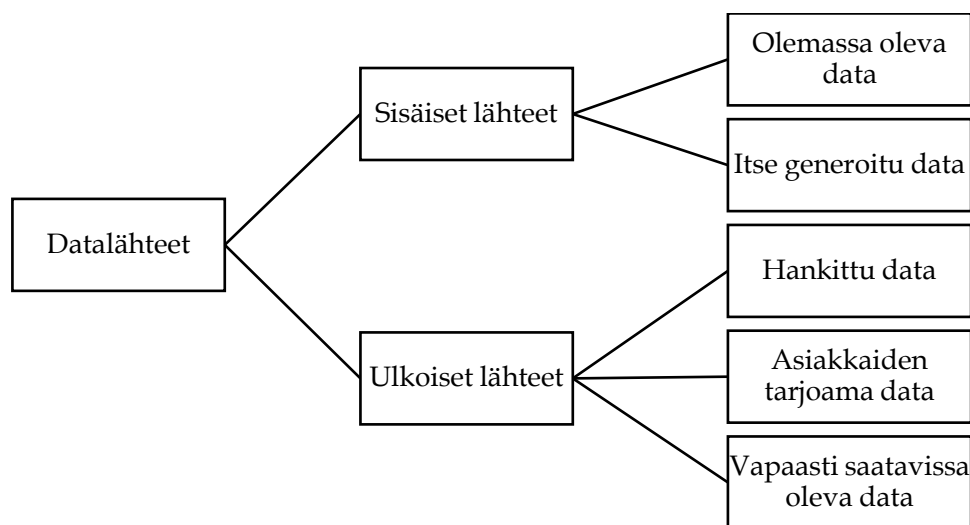
4.2 Mistä dataa kerätään?

Saatavilla olevan datan määrä on kasvanut valtavaksi, sen samalla muodostaen aiempaa rikkaamman ja monipuolisemman datajoukon (Miklosik & Evans, 2020). Markkinoinnin tekoälysovellukset tarvitsevat ennen kaikkea relevanttia ja laadukasta dataa toimiakseen toivotulla tavalla (Brock & von Wangenheim, 2019; Overgoor ym., 2019). Tätä dataa yritykset voivat kerätä niin sisäistä kuin ulkoisista datalähteistä (Campbell ym., 2020; Kibria ym., 2018; Kruhse-Lehtonen & Hofmann, 2020). Kuviossa 2 havainnoidaan tämän jaon mukaisesti, mistä sisäiset ja ulkoiset datalähteet muodostuvat.

Sisäinen data on yrityksen sisäisten prosessien kautta tuottamaa ja kerättyä dataa, joka on yritysten hallussa, mihin ulkopuolisilla ei ole suoraa pääsyä ilman erityistä lupaa (Montvilas, 2021). Sisäistä dataa voidaan kerätä mm. verkkosivustolta, mainonnanhallinnan- tai kanta-asiakasjärjestelmistä sekä kuluttajatutkimusten avulla (Campbell ym., 2020). Tällöin se voi olla esimerkiksi eri mainonnan kanavien performanssidataa, kanta-asiakasrekisteriin kerättyä dataa tai

verkkosivuston käyttäjien käytökseen perustuvaa dataa (Campbell ym., 2020). Koska yritykset ovat itse vastuussa keräämästään ja hallinnoimastaan sisäisestä datasta, on se tyypillisesti luotettavampaa ja tarkempaa kuin ulkoinen data (The Data Wiki, 2019).

Kuvio 2 esittää sisäisen datan rakentuvan yrityksen hallinnoimasta, jo olemassa olevasta datasta sekä sen itse tuottamasta ja keräämästä datasta. Olemassa oleva data on aktivoitavissa yrityksen eri tiedonhallintajärjestelmistä, mutta se ei kuitenkaan ole käytössä mitään tiettyä tarkoitusta varten juuri sillä hetkellä. Yrityksen itse generoima, jota voidaan kerätä esimerkiksi verkkosivustoseurannan tai asiakastutkimusten avulla tiettyä tarkoitusta varten (Hartmann ym., 2016).



KUVIO 2 Sisäiset ja ulkoiset datalähteet (Hartmann ym., 2016)

Ulkoisella datalla viitataan kaiken tyyppiseen dataa, jota on kerätty yrityksen ulkopuolelta (Montvilas, 2021). Se on jonkin toisen toimijan keräämää dataa, jonka käyttötarkoitus on eri kuin loppukäyttäjän (Ellram & Tate, 2016). Ulkoista dataa ovat esimerkiksi kausimuuttujat, kuten sää ja viikonpäivät, demografiset tiedot, kuten ikä ja koulutustaso sekä kolmannen osapuolen evästeiden kautta kerättävä verkkokäyttäytymiseen liittyvät tiedot (Campbell ym., 2020). Ulkoiset datalähteet tarjoavat huomattavaa potentiaalia suurten volyymien sekä monipuolisen ja rikkaan datan muodossa, mutta ne tuovat myös samalla mukanaan useita haasteita (Aaser & Mcelhaney, 2021). Yritysten tulee ymmärtää, minkälaista dataa on saatavilla ja mihin tarkoitukseen sitä halutaan kerätä, minkä lisäksi datan laadun määrittäminen, analysointi ja käsittely vaativat yrityksiltä entistä suurempia ponnistuksia (Aaser & Mcelhaney, 2021).

Ulkoinen data voi olla avointa ja julkisesti vapaasti hyödynnettävää tai yksityistä ja maksullista dataa (The Data Wiki, 2019). Avoin ulkoinen data määritellään dataksi, joka on julkista ja maksutonta tai kohtuudella hinnoiteltua, kone-luettavaa sekä kaikille saatavilla olevaa (Open Knowledge Foundation, n.d.). Suljettu ulkoinen data on jonkin toimijan tarjoamaa, usein monesta eri lähteestä koottua dataa, jota tarjotaan muille eteenpäin (Verhoeven, 2017). Esimerkiksi kolmannen osapuolen palveluntarjoajat, jotka eivät ole alkuperäisiä datan omistajia

vaan ovat keränneet dataa eri lähteistä ja myyvät sitä eteenpäin ovat tällaisia toimijoita (Aaser & Mcelhaney, 2021). Myös toisen osapuolen data, joka on esimerkiksi jonkin median omaa dataa, jota se myy eteenpäin, on ulkoista ja yksityistä (Schneider ym., 2017).

Hartmann ym. (2016) jakavat ulkoisen datan yrityksen itse hankkimaan, asiakkaiden ja kumppanien tarjoamaan sekä vapaasti saatavilla olevaan dataan. Hankitulla datalla viitataan datajoukkoihin, joita voidaan ostaa kolmannen osapuolen datan tarjoajilta. Asiakkaiden ja kumppanien tarjoama data pitää sisälleen yrityksen sidosryhmien tuottaman datan, joka ei ole saataville julkisesti muille toimijoille. Vapaasti saatavilla oleva data on julkisesti, ilman varsinaisia kustannuksia, kerättävää dataa esimerkiksi avoimista datalähteistä tai verkosta hakurobottien avulla.

4.3 Datan laatu ja sen edellytykset

Markkinoinnin tekoälysovellukset ovat yhtä toimivia kuin niille syötetty data, minkä takia datan laatu kriittinen tekijä niiden menestyksekkään implementoinnin kannalta (Hair & Sarstedt, 2021). Datan laadukkuus on kuitenkin erittäin kontekstisidonnainen käsite, minkä myötä sitä käsitellään kirjallisuudessa varsin monitahoisesti (Bertossi & Geerts, 2020). Vaikka markkinoinnin erilaiset tekoälyratkaisut vaativat toimiakseen laadukasta dataa, ei markkinoinnin akateeminen tutkimus tarjoa kovinkaan kattavaa keskustelua tämän ympärille (Hair & Sarstedt, 2021). Datan laatu omaa kuitenkin pitkän historian tietojenkäsittely- ja datatieteiden puolella, jonka odotetaan ottavan entistä merkittävämpää roolia myös markkinoinnin tutkimuksessa (Hair & Sarstedt, 2021).

Datan laatuun ja sen eri ulottuvuuksiin keskittynyt tutkimus sai 1990-luvulla nostetta, jolloin monet tutkijat esittivät näille erilaisia määritelmiä (Cai & Zhu, 2015). Wang ja Strong (1996) esittivät tällöin suosituksen määritelmän datan laadulle, esitellen sen muodostuvan datan sopivuudesta tiettyyn käyttötarkoitukseen, jota sen käyttäjät pystyvät arvioimaan. Lisäksi he esittivät myös datan laatu-ulottuvuudelle määritelmän, jonka mukaan se koostuu joukosta eri datan laadun attribuutteja, jotka edustavat yhtä laadun näkökulmaa. Knight ja Burn (2005) esittävät tutkimuksessaan yhteenvedon kirjallisuudessa yleisimmin esiintyvistä datan laadun ulottuvuuksista. Datan tarkkuus, johdonmukaisuus, turvallisuus sekä ajantasaisuus olivat tutkimuksen mukaan useimmiten käytettyjä datan laadun ulottuvuuksia. Datan laatu-ulottuvuuksien määrittelyyn ei kuitenkaan löydy yksimielisyyttä, minkä takia datan laatua käsitellään akateemisessa tutkimuksessa varsin vaihtelevasti (Gudivada, Ding ym., 2017).

Datan laadun ylläpitäminen ja sen tarkkailu tulee entistä haastavammaksi, kun monia eri digitaalisia datalähteitä ja -joukkoja joudutaan yhdistelemään keskenään (Hair & Sarstedt, 2021). Datan moninaisuus ja lukemattomien lähteiden määrä luovat samalla monia mahdollisuuksia datan hyödyntämiselle, mutta tekevät sen käsittelystä ja laadun tarkastelusta vaikeampaa (Gudivada, Ding, ym., 2017; Hair & Sarstedt, 2021). Datan käsittelyyn ja hallintaan liittyvät

eroavaisuudet, kuten strukturoimattoman datan konvertoimisessa tai poikkeamien havaitsemisessa, voivat vaikuttaa haitallisesti datan hyödyntämiseen sille määritellyssä käyttötarkoituksessa (Hair & Sarstedt, 2021). Vastaavanlaisten ongelmien tunnistaminen on erittäin haastavaa, kun datajoukot ovat peräisin useista eri lähteistä, joissa ei ole yhtenäisiä menetelmiä datan laadun varmistamiseksi (Hair & Sarstedt, 2021).

5 METODOLOGIA

5.1 Laadullinen tutkimusmenetelmä

Tutkimus voi olla luonteeltaan kartoittava, kuvaileva, ennustava, selittävä tai näiden yhdistelmä (Hirsijärvi ym., 2009; Malhotra, 2017). Kun tarkoituksena on saada syvällisempää ymmärrystä tutkimusaiheesta, jota ei ole tutkittu kovinkaan laajasti, sekä pyrkimyksenä selvittää, mitä kyseisessä ilmiössä tapahtuu, on kyseessä kartoittava tutkimus (Hirsijärvi ym., 2009; Malhotra, 2017). Kartoittava tutkimuksen etuna on sen joustavuus sekä mukautuvuus muutokseen, miksi myös tutkijan on oltava valmis muuttamaan tutkimuksen suuntaa ja sisältöä esiin tulevien tietojen ja oivallusten myötä (Malhotra ym., 2017). Tämän tutkimuksen tavoitteena on lisätä syvällistä ymmärrystä markkinoinnin tekoälysovellusten implementointiprosessissa hyödynnettävästä datasta, minkä myötä se on luonteeltaan kartoittava tutkimus.

Tämän tutkielman tutkimusmenetelmäksi valittiin kartoittava kvalitatiivinen eli laadullinen tutkimus, jonka tarkoituksena on luoda perusteellisia ja subjektiivisia havaintoja teorian rakentamiseksi (Crick, 2021). Laadullinen tutkimus ei ole olemukseltaan kovinkaan jäykkä menetelmä, vaan omaa muuttuvuuden piirteitä, kuten yllätyksellisyyttä ja poikkeavuutta (Eriksson & Kovalainen, 2016). Laadullista tutkimusta käytetään saavuttamaan yksityiskohtaista sekä rikasta ymmärrystä tutkittavasta aiheesta, ja sen tavoitteena on heijastaa laaja-alaisesti oikean elämän ilmiöitä, joissa totuuden voidaan nähdä olevan varsin moniulotteinen käsite (Crick, 2021; Hirsijärvi ym., 2009). Laadulliselle tutkimukselle ei löydy täysin omaa teoriaa tai paradigmaa, minkä takia sen kattava määrittely koetaan varsin haastavaksi (Metsämuuronen, 2011). Sitä vertaillaan ja asetetaan usein vastakkain kvantitatiivisen eli määrällisen tutkimuksen kanssa, vaikka nämä olisikin suotava nähdä toisiaan täydentävinä tutkimusmenetelminä (Malhotra ym., 2017; Metsämuuronen, 2011).

Laadullisen ja määrällisen tutkimuksen erot ovat melko suuria, minkä takia näistä toisen valitseminen päätutkimusmetodiksi on suotavaa (Metsämuuronen, 2011). Tutkimuksen metodien ja teorian valinta onkin laadukkaan ja onnistuneen tutkimuksen kokonaisuuden saavuttamisen kannalta kriittisiä tekijöitä (Eriksson & Kovalainen, 2016). Metodien valinnassa tulee kiinnittää huomiota ennen kaikkea tutkimuskysymysten asetteluun sekä niiden tarkempaan rajaukseen (Eriksson & Kovalainen, 2016). Laadullisen tutkimuksen avulla voidaan kehittää syvällisempää ymmärrystä monimutkaisista ja sensitiivisistä aiheista sekä selvittää, miten asiat toimivat käytännössä (Crick, 2021; Metsämuuronen, 2011). Se soveltuu tutkimusmenetelmäksi mallikkaasti esimerkiksi silloin kun kiinnostuksen kohteina ovat tutkittavien asioiden yksityiskohtaiset rakenteet ja merkitykset, eivätkä niiden yleiset jakaumat (Metsämuuronen, 2011).

Määrällisen tutkimuksen keskittyessä hypoteesien testaamiseen, tilastollisiin analyyseihin ja ilmiöiden selittämiseen, on laadullisen tutkimuksen

lähtökohtana tulkita ja ymmärtää tutkittavia ilmiötä perusteellisemmin (Eriksson & Kovalainen, 2016). Tutkimuksen tarkoituksena on tarkastella ja luoda näin ollen tietämystä tästä varsin uudesta tutkimusaiheesta, tässä tapauksessa datan syvällisemmästä ymmärryksestä markkinoinnin tekoälyn implementointiprosessissa. Laadullinen tutkimusmenetelmä soveltuu tämän aiheen tutkimiseen hyvin, sillä tutkielman tarkoituksena, tutkimuskysymyksiin pohjautuen, on lisätä ymmärrystä aiheesta, jota ei olla tutkittu kattavasti akateemisessa kirjallisuudessa (Hirsijärvi ym., 2009)

Laadullinen data on monimuotoista, sillä sitä ei voida määritellä vain yhdenlaiseksi datajoukoksi (Hair & Brunsveld, 2019; Puusa ym., 2020). Se määritellään tyypillisesti vastakkain asettamalla se määrällisen datan kanssa, jolloin laadullinen data nähdään datana, joka ei ole numeerista eikä mitattavaa (Eriksson & Kovalainen, 2016; Eskola & Suoranta, 1998). Laadulliselle datalle ominaista on sen tekstuaalinen, verbaalinen, äänellinen tai visuaalinen muoto, josta on mahdollista tehdä tulkintoja ja kuvauksia, ilman että huomion keskipiste olisi mittauksessa (Eriksson & Kovalainen, 2016). Laadullisessa tutkimuksessa aineiston koko on tyypillisesti vähäisempää verrattuna määrälliseen tutkimukseen, joka voi usein sisältää suuria määriä, esimerkiksi laajamittaisista kyselyistä kerättyä, dataa (Hair & Brunsveld, 2019). Luonteeltaan kartoittavan laadullisen tutkimuksen dataa voidaan kerätä esimerkiksi haastatteluista, kuvista, videoista tai havainnoimalla tapahtumia ja käytöstä (Hair & Brunsveld, 2019). Laadullinen data voi olla myös valmiista aineistoista tehtyjä muistiinpanoja, äänitteitä tai kenttämuistiinpanoja (Puusa ym., 2020)

Laadullinen tutkimus pitää sisällään useita eri aineistonkeruun menetelmiä, kuten haastattelut, havainnoinnin sekä kirjallisen tai visuaalisen materiaalin käytön (Hair & Brunsveld, 2019; Metsämuuronen, 2011). Hair ja Brunsveld (2019) esittävät laadullisen tutkimuksen aineistonkeruun menetelmille laajan jaon haastatteluihin ja havainnointiin. Havainnointi on sopiva valinta aineistonkeruun menetelmäksi, silloin kun tarkoituksena on tutkia ihmisten, tapahtumien tai asioiden käyttäytymistä objektiivisesti (Hair & Brunsveld, 2019; Metsämuuronen, 2011). Tyypillisesti havainnoinnin aineistonkeruun ajatellaan olevan usein visuaalista, mutta myös muita aisteja voidaan hyödyntää sen toteuttamisessa (Metsämuuronen, 2011). Valitsin tämän tutkimuksen aineistonkeruun menetelmäksi kuitenkin haastattelun, sillä se soveltuu hyvin monimutkaiselle tutkimusaiheelle, jossa syitä tietyn ilmiön tapahtumiselle halutaan tutkia (Hair & Brunsveld, 2019). Haastattelut voidaan jakaa strukturoituihin, puolistrukturoituihin ja avoimiin haastatteluihin, joista tutkimuksessani päädyin käyttämään puolistrukturoituja haastatteluja (Metsämuuronen, 2011). Puolistrukturoidut haastattelut soveltuvat varsinkin hieman heikommin tutkituille tutkimusaiheille, kuten tämän pro gradu -tutkielman aiheelle (Metsämuuronen, 2011).

5.2 Aineiston kerääminen

Tämän tutkimuksen aineistonkeruumenetelmäksi valittiin teemahaastattelu, joka on Suomessa varsin suosittu laadullinen aineistonkeruumenetelmä. Sen tarkoituksena on selvittää ennalta määritellyn kohdejoukon ajatuksia tutkittavasta aiheesta, ja siinä lähtökohtana on kyseisen tutkimuksen tutkimusongelma (Valli, 2018). Teemahaastattelussa aihealueet on tyypillisesti määritelty ennakkoon, mutta kysymysten muotoa ja järjestystä voidaan halutessa haastattelujen aikana muokata, mikä tekee haastattelusta luonteeltaan puolistruktruoidun (Hirsijärvi & Hurme, 2000).

Tutkimuksen teemahaastatteluihin valittujen haastateltavien valinnassa käytettiin harkinnanvaraista otantaa. Harkinnanvaraisessa otannassa kohde-ryhmä valitaan suuremmasta joukosta, jolloin osa potentiaalisista haastateltavista jätetään tietoisesti pois (Hirsijärvi ym., 2009). Tutkimuksessa pyrittiin selvittämään varsin rajatun ja kapean markkinoinnin osa-alueen asiantuntijoiden näkemyksiä tutkielman aiheesta, minkä Hair ja Brunsveld (2019) näkevät perusteltuna lähtökohtana harkinnanvaraisen otannan valitsemiseksi.

Tutkimuksen aineisto kerättiin Suomessa markkinoinnin tekoälysovellusten parissa työskenteleviltä asiantuntijoilta helmikuun 2022 aikana. Haastatteluihin osallistujia kartoitettiin hyödyntäen markkinoinnin alan verkostoja sekä eri sosiaalisen median alustoja. Potentiaalisia haastateltavia lähestyttiin viestin välityksellä, selvittäen heidän soveltuvuuttansa haastatteluihin sekä alustuen tutkimuksen tarkoitusta. Haastatteluun soveltuville lähetettiin lopulta kutsu Microsoft Teams -palvelussa toteutettuihin haastatteluihin.

Kaikki haastattelut toteutettiin Microsoft Teams -palvelun välityksellä etänä, mikä nykypäivänä koetaan erittäin tyypilliseksi tavaksi toteuttaa ihmisten välistä vuorovaikutusta. Haastatteluihin valikoituneet henkilöt olivat tuttuja kyseisen palvelun kanssa, minkä myötä etenivät alusta loppuun sujuvasti. Asiantuntijahaastattelut ovat tyypillisesti monipuolisia ja antoisia esiin tulevan informaation osalta, mikä ilmeni myös näissä haastatteluissa toteutuvan (Saunders ym., 2019).

Aineiston tallentamiseen käytettiin kahta nauhuria, jolla pyrittiin minimoimaan laitteisiin liittyvien teknisten sekä inhimillisten virheiden mahdollisuutta. Jokaista haastattelua seurasi litterointi, joka toteutettiin välittömästi haastattelun ajankohdan jälkeen. Litteroitu aineisto muodostui lopulta melko yhtenäiseksi, kun samankaltaisia teemoja havaittiin yhä enemmän, mitä pidemmälle haastatteluissa päästiin. Haastatteluissa esiintyvä informaatio saavutti näin ollen saturaatiopisteen, mikä on laadullisessa tutkimuksessa tavoiteltava elementti (Hirsijärvi ym., 2009).

Haastateltavien joukko koostui seitsemästä korkeakoulutetuista markkinoinnin teknologioiden ja tekoälysovellusten kanssa työskentelevistä asiantuntijoista, jotka ovat työskennelleet aiheiden parissa useita vuosia. Kaikkia haastateltavia yhdisti korkeakoulutausta, työskentely markkinointitoimistoissa sekä työn sijainti Suomessa. Koska tutkimusaineisto muodostui ainoastaan Suomessa

työskentelevistä markkinointitoimistojen asiantuntijoista, ei tutkimuksen tuloksia tai niistä koottuja johtopäätöksiä tule yleistää. Esimerkiksi globaalien teknologiajättien, kuten Google ja Applen, teknologiset resurssit asettaisivat täysin erilaiset lähtökohdat tutkimusaineistolle ja siitä tehtäville löydöksille.

Haastattelut alkoivat tutkimusaiheen tarkemmalla esittelyllä sekä haastattelun etenemisen läpikäymisellä. Myös haastatteluaineiston raportointiin, tallentamiseen ja haastateltavien yksityisyyteen liittyvät aiheet käytiin läpi ennen varsinaisen haastattelun aloittamista. Tutkimusaihe oli jokaiselle haastateltavalle erittäin tuttu, mikä loi hyvät lähtökohdat rikkaan haastatteluaineiston keruulle. Haastattelut etenivät haastattelurunkoa mukaillen, vaikkakin teemahaastattelulle tyypillisesti haastattelukysymysten järjestys ja muoto saattoi muuttua haastattelusta riippuen (Hirsijärvi & Hurme, 2000). Tutkimuksen teoriaosuudessa esiin tuodut aiheet datan ominaisuuksista ja sen roolista markkinoinnin tekoälysovellusten implementointiprosessissa ohjasivat haastatteluiden etenemistä ja sisältöä. Haastattelurungon aihealueet laadittiin näihin pohjautuen. Aiempi akateeminen tutkimus markkinoinnin tekoälysovelluksista ja niissä hyödynnettävistä datasta ei kuitenkaan tarjoa valmiita viitekehyksiä tutkimusaiheelle, minkä takia haastatteluihin jätettiin tietoisesti tilaa myös vapaammalle keskustelulle. Haastateltavia käsitellään H1-H7 lyhenteillä aineiston anonymiteetin säilyttämiseksi.

TAULUKKO 4 Haastattelujoukon esittely

Haastattelu	Tehtävänimike	Kesto	Toimipaikka	Päivämäärä	Toteutus-tapa
1	Senior Data Scientist	53:10	Helsinki	08.02.2022	Microsoft Teams
2	Director, Advanced Analytics	41:44	Helsinki	14.02.2022	Microsoft Teams
3	Chief Business Officer	46:06	Helsinki	15.02.2022	Microsoft Teams
4	Data Scientist	43:56	Helsinki	17.02.2022	Microsoft Teams
5	Head of MarTech	36:45	Helsinki	18.02.2022	Microsoft Teams
6	Data Scientist	53:44	Helsinki	18.02.2022	Microsoft Teams
7	Data Scientist	42:32	Helsinki	21.02.2022	Microsoft Teams

5.3 Sisällönanalyysi

Laadullisessa tutkimuksessa kerätyn aineiston analysoinnin tarkoituksena on tulkita, kuvailla ja ymmärtää tutkimuksen kiinnostuksen kohteeksi rajattua ilmiötä (Puusa ym., 2020). Tutkimuksen aineistonkeruumenetelmäksi valikoitui

sisällönanalyysin, sillä se soveltuu hyvin teemahaastatteluiden avulla kerätyn aineiston analysointiin (Hair & Brunsveld, 2019). Laadullisen tutkimuksen sisällönanalyysi on yleinen aineiston analysointimenetelmä, jota käytetään usein juuri haastatteluista kerätyn ja tekstimuotoon muutetun aineiston luokitteluun sekä analysointiin (Tuomi & Sarajärvi, 2018). Tutkijan tavoitteena sisällönanalyysissä on muodostaa mahdollisimman kokonaisvaltainen ja selkeä kuvaus aineistosta, etsien merkityssuhteita ja -kokonaisuuksia (Hair & Brunsveld, 2019; Puusa ym., 2020; Vilkka, 2021). Sisällönanalyysi auttaa laadullisen aineiston selkeässä ja tiiviissä jäsentämisessä, ja luo edellytyksiä aineistosta tehtäville laadukkaille johtopäätöksille (Tuomi & Sarajärvi, 2018).

Tutkijan on syytä päättää sisällönanalyysin toteuttamismenetelmä ennen aineistonkeruuta, mikä voidaan tyypillisesti jakaa teoria- ja aineistolähtöiseen sekä teoriaohjaavaan sisällönanalyysiin (Puusa ym., 2020; Tuomi & Sarajärvi, 2018; Vilkka, 2021). Teorialähtöisessä sisällönanalyysissä tutkimuksen teoria ohjaa vahvasti aineiston analysointia (Puusa ym., 2020; Vilkka, 2021). Sisällönanalyysin tavoitteena on tällöin tyypillisesti teorian testaaminen ja uudistaminen tutkimukseen valitussa kontekstissa (Vilkka, 2021). Aineistolähtöisen sisällönanalyysin tavoitteena on ymmärtää tutkittavien ajattelu- ja toimintatapoja tutkimalla kerättyä aineistoa ilman teoriaan nojaavaa mallia (Vilkka, 2021). Aineistolähtöisesti toteutettavassa sisällönanalyysissä keskitytään vahvemmin tutkimuksessa kerättyyn aineistoon, lukien sitä ilman teoriaan pohjautuvia ennako-oletuksia (Puusa ym., 2020). Siinä pyritään kuvaamaan tutkittavien merkitysmaailmaa, lopullisena tavoitteena ymmärtää heidän ajattelu- ja toimintatapoja (Vilkka, 2021). Teoriaohjaavassa sisällönanalyysissä niin teoria kuin aineisto ohjaavat tutkijan ajattelua (Tuomi & Sarajärvi, 2018). Teoriaohjaava sisällönanalyysi ei pohjaudu suoraan teoriaan, mutta teoria tukee analyysin etenemistä (Tuomi & Sarajärvi, 2018). Jos tutkimus keskittyy kartoittamaan uutta ilmiötä, eikä varsinaisesti testaamaan aiempaa teoriaa, on teoriaohjaava sisällönanalyysi tällöin sopiva menetelmä (Tuomi & Sarajärvi, 2018). Valitsin tutkimuksen analyysimenetelmäksi teoriaohjaavan sisällönanalyysin, koska tutkimusaiheesta löytyvä kirjallisuus tarjoaa tutkimukselle hyvän teoreettisen pohjan, muttei valmiita malleja.

Laadullisen aineiston analysoinnille on tyypillistä siitä löydettyjen havaintojen luokittelu ennalta määriteltyjen kategorioiden mukaan (Puusa ym., 2020). Luokittelun lisäksi aineistoa voidaan analysoida myös teemoittelun sekä tyypittelyn avulla (Tuomi & Sarajärvi, 2018). Tässä tutkimuksessa päädyin käyttämään teemoittelua, joka on laajalti hyödynnetty laadullisen aineiston analyysimenetelmä (Nowell ym., 2017). Teemoittelussa tutkimusaineistosta etsitään ja paikannetaan tutkimukselle olennaisia teemoja (Nowell ym., 2017).

Braun ja Clarke (2006) esittelevät teemoittelun olevan kuusivaiheinen prosessi, joka muodostuu aineistoon tutustumisesta, koodaamisesta, teemojen etsimisestä, teemojen arvioinnista, teemojen määrittelystä ja nimeämisestä sekä raportoinnista. Aineiston analysointi toteutettiin näitä vaiheita seuraten, aloittaen teemahaastatteluiden avulla kerättyyn aineistoon tutustumalla. Teemahaastatteluiden aineisto litteroitiin ensi alkuun kirjalliseen muotoon, jotta niihin tutustuminen ja koodaaminen onnisuisivat mahdollisimman sujuvasti. Tämän nähdään

olevan välttämätön osa teemoittelua, jos kerätty aineisto on verbaalisessa muodossa (Braun & Clarke, 2006). Litteroitu tutkimusaineisto luettiin tämän jälkeen kahteen otteeseen läpi, ja siitä tehtiin samalla muistiinpanoja listaten kiinnostavia ja esille nousevia aiheita. Aineiston tutustumisesta siirryttiin sen koodaamiseen, jossa tavoitteena on järjestellä rikasta ja moninaista tutkimusaineistoa yksinkertaisempaan muotoon, luokitellen sitä merkityksellisiin ryhmiin (Nowell ym., 2017). Koodaamista voidaan toteuttaa esimerkiksi värikoodeilla, tekstiin lisätyillä muistiinpanoilla sekä ajatuskarttojen ja taulukkojen avulla (Braun & Clarke, 2006). Sisällönanalyysin kolmannessa vaiheessa koodatusta aineistosta pyritään löytämään teemoja, jotka voivat olla joko aineisto- tai teorialähtöisesti muodostettuja (Nowell ym., 2017). Tämä sisällönanalyysi toteutettiin teoriaohjaavasti, jolloin teoria tukee, mutta ei ohjaa teemojen muodostamista. Tutkijan on otettava aktiivinen rooli teemojen etsimisessä sekä toteutettava tätä systemaattisesti (Nowell ym., 2017). Teemojen muodostamisen jälkeen sisällönanalyysissä siirryttiin arvioimaan edellisessä vaiheessa muodostettuja teemoja tarkemmin. Esimerkiksi jotkin teemat eivät välttämättä sisällä tarpeeksi dataa niiden uskottavuuden tukemiseksi, minkä lisäksi teemat voivat olla päällekkäisiä tai niitä voi olla tarpeen jakaa erilleen (Braun & Clarke, 2006). Teemojen arvioinnin lopuksi tutkimuksen aineistosta muodostettiin kolme pääteemaa, joita olivat strukturoidun ja strukturoimattoman datan hyödyntäminen, datalähteet ja niiden painotus sekä datan laatu ja sen merkitys. Edellä esitellyt pääteemat sisälsivät lisäksi alateemoja, jotka on esitelty tutkimuksen tuloksissa. Teemat saivat nimensä analyysin viidennessä vaiheessa, jossa on tärkeä muistaa, että nimen tulisi kuvata teemaa mahdollisimman selkeästi ja antaa siitä lukijalle paras mahdollinen kuva (Braun & Clarke, 2006).

6 TUTKIMUKSEN TULOKSET

6.1 Strukturoidun ja strukturoimattoman datan hyödyntäminen

Data voidaan jakaa sen rakenteen mukaan kahteen luokkaan, strukturoituun ja strukturoimattomaan dataan (Balducci & Marinova, 2018). Tässä luvussa tarkastellaan haastatteluissa esille nousseita teemoja markkinoinnin tekoälysovellusten implementointiprosessiin hyödynnettävän datan rakenteen osalta. Strukturoitu data on ominaisuuksiltaan standardoitua, eikä se voi samanaikaisesti edustaa eri ilmiöitä (Paschen ym., 2019). Strukturoimaton data taas ei ole ennalta määriteltyä, ja sille tunnusomaista on sen moniulotteisuus (Balducci & Marinova, 2018). Haastatteluissa selvitettiin, kuinka näitä kahta eri datajoukkoa hyödynnetään markkinoinnin tekoälysovellusten implementointiprosesseissa ja missä suhteessa. Lisäksi strukturoituun sekä strukturoimattomaan dataan liittyviä haasteita ja mahdollisuuksia tuodaan luvussa esille.

6.1.1 Strukturoitu data keskiössä

Markkinoinnin tekoälysovellusten implementoinnissa hyödynnettävä data voi olla rakenteeltaan niin strukturoitua kuin strukturoimatonta. Kuten jo aiemmin on esitetty, myös kirjallisuudessa tunnistetaan, että strukturoitua dataa hyödynnetään näistä kahdesta yleisemmin, sillä se on rakenteeltaan yksinkertaisempi kuvaus datasta ja vaivattomampi ottaa käyttöön. Tutkimukseen osallistuneet kuvailivat strukturoidun datan olevan selkeämpää ja helpommin hyödynnettävää. Lisäksi todettiin, että strukturoidun datan kanssa on vielä paljon työtä jäljellä, miksi sitä tulisi oppia vielä hyödyntämään entistä monipuolisemmin, ennen kuin siirrytään työskentelemään vahvemmin strukturoimattoman datan kanssa.

Kyl se pääasias pyörii siin strukturoidun datan keskellä. Meillä on tällä hetkellä kaikki se strukturoimattoman datan hyödyntäminen sillee ideatasolla. Strukturoitu data on helpommin hyödynnettävissä. Siitä on helpompi nähä, et miten sitä voidaan hyödyntää eri tilanteissa, kun strukturoimattoman kanssa ei oo semmosia valmiita esimerkkejä. (H4)

Se kaikki data on hyvin aggregoitua ja erittäin aseteltua, eli me ei käytetä semmosta rakenteetonta dataa oikeestaan tällä hetkellä. Algoritmien kyvykkyydet ja ylipäätään tekoälyn kyvykkyydet ei riitä semmoseen kontrolloimattomaan rakenteettoman data hyödyntämiseen. Siihen ei vielä oikeen paukut riitä oikeasti. Sen takii me tarvitaan koko ajan se rakenne sille datalle, ja selkeet raamit, ja ne algoritmit, joita käytetään, on tosi yksinkertaisia oikeasti. Kaikki vaatii viel niin paljon ihmisen ohjaamista. Ei sitä yritetä käyttää, koska niin paljon tekemistä helpommassa puolessa. (H5)

Vaikka strukturoitu data on yksinkertaisempaa ja helpommin hyödynnettävissä, haastateltavat kertoivat sen silti vaativan paljon datatyötä. Strukturoitu data on ennalta määritelty numeerinen kuvaus kiinnostuksen kohteena olevasta ilmiöstä, joka pitää sisällään ainoastaan yhden ainutlaatuisen palan tietoa (Balducci & Marinova, 2018). Haastateltavat myötäilivät tätä määritelmää, minkä lisäksi he

luonnehtivat strukturoitua dataa yksinkertaiseksi, helppokäyttöiseksi ja nimensä mukaisesti jonkin rakenteen omaavaksi dataksi. Strukturoitu data on yksilöteistä, eikä voi edustaa, strukturoimattoman datan lailla useampaa eri ilmiötä samanaikaisesti (Balducci & Marinova, 2018).

Strukturoitu data voi olla mm. digimainonnan mainosten klikkimääriä, asiakkaista kerättävää verkkosivustodataa tai transaktiotietoja (Balducci & Marinova, 2018; Kietzmann ym., 2018; Paschen ym., 2019). Haastattelussa esille nousseita esimerkkejä markkinoinnin tekoälysovellusten implementoinnissa hyödynnettävästä strukturoidusta datasta olivat mainonnanhallinnanjärjestelmistä kerättävät performanssitiedot, kuten klikit, näyttömäärät ja euromääräiset kustannukset. Myös ulkoiset muuttujat, kuten säätiedot tai kuluttajien luottamusindeksit, ovat tyypillisiä esimerkkejä strukturoidusta datasta.

Just vaikka digimainonnan performanssidata, kuten klikit, impressio ja kuinka paljon se on maksanut päivä- tai viikkotasolla. Sitten näitä jotain ulkopuolisia muuttujia, kuten sää, sademäärä, lumitilanne, kuluttajien luottamusindeksiä ja muita vastaavia. (H6)

Esimerkiks sivustoilta ja mainonnanhallinnanjärjestelmistä kerättävä data. Sit just vaikka jotkut laskutustiedot ja vastaavat. Kyllähän sitä löytyy vaikka ja mistä, ja suuri osa, mitä käytetään, on just strukturoitua dataa. (H7)

6.1.2 Strukturoimattoman datan potentiaali

Strukturoimaton data voi edustaa useita eri ilmiötä samanaikaisesti, eikä se strukturoidun datan lailla omaa ennalta määriteltyä numeerista muotoa (Hair & Sarstedt, 2021). Se voi olla esimerkiksi videota, ääntä tai kuvaa (Hair & Sarstedt, 2021). Strukturoimatonta dataa joudutaan käsittelemään huomattavasti enemmän kuin strukturoitua dataa, minkä haastateltavat näkivät suurena haasteena sen hyödyntämiselle markkinoinnin tekoälysovellusten implementointiprosessissa sekä tekoälyratkaisuissa ylipäätään. Tämä koetaan myös Hairin ja Sarstedtin (2021) mukaan tyypillisenä ongelmana markkinoinnin teknologioiden kanssa työskenteleville.

Strukturoimattoman datan hyödyntäminen koettiin ennemmin tulevaisuuden mahdollisuutena kuin markkinoinnin tekoälysovellusten implementointiprosessin polttoaineena tällä hetkellä. Haastattelussa painotettiin strukturoimattoman datan olevan trendikäs ja suosittu aihe markkinoinnissa, ja sen uskotaan omaavan paljon potentiaalia lähitulevaisuudessa. Vaikka akateeminen kirjallisuus tarjoaa samankaltaisia näkymiä strukturoimattoman datan hyödyntämiselle markkinoinnin tekoälysovelluksissa, on markkinointitoimistoissa työskentelevien asiantuntijoiden antama kuva siitä tätäkin etäisempi. (Balducci & Marinova, 2018; Hair & Sarstedt, 2021). Strukturoimattoman datan uskottiin omaavan kuitenkin suurta potentiaalia lähitulevaisuudessa, kunhan kyvykkyydet sen käsittelyyn saadaan sen käsittelyä vaativiksi.

Meillä tällä hetkellä kaikki se strukturoimattoman datan hyödyntäminen on paljolti ideatasolla. Strukturoimattomassa datassa se ongelma on just siinä, et miten sitä saadaan strukturoitua ennen kun päästään käyttämään. Ne toimiviksi todistetut sovellukset strukturoimattoman datan osalta on melko vähäisiä vielä tällä hetkellä. Tuntuu

kans et siinä on semmonen pöhinävaihe. Ideoit näkyy paljon ja sitä, et me voitais tehdä tälle ja tälle, minkä jälkeen rakennetaan järjestelmää tälle. Ei kuitenkaan semmosii toimivii sovelluksii tuu kauheesti mieleen. (H4)

Vaikka strukturoimattoman datan määrä on kasvanut globaalisti generoidusta datan kokonaismäärästä 80-90 % tasolle, asettaa sen vaatima käsittely huomattavia haasteita (Hair & Sarstedt, 2021). Strukturoimattomalle datalle ominaiset piirteet, kuten sen moniulotteisuus ja heterogeenisyys luovat haastateltavien mielestä samalla mahdollisuuksia, mutta myös huomattavia haasteita sen hyödyntämiselle markkinoinnin tekoälysovellusten implementointiprosessiin. Strukturoidun datan rooli on näin ollen niissä usein selkeästi suurempi. Haastateltavat kokivat strukturoimattoman datan haasteiden löytyvän sen työläessä ja haastavassa käsittelyssä. Sitä joudutaan konvertoimaan koneluettavaksi strukturoituun muotoon ennen käyttöönnottoa.

Jos miettii tällösi hyvin määriteltyjä projekteja, joita pystytään myymään valmiina tuotteina tai analyysiprojekteina, niin niistä menee joku 30 % datan käsittelyyn. Jos sitä ruvetaan penkomaan strukturoimatonta dataa, niin se datan määrittäminen on niin paljon työläämpää, minkä nään olevan tässä se suurin ongelma. (H4)

Yleensä strukturoimatonta dataa joutuu konvertoimaan. Varmaan poikkeuksetta tarvii muuttaa semmoiseen koneluottavaan muotoon. Yleensä et voi syöttää samalle koneelle kuvia, tekstiä ja numeroita. Siinä tarvii tehdä aika paljon datan käsittelyä ennen sitä ja tosi harvoin me tehdään. (H2)

Markkinoinnin tekoälysovellusten implementointiprosesseissa hyödynnettävä data on siis haastateltavien näkemysten mukaan pääosin strukturoitua, mutta strukturoimattoman datan uskotaan omaavaan suurta potentiaalia ja tulevan vahvemmin osaksi markkinoinnin tekoälysovelluksia ja niiden implementointia. Strukturoimaton data tarjoaa paljon mahdollisuuksia tulevaisuuteen markkinoinnissa, minkä takia siitä on syytä olla kiinnostunut ja ajan tasalla (Balducci & Marinova, 2018; Hair & Sarstedt, 2021).

Ainakin jos mietitään markkinoinnin kenttää, niin kyl se on aika paljolti strukturoitua vielä. Esimerkiks nää budjettioptimaattorit, joita on melko paljon markkinoilla, niin ne perustuu ihan puhtaasti siihen ettei siel oo muita muuttujia. Mutta pakko sanoo et se on tosi kovassa nousussa, et optimoidaan sitä strukturoimatonta dataa, koska sehän on se mielenkiintoinen juttu, kun me päästää kiinni niihin, mitä me markkinoinnissa halutaankin. (H3)

Ideoita kuitenkin pyörii paljon strukturoimattoman datan hyödyntämisessä. Ehkä ne suurimmat ideat on siinä, että miten sitä pystyttäis järkeviten strukturoitua, ja miten siitä saatais parhaiten tuotua insightteja. Kaikenlaisii kehitysideoita on strukturoimatomaan dataan liittyen ollut pöydällä, mut en ainakaan ite oo päässy vielä tekemään sen kanssa töitä. (H4)

Markkinointitoimistojen asiantuntijat muistuttivat tekoälysovellusten ratkovan erittäin rajattuja ja kapeita liiketoiminnallisia ongelmia, joiden kehittäminen vaatii usein varsin yksinkertaistettua strukturoitua dataa. Strukturoimattoman datan roolin uskotaan vahvistuvan lähitulevaisuudessa, mutta vielä tämänhetkiset tekoälyratkaisut painottuvat vielä vahvasti strukturoidun datan ympärille.

6.2 Datalähteet ja niiden painotus

Tässä luvussa esitellään yritysten hyödyntämiä datalähteitä markkinoinnin tekoälysovellusten implementointiprosessissa. Yritysten keräämä data voidaan tyyppillisesti jakaa sen lähteen mukaan sisäiseen sekä ulkoiseen dataan (Campbell ym., 2020; Kruhse-Lehtonen & Hofmann, 2020). Markkinoinnin akateemisessa tutkimuksessa molempien datalähteiden hyödyntäminen tunnistetaan tärkeäksi, mutta oman sisäisen datan käyttö painottaa ulkoista dataa yleisemmin. Ulkoisen datan nähdään omaavan paljon potentiaalia sen volyymien sekä rikkaan sisällön johdosta, mutta samalla nämä tuovat esiin myös monia haasteita (Aaser & Mcelhaney, 2021). Näitä teemoja nousi myös teemahaastattelussa esille, vahvistaen näkemystä sisäisen datan roolista markkinoinnin tekoälysovellusten implementointiprosessissa.

6.2.1 Sisäisen datan vahva rooli

Markkinoinnin tekoälysovellusten implementointiprosessiin kerättävä data on aina riippuvainen sen käyttötarkoituksesta (Campbell ym., 2020; Overgoor ym., 2019). Haastattelussa ilmeni, että näissä prosesseissa hyödynnettävää dataa kerätään niin sisäisistä kuin ulkoisista datalähteistä, joista sisäisen datan rooli nähtiin kuitenkin selkeästi vahvemaksi. Sisäinen data tarjoaa useammin valmiita, toisiinsa täsmättäviä datajoukkoja, joita on helpompi käyttää, minkä lisäksi sen arvon nähtiin muodostuvan myös siitä, että sitä ymmärretään ja tunnetaan keskimäärin paremmin kuin ulkoista. Markkinoinnin tekoälysovelluksen toimivuus perustuu voimakkaasti sille syötettyyn dataan ja sen laatuun, minkä takia on ymmärrettävää, että data helpompi ymmärrettävyys ja nopeampi käyttövalmius suosivat sisäisen datan hyödyntämistä (Brock & von Wangenheim, 2019; Overgoor ym., 2019; Stone ym., 2020).

Se on aina paljon hankalampaa se jonkun ulkoisen datan integroiminen, tämmösiin projekteihin tai tekoälysovelluksiin. Mun mielestä on huomattavasti helpompaa käyttää sisäistä dataa. Tietyllä tavalla sisäinen data on datasettejä, jotka on koostettu usein samalla tavalla ja on valmiiksi täsmättävis toisiinsa. Ulkoista dataa joutuu aina tarkastelemaan vähän tarkemmin ja miettimään, et miten se mätsää siihen sisäiseen dataan. (H4)

Se on enemmän sitä sisäistä omaa dataa. Sehän myös kertyy silleen helpommin, et jotkut asiakkaat ylläpitää vaan esim. jotain däsäreitä. Enemmän on sitä sisäistä dataa. Ulkoisella datalla useimmiten enemmänkin rikastetaan sisäistä dataa, eikä vaan käytetä yksittäin. (H6)

Sisäistä dataa voidaan kerätä monista eri yrityksen omista lähteistä, joista Campbell ym. (2020) mainitsee mm. oman verkkosivuston, myyntitiedot sekä mainonnan eri performanssimittarit. Haastattelussa sisäisen datan lähteiksi nostettiin useaan otteeseen yrityksen omat verkkosivut, mobiilisovellukset, kanta-asiakas-, CRM- sekä mainonnanhallinnanjärjestelmät. Myös yrityksen mediainvestoinnit ja asiakaslaskutustiedot nousivat haastattelussa moneen otteeseen

esille. Kuten Overgoor ym. (2019) muistuttavat, tämän todettiin aina olevan lii-
ketoiminnasta ja tekoälysovelluksen tarkoituksesta riippuvaista.

Riippuu paljon siitä, et millainen se yritys on. Jos on vaikka paljon kivijalkaa, niin sitä kerätään paljon jollain kanta-asiakaskortilla. Sitten totta kai oma sivusto, varsinkin jos on verkkokauppa, niin se on ihan primääri lähde. Kyl mä nään et myös B2B-puolel oma sivusto voi olla tosi tärkeä ja arvokas vaik ei oiskaan verkkokauppaa. Yhä enenevässä määrin myös tietenkin mobiili. Semmoset yritykset, joilla on oma mobiili-sovellus, niin ne kerää sieltä dataa. (H3)

Meiän projekteissa varmasti yleisimmät on eri mainonnanhallinnanjärjestelmät ja tietenkin, jos markkinointii tehdään meidän kautta, niin meidän läpi menevä laskutus. Sitten myös asiakkaan oma CRM, missä on sitten esim. asiakkuusmarkkinointiin ja muuhun liittyvää tietoa. Lisäks myös markkinoinnin automaation työkalut. (H2)

Apit, joista löytyy rajat, joista logataan mitä asioita katsot sieltä. Se lähettää eventtejä esim. Googlen Firebaseen, mistä näkee, kuinka paljon tapahtumia apissa tapahtuu. Myös minkä käyttöjärjestelmän käyttäjä ollaan jne. Se on kanava, minkä kautta lähdetään mainostamaan eri tuotteita ja tuodaan personointia. (H1)

Asiakasdatan kerääminen yrityksen omaan asiakasrekisteriin verkkosivujen kirjautumisten sekä kanta-asiakassovellusten ja -korttien avulla koettiin erittäin tärkeäksi sekä ajankohtaiseksi teemaksi. Yritykset ovat alkaneet suosimaan sisäänkirjautumista entistä vahvemmin, sillä kuluttajien tietoisuus ja kriittisyys datankeruuta kohtaan on tehnyt siitä vaikeampaa (Martin & Murphy, 2017). Tämän lisäksi kolmannen osapuolen evästeet tulevat poistumaan lähitulevaisuudessa, mikä tulee myös osaltaan vaikeuttamaan datan keräämistä ja mainonnan ratkaisuja (Bump, 2021). Tämän uskottiin tulevan vaikuttamaan datan keräämisen ja sen käytettävyyden mahdollisuuksiin, mikä johtaa asiakasdatan keräämisen merkityksen korostumiseen. Asiakasdataa kerätään myös kivijalkamyymälöistä vahvistamaan yritysten asiakasrekistereitä esimerkiksi kanta-asiakaskorttien ja -sovellusten avulla.

Tosi moni on evästeiden katoamisen takia alkanut panostamaan omiin sivustoihin, omiin kirjautumisiin ja vastaaviin. He keräävät sillä omaa asiakasrekisteriä, jossa he pystyvät tunnistamaan selain ID:tä sen perusteella, että olet ensimmäisenä ollut ei kirjautunut ja sitten kirjautunut. (H1)

Ehdottomasti semmonen mikä kasvaa, on asiakasrekisterien merkitys. Jos miettii keksien kuolemaa, niin kanta-asiakasjärjestelmien kehittäminen ja se et kaikilla ois se oma kanta, niin sen merkitys on kasvanut ja kasvaa edelleen paljon. (H6)

Nyt on viimeisimmätkin lähteneet kanta-asiakaskuvioon. Pyritään keräämään kivijalasta ymmärrystä, siitä mitä ihmiset ostaa, jotta yritykset pystyy lähtee rakentamaan valikoimaa ja asiakasrekisteriä. (H3)

Sisäisen datan rooliin koettiin myös vahvistuvan kolmannen osapuolen evästeiden katoamisen myötä. Juuri tähän nähtiin edellä mainittujen asiakasrekisteriä kasvattavien datankeruumenetelmien, kuten verkkosivuston kirjautumisien ja kanta-asiakasjärjestelmien olevan ratkaisuja. Oman asiakaskunnan tuntemuksen merkityksen koettiin lisäksi kasvavan tämän myötä, minkä todettiin vahvistavan oman sisäisen datan keräämisen ja hallinnoinnin tärkeyttä niin markkinoinnin

tekoälysovellusten implementoinnissa kuin muissa datan hyödyntämiseen liittyvissä prosesseissa.

Kylmä uskon et sisäisen datan merkitys tulee olemaan ainakin yhtä merkittävä, kun se nytkin on. Kun pohtii sitä omaan kantaan liittyvän datan merkitystä, niin se on just semmosta insighttiä, jota ei muualta pystykään saamaan. Se on suures roolissa. (H4)

Esim. Google ja FB, niin niillä on tosi iso rooli yrityksillä, siin mieles, et ne omistaa dataa. Siihenhän niiden liiketoiminta paljolti perustuu. Kyl se pienemmilläkin on tosi tärkeätä, et ne tietää asiakkaasta mahdollisimman paljon ja pystyy hyödyntämään sitä dataa. Tulee sen rooli varmasti kasvamaan. (H2)

Datan kerääminen ja omistajuus on ollut jo pitkän aikaa liiketoimintamalli, jolla esimerkiksi Google, Apple, Amazon ja Facebook ovat saavuttaneet ylivoimaista markkinavoimaa (Kruhse-Lehtonen & Hofmann, 2020). Vaikka suurin osa yrityksistä ei oletetusti pysty vastaavia datamääriä keräämään, uskoivat tutkimukseen osallistuneet oman sisäisen datan roolin vahvistuvan entisestään. Oma data koetaan arvokkaaksi resurssiksi, jonka keräämiseen on olemassa monia keinoja, kuten verkkosivustoseurannat ja kanta-asiakasjärjestelmät.

6.2.2 Ulkoisen datan rikastava rooli

Ulkoisen datan hyödyntäminen markkinoinnin tekoälysovellusten implementoinnissa koettiin haastateltavien mielestä vielä tällä hetkellä melko vähäiseksi verrattuna sisäiseen dataan. Sen roolin todettiin useampaan otteeseen haastattelussa olevan sisäistä dataa rikastava. Ulkoinen data on yrityksille vierasta, minkä takia aluksi on saatava perusymmärrys siitä, mitä dataa on saatavilla ja mistä lähteistä (Aaser & Mcelhaney, 2021). Lisäksi ulkoisen datan hyödyntäminen voi edellyttää muutoksia olemassa oleviin tiedonhallintajärjestelmiin sekä perehtymistä kerättävää dataa koskeviin tietosuojasäännöksiin (Aaser & Mcelhaney, 2021). Ulkoisen datan käsittely sekä sen tunteminen koettiin sisäistä dataa haastavammaksi. Ulkoinen data voi kuitenkin olla tärkeä hyödynnettävä resurssi esimerkiksi yrityksille, jotka eivät omista tarvittavia datamääriä omaa dataa.

Ulkoista dataa voidaan kerätä lukuisista avoimista ja yksityisistä datalähteistä, mikä tekee siihen perehtymisestä sekä käsittelystä sisäistä dataa työläämpää (Aaser & Mcelhaney, 2021). Tyypillisimpänä avoimena ulkoisen datan lähteenä haastattelussa tuotiin esille Tilastokeskus, joka tarjoaa yrityksille avointa tilastotietoa koko Suomen laajuudella tutkimusta ja päätöksentekoa varten (Tilastokeskus, n.d.). Myös monet haastateltavat mainitsivat Ilmatieteenlaitoksen, Traficom ja Maanmittauslaitoksen olevan usein hyödynnettäviä ulkoisia datalähteitä. Avointen ulkoisen datalähteiden hyödyntäminen markkinoinnin tekoälysovellusten implementoinnissa verrattuna yksityisiin korostui haastatelussa toistuvasti.

Just Ilmatieteenlaitos ja Tilastokeskus ovat hyviä ulkoisen datan lähteitä. Niihin suorat API-rajapinnat, ja niistä löytyy oikein kivasti näihin prosesseihin hyödynnettävää dataa. (H2)

Yksi parhaita lähteitä Suomen tasolla on mm. Tilastokeskus, joka on yksi iso aarreaitta, mistä löytyy vaikka mitä. Se on yksi hyvä paikka, josta löytyy taustatietoa, jota käyttää ohjaamaan oikeeseen suuntaan. Myös Traficomilta on yks hyvä, mistä löytyy liikennemittauspisteitä, kelikameroita, jne. Niistä voi löytää paljon muuta ulkoista dataa. (H1)

Vaikka haastatteluissa ilmeni, että markkinoinnin tekoälysovellusten implementoinnissa hyödynnettävät ulkoiset datalähteet painottuvat useammin avoimiin datalähteisiin, voidaan dataa kerätä myös yksityisten datalähteiden kautta rikastamaan jo olemassa olevaa dataa (Aaser & Mcelhaney, 2021). Esimerkiksi kolmannen osapuolen palveluntarjoajat, jotka keräävät dataa eri lähteistä ja myyvät sitä eteenpäin, eivätkä ole alkuperäisiä datan omistajia, ovat esimerkkejä tästä (Aaser & Mcelhaney, 2021). Kolmannen osapuolen evästeiden hyödyntäminen verkkosivustovierailijoiden seuraamiseen, käyttäjäkokemuksen kehittämiseen sekä mainonnan kohdentamiseen, ovat jo pitkään olleet laajasti käytetty datankeruumenetelmä (Bump, 2021). Erinäisten tietosuojahaasteiden myötä, niistä ollaan kuitenkin luopumassa lähitulevaisuudessa, mikä tulee vaikuttamaan datan saatavuuteen ja hyödyntämiseen (Bump, 2021). Haastateltavat uskovat, että yritysten tulee viimeistään nyt pohtia tarkkaan, mistä ne keräävät tarvittavan datan niin markkinoinnin tekoälysovelluksiin kuin muihin markkinoinnin tehtäviin.

Semmosena, jona me tunnetaan se nyt, niin se tulee loppumaan kokonaan. En usko et semmosta tulee enää. Mut sit me tullaan varmasti menemään semmoseen luotettavampaan malliin, jossa se datan laatu tulee olemaan parempaa. Sitä tulee olemaan vähemmän, eli se haaste tässä tulee. Mut uskon voimakkaasti et se laatu tulee olemaan parempi, sillä yleensä siel tulee olemaan suostumukset takana. (H3)

No käytännössä tässä seuraavina vuosina sitten mietitään mikä ratkaisu on ja mennään vaan kovempiin tunnisteesiin, et mennään tunnistamaan yhä enemmän käyttäjistä sähköposteilla ja puhelinnumeroilla. Eli pyritään saamaan se käyttäjä kirjautumaan, jotta evästeitä ei tarvita. Ja siihen täs kaikki tulee menemään, siihen vahvempaan kirjautumiseen. (H5)

Ulkoisen datan roolin uskotaan jo muuttuneen tietosuojasäännösten ja eväste-
muutosten myötä. Dataa koetaan saavan aiempaa niukemmin GDPR:n voimaantulon jälkeen ja tämän trendin nähdään vain voimistuvan. Tähän yritysten on haastateltavien mielestä myös varauduttava esimerkiksi omilla datankeruujärjestelyillä ja erilaisilla markkinoinnin innovaatioilla.

Dataa on yhä vähemmän saatavilla, eikä keksien kuolema mitenkään helpota tätä asiaa yritysten näkökulmasta. Muutenkin ihmisten tietämys näistä tietosuoja-asioista on aika hyvällä mallilla, joten ne on paljon varovaisempia kuin joskus ennen. Kyl näihin muutoksiin pitää yrittää keksiä omia ratkaisuja eikä jäädä vaan paikalleen valittelemaan. (H7)

Kyl se on jo muuttunut silloin kun GDPR tuli voimaan. Ei dataa oo enää niin helposti saatavilla. Kolmannen osapuolen evästeiden lähtemisen myötä sitä aiemmin saatavilla olevaa dataa olis hyvä jotenkin hyvä miettiä. Varmasti esim. omien asiakasrekisterien kerääminen kaikkien kirjautumisten kautta on yks mikä on nähty siihen keinoon. (H7)

Markkinoinnin tekoälysovelluksissa hyödynnettävä data, kuten aiemmin mainittu, on aina tilanteesta riippuvaista, miksi täsmällistä määrittelyä sisäisen ja ulkoisen datan käytön suhteen ei ole tarvetta tehdä. Haastatteluista pystyi kuitenkin selkeästi vetämään johtopäätöksen, että ulkoinen data on tällä hetkellä vahvasti yritysten sisäistä dataa rikastavassa roolissa, eikä datan kerääminen ja käyttö useimmiten pohjautu siihen. Tietosuojasäännökset, kuten GDPR, ovat jo vaikuttaneet ulkoisen datan keräämiseen ja hallinnointiin, minkä koettiin entisestään vahvistavan oman datan arvoa. Samalla kuitenkin ulkoisen datan käyttäminen nähtiin alihyödynnettynä.

6.3 Datan laatu ja sen merkitys

Datan laatu ja sen merkitys markkinoinnin tekoälysovelluksissa on teema, jonka Hair ja Sarstedt (2021) nostavat tutkimuksessaan esille. Aiheen käsittely jää markkinoinnin kirjallisuudessa kuitenkin melko kapeaksi, minkä lisäksi datan laatu on käsitteenä varsin kontekstisidonnainen, sillä datan laadulle ja sen ulottuvuuksille ei tarjota vain yhtä tiettyä määritelmää (Bertossi & Geerts, 2020; Hair & Sarstedt, 2021). Tämä heijastui selkeästi useaan otteeseen haastatteluista, joissa pyrittiin selvittämään, minkä markkinoinnin tekoälysovellusten asiantuntijat kokevat laadukkaaksi dataksi, mitkä ovat sen laatuun vaikuttavia tekijöitä ja löytyykö laadukkaan datan keräämiselle ja käsittelyyn yhteisesti ymmärrettyjä parhaita käytäntöjä.

6.3.1 Datan laadun kontekstisidonnaisuus

Datan laatu oli haastatteluissa keskustelua herättänyt käsite, sillä useimmat osallistujista kokivat sen olevan erittäin kontekstisidonnaista, kuten myös akateemisessa kirjallisuudessa esitetään (Bertossi & Geerts, 2020). Yksi haastateltavista totesi laadukkaan datan olevan käsite, jota lähinnä löytää vain koulun penkiltä ja erilaisista harjoitustehtävistä, eikä niinkään tosielämän tilanteista. Työelämän käytännön tehtävissä käsiteltävä data on lähes aina jossain määrin vinoutunutta, minkä takia laadukkaan datan käsite kuulosti hieman vieraalta monille haastateltaville. Akateeminen kirjallisuus tarjoaa monia määrittelyjä laadukkaalle datalle, mutta ei vastaavasti usein huomioi haastatteluissa ilmenneitä oikean elämän olosuhteita (Cai & Zhu, 2015)

Vähän semmonen näkemys, et laadukasta ja puhdasta dataa löytää vaan jostain koulun penkiltä ja erilaisist harjoitustehtävistä. Oikeen maailman data on aina jossain määrin vinoutunutta. Näen puhtaan datan enemmän semmosena toivekuvana. Lähimmäs sitä pääsee ehkä esim. digimainonnan järjestelmien kautta, joissa data kirjautuu automaattisesti. (H4)

Tästä huolimatta vastauksista löydettiin useampia teemoja datan laadun määrittelemiseksi, joista yksi nousi yli muiden. Haastateltavat kokivat laadukkaan datan olevan yhteensopivaa sille määriteltyä käyttötarkoitusta varten. Kyseinen

ajatus saa tukea myös kirjallisuudesta, jossa datan laadun uskotaan määrittävän juuri sen käyttötarkoitukseen sopivuudesta (Gudivada, Ding ym., 2017). Myös käyttövalmiin, ajallisesti kohdennettavissa olevan, tarkan sekä ajantasaisen datan uskottiin kuvaavan hyvin laadukasta datajoukkoa.

Kylhän periaattees kaikki data on laadukasta, mikäli se on oikein. Jos halutaan määritellä, et mikä on laadukkuuden määritelmä, et onks se esim. kuinka tarkkaa tai muuta, niin se ehkä rajaa laadukkuutta siihen, et tunnistaa esim. kaikki asiakkaan toimenpiteet verkkosivulla kohti myyntiä tai on muuten niitä tapahtumia kuvaavaa. Vois sanoo et laadukas data on sitä, minkä avulla me saadaan tarkemmin sen asiakkaan käytöksestä tietoa. (H2)

Semmosta, mikä on käyttövalmista. Ei tarvii tehä siistimisiä. Datan luokitteluun menee paljon aikaa, niin ois kiva jos data ois helposti nähtävissä. Sit myös, että data on ajallisesti kohdennettavissa. Jos vaikka kampanjan päättymispäivää ei ole, niin datassa saattaa olla et 30 000 € ulkomainonta kampanjaa yhdelle päivällä vaik se ois ollu viikkoja päällä. Silloin data ei kuvaa todellisuutta. (H6)

Laadukkaaksi koetun datan nähdään kuvaavan parhaalla mahdollisella tavalla sitä esittävää joukkoa ja sopivan sen käyttötarkoitusta varten. Laadukkuus on abstrakti ja tilanteesta riippuva käsite, eikä sitä ollut totuttu käyttämään markkinoinnin tekoanalysoveluksiin liittyvissä käytännön työtehtävissä samalla tavalla kuin akatemisessä kirjallisuudessa.

6.3.2 Datan laatuun vaikuttavat tekijät

Yhteensopivuus jotain tiettyä käyttötarkoitusta varten, on melko laaja ja yleinen määritelmä datan laadulle. Jotta tähän saataisiin konkreettisempaa näkemystä, mitataan datan laatua usein datan eri laatu-ulottuvuuksien, kuten tarkkuuden, jatkuvuuden ja ajantasaisuuden kautta (Gudivada, Ding ym., 2017). Haastateltavat nostivat datan laatuun useimmiten vaikuttavina tekijöinä esille sen käyttötarkoituksen yhteensopivuuden lisäksi saatavuuden, määrittelyn, ajantasaisuuden, tarkkuuden, jatkuvuuden ja johdonmukaisuuden. Heidän mukaansa data tulisi olla helposti ja nopeasti saatavilla sitä tarvitseville. Sen pitäisi olla selkeästi määriteltyä ja dokumentoitua, jotta sen käyttäjät ja joskus tuottajat ymmärtävät, mistä datajoukko muodostuu. Dataa säilytetään nykyään pitkiä aikajaksoja, minkä takia sen ajantasaisuus on myös merkittävä tekijä, joka voi vaikuttaa sen laatuun (Davenport ym., 2020).

No enemmänkin se ongelma on vaan et se ei oo yhdessä paikassa ja nopeasti saatavissa. Se voi olla niin monessa paikassa. Voi olla semmosii yli yön ajoja, eikä saada semmosta reaaliaikasta profiilia muodostettua ja voi profiili voi olla vaikka viikon vanha. (H5)

Varmaan semmonen selkeys ja määrittely. Jos normaalin työn ohessa tulee sivutuotteenä dataa, eikä tiedä, et se tulee johonkin tarkoitukseen. Jos on pyydetty dataa tietyl tasolla, kuten viikkotasolla, niin voidaan saadakin se jossain ihmemuodossa. (H6)

Vaikka haastateltavat olivat yksimielisiä siitä, että data ei juuri koskaan ole täydellinen kuvaus tietystä joukosta, niin sen tulee silti olla mahdollisimman tarkkaa. Riittävää tarkkuutta ei koettu relevantiksi määrittää, sillä se on tilanteesta

riippuvaa, mutta tämä koettiin silti tärkeäksi laatutekijäksi. Jatkuvia ja johdonmukaisia prosesseja datan keräämisessä ja hallinnoinnissa painotettiin vahvasti. Näiden uskotaan luovan hyviä edellytyksiä mahdollisimman laadukkaaseen datan hallintoiintiin.

Datan laadukkuus on syntynyt siitä sen prosessin määrittelystä, millä tavalla sitä dataa kerätään. Jos on johdonmukaisesti ja järkevästi kerätty, niin sillä päästään siihen mahdollisimman laadukkaaseen tulokseen aina. Jos kerätään aina samalla tavalla, niin sil pästään lähemmäksi sitä laadukkuutta. Järjestelmällisyys on myös todella tärkeä. Vaikka jotain jäis keräämättä, niin siitäkin on enemmän hyötyä, jos meil on vaan osa datasta kerättyä, jos se on kerätty järjestelmällisesti ja säännönmukaisesti. Et sitä keräämistä tehdään jatkuvasti ja samalla tavalla pitkän ajanjakson aikana. (H4)

Edellä esiteltyjen laatutekijöiden lisäksi, yhtenä yleisimpänä datan laatuun vaikuttavana tekijänä nähtiin ihminen välikätenä. Datavirheet johtuvat monesti ihmellisistä tekijöistä, kuten näppäilyvirheistä, joita väkisinkin tulee eteen, jos dataa kirjataan esimerkiksi käsin.

Sitten jotain aivan ihmeellisiä Excel-ongelmia, kun on vaikka joku väärä merkki siellä, jolloin tuhannet muuttuu vaikka desimaaleiksi. (H6)

Ainaki se ettei se oo kenenkään ihmisen käsin kirjaamaa missään vaiheessa. Hauska käsittekin semmonen kuin artesaanidata, mikä tarkoittaa just jonkun ihmisen kirjaamaa dataa. Siihen harvemmin voi ihan täysin luottaa. (H4)

Datan laatuun vaikuttavat tekijät ovat varsin kontekstisidonnaisia sekä subjektiivisia elementtejä, joita käsitellään myös akateemisessa kirjallisuudessa vastaavalla tavalla (Gudivada, Ding, ym., 2017). Myös haastateltavat tiedostivat ja toivat tämän esille. Haastateltavien kesken havaittiin kuitenkin huomattavasti yhtäläisyyksiä, joita on esitelty. Nämä havainnot tuovat esiin mielenkiintoisia ja kaihuttuja vastauksia markkinoinnin tekoälysovellusten asiantuntijoilta, sillä markkinoinnin kirjallisuus ei tarjoa kovinkaan kattavaa vastausta määrittelemään datan laadukkuutta ja siihen vaikuttavia tekijöitä (Hair & Sarstedt, 2021).

Yrityksen keräämää sekä hallinnoimaa dataa käsitellään aina jollain tavalla, riippumatta siitä, onko se strukturoitua tai strukturoimatonta (Hair & Sarstedt, 2021). Tutkimukseen osallistuneet kertoivat, että luokittelua tehdään lähes aina datalle ennen kuin sitä otetaan käyttöön niin strukturoidun kuin strukturoimatoman datan tapauksessa.

Kylhän data vaatii aina jotain käsittelyä. Tottakai sitä on automatisoitavissa, mutta tietynlaista luokittelua tarvii aina tehdä datalla. Meil voi olla, vaikka kampanjan nimet, mut meidän pitää pystyy luokittelemaan ne esim. kampanjaryhmiin. Se yksittäin kampanja voi olla liian tarkka, eikä sil tasol voida oikein mitata. Tietynlaista luokittelu pitää melkein aina tehdä. (H2)

Tutkimukseen osallistuneiden joukosta ei tunnustettu muita datan käsittelyyn yleistettävissä olevia menetelmiä, joita tyypillisesti käytettäisiin toistuvasti. Datan laadun varmistamisen uskottiin olevan syvemmillä organisaation ja asiantuntijoiden määrittelemissä prosesseissa, mutta mm. datan visuaalinen vertailu

sekä datan poikkeavaisuuksien havainnointi algoritmeja käyttämällä tuotiin esille tyypillisinä datan laadun tarkkailuun liittyvinä toimenpiteinä.

Joidenkin isojen datamassojen kanssa joudutaan käyttämään joskus outliner detection -algoritmeja. Sit ihan arkipäiväisissä keisseissä semmonen visuaalinen datan tarkistaminen, et näkyys siel jotain hassuja piikkejä. Sit niiden visuaalinen vertailu, jossa niitä käppyröitä tai viivakuvaajii päällekkäin vertaillaan. Sielt aika usein saattaa bongata jonkilaisia kirjausvirhetä. Jos vaikka jollain viikolla on paljon impressioita ja seuraaval viikolla tuplamäärä kustannuksia, niin voidaan havaita näit kirjausvirheitä. Riippuu paljon kuitenkin siitä et millasta dataa hyödyntää. (H4)

Markkinoinnin tekoälysovelluksissa käytettävän datan käsittely ilmeni siis tyypillisesti luokitteluna ja yleisinä datavirheiden tarkasteluna, esimerkiksi visuaalista vertailua hyödyntäen, vaikkakin huomionarvoista on myös muistaa, että se on täysin datajoukosta riippuvaista. Data on markkinoinnin tekoälysovelluksissa tyypillisesti strukturoitua, joka ei strukturoimattoman lailla vaadi yhtä työläitä ja monimutkaisia toimenpiteitä käyttöönottettavaksi (Hair & Sarstedt, 2021).

6.3.3 Datan laadun merkitys

Data on markkinoinnin tekoälysovellusten perusta ja sen laatu luo niiden toimivuudelle edellytykset (Overgoor ym., 2019). Datan laadun varmistamiseksi, on edellä mainitun mukaisesti tärkeintä huolehtia, että se on relevanttia ja sopii määriteltyyn käyttötarkoitukseen. Vaikka tutkimus tarjoaa vahvan ja selkeän kannan tekoälysovellusten toimivuuden ja datan laadun korrelaatioon, selvitetiin tutkimuksessa myös haastateltavien ajatuksia siihen liittyen.

Tutkimukseen osallistuneet olivat yksimielisiä siitä, että tekoälysovellukselle syötetty data tuottaa myös sen laadun kaltaisia tuloksia. He mainitsivat, ettei tekoälysovellus voi koskaan olla laadukkaampi kuin sille syötetty datajoukko. Tämä kiteytettiin usean haastateltavan kommentteilla ”garbage in, garbage out”, jolla viitataan heikkolaatuisen datan tuottavan heikkolaatuisia tuloksia. Laadun merkityksen koettiin olevan erittäin suuri ja onnistumisen määrittävä tekijä.

Se luo sen tarkkuuden siihen malliin. Se tarkentaa niitä meidän malleja. Niiden todennäköisyydet osua oikeeseen ovat paljon parempia, kun saadaan laajempi ja parempi datasetti. (H5)

Garbage in, garbage out. Eli datan laatu määrittää semmosen ylärajan sille et kuinka laadukas analyysi tai sovellus pystytään tekemään. Lopputulokset analyysistä tai algoritmista ei voi koskaan olla laadukkaampia, kuin mitä sinne syötetään sisään. Käytännössä sen datan laatu määrittää, et kuinka hyviä tuloksia voidaan ees teoreettisesti saada. Eli merkitys on erittäin suuri. (H4)

Kuten jo aiemmin tutkimustuloksissa on nostettu esille, haastatteluissa vahvasti painottunut teema oli, ettei data ole lähes koskaan täydellistä tosi maailman tilanteissa. Tärkeimpänä elementtinä läpi haastattelujen tuotiin esille se, että kerätään relevanttia ja haluttua joukkoa luotettavasti kuvaavaa dataa, joka luo hyvät edellytykset tekoälysovellusten implementoinnille. Markkinoinnin akateeminen kirjallisuus toki käsittelee dataa sekä sen laatua ja tämän merkitystä, antaen

vastauksia tekoälysovellusten ja datan keskinäiselle suhteelle (Hair & Sarstedt, 2021; Overgoor ym., 2019). Tätä tarkastelua tehdään kuitenkin useimmiten melko pintapuolisesti, minkä takia siihen syventyminen teemahaastatteluihinsa oli mielenkiintoista ja suotavaa.

Noh garbage in, garbage out. Se on ihan perustavanlaatuisen. Jos heität jotain ihan väärää dataa, niin sun tuloksetkin on väärät. Jos se ei oo kunnossa niin... Datan ei tarvii olla 100 % kunnossa. Kyl sen pitää olla tietyllä tasolla, jotta tuloksiin voi luottaa. (H2)

Markkinoinnin tekoälysovellusten implementointiprosessin laadukas toteuttaminen vaatii siis laadukasta dataa hyödynnettäväksi. Datan laatu on haastattelusta havainnoitujen teemojen valossa merkittävä, mutta sen ei silti tarvitse olla oppikirjamaisesti täydellistä toimiakseen. Tämän koettiin olevan lähes mahdotonta, sillä datasta löytyy aina jotain puutteita, mutta se ei silti tarkoita, että sen laatu olisi heikko. Markkinoinnin tekoälysovellukset antavat ratkaisuja monesti erittäin rajattuihin ongelmiin, joiden toteutus on aina tilanteesta riippuvainen. Tämän johdosta myös laadukkaan datan määrittäminen on erittäin kontekstisidonnaista, vaikkakin edellä esiteltyjä yleisesti hyväksi nähtyjä lainalaisuuksia voidaan nostaa esille.

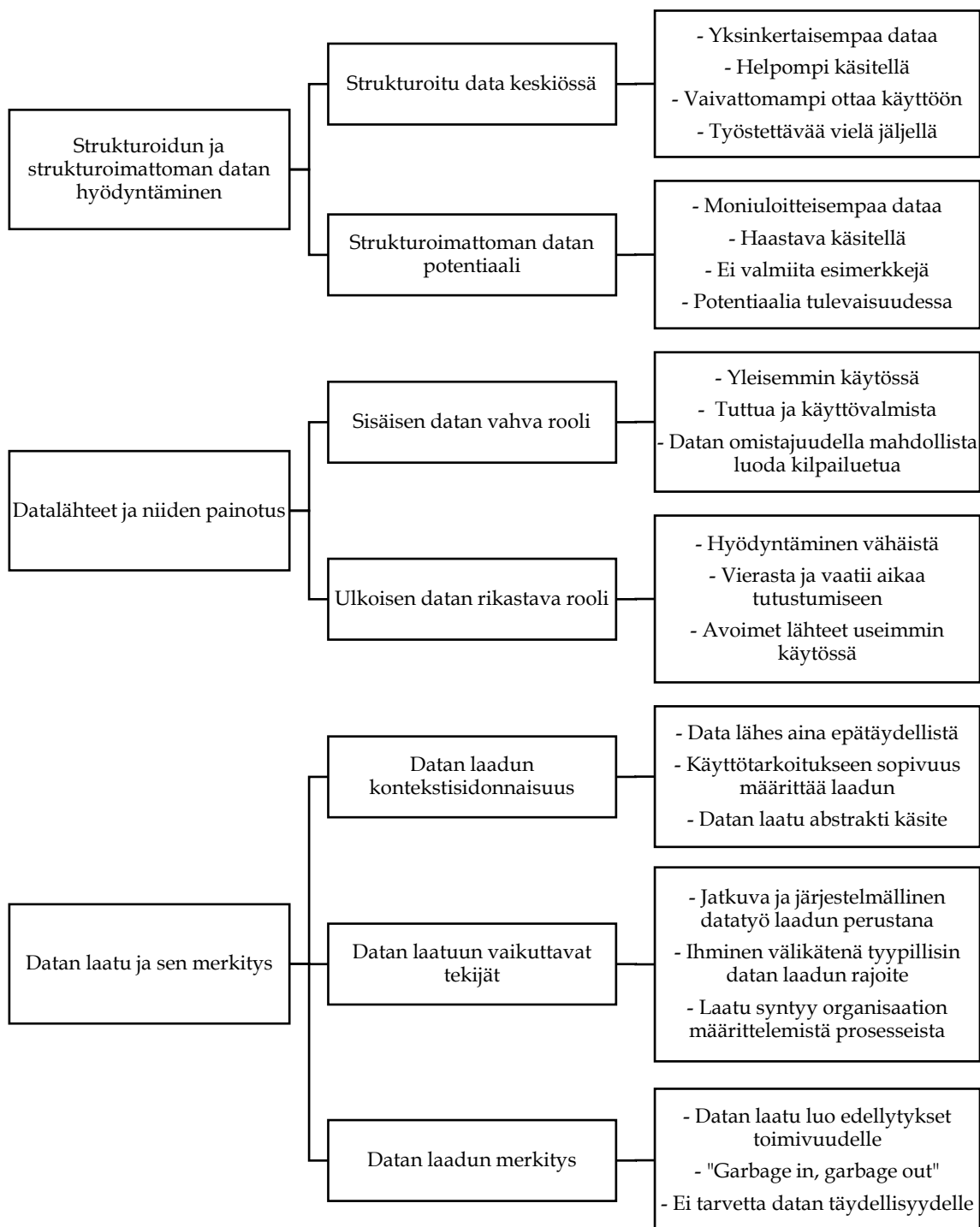
6.4 Tulosten yhteenveto

Kuvio 3 havainnollistaa tutkimustuloksissa esiteltyjen pää- ja alateemojen keskinäisiä suhteita sekä teemojen muodostamisessa hyödynnettyjä koodeja. Näistä ensimmäisenä esitelty pääteema on strukturoidun ja strukturoimattoman datan hyödyntäminen, joka muodostuu kahdesta alateemasta. Alateemat kuvaavat strukturoidun ja strukturoimattoman datan erilaisia rooleja markkinoinnin tekoälysovellusten implementointiprosessissa. Strukturoitu data on yksinkertaisempaa sekä helpompaa käsitellä ja näin ollen vaivattomampi ottaa käyttöön. Tämä tekee siitä huomattavasti strukturoimatonta dataa käytetympää. Strukturoimatonta dataa on haastavampi ottaa käyttöön sen moniulotteisen rakenteen takia. Sillä koetaan olevan suurta potentiaalia tulevaisuudessa, mutta tällä hetkellä strukturoimattoman hyödyntäminen markkinoinnin tekoälysovelluksissa on kuitenkin vähäistä.

Datalähteet ja niiden painotus on tutkimustulosten ja kuvion toinen pääteema, joka sisältää kaksi alateemaa, joita ovat sisäisen datan vahva rooli sekä ulkoisen datan rikastava rooli. Sisäinen data on useimmiten tuttua ja käyttövalmista dataa, minkä takia se on ulkoista dataa yleisemmin käytössä. Huomiota herättävä havainto oli myös se, että yrityksen on mahdollista rakentaa datan omistajuudella kilpailuetua, mikä on vahvasti yhteydessä sisäisen datan arvoon. Ulkoinen data on puolestaan vierasta ja vaatii aikaa tutustumiseen, miksi sen hyödyntäminen on selkeästi vähäisempää. Tyypillisinä ulkoisen datan lähteinä toimivat useimmiten avoimet datalähteet.

Kolmas pääteema, datan laatu ja sen merkitys, rakentui kolmesta alateemasta, jotka havainnoivat laadun kontekstisidonnaisuutta, siihen vaikuttavia tekijöitä sekä sen varsinaista merkitystä. Datan laatu koetaan varsin abstraktiksi käsitteeksi, mutta samalla myös kriittiseksi osaksi markkinoinnin tekoälysovellusten toimintaa. Datan sopivuus tiettyyn käyttötarkoitukseen mielletään kuitenkin yleisesti parhaiten laatua kuvaavaksi tekijäksi. Datan laatutekijöitä tarkemmin tarkasteltaessa havaitaan, että jatkuva ja järjestelmällinen datatyö sekä organisaation määrittelemät selkeät prosessit luovat parhaat edellytykset laadukkaalle datalle. Yksi yleisimmistä laadun epätasaisuuteen vaikuttavista tekijöistä ilmeni olevan ihminen välikätenä. Lisäksi datan laadun merkitys on kiistaton markkinoinnin tekoälysovellusten implementointiprosessissa. Datan ei silti tarvitse olla täydellistä ollakseen laadukasta.

Taulukko 5 vetää puolestaan yhteen edellä esitetyt tutkimustulokset pääteemojen tasolla. Taulukkoon on tiivistetty tutkimustulosten tärkeimpiä havainnot kolmen pääteeman kesken.



KUVIO 3 Tutkimuksen teemoittelun tulokset ja niiden suhteet

TAULUKKO 5 Tutkimustulosten yhteenveto

TEEMA	YHTEENVETO
Strukturoidun ja strukturoimattoman datan hyödyntäminen	<p>Markkinoinnin tekoälysovellusten implementointiprosessissa hyödynnettävä data painottuu tällä hetkellä vahvasti strukturoidun datan ympärille.</p> <p>Tyypillisiä esimerkkejä strukturoidusta datasta ovat mm. digimainonnan performanssidata, kuten klikit ja näyttökerrat.</p> <p>Strukturoitua dataa hyödynnetään huomattavasti strukturoimatonta enemmän, sillä se vaatii vähemmän datatyötä ja on yksinkertaisempi ottaa käyttöön.</p> <p>Vaikka strukturoimattoman datan hyödyntäminen on vielä vähäisempää sen moniulotteisen luonteen ja työlään käsitteilyn takia, nähdään sen omaavan suurta potentiaalia tulevaisuutta ajatellen.</p>
Datalähteet ja niiden painotus	<p>Yritysten sisäiset datalähteet, kuten oman asiakaskannan tai verkkosivustolta kerätty data, ovat markkinoinnin tekoälysovelluksissa käytettävän datan perusta.</p> <p>Asiakasdatan kerääminen eri kirjautumismenetelmien kautta sekä erilaiset kanta-asiakkuusjärjestelmät ovat esimerkkejä viime aikoina esillä olleista menetelmistä, joilla on pyritty vahvistamaan yrityksen sisäisen datan asemaa.</p> <p>Ulkoisen datan, kuten sää- tai sijaintitietojen, rooli on useimmiten sisäistä dataa rikastava, eikä sitä käytetä samoissa määrin kuin sisäistä.</p> <p>Ulkoisen data on yrityksille vieraampaa, ja voi näin ollen vaatia yritykseltä huomattavasti suurempia ponnistuksia siihen tutustumisen ja käsittelyn osalta.</p>
Datan laatu ja sen merkitys	<p>Datan laatu on varsin kontekstisidonnainen käsite, jota parhaiten määrittää datan yhteensopivuus sille määritellyä tarkoitusta varten.</p> <p>Markkinoinnin tekoälysovellukset pohjautuvat vahvasti niille syötettyyn dataan, minkä takia datan laadulla on kriittinen rooli näiden toimivuudessa.</p> <p>Laadukasta ja puhdasta dataa ei koeta löytyvän oikean elämän tilanteista, mutta sitä kohti voidaan kuitenkin pyrkiä tämä asia tiedostaen.</p> <p>Vaikka datan laadun määrittäminen on tilanteesta riippuvaista, koettiin datan saatavuuden, selkeän määrittelyyn, ajantasaisuuden, tarkkuuden, jatkuvuuden ja johdonmukaisuuden luovan edellytyksiä sen laadulle.</p>

7 JOHTOPÄÄTÖKSET JA ARVIONTI

7.1 Tulosten tarkastelu ja johtopäätökset

Tekoälyn hyödyntäminen markkinoinnissa on saanut suurta suosiota viime vuosien aikana niin liiketoiminnan kuin akateemisen tutkimuksen puolella (Borges ym., 2021; Davenport, 2018; Kopalle ym., 2021). Markkinoinnin tekoälysovelluksia ja niiden roolia on tutkittu verrattain aktiivisesti, mutta tekoälysovellusten varsinaista implementointia ja niissä hyödynnettävän datan tutkiminen on jäänyt olemattomaksi (Borges ym., 2021; Overgoor ym., 2019; Pappas ym., 2018). Markkinointi on nykypäivänä monella tapaa datalähtoisempää kuin aiemmin, minkä takia markkinoinnin tutkimuskenttä kaipaa entistä runsaampaa syventymistä datan tutkimiseen (Hair & Sarstedt, 2021). Varsinkin markkinoinnin tekoälyyn liittyvä kirjallisuus tarvitsee enemmän dataan keskittyvää tutkimusta, jota toteutetaan enemmän tietojenkäsittely- ja datatieteiden puolelle, joiden odotetaan olevan entistä aktiivisemmassa roolissa myös markkinoinnin tutkimuskentällä (Hair & Sarstedt, 2021). Datan merkitys markkinoinnin tekoälysovellusten implementointiprosessissa on merkitsevä, sille niiden toimivuus pohjautuu vahvasti syötettyyn dataan (Campbell ym., 2020). Näistä syistä tutkielman aiheen rajaaminen markkinoinnin tekoälysovellusten implementointiprosessissa hyödynnettävän datan tutkimiseen on varsin perusteltua.

Tämän tutkimuksen tavoitteena oli selvittää minkälaiset lainalaisuudet markkinoinnin tekoälysovellusten implementointiprosessissa hyödynnettävään dataan pätevät. Tämän toteuttamiseksi, tutkimukseen kerättiin laadullinen aineisto seitsemän puolistrukturoidun temahaastattelun avulla, mikä analysoitiin sisällönanalyysiä hyödyntäen. Teemoittelu on yksi sisällönanalyysin muoto, jolla tutkija voi etsiä aineistosta kumpuavia teemoja (Nowell ym., 2017). Aineiston sisällönanalyysi toteutettiin Braunin ja Clarken (2006) esittelemän kuusivaiheisen teemoitteluprosessin vaiheita seuraten, jonka lopputulemana aineistosta muodostettiin kolme pääteemaa. Tutkielman viimeisessä luvussa esitellään tutkimuksen teoreettiset johtopäätökset, suositukset liikkeenjohdolle, tutkimuksen arviointi sekä jatkotutkimusehdotukset.

Tekoälyteknologia ja sen mahdollistamat ratkaisut eivät ole aina selkeitä markkinoinnin asiantuntijoille, ja tekoälysovellusten kuvitellaan aika ajoin olevan lähempänä yleistä, kaiken kattavaa tekoälyä, kuin erittäin rajattuun ongelmaan kehitettyä teknologista ratkaisua. Tätä tekoälyteknologian tasoa ei kuitenkaan olla vielä saavutettu, sillä tekoälyratkaisut ovat lähes aina yksinkertaistettuja ratkaisuja erittäin tiukasti rajattuihin liiketoiminnallisiin ongelmiin (Kaplan & Haenlein, 2019; Wirth, 2018). Markkinoinnin tekoälysovellusten asiantuntijat peräänkuuluttavat aiempaa parempaa perehtymistä tekoälyteknologioihin sekä dataan liittyen, jotta yhteistyö eri osastojen välillä olisi entistä saumattomampaa. Hairin ja Sarstedtin (2021) kehotukset tiiviimmästä poikkitieteellisestä

yhteistyöstä markkinoinnin ja tietojenkäsittely- sekä datatieteiden välillä tukevat tätä toivetta.

Strukturoimattoman datan määrä kasvanut nopeaan tahtiin ja saavuttanut jo 80-90 % osuuden kaikesta globaalisesti generoidusta datasta (Hair & Sarstedt, 2021). Tästä huolimatta markkinoinnin tekoälysovellusten implementoinnissa hyödynnettävä data on pääosin strukturoitua, sillä sen käsittely on huomattavasti helpompaa kuin strukturoimattoman. Vaikka strukturoitu data on strukturoimatonta yksinkertaisempaa, ei tämä tarkoita, ettei myös se vaatisi lähes aina jonkinlaista datatyötä. Strukturoidussa datassa uskotaan olevan vielä paljon opittavaa ja kehitettävää jäljellä, minkä takia strukturoimattoman datan hyödyntämiseen ei ole vielä siirrytty vahvemmin. Lisäksi teknologiset kyvykkyydet eivät myöskään tue strukturoimattoman datan hyödyntämistä tarpeeksi kattavasti tällä hetkellä, jotta sen käyttöönotto olisi tarpeeksi vaivatonta. Akateeminen tutkimus antaa tukea strukturoidun ja strukturoimattoman datan hyödyntämisen edellä esitetyle painotukselle markkinoinnin tekoälysovellusten implementoinnissa, mutta empiiriset tutkimustulokset tuovat tähän selkeästi vahvemman kannan strukturoimattoman datan haastavasta käytettävyydestä (Balducci & Marinova, 2018; Hair & Sarstedt, 2021). Tämä tutkimus tuo siis voimakkaampaa painotusta strukturoimattoman datan hyödyntämiselle, mutta myötäilee akateemisessa kirjallisuudessa strukturoimattoman datan potentiaalia tulevaisuudessa, sen suurten volyymien sekä moniulotteisen rakenteen takia (Balducci & Marinova, 2018; Hair & Sarstedt, 2021).

Markkinoinnin tekoälysovellusten implementoinnissa hyödynnettävän datan lähteitä tarkastellessa, sisäisten datalähteiden rooli korostui huomattavasti ulkoisia vahvemmin. Yrityksen sisäinen data on usein selkeämmin organisoitua ja paremmin täsmättävää, minkä lisäksi sitä on helpompi ymmärtää sekä nopeampi ottaa käyttöön. Tämä on merkittävää, sillä markkinoinnin tekoälysovellusten implementointiprosesseissa hyödynnettävä data luo edellytykset sen toimivuudelle, minkä takia datan helppokäyttöisyys ja ymmärrettävyys suosivat sisäisen datan käyttöä (Brock & von Wangenheim, 2019; Overgoor ym., 2019; Stone ym., 2020). Toisin kuin sisäisen datan, ulkoisen datan rooli nähdään monesti sisäistä dataa rikastavana, eikä niinkään markkinoinnin tekoälysovelluksille perustaa luovana. Ulkoinen data on yrityksille sisäistä vieraampaa, mikä nähdään myös akateemisessa kirjallisuudessa muodostavan haasteita datan ymmärtämiselle sekä laadun varmistamiselle (Aaser & Mcelhaney, 2021). Ulkoisen datan rooli on riippuvainen kuitenkin yrityksen datarakenteesta. Jos yritys ei omista tarpeeksi kattavaa omaa kantaa, voi ulkoisen datan vahvempi hyödyntäminen olla tällöin aiheellista.

Datan laadun koetaan muodostavan perustan ja reunaehdot markkinoinnin tekoälysovellusten toimivuudelle, ja näin ollen menestykselle. Nämä empiiriset havainnot saavat myös vahvaa tukea akateemisesta kirjallisuudesta, joka osaltaan kuitenkin jättää aiheen tutkimisen varsin kapeaksi. Tästä johtuen Hair ja Sarstedt (2021) kehottavat markkinoinnin tutkijoita olemaan aiempaa poikkitieteellisempiä ja integroimaan esimerkiksi tietojenkäsittely- ja datatieteistä löytyvää kirjallisuutta datatutkimukseen. Datan laadusta keskustellessa,

markkinoinnin tekoälysovellusten asiantuntijat toivat vahvasti esille, että toisin kuin kirjallisuudessa annetaan ymmärtää, data on tosielämän tilanteissa aina jollain tapaa vinoutunutta ja epätäydellistä. Datan laadun nähtiin olevan erittäin kontekstisidonnainen käsite, minkä takia sen selkeä määrittely voi olla haastavaa. Kontekstisidonnaisuudesta huolimatta, niin tutkimusaineistosta kuin kirjallisuudesta kumpuaa näkemys, että laadukas data on tiettyyn käyttötarkoitukseen sopivuuden mittari (Gudivada, Ding, ym., 2017). Tämän lisäksi tutkimusaineistosta kuitenkin nousi selkeitä markkinoinnin tekoälysovelluksissa hyödynnettävän datan laatuun vaikuttavia tekijöitä, joita olivat saatavuus, selkeä määrittely, ajantasaisuus, tarkkuus, jatkuvuus ja johdonmukaisuus.

Teoreettisten johtopäätösten lisäksi, tutkimuksen tuloksista muodostettiin myös seuraavia liikkeenjohdollisia suosituksia. Markkinoinnin tekoälysovellusten implementointiprosessissa hyödynnettävään datan valikoimisessa on erittäin tärkeä keskittyä ymmärtämään varsinainen liiketoiminnallinen ongelma, johon tekoälyratkaisua ollaan ottamassa käyttöön. Tekoälysovellukset ovat tyypillisesti ratkaisuja erittäin rajattuun ongelmaan, minkä takia relevantin datan hyödyntäminen on ensisijaisen tärkeää. Datan laadun nähdäänkin rakentuvan datan sopivuudesta määriteltyyn tehtävään.

Datan tulee, edellä mainitun mukaisesti, olla relevanttia, mutta sen laatuun vaikuttavat myös muut tekijät. Yritysten olisi tärkeä rakentaa selkeät prosessit datan käsittelylle, jotka ovat luonteeltaan jatkuvia ja mahdollisimman automatisoituja. Datan laatu on suurilta osin juuri yrityksen sisäisten prosessien lopputulema, joka perustuu tasaiselle ja järjestelmälliselle datatyölle.

Tutkimuksen tulokset osoittavat, että data on markkinoinnin tekoälysovellusten implementoinnissa tyypillisesti strukturoitua ja sisäisistä lähteistä kerättyä. Tämä ei kuitenkaan tarkoita, että se olisi sitä myös tulevaisuudessa. Varsinkin strukturoimattoman datan hyödyntämisestä ja sen potentiaalista on puhuttu paljon viime aikoina, vaikka varsinaisia liiketoiminnan käytännön esimerkkejä ei ole vielä juurikaan nähty. Vaikka lyhyen aikavälin katseet olisivat suunnattu tämän hetken prosesseihin ja strukturoidun datan parhaaseen mahdolliseen hyödyntämiseen, on strukturoimattoman datan roolia syytä seurata.

7.2 Tutkimuksen arviointi

Akateeminen tutkimus sisältää useimmiten jonkinlaisia rajoitteita ja haasteita, joita myös tässä tutkimuksessa voidaan havaita. Tämän pro gradu -tutkielman tutkimusaihe jo itsessään rajoittaa osaltaan tutkimusta, sillä akateeminen kirjallisuus on vasta viime vuosina ottanut tarkemmin kantaa markkinoinnin tekoälyratkaisuihin. Tutkimuksen tavoitteena oli lisätä ymmärrystä markkinoinnin tekoälysovellusten implementointiprosessissa hyödynnettävästä datasta, jota ei vastaavassa muodossa ole akateemisessa kirjallisuudessa tutkittu. Tästä syystä tarkoituksena ei ollutkaan muodostaa yleistettävissä olevaa viitekehystä, vaan luoda ja syventää tietämystä verrattain vähän tutkitusta aiheesta.

Myös tutkielman empiirinen osio asettaa rajoitteita tutkimukselle. Haastateltujen asiantuntijoiden lukumäärä oli riittävä muodostamaan selkeät teemat kerätystä aineistosta, mutta tutkimustulosten yleistämiseen ei kuitenkaan riittävä. Teemahaastatteluissa tutkija itse ohjaa haastatteluiden kulkua, jolloin niiden eteneminen ja sisältö on hänen vastuullaan. Tämä voi asettaa varsinkin kokemattomammalle tutkijalle haasteita tutkimusaineiston keräämiseen. Tähän liittyy vahvasti myös tutkijaan liittyvä subjektiivisuus, joka väistämättä on osa tutkimusprosessia. Lisäksi kaikki tutkimukseen osallistuneet toimivat työtehtävissä eri markkinointitoimistoissa Suomen sisällä, mikä osaltaan rajaa tutkimuksen maantieteellisesti pienelle alueelle. Toisaalta markkinointitoimistossa työskentelevien työtehtävät ohjautuvat paljolti asiakkaiden mukaan, jolloin markkinoinnin tekoälysovellusten asiantuntijat ovat tekemisissä monien eri toimialojen kanssa. Tutkimus ei ole täysin yleistettävissä myöskään siitä syystä, että tekoälysovelluksiin löytyvät resurssit ovat esimerkiksi teknologiajäteillä, kuten Googlella ja Applella, täysin eri luokkaa. Näin ollen myös tutkimuksen tulokset ja johtopäätökset saisivat eri kontekstissa erilaisia lopputulemia.

7.3 Jatkotutkimusehdotukset

Markkinoinnin akateeminen kirjallisuus tarkastelee tekoälysovellusten käyttöä ja niiden roolia markkinoinnin tukena melko kattavasti, mutta samalla tekoälysovellusten implementointiprosessi ja niissä hyödynnetyn datan rooli jää kuitenkin varsin pintapuoliseksi. Tämä tutkimus antaa katsauksen tekoälysovellusten implementoinnissa hyödynnettävään datan ympärille, ja luo näin ollen syvempää näkemystä aiheesta. Tutkimusaihe tarvitsee kuitenkin lisätutkimusta, sillä tekoälysovellusten ja niiden implementoinnissa hyödynnettävän datan tutkiminen on markkinoinnin akateemisessa kirjallisuudessa erittäin ohutta.

Tutkimuksen toteuttaminen toi esille muutamia jatkotutkimusehdotuksia markkinoinnin tekoälysovelluksissa hyödynnettävän datan ympärille, joista yhtenä esimerkkinä datan käyttöönottamisen parhaiden käytäntöjen ja prosessin tutkiminen. Myös tekoälysovellusten implementointiprosessia tulisi tutkia aiempaa kattavammin markkinoinnin kirjallisuudessa. Tutkimus tästä aiheesta on erittäin vähäistä, vaikka kyseessä on merkittävä osa uuden teknologian hyödyntämistä. Esimerkiksi implementointiprosessin etenemisen ja sen vaiheiden tutkiminen olisi tästä syystä mielekästä kattavamman käsityksen muodostamiseksi. Tutkimusaihe voitaisiin kohdistaa myös esimerkiksi suurempiin teknologiayrityksiin ja niiden asiantuntijoihin, jotka omaavat täysin erilaisia resursseja markkinoinnin tekoälysovellusten ja käytettävissä olevan datan osalta.

LÄHTEET

- Aaser, M., & Mcelhaney, D. (2021). *Harnessing the power of external data*. 7. <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/harnessing-the-power-of-external-data?cid=other-eml-alt-mip-mck&hdpid=ab36470a-d237-4eeb-ad67-d33fc1ed4db5&hctky=11707914&hlkid=16b6f6236a9e4fbaa3d87152e6ae9739>
- Adnan, K., & Akbar, R. (2019). An analytical study of information extraction from unstructured and multidimensional big data. *Journal of Big Data*, 6(1), 91. <https://doi.org/10.1186/s40537-019-0254-8>
- Alsultanny, Y. (2011). Selecting a suitable method of data mining for successful forecasting. *Journal of Targeting, Measurement and Analysis for Marketing*, 19(3/4), 207–225. <https://doi.org/10.1057/jt.2011.21>
- Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*, 24(1), 1–7. <https://doi.org/10.1016/j.iemeen.2017.06.002>
- American Marketing Association. (2020). *What is Digital Marketing?* <https://www.ama.org/pages/what-is-digital-marketing/>
- Ascarza, E., Ross, M., & Hardie, B. G. S. (2021). Why You Aren't Getting More from Your Marketing AI? *Harvard Business Review*, 99(4), 49–54.
- Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557–590. <https://doi.org/10.1007/s11747-018-0581-x>
- Barrett, A. M., & Baum, S. D. (2017). A model of pathways to artificial superintelligence catastrophe for risk and decision analysis. *Journal of Experimental and Theoretical Artificial Intelligence*, 29(2), 397–414. <https://doi.org/10.1080/0952813X.2016.1186228>
- Bertossi, L., & Geerts, F. (2020). Data Quality and Explainable AI. *Journal of Data and Information Quality*, 12(2). <https://doi.org/10.1145/3386687>
- Borges, A. F. S., Laurindo, F. J. B., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, 57, 1–16. <https://doi.org/10.1016/j.ijinfomgt.2020.102225>
- Braun, V., & Clarke, V. (2006). Applied Qualitative Research in Psychology. *Applied Qualitative Research in Psychology*, 3, 77–101. <https://doi.org/10.1057/978-1-137-35913-1>
- Brock, J. K. U., & von Wangenheim, F. (2019). Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review*, 61(4), 110–134. <https://doi.org/10.1177/1536504219865226>
- Bump, P. (2021). *The Death of the Third-Party Cookie: What Marketers Need to Know*

- About Google's 2022 Phase-Out. HubSpot.
<https://blog.hubspot.com/marketing/third-party-cookie-phase-out>
- Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, 14, 1-10.
<https://doi.org/10.5334/dsj-2015-002>
- Caliskan, A., Özkan Özen, Y. D., & Ozturkoglu, Y. (2020). Digital transformation of traditional marketing business model in new industry era. *Journal of Enterprise Information Management*, 34(4), 1252-1273.
<https://doi.org/10.1108/JEIM-02-2020-0084>
- Campbell, C., Sands, S., Ferraro, C., Tsao, H. Y. (Jody), & Mavrommatis, A. (2020). From data to action: How marketers can leverage AI. *Business Horizons*, 63(2), 227-243. <https://doi.org/10.1016/j.bushor.2019.12.002>
- Chaffey, D., & Ellis-Chadwick, F. (2019). *Digital Marketing: Strategy, Implementation & Practice* (7.). Pearson.
- Chaffey, D., & Patron, M. (2012). From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. *Journal of Direct, Data and Digital Marketing Practice*, 14(1), 30-45.
<https://doi.org/10.1057/dddmp.2012.20>
- Chaffey, D., & Smith, P. (2013). Emarketing Excellence: Planning and Optimizing Your Digital Marketing. In *Emarketing Excellence*.
<https://doi.org/10.4324/9780203082812>
- Crick, J. M. (2021). Qualitative research in marketing: what can academics do better? *Journal of Strategic Marketing*, 29(5), 390-429.
<https://doi.org/10.1080/0965254X.2020.1743738>
- Davenport, T. (2018). *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. The MIT Press.
- Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48, 24-42. <https://doi.org/10.1007/s11747-019-00696-0>
- Davenport, T., & Ronanki, R. (2018). Artificial Intelligence for the Real World. *Harvard Business Review*, 1-10.
- De Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K. U., & von Wangenheim, F. (2020). Artificial Intelligence and Marketing: Pitfalls and Opportunities. *Journal of Interactive Marketing*, 51, 91-105.
<https://doi.org/10.1016/j.intmar.2020.04.007>
- Desai, V. (2019). Digital Marketing A Review. *International Journal of Trend in Scientific Research and Development*, 196-200.
- Dwivedi, Y. K. (2016). Pay-per-click advertising: A literature review. *Marketing Review*, 16(2), 183-202.
- Dwivedi, Y. K., Rana, N. P., Slade, E. L., Singh, N., & Kizgin, H. (2020). Editorial introduction: Advances in theory and practice of digital marketing. *Journal of Retailing and Consumer Services*, 53(August 2019), 2018-2021.
<https://doi.org/10.1016/j.jretconser.2019.101909>
- Ellram, L. M., & Tate, W. L. (2016). The use of secondary data in purchasing and

- supply management (P/SM) research. *Journal of Purchasing and Supply Management*, 22(4), 250–254. <https://doi.org/10.1016/j.pursup.2016.08.005>
- Eriksson, P., & Kovalainen, A. (2016). *Qualitative Methods in Business Research*. Sage.
- Eskola, J., & Suoranta, J. (1998). *Johdatus laadulliseen tutkimukseen* (1.). Kustannusosakeyhtiö Vastapaino.
- Feng, C. M., Park, A., Pitt, L., Kietzmann, J., & Northey, G. (2021). Artificial intelligence in marketing: A bibliographic perspective. *Australasian Marketing Journal*, 29(3), 252–263. <https://doi.org/10.1016/j.ausmj.2020.07.006>
- Frösén, J., Luoma, J., Jaakkola, M., Tikkanen, H., & Aspara, J. (2016). What counts versus what can be counted: The complex interplay of market orientation and marketing performance measurement. *Journal of Marketing*, 80(3), 60–78. <https://doi.org/10.1509/jm.15.0153>
- Gill, K. S. (2016). Artificial super intelligence: beyond rhetoric. *AI and Society*, 31(2), 137–143. <https://doi.org/10.1007/s00146-016-0651-x>
- Gudivada, V. N., Apon, A., & Ding, J. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *International Journal on Advances in Software*, 10(1), 1–20. <https://www.researchgate.net/publication/318432363>
- Gudivada, V. N., Ding, J., & Apon, A. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *International Journal on Advances in Software*, 10(1), 1–20. <https://www.researchgate.net/publication/318432363>
- Gupta, S., Leszkiewicz, A., Kumar, & V, Bijmolt, T., & Potapov, D. (2020). Digital Analytics: Modeling for Insights and New Methods. *Journal of Interactive Marketing*, 51, 26–43. <https://doi.org/10.1016/j.intmar.2020.04.003>
- Haghighatnia, S., Abdolvand, N., & Rajae Harandi, S. (2018). Evaluating discounts as a dimension of customer behavior analysis. *Journal of Marketing Communications*, 24(4), 321–336. <https://doi.org/10.1080/13527266.2017.1410210>
- Hair, J. F., & Brunsveld, N. (2019). Essentials of business research methods. In *Essentials of Business Research Methods*. <https://doi.org/10.4324/9780429203374>
- Hair, J. F., & Sarstedt, M. (2021). Data, measurement, and causal inferences in machine learning: opportunities and challenges for marketing. *Journal of Marketing Theory and Practice*, 29(1), 65–77. <https://doi.org/10.1080/10696679.2020.1860683>
- Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations and Production Management*, 36(10), 1382–1406. <https://doi.org/10.1108/IJOPM-02-2014-0098>
- Hirsijärvi, S., & Hurme, H. (2000). *Tutkimushaastattelu: Teemahaastattelun teoria ja käytäntö*. Yliopistopaino.

- Hirsijärvi, S., Remes, P., & Sajavaara, P. (2009). *Tutki ja kirjoita* (15th ed.). Tammi.
- Huang, M. H., Rust, R., & Maksimovic, V. (2019). The Feeling Economy: Managing in the Next Generation of Artificial Intelligence (AI). *California Management Review*, 61(4), 43–65. <https://doi.org/10.1177/0008125619863436>
- Huang, M. H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>
- Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50. <https://doi.org/10.1007/s11747-020-00749-9>
- Hughes, T., & Vafeas, M. (2019). Marketing Agency/Client Service-For-Service Provision in an Age of Digital Transformation. *Journal of Business-to-Business Marketing*, 26(3–4), 265–280. <https://doi.org/10.1080/1051712X.2019.1611080>
- Hurwitz, J., Kaufman, M., & Bowles, A. (2015). *Cognitive Computing and Big Data Analytics*. Wiley.
- Järvinen, J. (2016). The Use of Digital Analytics for Measuring and Optimizing Digital Marketing Performance. In *Dissertation*.
- Järvinen, J., Tollinen, A., Karjaluoto, H., & Jayawardhena, C. (2012). Digital and Social Media Marketing Usage in B2B Industrial Section. *Marketing Management Journal*, 22(2), 102–117.
- Järvinen, J., Töllinen, A., Platzer, E., & Karjaluoto, H. (2012). Web Analytics and Social Media Monitoring in Industrial Marketing: Tools for Improving Marketing Communication Measurement. *Proceedings of the 41st Academy of Marketing Science Annual Conference*, 477–486.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Kannan, P. K., & Li, H. "Alice." (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, 34(1), 22–45. <https://doi.org/10.1016/j.ijresmar.2016.11.006>
- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25. <https://doi.org/10.1016/j.bushor.2018.08.004>
- Key, T. M. (2017). Domains of Digital Marketing Channels in the Sharing Economy. *Journal of Marketing Channels*, 24(1–2), 27–38. <https://doi.org/10.1080/1046669X.2017.1346977>
- Kibria, M. G., Nguyen, K., Villardi, G. P., Zhao, O., Ishizu, K., & Kojima, F. (2018). Big Data Analytics, Machine Learning, and Artificial Intelligence in Next-Generation Wireless Networks. *IEEE Access*, 6, 32328–32338. <https://doi.org/10.1109/ACCESS.2018.2837692>
- Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer

- journey. *Journal of Advertising Research*, 58(3), 263–267. <https://doi.org/10.2501/JAR-2018-035>
- Klaus, P., & Zaichkowsky, J. (2020). AI voice bots: a services marketing research agenda. *Journal of Services Marketing*, 34(3), 389–398. <https://doi.org/10.1108/JSM-01-2019-0043>
- Knight, S. A., & Burn, J. (2005). Developing a framework for assessing information quality on the World Wide Web. *Informing Science Journal*, 8, 159–172. <https://doi.org/10.28945/493>
- Kopalle, P. K., Gangwar, M., Kaplan, A., Ramachandran, D., Reinartz, W., & Rindfleisch, A. (2021). Examining artificial intelligence (AI) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing*, 1–19. <https://doi.org/10.1016/j.ijresmar.2021.11.002>
- Kruhse-Lehtonen, U., & Hofmann, D. (2020). How to Define and Execute Your Data and AI Strategy. *Harvard Data Science Review*, 2.3. <https://doi.org/10.1162/99608f92.a010feeb>
- Kulkarni, K. K., Kalro, A. D., Sharma, D., & Sharma, P. (2020). A typology of viral ad sharers using sentiment analysis. *Journal of Retailing and Consumer Services*, 53, 1–10. <https://doi.org/10.1016/j.jretconser.2019.01.008>
- Li, J., Wei, Y., & Chiang, Y. H. (2020). Bubbles or cycles? Housing price dynamics in China's major cities. *International Journal of Strategic Property Management*, 24(2), 90–101. <https://doi.org/10.3846/ijspm.2019.11535>
- Ma, L., & Sun, B. (2020). Machine learning and AI in marketing – Connecting computing power to human insights. *International Journal of Research in Marketing*, 37(3), 481–504. <https://doi.org/10.1016/j.ijresmar.2020.04.005>
- Malhotra, N. K., Nunan, D., & Birks, D. F. (2017). *Marketing Research: An Applied Approach* (5th ed.). Pearson. <https://web-a-ebscobhost-com.ezproxy.jyu.fi/ehost/ebookviewer/ebook/bmxlymtfXzE1MzEyODBFX0FOO?sid=aa45908a-08ed-4833-8188-d9b47fbd7972%40sessionmgr4008&vid=0&format=EB&rid=1>
- Martin, K. D., & Murphy, P. E. (2017). The role of data privacy in marketing. *Journal of the Academy of Marketing Science*, 45(2), 135–155. <https://doi.org/10.1007/s11747-016-0495-4>
- Metsämuuronen, J. (2011). *Tutkimuksen tekemisen perusteet ihmistieteissä*. Booky.fi.
- Miklosik, A., & Evans, N. (2020). Impact of Big Data and Machine Learning on Digital Transformation in Marketing: A Literature Review. *IEEE Access*, 8, 101284–101292. <https://doi.org/10.1109/ACCESS.2020.2998754>
- Miklosik, A., Kuchta, M., Evans, N., & Zak, S. (2019). Towards the Adoption of Machine Learning-Based Analytical Tools in Digital Marketing. *IEEE Access*, 7, 85705–85718. <https://doi.org/10.1109/ACCESS.2019.2924425>
- Mitchell, T. (1997). *Machine Learning*. McGraw-Hill.
- Montvilas, T. (2021). Understanding The External Data Revolution. *Forbes*. <https://www.forbes.com/sites/forbesbusinessdevelopmentcouncil/2021/06/25/understanding-the-external-data-revolution/?sh=6123425d150a>

- Nair, K., & Gupta, R. (2020). Application of AI technology in modern digital marketing environment. *World Journal of Entrepreneurship, Management and Sustainable Development*, 17(3), 318–328. <https://doi.org/10.1108/WJEMSD-08-2020-0099>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 16(1), 1–13. <https://doi.org/10.1177/1609406917733847>
- Open Knowledge Foundation. (n.d.). *Open Knowledge Foundation*.
- Overgoor, G., Chica, M., Rand, W., & Weishampel, A. (2019). Letting the computers take over: Using AI to solve marketing problems. *California Management Review*, 61(4), 156–185. <https://doi.org/10.1177/0008125619859318>
- Pan, B., Xiang, Z., Law, R., & Fesenmaier, D. R. (2011). The Dynamics of Search Engine Marketing for Tourist Destinations. *Journal of Travel Research*, 50(4), 365–377. <https://doi.org/10.1177/0047287510369558>
- Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. *Information Systems and E-Business Management*, 16(3), 479–491. <https://doi.org/10.1007/s10257-018-0377-z>
- Paschen, J., Kietzmann, J., & Kietzmann, T. C. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of Business and Industrial Marketing*, 34(7), 1410–1419. <https://doi.org/10.1108/JBIM-10-2018-0295>
- Puusa, A., Juuti, P., & Aaltio, I. (2020). *Laadullisen tutkimuksen näkökulmat ja menetelmät*. Gaudeamus.
- Royle, J., & Laing, A. (2014). The digital marketing skills gap: Developing a Digital Marketer Model for the communication industries. *International Journal of Information Management*, 34(2), 65–73. <https://doi.org/10.1016/j.ijinfomgt.2013.11.008>
- Rust, R. T. (2020). The future of marketing. *International Journal of Research in Marketing*, 37(1), 15–26. <https://doi.org/10.1016/j.ijresmar.2019.08.002>
- Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research Methods for Business Students* (8th ed.). Pearson.
- Saura, J. R. (2021). Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *Journal of Innovation and Knowledge*, 6(2), 92–102. <https://doi.org/10.1016/j.jik.2020.08.001>
- Schneider, M. J., Jagpal, S., Gupta, S., Li, S., & Yu, Y. (2017). Protecting customer privacy when marketing with second-party data. *International Journal of Research in Marketing*, 34(3), 593–603. <https://doi.org/10.1016/j.ijresmar.2017.02.003>
- Shah, D., & Murthi, B. P. S. (2021). Marketing in a data-driven digital world: Implications for the role and scope of marketing. *Journal of Business Research*, 125, 772–779. <https://doi.org/10.1016/j.jbusres.2020.06.062>
- Shah, N., Engineer, S., Bhagat, N., Chauhan, H., & Shah, M. (2020). Research

- Trends on the Usage of Machine Learning and Artificial Intelligence in Advertising. *Augmented Human Research*, 5(1), 1-15. <https://doi.org/10.1007/s41133-020-00038-8>
- Shankar, V. (2018). How Artificial Intelligence (AI) is Reshaping Retailing. *Journal of Retailing*, 94(4), 6-11. [https://doi.org/10.1016/s0022-4359\(18\)30076-9](https://doi.org/10.1016/s0022-4359(18)30076-9)
- Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 5(4), 13-22. <https://doi.org/10.1109/EITech.2016.7519646>
- Siau, K. L., & Yang, Y. (2017). Impact of Artificial Intelligence, Robotics, and Machine Learning on Sales and Marketing. *MWAIS 2017 PROCEEDINGS*, 48. [https://doi.org/10.1016/S0008-6215\(00\)80957-X](https://doi.org/10.1016/S0008-6215(00)80957-X)
- Spiller, L., & Tuten, T. (2015). Integrating metrics across the marketing curriculum: The digital and social media opportunity. *Journal of Marketing Education*, 37(2), 114-126. <https://doi.org/10.1177/0273475315587103>
- Stone, M., Aravopoulou, E., Ekinci, Y., Evans, G., Hobbs, M., Labib, A., Laughlin, P., Machtynger, J., & Machtynger, L. (2020). Artificial intelligence (AI) in strategic marketing decision-making: a research agenda. *The Bottom Line*, 33(2), 183-200. <https://doi.org/10.1108/BL-03-2020-0022>
- Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An Introduction. In *MIT Press* (2nd ed.).
- Syam, N., & Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial Marketing Management*, 69(January), 135-146. <https://doi.org/10.1016/j.indmarman.2017.12.019>
- Taiminen, H. M., & Karjaluo, H. (2015). The usage of digital marketing channels in SMEs. *Journal of Small Business and Enterprise Development*, 22(4), 633-651. <https://doi.org/10.1108/JSBED-05-2013-0073>
- Tariq, M. U., & Abonamah, A. A. (2021). Proposed Strategic Framework for Effective Artificial Intelligence Adoption in Uae. *Academy of Strategic Management Journal*, 20(2), 1-14.
- The Data Wiki. (2019). *Internal Data*. <https://wiki.atlan.com/internal-data/>
- Thontirawong, P., & Chinchachokchai, S. (2021). Teaching Artificial Intelligence and Machine Learning in Marketing. *Marketing Education Review*, 31(2), 58-63. <https://doi.org/10.1080/10528008.2021.1871849>
- Tilastokeskus. (n.d.). *Avoim data*. <https://www.stat.fi/org/avoindata/index.html>
- Tuomi, J., & Sarajärvi, A. (2018). *Laadullinen tutkimus ja sisällönanalyysi*. Tammi.
- Valli, R. (2018). *Ikkunoita tutkimusmetodeihin 1: Metodien valinta ja aineiston keruu: virikkeitä aloittelevalle tutkijalle* (5.). PS-kustannus.
- Verhoeven, R. (2017). *Capitalizing on external data is not only an outside-in concept*. Compact_2. <https://www.compact.nl/articles/capitalizing-on-external-data-is-not-only-an-outside-in-concept/>
- Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International*

- Journal of Information Management Data Insights*, 1, 1–8.
<https://doi.org/10.1016/j.jjime.2020.100002>
- Vilkka, H. (2021). *Tutki ja kehittä* (5th ed.). PS-kustannus.
- Vlačić, B., Corbo, L., Costa e Silva, S., & Dabić, M. (2021). The evolving role of artificial intelligence in marketing: A review and research agenda. *Journal of Business Research*, 128(March 2020), 187–203.
<https://doi.org/10.1016/j.jbusres.2021.01.055>
- Waa. (2008). Web Analytics Definitions. *Web Analytics Association*, 1–32.
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33.
<https://doi.org/10.1080/07421222.1996.11518099>
- Wedel, M., & Kannan, P. K. (2016). Marketing Analytics for Data-Rich Environments. *Journal of Marketing*, 80(6), 97–121.
<https://doi.org/10.1509/jm.15.0413>
- Wirth, N. (2018). Hello marketing, what can artificial intelligence help you with? *International Journal of Market Research*, 60(5), 435–438.
<https://doi.org/10.1177/1470785318776841>
- Wymbs, C. (2011). Digital marketing: The time for a new “academic major” has arrived. *Journal of Marketing Education*, 33(1), 93–106.
<https://doi.org/10.1177/0273475310392544>