

Samuli Ala-Nissilä

# SUOSITTELUJÄRJESTELMIEN EETTISET HAASTEET



JYVÄSKYLÄN YLIOPISTO  
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA  
2021

# TIIVISTELMÄ

Ala-Nissilä, Samuli

Suosittelujärjestelmien eettiset haasteet

Jyväskylä: Jyväskylän yliopisto, 2021, 30 s.

Tietojärjestelmätiede, kandidaatintutkielma

Ohjaaja: Taipalus, Toni

Suosittelujärjestelmät ovat laajasti käytettyjä työkaluja suurien tietomäärien hallintaan ja merkittävässä roolissa monissa digitaalisissa palveluissa, kuten Netflixissä ja Amazonissa. Tämä systemaattisena kirjallisuuskatsauksena toteutettu kandidaatintutkielma toimii hyvänä aloituskohtana suosittelujärjestelmien eettisten haasteiden kirjajaan ja suhteellisen tuoreeseen tutkimukseen. Aiheen tutkimuksesta havaittiin, että niiden toiminnassa ja käytössä on eettisiä haasteita, jotka liittyvät yksityisyyteen, suosittelujen perustelemattomuuteen, algoritmien sulkeutuneisuuteen, suosittelujen eettisyyteen, sosiaalisiin vaikutuksiin, käyttäjän autonomiaan ja reiluuteen. Näillä haasteilla todettiin kirjallisuudessa erilaisia vaikutuksia suosittelujärjestelmien eri sidosryhmille. Lisäksi tutkielmassa esitellään havaittuja ratkaisuehdotuksia kyseisiin haasteisiin.

Asiasanat: suosittelujärjestelmät, etiikka, tekoäly, koneoppiminen, algoritmit

## **ABSTRACT**

Ala-Nissilä, Samuli

Ethical challenges of recommender systems

Jyväskylä: University of Jyväskylä, 2021, 30 pp.

Information Systems, Bachelor's Thesis

Supervisor: Taipalus, Toni

Recommender systems are widely adopted tools for efficient information retrieval. They have gained a notable role in many digital services such as Netflix and Amazon. This Bachelor's thesis was carried out as a systematic literature review and it serves as a great starting point for the diverse and relatively new stream of research focused on different ethical challenges of recommender systems. Observed ethical challenges were related to privacy, opaqueness of algorithms, social effects, fairness, user autonomy, and explainability and ethicalness of recommendations. In the literature these challenges were noted to have different kinds of effects for the different stakeholders involved. In addition, some observed solutions proposed in the literature are showcased.

Keywords: recommender systems, ethics, artificial intelligence, machine learning, algorithms

## TAULUKOT

TAULUKKO 1 Havaittujen eettisten haasteiden pääpiirteet .....	12
---	----

# SISÄLLYS

TIIVISTELMÄ

ABSTRACT

TAULUKOT

1	JOHDANTO.....	6
2	SUOSITTELUJÄRJESTELMÄT .....	9
	2.1 Määritelmä, käyttötapauksia ja merkitys.....	9
	2.2 Menetelmät .....	10
	2.3 Suosittelevjärjestelmät ja etiikka .....	11
3	SUOSITTELUJÄRJESTELMIEN EETTISET HAASTEET .....	12
	3.1 Suosittelevjen perustelemattomuus.....	13
	3.2 Algoritmien sulkeutuneisuus .....	14
	3.3 Reiluus.....	15
	3.3.1 Reiluus käyttäjien ja suosittelukohteiden näkökulmista.....	16
	3.3.2 Reiluus useammalle sidosryhmälle samanaikaisesti .....	17
	3.4 Sosiaaliset vaikutukset.....	18
	3.5 Käyttäjän autonomia .....	19
	3.6 Suosittelevjen eettisyys.....	21
	3.7 Yksityisyys.....	22
	3.8 Pohdintaa ja huomioita tuloksista.....	23
4	YHTEENVETO .....	26
	LÄHTEET .....	28

# 1 JOHDANTO

Suuri osa tämän päivän verkkokaupoista, sosiaalisista medioista, uutissivustoista ja muista digitaalisista palveluista hyödyntävät suosittelujärjestelmiä tarjotakseen osuvampia suositteluja käyttäjille, jonka myötä palvelu koettaisiin mielekkäämmäksi käyttää. Nämä yksilöllistetyt suosittelut voivat olla palvelusta riippuen esimerkiksi tuotteita, mainoksia, uutisia, elokuvia, muiden käyttäjien julkaisuja tai muita palveluja. Käyttäjistä kerätty tieto ja data mahdollistaa yksilöityjen suosittelujen suodattamisen suuresta määrästä informaatiota. Suosittelujärjestelmien mahdollistaman ajansäästön vuoksi ne ovat hyviä työkaluja informaatiotulvaa vastaan.

Suosittelujärjestelmien käyttöön ja toimintaan kuitenkin liittyy eettisiä haasteita, joista useat liittyvät niiden datavetoisuuteen tai tehtyjen suosittelujen suoriin ja epäsuoriin vaikutuksiin. Nämä vaikutukset voivat olla merkitykseltään muun muassa poliittisia, taloudellisia, kulttuurillisia tai henkilökohtaisia. Esimerkkejä näistä ovat suositteluina tarjottujen valeuutisten merkittävä rooli Yhdysvaltain vuoden 2016 presidentinvaaleissa (Allcott & Gentzkow 2017), suosittelujärjestelmän mahdollinen epäreilu vaikutus verkkokaupan tuotteiden kilpailuun näkyvyydestä (Shestra & Yang, 2019; Polonioli, 2020) sekä huomattamaton käyttäjään vaikutus (Seaver, 2019).

Aiheen tutkimus on suhteellisen uutta ja sen useat tutkimukset keskittyvät vain jonkin tietyn toimialan suosittelujärjestelmien tiettyihin eettisiin haasteisiin tai tarkastelevat niitä vain tietyn näkökulman kautta. Tästä syystä näitä haasteita yhteen kokoavat yleiskatsaukset ovat oleellisia. Tämä tutkielma ei kuitenkaan pyri perustavanlaatuisen kaikkia mahdollisia eettisiä haasteita ja ratkaisuehdotuksia kokoavaan lopputulokseen, vaan se pyrkii toimimaan johdatuksena suosittelujärjestelmiin ja niiden kirjajaan eettisten haasteiden aiheeseen käsittelemällä lähdeaineistossa painotetuimmat eettiset haasteet ja niiden yhteydessä mahdollisesti esitetyt ratkaisuehdotukset.

Tutkimuskysymykseni ovat: *"Mitä eettisiä haasteita suosittelujärjestelmät tänä päivänä kohtaavat?"* sekä *"Millaisia ratkaisuja havaituille haasteille on esitetty?"*. Näiden kysymysten avulla kandidaatintutkielmani pyrkii antamaan yleiskuvan aiheen tutkimuksessa käsitellyistä suosittelujärjestelmien eettisistä haasteista.

Tutkielma on toteutettu narratiivisena systemaattisena kirjallisuuskatsauksena Templierin ja Parén (2015) ohjeita noudattaen. Narratiiviset systemaattiset kirjallisuuskatsaukset ensisijaisesti kokoavat olemassa olevaa tieteellistä kirjallisuutta ja tarjoavat lukijoille selkokielisen yhteenvedon tutkittavan aiheen nykytilasta. Ne eivät siis esitä mitään uutta tai kritisoi aikaisempaa kirjallisuutta, vaan säästävät hajautuneesta aiheesta kiinnostuneen lukijan aikaa ja resursseja. (Templier & Paré, 2015.) Näistä syistä kyseinen menetelmä on luonnollinen valinta tässä tutkielmassa esitetyille tutkimuskysymyksille.

Tietokannat, joita käytettiin lähdeaineiston hankintaan ovat ACM Digital Library, IEEE Xplorer ja Scopus. Tutkimuskysymyksiin vastaavan lähdeaineiston haussa hyödynnettiin seuraavaa hakulauseetta: (*recommend\** OR *personaliz\**) AND (*system\** OR *algorithm\** OR "artificial intelligence" OR "machine learning") AND (*ethic\** OR *moral\**). Hakulauseen tuottamia suuria tulosmääriä rajattiin Scopus ja IEEE Xplorer -tietokannoissa asettamalla hakulauseen ensimmäinen sulku koskemaan vain otsikoita ja viimeinen sulku tiivistelmiä. ACM Digital Library -tietokannassa koko hakulause rajattiin tiivistelmiin.

Tietokantojen tuloksista valittiin otsikoiden ja tarpeen vaatiessa tiivistelmien perusteella artikkelit, jotka mahdollisesti vastasivat tutkimusongelmiin. Tämä johti 43 artikkelin joukkoon, jota kavennettiin edelleen perehtymällä jokaisen tiivistelmään sekä tarvittaessa johdantoon ja sisällysluetteloon. Lisäksi muutamia artikkeleita valikoitui takaisinpäin suuntautuvalla lumipallomenetelmällä. Näin päädyttiin 20 artikkelin joukkoon, johon perehdyttiin tarkemmin ja jonka avulla pyrittiin vastaamaan tutkimuskysymyksiin. Julkaisijoiden arvioinnissa hyödynnettiin Julkaisufoorumia ja aineisto täyttää sen puolesta vähintään perustason. Aiheen tutkimuksen ollessa suhteellisen tuoretta ja siten vähäistä, rajaaminen ainoastaan vertaisarvioituihin ja paljon viitattuihin artikkeleihin olisi ollut ongelmallista. Lisäksi pyrittiin sisällyttämään mahdollisimman monia näkökulmia vastattaessa tutkimusongelmaan, jota edellä mainittu rajaaminen ei olisi myöskään helpottanut.

Tutkielmassa on johdannon lisäksi kaksi sisältöluvua. Näistä ensimmäinen määrittelee suosittelujärjestelmät ja niiden merkityksen sekä esittelee niiden hyödyntämiä menetelmiä ja niiden aikaisemman etiikan tutkimuksen. Toisessa sisältöluvussa esitellään kohdittain lähdeaineistossa havaitut suosittelujärjestelmien eettiset haasteet ja niille esitetyt ratkaisut. Näiden lisäksi saman sisältöluvun lopussa on yleistä pohdintaa liittyen muun muassa käsitelyihin eettisiin haasteisiin ja ratkaisuehdotuksiin. Tutkielma päättyy yhteenvedoon, jossa tiivistetysti kerrataan tutkielman ongelmanasettelu ja tavoitteet, esitetään tutkimustulokset ja niiden rajoitukset, tutkielman kontribuutio, vertailu muihin tutkimuksiin ja jatkotutkimusaiheet.

Yleisimmät yläkäsitteet, joihin liittyen haasteita ja ratkaisuehdotuksia löydettiin, olivat yksityisyys, suosittelujen perustelemattomuus, algoritmien sulkeutuneisuus, suositusten eettisyys, sosiaaliset vaikutukset, käyttäjän autonomia ja reiluus. Haasteina nämä kaikki eivät olleet täysin uusia ja niille löytyi erilaisia ratkaisuehdotuksia, joissa esimerkiksi annetaan käyttäjille enemmän mahdollisuuksia vaikuttaa suosittelualgoritmin toimintaan tai mahdollistetaan suositteluille perustavampia perusteluita (Calero, Valdez, Ziefle & Verbert, 2016;

Sharma, 2016). Eettisistä haasteista voitiin yleisesti todeta, että ne voivat muun muassa vaikuttaa suosittelujärjestelmien käyttäjien ajatusmaailman muovautumiseen, turvallisuuteen, koettuun luottamukseen järjestelmää kohtaan ja toteutuneeseen oikeudenmukaisuuteen. Tutkielman kontribuutioina ovat kokoa-va katsaus ja tulokset, jotka ovat pitkälti linjassa Milanon, Taddeon ja Floridin (2020) samasta aiheesta tehdyn tutkimuksen tulosten kanssa.



## 2 SUOSITTELUJÄRJESTELMÄT

Tässä luvussa esitellään mitä suosittelujärjestelmät ovat, missä niitä käytetään ja miten ne toimivat, eli millaisia eri menetelmiä ne voivat käyttää saavuttaakseen tavoitteensa.

### 2.1 Määritelmä, käyttötapauksia ja merkitys

Sacharidis (2020) määrittelee suosittelujärjestelmät tekoälyn sovellutuksina, jotka havaitsevat mieltymysmalleja suurissa ihmisen käytöksen jättämissä jäljissä. Ne siis yksilöllistävät monien digitaalisten palveluiden käyttäjäkokemusta suodattamalla suuresta määrästä tietoa sen, mikä kiinnostaa käyttäjää todennäköisemmin. Tämä on mahdollista hyödyntämällä tietoa käyttäjän aikaisemmista toimista palvelussa, käyttäjän itse ilmaisemia mieltymyksiä, kolmannelta osapuolelta saatua käyttäjään liittyvää dataa tai jopa älypuhelinien sensoreista saatua dataa (Seaver, 2019). Suositteluja voivat olla esimerkiksi tuotteet, hakukoneen tulokset, sosiaalisen median julkaisut, uutiset, muut käyttäjät sekä kaikenlaiset avustukset, jotka pyrkivät edesauttamaan käyttäjän päätöksentekoa. Käyttötapaukset eivät kuitenkaan rajoitu näihin, vaan suosittelujärjestelmiä voidaan soveltaa moneen. Jatkossa tässä tutkielmassa näihin suosittelujen moninaisesti sisältöihin viitataan yleisesti suosittelukohteina. Suosittelujärjestelmien pääasiallisiksi sidosryhmiksi Sacharidis (2020) nimeää suosittelujen vastaanottajat, suosittelukohteiden tarjoajat ja järjestelmän omistajat.

Suosittelujärjestelmät ovat käytössä hyvin laaja-alaisesti ja merkittävässä asemassa tämän päivän yhteiskunnissa. Suurin osa suosituista digitaalisista palveluista kuten Netflix ja Amazon hyödyntävät niitä merkittävänä osana heidän palveluitaan. Ne ovat merkittäviä kyseisille yrityksille, koska niiden tarjonta on niin laajaa, että ilman suosittelujärjestelmiä asiakkaiden olisi huomattavasti vaikeampaa löytää heidän etsimiänsä tuotteita tai elokuvia, mahdollisesti jopa sellaisia, joita he eivät alun perin oivaltaneet edes etsiä. Suosittelujärjestelmät harvoin enää ovat vain käyttöliittymän irrallisia osia suurissa digitaalisissa

palveluissa, vaan suosittelujen lisäksi niitä voidaan käyttää kaiken muunkin käyttäjän näkemän mukauttamiseen. Esimerkiksi Netflix yksilöllistää etusivun elokuvatarjonnan lisäksi myös kategoriat sekä elokuvien kansikuvat. (Seaver, 2019.) Ne sitouttavat käyttäjiä entistä paremmin palveluihin, mikä puolestaan näkyy tulojen kasvamisena joko myynnin, mainostulojen tai muun rahavirran muodossa (Ricci, Rokach & Shapira, 2011; Seaver, 2019).

Aikaisempi suosittelujärjestelmien tutkimus on painottunut paljon tekniiseen näkökulmaan ja suosittelujen tarkkuuteen, eli siihen kuinka oleellisia algoritmien tarjoamat suosittelukohteet ovat. Nykyisin kuitenkin on tunnistettu, että suosittelujärjestelmien laatu muodostuu myös suosittelujen monipuolisuudesta, yllätyksellisyydestä, kontekstin huomioimisesta ja vuorovaikutuksellisuudesta. (Harambam, Helberger & van Hoboken, 2018; Calero Valdez ym., 2016.) Näiden lisäksi suosittelujen perustelut ja suosittelujärjestelmän toiminnan avoimuus, eli taustalla vaikuttavien eri tarkoituksien julkituonti, tulisi ottaa huomioon yhteiskunnallisen arvon ja luottamuksen lisäämiseksi (Harambam ym., 2018).

Näiden järjestelmien merkitys on suuri, sillä ne vaikuttavat mieltymysten muodostumiseen, sosiaaliseen vuorovaikutukseen ja rajaavat niiden käyttäjille näkyviä vaihtoehtoja (Milano, Taddeo & Floridi, 2021; Tang & Winoto, 2016). Niiden on jopa esitetty vaikuttavan vahingollisesti yleiseen poliittiseen keskusteluun, muodostaen kaikukammioita (Mittelstadt, 2016). Kaikukammiot ja muita sosiaalisia vaikutuksia esitellään tarkemmin alaluvussa 3.4. Ihmisten yleinen suhtautuminen ja etenkin luottamus suosittelujärjestelmiä kohtaan on kärsinyt johtuen huolista, jotka usein liittyvät suosittelujärjestelmien toiminnan sala-myhkäisyyteen ja monimutkaisuuteen (Sacharidis, 2020).

## 2.2 Menetelmät

On olemassa erilaisia menetelmiä, joihin suosittelujärjestelmät voivat pohjautua. Ne eroavat toisistaan hyödyntämällä erilaista dataa ja erilaisia algoritmeja. Tässä alaluvussa esitellään yhteistoiminnallisen, sisältöpohjaisen, tietämyspohjaisen, demografisen suodatuksen ja hybridimenetelmien toimintaperiaatteet antaa johdatuksen suosittelujen generoimiseen.

Yhteistoiminnallinen suodatus (engl. collaborative filtering) on yksi yleisimpiä suosittelujärjestelmien hyödyntämiä menetelmiä (Souali, El Afia & Faizi, 2011). Sen toimintaperiaate perustuu oletukseen siitä, että mitä samankaltaisempi suosittelun antanut henkilö on sen vastaanottajan kanssa, sitä todennäköisemmin annettua suositusta noudatetaan (Shestra & Yang, 2019). Esimerkiksi läheisen naapurin tai ystävän suositteluita kuunnellaan todennäköisemmin kuin ventovieraan. Yhteistoiminnallinen suodatus automatisoi tämän prosessin tarjoamalla suosittelukohteita käyttäjälle sen perusteella, millaisia arvioita ne ovat saaneet samankaltaisilta käyttäjiltä (Polonioli, 2020). Tähän menetelmään perustuvat suosittelujärjestelmät voivat jakaa käyttäjiä ryhmiin heidän mieltymystensä perusteella ja suositella samankaltaisille käyttäjille samoja tuotteita.

Tätä kutsutaan käyttäjäperustaiseksi yhteistoiminnalliseksi suodattamiseksi. (Shestra & Yang, 2019.) Tätä menetelmää voidaan käyttää myös suosittelukohdeperustaisesti, jolloin se etsii tarkastelussa oleville kohteille samankaltaiset käyttäjän jo arvioimat kohteet, ja vertaa näitä keskenään arvioidakseen tarkastelussa olevien kohteiden hyödyllisyyden (Shestra & Yang, 2019).

Sisältöpohjaiseen suodatukseen (engl. content-based filtering) pohjautuvat menetelmät oppivat suosittelemaan samankaltaisia asioita, kuin mistä käyttäjä on osoittanut pitävänsä aikaisemmin (Ricci ym., 2011; Polonioli, 2020). Ne ovat siis yksinkertaisempia kuin yhteistoiminnalliseen suodatukseen perustuvat menetelmät. Yksi niiden haittapuoli on oletus käyttäjän kiinnostusten muuttumattomuudesta ajan myötä.

Muitakin menetelmiä on, kuten tietämuspohjainen ja maantieteellinen suodatus sekä useampia menetelmiä yhdistelevät hybridimenetelmät. Useampien menetelmien yhdistämisellä pyritään tasapainoittamaan eri yksittäisten menetelmien hyviä ja huonoja puolia (Ricci ym., 2011). Tietämuspohjainen suodatus tarjoaa suositteluja perustuen tietoon käyttäjien tarpeesta ja suosittelukohteiden vastaavuudesta. Demografinen suodatus puolestaan luokittelee käyttäjäluokat heistä tiedettyjen ominaispiirteiden mukaan (kuten ikä, sukupuoli, kulttuuri, mieltymykset ja sijainti) ja jakaa suosittelut luokkakohtaisesti. (Souali ym., 2011.)

### 2.3 Suosittelujärjestelmät ja etiikka

Etiikka on tutkimusala, joka tutkii moraalialia, eli käsityksiä ja käyttäytymissääntöjä siitä, mikä on hyvää ja paha, oikein tai väärin. Souali ym. (2011) määrittelevät etiikan filosofian osa-alueena, joka on sääntöjen ja toimien kokoelma. Tämä kokoelma pyrkii osoittamaan, miten ihmiset käyttäytyvät ja vuorovaikuttavat keskenään heidän ympäristössään (Souali ym., 2011).

Tässä tutkielmassa suosittelujärjestelmien haasteita tarkastellaan tietokone- ja informaatioetiikan kautta. Tietokone- ja informaatioetiikka on sovelletun etiikan osa-alue, joka tutkii informaatioteknologian eettisiä ja sosiaalisia vaikutuksia yksilöihin ja yhteiskuntaan (Heersmink, van Den Hoven, van Eck & van Den Berg, 2011). Tutkielman painopiste ei siis ensisijaisesti ole sen määrittelyssä, ovatko jotkin asiat itsessään oikein vai väärin, vaan aiheen tutkimuksessa käsiteltyjen vaikutuksien ja huolien tarkastelussa.

Suosittelujärjestelmien etiikan tutkimuksen historia on lyhyt, ja tästä syystä aiheen tutkimusta on suhteellisen vähän. Milano ym. (2020) esittävät aiheen tutkimuksen vähäisyydelle kaksi mahdollista syytä, joista ensimmäinen on kyseisen teknologian suhteellinen uutuus: suosittelujärjestelmien käyttö on yleistynyt vasta 1990-luvulla. Toinen mahdollinen syy on suosittelujärjestelmien algoritmien kehityksen salamyhkäisyys (Milano ym., 2020). Salamyhkäisyydellä tarkoitetaan vaitonaisuutta algoritmien tarkkojen toimintatapojen suhteen, sillä ne ovat merkittävä kilpailuvaltti yritysten välisessä kilpailussa ja tämän vuoksi liikesalaisuuksia (Harambam ym., 2018; Mittelstadt, 2016).

### 3 SUOSITTELUJÄRJESTELMIEN EETTISET HAASTEET

Suosittelujärjestelmien toimintaan liittyvät eettiset haasteet, joihin lähdeaineistossa kohdattiin, tullaan pääpiirteittäin ja kohdittain esittämään tässä pääluvussa. Näiden yhteydessä esitellään myös lähdeaineistossa esitettyjä ratkaisuehdotuksia. Seuraavassa taulukossa on yhteenveto havaituista eettisistä haasteista ja niiden pääpiirteistä (taulukko 1).

TAULUKKO 1 Havaittujen eettisten haasteiden pääpiirteet

Eettinen haaste	Pääpiirteet
Suosittelujen perustelemattomuus	Suositteluun johtaneita tekijöitä ei viestitä käyttäjälle.
Algoritmien sulkeutuneisuus	Suosittelujärjestelmän algoritmin toiminta ei ole avointa, jolloin sen toiminnan luotettavuus ja oikeudenmukaisuus eivät ole helposti tulkittavissa.
Reiluus	Suosittelujärjestelmän toiminta tai suosittelut voivat olla epäoikeudenmukaisia yhtä tai useampaa osapuolta kohtaan.
Sosiaaliset vaikutukset	Käyttäjät ovat alttiita ilmiöille, kuten informaatikuplille ja kaikukammioille, joissa heidän vastaanottamansa tieto on yksipuolisempaa.
Käyttäjän autonomia	Suosittelujärjestelmät voivat olla koukuttavia ja käyttäjien huomaamatta vaikuttaa mieltymysten ja ajatusten muovautumiseen. Käyttäjillä kyseenalaiset vaikutusmahdollisuudet ja oikeudet.
Suosittelujen eettisyys	Suosittelujärjestelmä epäonnistuu, ei yritä tai ei pysty vastaamaan tapauskohtaisiin odotuksiin eettisestä sisällöstä.
Yksityisyys	Käyttäjistä kerätään monenlaista tietoa ilman tiedostettua suostumusta. Käyttäjät ovat alttiita erilaisille tavoille, joiden myötä heidän tietonsa voivat vuotaa.

### 3.1 Suosittelevien perustelemattomuus

Osa suosittelujärjestelmistä ei anna käyttäjille minkäänlaisia tai vain hyvin vavajaisia perusteluja siitä, minkä takia he saavat tiettyjä suositteluja. Tässä alaluvussa käsitellään suosittelujärjestelmien tarjoamia perusteluita käyttäjälle tehdyille suositteluille, sekä millaisia vaikutuksia niiden sisällyttämisellä ja poisjättämisellä voi olla. Eroa tämän ja algoritmien sulkeutuneisuutta käsittelevän luvun välillä voi auttaa hahmottamaan seuraava analogia: suosittelujen perustelut ovat kakkureseptin ainekset, algoritmi on reseptin vaiheittaiset ohjeet siitä, mitä aineksille tehdään ja lopullinen suosittelu on valmis kakku. Tässä luvussa pääasiallisesti siis käsitellään, että mitä aineksia kakkuun on pistetty, eikä niinkään mitä niille on tehty.

Syitä käyttäjälle tehdyille yksittäiselle suosittelulle voivat olla esimerkiksi käyttäjän oma aikaisempi toiminta palvelussa, palvelulle ilmaistut preferenssit, muiden käyttäjien toiminta tai suosittelujärjestelmän tarjoajan omat intressit, kuten tietyn tuotteen myynnin edistäminen tai muu sponsoroitu sisältö. Suosittelevien järjestelmän toiminta on läpinäkyvämpää, kun se tarjoaa käyttäjän halutessa selitykset sen tekemille suositteluille. Eräs esimerkki algoritmin suosittelulle tarjoamasta perustelusta on Netflixin ”Koska katsoit tämän: <elokuvan nimi>”, joka mainitsee käyttäjän aikaisemmin katsoman elokuvan nimen perusteluna sen tekemille elokuvasuositteluille. Toinen on Amazonin verkkokaupan ”Asiakkaat, jotka ostivat tämän, ostivat myös <tuotesuosittelut>”. Suosittelevien järjestelmän ilmoittama luku käyttäjälle siitä, että kuinka suurella luottamuksella se arvioi käyttäjän pitävän yksittäisistä suositteluistaan, on yksi tapa viestiä käyttäjälle suosittelujen takaisista arvioista, vaikkei kovin perustavanlaatuinen. Tavat eivät kuitenkaan rajoitu esiteltyihin, vaan niitä on useampia.

Suosittelujen perusteluilla on todettu positiivisia vaikutuksia, kuten lisääntynyt käyttäjien luottamus kyseistä suosittelujärjestelmää kohtaan (Tintarev & Masthoff, 2007). Vaikka suosittelujärjestelmä olisi muuten eettinen, mutta se ei ole millään tapaa avoin toiminnastaan eikä perustele tekemiään suositteluja, niin käyttäjä ei voi olla varma sen eettisyydestä, johtaen mahdollisesti epäluottamukseen järjestelmää kohtaan tai jopa käytön lopettamiseen. Luottamuksen lisäksi on muita yleisiä vaikutuksia, joita algoritmien perustelut voivat pyrkiä saamaan aikaan: ne näyttävät miten järjestelmä toimii, auttavat käyttäjiä tekemään hyviä päätöksiä, vakuuttavat käyttäjiä, lisäävät käyttäjätyytyväisyyttä ja antavat mahdollisuuden käyttäjälle ilmaista, jos järjestelmä on väärässä (Tintarev & Masthoff, 2007). Jos suosittelun sisältö on painoarvoltaan merkittävä, kuten poliittinen mediasisältö, niin algoritmien läpinäkyvyyttä voidaan vaatia käyttäjien oikeutena (Mittelstadt, 2016).

Suosittelujen perusteluun liittyy kuitenkin haasteita, jotka koskevat niiden mahdollisia haitallisia vaikutuksia sekä kompromisseja, mitä niiden mukaanoton myötä saattaa joutua tekemään. Polonioli (2020) esitti, että suosittelujen perustelevien saattaa vahvistaa epäoikeudenmukaisuutta algoritmin toiminnassa. Esimerkiksi tuotteen suosittelun perustelevien sen yleisellä suosiolla

saattaa lisätä sen haluttavuutta, mikä voi johtaa itseään vahvistavaan kehään, jonka seurauksena tuotetta suositellaan jatkuvasti entistäkin enemmän. Tällöin suosittelun perustelu on johtanut tilanteeseen, jossa voittaja vie kaiken, ja algoritmin tarjoamien suosittelujen monipuolisuus kärsii. (Polonioli, 2020.) Tämänkaltaista suosittelujärjestelmien algoritmien taipumusta epäoikeudenmukaisuuteen käsitellään tarkemmin reiluutta tarkastelevassa luvussa 3.3. Suosittelujen perustelujen generoiminen voi olla suosittelujärjestelmän menetelmästä riippuen laskennallisesti liian monimutkaista sekä hidasta (Mittelstadt, 2016; Paraschakis, 2017) sekä se saattaa vaikuttaa myös suosittelun tarkkuuteen negatiivisesti (Shulman & Wolf, 2020).

Sacharidis (2020) pisti merkille kaksi puutetta suosittelujärjestelmien perusteluiden tutkimuksessa. Näistä ensimmäisessä vain yksi sidosryhmä eli käyttäjän otetaan huomioon, vaikka kaikkien sidosryhmien luottamusta tulisi kasvattaa. Toisena puutteena vain yksittäinen suosittelu perustellaan kerrallaan sen sijaan, että algoritmin toimintaa seurattaisiin myös pitkäjänteisesti ja tästä seurannasta olisi johdettavissa tulkintoja algoritmin toiminnasta. (Sacharidis, 2020.)

### 3.2 Algoritmien sulkeutuneisuus

Suuri osa tämän päivän suosittelujärjestelmien algoritmeista ovat niin sanottuja mustia laatikoita (engl. black box), joista on vaikea tulkita tai selvittää yksittäisten perusteluiden syiden lisäksi yleistä kuvaa suosittelujen muodostamisen prosessista. Ne saavat syötteenä dataa käyttäjistä sekä suosittelukohteista ja antavat ulostulona suositteluja, mutta suosittelujen muodostamisen prosessi itsessään ei yleensä ole käyttäjille tai muille nähtävissä. (Paraschakis, 2017.) Sacharidis (2020) määritteli algoritmien sulkeutuneisuuden vastakohtana, eli läpinäkyvyyden, ihmisen ymmärrettävissä olevana tulkintana järjestelmän sisäisestä toiminnasta. Algoritmien sulkeutuneisuus on haitallista, koska tällöin on vaikeaa selvittää sen tekemien suosittelujen syitä ja alkuperäisiä tarkoitusperiä sekä algoritmin mahdollisia taipumuksia epäoikeudenmukaisuuksiin ja alttiutta manipulointiin (Paraschakis, 2017). Tarkoituserillä tarkoitetaan sitä, että millä tavoin suosittelujärjestelmän eri sidosryhmien tai käyttäjien intressit pyrkivät vaikuttamaan sen toimintaan. Epäoikeudenmukaisuuksilla tarkoitetaan algoritmin epäreilua toimintaa.

Algoritmien sulkeutuneisuudelle voi olla useita eri syitä. Niiden tehokkuus on merkittävä kilpailuvaltti useille digitaalisia palveluita tarjoaville yrityksille, kuten Facebookille, Netflixille ja Amazonille. Tämän vuoksi ne ovat tarkoin varjeltuja liikesalaisuuksia, sillä algoritmien toiminnan liika paljastaminen saattaa altistaa ne kopioinnille kilpailijoiden toimesta. (Harambam ym., 2018; Mittelstadt, 2016.) Muut syyt liittyvät käyttäjien yksityisyyden turvaamiseen ja turvallisuuteen (Mittelstadt, 2016).

Harambam ym. (2018) esittivät, että teknisten vaikeuksien lisäksi algoritmien avoimuuden lisäämistä vaikeuttaa niiden kehitystyölle ominaiset piirteet. Heidän mukaansa algoritmit ovat monimutkaisia, koska ne ovat monien väliintulojen sekä erilaisten ihmisten tuotteita, joilla on eriävät kiinnostuksen kohteet, arvot ja päämäärät. Tämän lisäksi kehitystyössä tulee ottaa huomioon, että algoritmien pitää voida olla vuorovaikutuksessa eri sidosryhmien ja muiden algoritmien kanssa. Niitä myös tulee voida käyttää erilaisissa alustoissa. (Harambam ym., 2018.) Nämä piirteet vaikeuttavat kehitettävän algoritmin läpinäkyvyyden määrittelyä ja toteutusta. Tulevaisuudessa itseään kehittävien algoritmien läpinäkyvyys voi olla jopa mahdotonta, sillä ne saattavat ylittää ihmisen kapasiteetin ymmärtää niiden logiikkaa. (Harambam ym., 2018.)

Algoritmien sulkeutuneisuudelle tuli lähdeaineistosta ilmi erilaisia ratkaisuehdotuksia. Ensimmäinen näistä on sääntelyvirasto, joka lainsäädännön keinoin pyrkii valvomaan tekoälyyn pohjautuvien algoritmien eettisyyttä (Erdélyi & Goldsmith, 2018; Mittelstadt, 2016). Toisena ratkaisuna on menetelmät, jotka sallivat sulkeutuneiden suosittelujärjestelmien toiminnan ulkopuolisen tarkastelun. Eräs tällainen metodi kokeilee millä tavoin ja kuinka paljon käyttäjien erilaiset toimintahistoriat vaikuttavat suosittelujärjestelmän tarjoamiin suositteluihin (Lonjarret, Robardet, Plantevit, Auburtin & Atzmueller, 2020). Näin se luo käsityksen siitä, miten taustalla olevat algoritmit toimivat. Se muun muassa tunnistaa käyttäjän preferenssien ja toimien järjestyksen vaikutuksen tarjottujen suosittelujen kokonaisuuteen, mistä voidaan edelleen päätellä algoritmin yleisesti tekemien suosittelujen syitä. (Lonjarret, ym., 2020.)

### 3.3 Reiluus

Sacharidis (2020) määritteli suosittelujärjestelmän reiluuden sen algoritmin oikeudenmukaisuutena sekä suosittelujen monimuotoisuuden, oleellisuuden ja vaikutusten huomioimisena. Shestra ja Yang (2019) erittelivät reiluuden kolmeen osaan käyttäjille ja ryhmille, suosittelukohteille ja samanaikaisesti useille sidosryhmille osoitettuna reiluutena. He totesivat suosittelujärjestelmien reiluuden tutkimuksen olevan haastavaa ja monimutkaista johtuen siitä, että suosittelujärjestelmät usein koostuvat useammista osista, niiden on tasapainoteltava useampien päämäärien välillä ja niiden toteutuneen reiluuden arvioiminen on haasteellista (Shestra & Yang, 2019). Vaikka algoritmien tulo yleisesti on ehkäissyt ja paljastanut epäoikeudenmukaisuuksia ihmisille saatavissa olevassa tiedossa, on se myös todistetusti tuonut uusia epäoikeudenmukaisuuksia, joilla on ollut tahattomia ja vahingollisia vaikutuksia (Barocas & Selbst, 2016). Syy tällaiseen suosittelujärjestelmän toteuttamiseen diskriminaatioon voi juontaa itse algoritmiin tai sen hyödyntämään dataan.

Seuraavassa alaluvussa käsitellään suosittelujärjestelmien reiluutta yksittäisten sidosryhmien näkökulmasta. Sitä seuraavassa alaluvussa käsitellään reiluutta useamman sidosryhmän samanaikaisesti toteutuvan reiluuden näkökulmasta.

### 3.3.1 Reiluus käyttäjien ja suosittelukohteiden näkökulmista

Yksittäisille käyttäjille ja suosittelukohteiden tarjoajille on olemassa erilaisia haasteita liittyen suosittelujärjestelmien reiluuteen, joiden lähteet voivat juontaa joko algoritmiin tai sen käyttämään dataan. Joidenkin algoritmien on osoitettu vahvistavan sen käyttämässä datassa olevia ennakkoluuloja ja vääristymiä, jotka voivat liittyä esimerkiksi sukupuoleen, kulttuuriin tai etniseen ryhmään (Shestra & Yang, 2019). Tätä kutsutaan systemaattiseksi vääristymäksi (engl. systematic bias), ja siitä voi olla haittaa esimerkiksi työnhaussa, kun työnantajan suosittelujärjestelmä arvioi ehdokkaiden soveltuvuutta ja järjestelmä syrjii ulkomaalaisen nimen omaavia ehdokkaita datasta opittujen ja vahvistuneiden ennakkoluulojen vuoksi. Systemaattiseen syrjintään tulisi yrittää puuttua suosittelujärjestelmää varten hyödynnettävän datan tiedonlouhinnassa huomioimalla sen mahdollisuus diskriminaatioon. Tiedonlouhinnalla tarkoitetaan käytötarkoituksen kannalta oleellisen tiedon löytämistä suurista tietomassoista. Kun diskriminaatiota havaitaan, niin sitä voidaan korjata. (Paraschakis, 2017.)

Suosittelujärjestelmien epäreiluus voi myös olla niiden algoritmeihin tarkoituksellisesti sisäänrakennettua, kuten hintadiskriminoinnissa ja A/B-testauksessa. Hintadiskriminoinnissa hotellihuoneiden hinnat muuttuvat sen mukaan kuka niitä tarkastelee, eli hinnat perustuvat algoritmin arvioon suurimmasta määrästä rahaa, jonka asiakas on valmis maksamaan. On myös mahdollista, että suosittelujärjestelmän ylläpitäjä suorittaa A/B-testauksia, joissa vain osa järjestelmän käyttäjistä altistetaan kehitteillä olevalle algoritmille selvittääkseen sen tehokkuuden suhteessa aikaisempaan algoritmiin. (Paraschakis, 2017.) Paraschakis (2017) rinnasti tämän lääketieteellisiin kokeiluihin, joissa uusien metodien kokeilu saattaa johtaa arvaamattomiin sivuvaikutuksiin ja harmiin, ja suurimmassa osassa tapauksia käyttäjät ovat tässä mukana heidän tietämättään.

Etenkin suosittelukohteiden puolesta oleellinen haaste, joka liittyy algoritmin tahattomaan epäoikeudenmukaisuuteen, on huomiovääristymä (engl. observation bias). Huomiovääristymässä käyttäjiltä positiivisen reaktion saaneet suosittelut saavat suosittelujärjestelmältä enemmän näkyvyyttä, mikä puolestaan johtaa entistä suurempaan näkyvyyteen (Shestra & Yang, 2019). Tällöin kyseessä on itseään ruokkiva silmukka, jossa suosituimmat suosittelukohteet saavat suhteessa enemmän näkyvyyttä verrattuna muihin kohteisiin, mikä samanaikaisesti myös pitää enemmän näkyvyyttä saaneet kohteet näkyvänä.

Toinen mahdollinen diskriminaatio on suosiovääristymä (engl. popularity bias), jossa jopa suhteellisen pienenkin ryhmän osoittama suosio voi nostaa jotkin kohteet yleisesti suositelluimmiksi (Polonioli, 2020). Vaikutus on siis sama kuin huomiovääristymässä, eli jotkin suosittelukohteet saavat epäoikeutetusti enemmän näkyvyyttä ja muut jäävät niiden varjoon. Tämä on haitallista esimerkiksi silloin, kun suosittelukohteiden tarjoajat kilpailevat näkyvyydestä keskenään. Syy ilmiölle on kuitenkin eri, sillä huomiovääristymässä epäonnis-



tutaan huomioimaan suosittelukohteelle lisätyn näkyvyyden vaikutusta sen suosioon, kun taas suosiovääristymässä pienen ryhmän osoittama suosio arvioidaan virheellisesti koskemaan todellista suurempaa yleisöä.

Suosittelujärjestelmien epäreiluuteen on esitetty monenlaisia ratkaisuja, jotka pyrkivät lisäämään suosittelujen monipuolisuutta sekä mahdollistamaan niiden reiluuden tarkastelun sitä mittaavien mittareiden avulla (Shestra & Yang, 2019). Yleinen haaste ratkaisussa kuitenkin on, että on tehtävä jonkinasteinen kompromissi suosittelujärjestelmän koetun hyödyllisyyden ja sen reiluuden välillä (Shestra ja Yang, 2019; Kermany, Zhao, Yang, Wu & Pizzato, 2020). Ratkaisuina Paraschakis (2017) painotti käyttäjän vuorovaikutusmahdollisuuksien lisäämistä, suosittelujen perusteluja, järjestelmän avoimuutta sekä algoritmien reiluutta valvovia sääntelyvirastoja. Vuorovaikutuksen mahdollisuus voi esimerkiksi tarkoittaa sitä, että käyttäjä voi itse valita haluaako hän mahdollisimman yksilöllistettyjä suositteluja vai haluaako hän olla osallisena uusien algoritmien kokeilussa.

### 3.3.2 Reiluus useammalle sidosryhmälle samanaikaisesti

Lähdeaineistosta tuli ilmi hiljattain suosiotaan kasvattanut lähestymistapa, jossa korostetaan suosittelujärjestelmien eettisyyden tarkastelua useampien sidosryhmien ja näkökulmien kautta samanaikaisesti pelkän käyttäjän sijaan (Milano, Taddeo & Floridi, 2021; Karpati, Najjar & Ambrossio, 2020). Siinä pyritään tunnistamaan suosittelujärjestelmän kannalta kaikki oleelliset sidosryhmät, joista yleisimmät Milanon ym. (2021) mukaan ovat käyttäjät, palveluntarjoajat, järjestelmä ja yhteiskunta. Tämä lähestymistapa ottaa järjestelmällisesti huomioon haasteiden sekä välittömät että tulevat seuraukset ja se myös auttaa ymmärtämään kokonaisvaltaisemmin eettiset, yhteiskunnalliset ja lainsäädännölliset puolet, jotka jäävät vähemmälle huomiolle käyttäjäkeskeisessä tutkimuksessa (Karpati ym., 2020).

Tätä lähestymistapaa voidaan hyödyntää yleisessä suosittelujärjestelmien kehitystyössä linjaamalla näkökulmat, joiden kannalta järjestelmän toimien seurauksia tarkastellaan (Milano ym., 2021). Esimerkiksi kun tuotteen suosittelu johtaa käyttäjän ostopäätökseen, niin siitä hyötyvät käyttäjä, tuotteen valmistaja ja verkkokauppa. Jos tuote kuitenkin on ominaisuuksiltaan ympäristölle haitallinen, niin suosittelun toteutunut vaikutus yhteiskunnan kestäväen kehityksen kannalta on negatiivinen.

Suosittelujärjestelmiä, jotka ovat kehitetty tavoittelemaan samanaikaisesti useampien sidosryhmien reiluutta, kutsutaan monisidosryhmäsuosittelujärjestelmiksi (engl. multi-stakeholder recommender system). Ne tavoittelevat kompromissia eri suosittelujärjestelmän sidosryhmien tavoitteiden välillä, huomioimalla samanaikaisesti käyttäjille kiinnostavien ja oleellisten suosittelujen tarjonnan, sekä häntäpäin vähemmän suosittu suosittelukohteet (Kermany ym., 2020). Ne ovat yleisiä palveluissa, joissa on useamman puolen markkinat (engl. multisided market), kuten Spotifyn käyttäjät, sisällöntuottajat ja palvelun omistajat (Milano ym., 2021). Ratkaisuja useamman puolen markkinoiden haasteisiin

on kehitetty, joista yksi esimerkki on Kermanyn ym. (2020) esittelemä algoritmi, joka pyrkii ottamaan kaikki osapuolet huomioon suhteellisen pienellä suosittelujen oleellisuuden menetyksellä.

### 3.4 Sosiaaliset vaikutukset

Meillä kaikilla on jonkinasteinen taipumus hyväksyä omaa aikaisempaa käsitystä vahvistava tieto suhteellisesti helpommin verrattuna sitä vastustavaan. Tämä voi tapahtua esimerkiksi hakemalla valikoivasti tietoa tai lopettamalla objektiivinen tiedonhaku heti, kun valmiiksi oletettu johtopäätös on saavutettu. (Polonioli, 2020.) Tätä kutsutaan vahvistusharhaksi (engl. confirmation bias). Suuremmat ilmiöt, joita taipumus vahvistusharhaan voi saada aikaan yhdessä suosittelujärjestelmien liian yksilöllistettyjen sisältöjen kanssa, ovat herättäneet huolta käyttäjissä ja saaneet paljon huomiota tutkijoilta (Polonioli, 2020; Harambam ym., 2018; Harambam, Bountouridis, Makhortykh & van Hoboken, 2019; Lonjarret ym., 2020; Paraschakis, 2017). Näitä paljon huomiota kerryttäneitä ilmiöitä ovat olleet erityisesti informaatiokuplat (engl. filter bubble) ja kaikukammiot (engl. echo chamber).

Mediasisältöjen liika yksilöllistäminen voi johtaa tilanteeseen, jossa niiden kuluttaja ei enää kohtaa eriäviä mielipiteitä omistaan, vahvistaen vahvistusharhan vaikutusta. Etenkin suosittelujärjestelmät, jotka välittävät tietoa kuten uutisia ja sosiaalisen median sisältöä, voivat luoda niin sanottuja informaatiokuplia (Paraschakis, 2017). Tästä syystä Harambam ym. (2018) olivat sitä mieltä, että liian suuri suosittelujen tarkkuus ei ole hyväksi varsinkin, jos niiden painoarvo on yhteiskunnallisesti merkittävä, sillä erilaisten ideoiden ja mielipiteiden huomioiminen on oleellinen osa demokraattista mediaa.

Kaikukammioita korostetaan yhtenä merkittävänä haasteena (Mittelstadt, 2016; Polonioli, 2020). Kaikukammioita voi syntyä, kun ihmiset tietoisesti tai tiedostamatta yksilöllistettyjen suosittelujärjestelmien avustamana hakeutuvat internetin yhteisöihin, joissa he eivät kohtaa vastakkaisia mielipiteitä (Polonioli, 2020). Oleellinen osa ilmiötä on nämä yhteisöt, joiden jäsenet tukevat toisiaan ja näin vahvistavat keskenään jaettuja mielipiteitä. Web 2.0:n tulo on mahdollistanut yhtäaikaisesti hyvin pienilukuisten, samanhenkisten ja ympäri maailmaa hajautuneiden yhteisöjen muodostuksen, joiden ajatusmaailmat ja kiinnostuksen kohteet voivat olla erittäin uniikkeja. Äärimmäisissä tapauksissa esimerkiksi jokin salaliittoteoria voi toimia yhdistävänä tekijänä. Yhdistävä tekijä voi myös olla hyvin tavallinen, kuten jokin arkinen kiinnostuksen kohde, jolloin ilmiö voi ilmetä vain lievästi.

Milano ym. (2021) esittivät Bradshawin ja Howardin (2019) osoittaneen, että lähes ainoastaan omia aikaisempia käsityksiä vahvistavien sisältöjen kuluttaminen, ja siihen liittyvä edellisessä tekstikappaleessa käsitelty ryhmytyminen, voivat syventää eriävien mielipiteiden omaavien ihmisten välistä kuilua entisestään. Tätä kutsutaan polarisaatioksi, joka voi ilmetä esimerkiksi poliittisten ääripäiden suosion kasvuna. Tällöin ääripäillä on itseään vahvistava vaiku-

tus ja vastakkaiset mielipiteet voivat tuntua entistäkin kaukaisemmilta, hankaloittaen poliittista diskurssia ja demokratiaa yleisesti (Mittelstadt, 2016; Harambam ym., 2018). Kaikukammioita ja polarisaatiota käsittelevistä keskusteluista on kuitenkin havaittavissa moraalista paniikkia, ja aiheen tämänhetkinen empiirinen tutkimus ehdottaakin hälyttävän sävyisen keskustelun lieventämistä (Harambam ym., 2018).

Yleisiä ehdotettuja ratkaisuja edellä mainittuihin sosiaalisiin vaikutuksiin ovat olleet tiedon monipuolistaminen, käyttäjän laajemmat vaikutusmahdollisuudet, suosittelujen perustelut, toiminnan läpinäkyvyys ja algoritmien sääntelyvirasto, joka valvoisi ja sääntelisi algoritmeja (Mittelstadt, 2016). Esimerkkinä erityisesti käyttäjän vaikutusmahdollisuuksiin ja toiminnan läpinäkyvyyteen perustuvasta ratkaisusta Harambam ym. (2018) ehdottivat uutissivustojen suosittelujärjestelmille niin sanottuja ”suositteluhenkilöitä” (engl. recommender personae). Suositteluhenkilöt ovat ennakoon säädettyjä vaihtoehtoisia algoritmeja, joiden välillä käyttäjien on intuitiivisempaa valita perustuen käyttäjän tunnetilaan ja tavoitteisiin. Jos käyttäjä valitsee suositteluhenkilökseen esimerkiksi löytöretkeilijän, niin hän saa enemmän muun aiheisia uutisia kuin aikaisemmin hänen suositteluhistoriassaan. Diplomaatti puolestaan suosittelee uutisia ulkomailta, ja asiantuntija suosittelee saman aiheisia uutisia kuin mitä käyttäjä on lukenut aikaisemminkin. (Harambam ym., 2018.) Näiden esimerkkien lisäksi artikkelissa esitetään muitakin suositteluhenkilöitä. Tämä ratkaisuehdotus on esitetty uutissivustojen kontekstissa, mutta sen perusidea voisi soveltaa muunkinlaisissa palveluissa, kuten sosiaalisessa mediassa.

### 3.5 Käyttäjän autonomia

Tässä alaluvussa käsitellään käyttäjien alttiutta vaikuttamiselle, suosittelujärjestelmiä loukkuina, tiedostettua suostumusta, vaikutusmahdollisuuksia ja oikeuksia. Käyttäjän autonomia liittyy tässä yhteydessä yllä mainittujen lisäksi siihen, että missä määrin käyttäjät uskovat itse olevansa ajatustensa muovautumisen takana verrattuna todelliseen ulkopuolisten vaikuttimien osuuteen. Näistä mahdollisista vaikuttimista käsitellään tässä yhteydessä vain suosittelujärjestelmiä.

On olemassa suosittelujärjestelmien toimia, jotka ovat laillisia, mutta eettisesti kyseenalaisia. Nämä tapahtuvat usein käyttäjien tietämättä, eli suostumus niihin on joko piilotettu palveluehtoihin tai sitten lupaa ei kysytä lainkaan, eikä jälkeenpäinkään välttämättä tiedoteta mitään. (Paraschakis, 2017.) Esimerkki tällaisesta toimesta on A/B-testaus (ks. alaluku 3.3.1). Käyttäjillä ei liian usein ole mahdollisuuksia saada tietää tämän kaltaisista tai edes muista eettisesti hyväksyttävistäkään taustatekijöistä, jotka ohjaavat suosittelujärjestelmän toimintaa. Harambamin ym. (2018) mukaan tavallisilla käyttäjillä onkin vain vähän tietoa siitä, että miten ja missä määrin erilaiset algoritmit muovaavat kaikkea heidän näkemäänsä tietoa. Lisäksi vaikutusmahdollisuudet algoritmin toimintaan voivat olla hyvin rajalliset tai olemattomat, jolloin heillä ei ole mahdolli-

suuksia hyödyntää algoritmeja siten, että ne tavoittelisivat ensisijaisesti heidän tai yhteiskunnan tavoitteita. (Harambam ym., 2018.)

Edellä mainitut seikat asettavatkin käyttäjät erityisen alttiiksi monenlaiselle manipuloinnille, kuten tuuppimiselle (engl. nudging) tietyn ideologisen, moraalisen tai poliittisen sisällön kuluttamiseen (Harambam ym., 2018). Kuten on aiemmin todettu, suosittelujärjestelmät voivat vaikuttaa käyttäjän mieltymysten muodostumiseen sekä sosiaaliseen vuorovaikutukseen, joka tekee niistä oivallisen alustan myös manipulointiin (Milano ym., 2021; Tang & Winoto, 2016). Seaver (2019) esittääkin, että vastoin useimpien käyttäjien omia uskomuksia he eivät itse täysin hallitse heidän omien mieltymysten ja ajatusten muodostumista, vaan suosittelujärjestelmät voivat muovata niitä johonkin suuntaan hiljattain ja huomaamattomasti.

Sen lisäksi, että suosittelujärjestelmät voivat olla merkittäviä vaikuttimia, niin Seaver (2019) huomioi myös niiden suhteellisen helpon käyttöön ajautumisen ja koukuttavuuden. Suosittelujärjestelmiä voi olla vaikea välttää tämän päivän yhteiskunnassa johtuen niiden laajasta levinneisyydestä ja siitä, että niiden koko ympäristö ja infrastruktuuri saattaa ajaa niiden käyttöön. Tämän lisäksi kyseiset järjestelmät kehittyvät jatkuvasti käyttäjien säilyttämisen suhteen. (Seaver, 2019.) Suosittelualgoritmien kehityksessä ennen pitkää huomattiin, että suuremmalla suosittelujen tarkkuudella ei enää saavutettu parempaa käyttäytyväisyyttä, joten siirryttiin suosittelujen tarkkuuden mittauksesta käyttäjien pysyvyyden mittaukseen, jonka avulla käyttäjä pyritään saamaan hyödyntämään palvelua mahdollisimman paljon. Käyttäjien kaikkien mahdollisten toimien tulkitseminen epäsuorana palautteena järjestelmän tehokkuudesta mahdollistaa niiden kehittämisen jatkuvasti koukuttavampaan suuntaan, sillä tehtyjen muutosten seuraukset ovat tällöin paremmin tulkittavissa. (Seaver, 2019.) Epäsuoraa palautetta esimerkiksi ovat palvelun käyttöajat, avauskerrat ja yksittäisen suosittelun tarkastelun kesto. Käyttäjiä kiinnostavien, mutta epäeettisten suosittelujen tahaton tai tahallinen tarjoaminen voi auttaa säilyttämään käyttäjiä. Vaikka tämä on epäeettistä jo itsessään, niin sillä voi olla muitakin merkittäviä seurauksia. Esimerkiksi Allcott ja Gentzkow (2017) osoittivat valeutisten olleen merkittävässä roolissa Yhdysvaltain vuoden 2016 presidentinvaaleissa, jotka johtivat Donald Trumpin valintaan.

Johtuen suosittelujärjestelmien edellä käsitellyistä vaikeasti vältettävyydestä ja kasvavasta koukuttavuudesta Seaverin (2019) mukaan niitä tulisi käsitellä loukkuina, sillä se tuo paremmin esiin niihin liittyvät sosiotekniset huolet. Toisin sanoen se auttaa käsittämään sitä, miten ne toimivat, minkälaiseen ajatteluun ne pohjautuvat sekä miten niitä voitaisiin kritisoida. Loukuilla ajattelu olisi parempi lähestymistapa sen sijaan, että vaihtoehdot rajattaisiin vain täysin vapaaehtoiseen ja pakotettuun suosittelujärjestelmien käyttöön, sillä eettisesti kyseenalaisilla toimilla on tapana sijaita kahden ääripään välisellä harmaalla alueella. (Seaver, 2019.)

Loukuttamisella, käyttäjän alttiudella vaikuttamiselle ja käyttäjän tietämättömyydellä häntä hallitsevista infrastruktuureista Seaver (2019) argumentoi, että onko käyttäjä todella täysin autonominen. Polonioli (2020) on samalla kan-

nalla sen suhteen, että jos käyttäjä ei ole tietoinen informaatiokuplaa aiheuttavan suodattimen olemassaolosta, niin se heikentää hänen autonomiaansa. Vaikka suodattimien olemassaolo tiedostettaisiin, niin puuttuvan tiedon selvittäminen olisi silti haastavaa. (Polonioli, 2020.)

Harambam ym. (2019) osoittivat käyttäjien olleen huolissaan suosittelujärjestelmien taustalla olevista tavoitteista, datan väärinkäytöstä ja vaikutusmahdollisuuksien riittämättömyydestä. Näiden lisäksi käyttäjät ilmaisivat, että he tunsivat menettäneensä itsenäisyytensä ja valinnanmahdollisuuksiaan. Esimerkiksi kun suosittelujärjestelmät valitsivat ennakkoon uutiset, jotka käyttäjät olivat oletettavasti kiinnostuneita lukemaan, niin he kokivat heidän oman kriittisen ajattelunsa ja luovuutensa ohitetuiksi. (Harambam ym., 2019.)

Paraschakis (2017) uskoi, että mahdollisuus vaikuttaa suosittelujen yksilöllistämiseen on käyttäjien oikeus ja tämä voisi tapahtua esimerkiksi algoritmin kahden vaihtoehdoisen tilan kautta. Näistä ensimmäinen tila olisi valittu oletuksena ja se olisi yksilöimätön, tarjoten vain yleiseen dataan pohjautuvia suositteluja. Toinen olisi kohdennettujen suosittelujen tila, joka tarjoaisi kaikista osuvimmat suosittelut, jotka perustuisivat itse käyttäjästä kerättyihin tietoihin. (Paraschakis, 2017.) Myös Harambam ym. (2019) pitivät käyttäjien erilaisia vaikutusmahdollisuuksia suosittelujärjestelmissä hyvänä lähestymistapana, jotka lisäisivät suosittelujärjestelmän toiminnan selitettävyyttä, avoimuutta, luotettavuutta ja myös käytettävyyttä. Nämä tavoiteltavat piirteet auttaisivat käyttäjän puoleisessa ”loukkujen” hallinnassa, jolla olisi käyttäjän autonomiaa vahvistava vaikutus. He osoittivat myös käyttäjien itse haluavan lisää vaikutusmahdollisuuksia (Harambam ym., 2019).

### 3.6 Suosittelujen eettisyys

Suosittelut voivat olla epäeettisiä useista eri syistä. Suosittelu voi olla epäeettinen, jos suosittelujärjestelmä epäonnistuu, tai ei edes yritä ottaa huomioon esimerkiksi sen vastaanottajan ikää, kulttuuria, uskontoa, vakaumusta tai muuta oleellista tietoa. Avainhaaste eettisissä suositteluissa onkin, että käsitys eettisestä sisällöstä voi vaihdella henkilöittäin (Karpati ym., 2020). Eettisten suosittelujen tarjoaminen voi olla hankalaa tai jopa mahdotonta, jos suosittelujärjestelmällä ei ole riittävästi tietoa käyttäjästä, tai palvelua käytetään jonkin toisen henkilön tunnuksilla. Tällöin käyttäjäprofiilin tiedot eivät sovellu todelliselle käyttäjälle. Paraschakiksen (2017) mukaan useiden suosittelujärjestelmien suositteluita ei millään tapaan suodateta eettisesti kyseenalaisista tai sopimattomista suosittelukohteista. Yksi syy tähän voi olla yhteistoiminnallisen suodatuksen suositteluille ominainen yllätyksellisyys, joka asettaa suosittelujen eettisyyden takaamiselle merkittävän haasteen (Paraschakis, 2017). Epäeettisillä suositteluilla voi olla vakavia seurauksia. Esimerkiksi Tang ja Winoto (2016) totesivat useiden lähteiden avulla aikaisempien tutkimuksien osoittaneen, että etenkin nuorempien lasten altistuminen väkivaltaisille elokuville ja videopeleille saattaa

olla yhteydessä aggressiiviseen käytökseen ja haitallisiin asenteisiin, kuten väkivallalle turtumiseen.

Tang ja Winoto (2016) esittivät, että suosittelukohteiden eettistä soveltuvuutta voitaisiin arvioida kartoittamalla niiden mahdollisesti epäeettiset piirteet ja vertaamalla näiden piirteiden yhteensopivuutta käyttäjän piirteiden, kuten iän, sukupuolen ja uskonnon kanssa. Heidän ratkaisussaan suosittelujärjestelmään on sisäänrakennettu eettisiä sääntöjä, ja käyttäjillä on mahdollisuus vaikuttaa määrittelyyn siitä, minkä tyyppiset suosittelut eivät ole sopivia. Souali ym. (2011) puolestaan esittivät ratkaisuehdotuksena verkkokauppoja varten automaattisesti suosittelujen eettisyyden huomioivan suosittelujärjestelmän, joka oppii ja päätelee käyttäjien toimista ilman käyttäjien suoraa vaikutusta. Se hyödyntää eettistä katalogia ja käyttäjien vähäisiä rekisteröintitietoja muodostaakseen eettisiä suositteluja (Souali ym., 2011). Paraschakis (2017) kuitenkin kritisoi tätä lähestymistapaa sen riippuvuudesta käyttäjien itse määrittelemään demograafiseen tietoon, joka saatetaan jättää pois tai olla väärennettyä yksityisyshuolien vuoksi. Lisäksi hän piti täysin automatisoitua sensuuria itsessään eettisesti ongelmallisena (Paraschakis, 2017).

### 3.7 Yksityisyys

Viimeaikaiset julkiset keskustelut liittyen algoritmeihin ja massadataan (engl. big data) ovat tuoneet käyttäjille huolia muun muassa siitä, että heistä kerättyä tietoa väärinkäytettäisiin joko epäeettisesti tai laittomasti (Harambam ym., 2019). Näihin huoliin on hiljattain puututtu päättäjien puolesta Euroopan unionin tietosuojasetuksen keinoin, joka astui voimaan toukokuussa 2018. Se on omalta osaltaan muun muassa parantanut käyttäjien oikeuksia sekä asettanut heidän dataansa säilyttävälle yrityksille velvollisuuksia, jotka suojaavat käyttäjien yksityisyyttä. Siitä huolimatta lähdeaineistosta tunnistettiin yksityisyyden kannalta edelleen oleellisia eettisiä haasteita.

Suosittelujärjestelmän käyttäjäprofiileissa oleva data käyttäjien mieltymyksistä ja aikaisemmasta toiminnasta mahdollistaa suosittelujen muodostamisen (Souali ym., 2011; Paraschakis, 2017). Sitä kuitenkin usein kerätään ilman käyttäjän tiedostettua suostumusta (Paraschakis, 2017). Tiedostetun suostumuksen puuttuminen yleensä merkitsee sitä, että suostumus on piilotettu palveluehtoihin, joita suurin osa käyttäjistä ei lue tai ymmärrä (Obar, 2015). Mahdollistaakseen entistäkin tarkemmat suosittelut, dataa saatetaan kerätä myös seurantaevästeillä, linkitetystä avoimesta datasta, sosiaalisista verkoista sekä tahoilta, jotka keräävät ja jälleenmyyvät käyttäjien toimista kerättyä dataa (Paraschakis, 2017). Käyttäjiin liittyvää tietoa voidaan siis kerätä ja hyödyntää useista paikoista mahdollisesti heidän sitä tiedostamatta, mikä heikentää heidän yksityisyyttään. Käyttäjät eivät välttämättä ole edes tietoisia kaikista tavoista, joilla tietoa ylipäätään on mahdollista kerätä. Eettisesti tämä on ongelmallista varsinkin, jos käyttäjä haluaisi suojella yksityisyyttään, mutta se on edellä mainittujen mahdollisten syiden vuoksi liian vaikeaa.

Paraschakiksen (2017) mukaan Friedman, Knijnenburg, Vanhecke, Martens ja Berkovsky (2015) esittivät kirjassaan kolme yleisintä tapaa, miten suosittelujärjestelmissä tapahtuu yksityisyyden loukkauksia. Ensimmäisessä tavassa itse suosittelujärjestelmä ja sitä ylläpitävät henkilöt eivät onnistu vastaamaan käyttäjien odotuksiin yksityisyydestä. Toisessa tavassa suosittelujärjestelmien käyttäjät tekevät luonteeltaan henkilökohtaisia päätelmiä muiden käyttäjien vastaanottamista suositteluista. (Friedman ym., 2015.) Tämä osoittaa sen, että itse suositteluitakin tulisi kohdella luonteeltaan henkilökohtaisina. Kolmannessa tavassa hakkerit ovat onnistuneet tunnistamaan käyttäjiä suosittelujärjestelmän anonymisoidusta tietoaaineistosta, joka on saatettu julkaista esimerkiksi suosittelujärjestelmiä kehittävää tutkimustyötä varten (Friedman ym., 2015). Tätä käyttäjistä vuotanutta tietoa, kuten käyttäjäprofiileja ja päätelmiä, voidaan väärinkäyttää verkkourkintaan ja manipulointiin (Paraschakis, 2017). Myös useissa muissa suosittelujärjestelmien eettisissä haasteissa ratkaisuehdotuksena esitetty läpinäkyvyys voi liiallisesti tai muuten väärin toteutettuna heikentää käyttäjien yksityisyyttä (Mittelstadt, 2016).

Ratkaisuna käyttäjien yksityisyyden takaamiseksi Paraschakis (2017) ehdotti suosittelujärjestelmien arkkitehtuureja, jotka ovat varta vasten kehitetty suojaamaan yksityisyyttä. Tällaiset arkkitehtuurit pyrkivät puuttumaan ensimmäiseen kolmesta Friedmanin ym. (2015) esittelemistä yleisimmistä tavoista yksityisyyden loukkauksiin suosittelujärjestelmissä, joka oli yksityisyyttä suojelematon suosittelujärjestelmä. Näiden arkkitehtuurien haittapuolena yleensä kuitenkin on se, että niiden tekemien suosittelujen tarkkuus kärsii paremman yksityisyyden takaavan arkkitehtuurin vuoksi. Suosittelujen tarkkuuden ja käyttäjien yksityisyyden tasapainottaminen onkin ollut merkittävä haaste suosittelujärjestelmien kehityksessä (Karpati ym., 2020). Toisena ratkaisuna Paraschakis (2017) esitti graafiset käyttöliittymät, jotka mahdollistavat käyttäjille heidän oman datansa tarkastelun ja hallinnan.

### 3.8 Pohdintaa ja huomioita tuloksista

Useammassa artikkelissa mainittiin, että erityisesti suosittelujärjestelmien yksityisyyttä olisi aiheen aikaisemmassa tieteellisessä kirjallisuudessa käsitelty paljon, mutta lähdeaineistoksi valikoitunut kirjallisuus ei tuntunut heijastavan tätä väitettä, vaan lähes kaikkia muita yläkäsitteitä käsiteltiin enemmän. Tämä voi johtua siitä, että yksityisyyteen liittyviä eettisiä haasteita on enemmän käsitelty yleisemmällä tasolla eikä niinkään erityisesti suosittelujärjestelmien näkökulmasta.

Lähdeaineistosta tuli ilmi eräs eettinen haaste, joka mainittiin muutamassa artikkelissa lyhyesti, mutta jota ei käsitelty sen tarkemmin. Kyseinen haaste oli vastuuvollisuus, joka määrittää sen, että kuka on missäkin tapauksessa vastuussa, jos suosittelujärjestelmän toiminnalla on haitallisia vaikutuksia. Esimerkki vastuuvollisuuteen liittyvästä eettisestä haasteesta voisi olla se, että

pitäisikö suosittelujärjestelmää pitää vastuullisena, jos se suosittelee muiden käyttäjien siihen lisäämää haitallista sisältöä (Paraschakis, 2017).

Tärkeää on myös huomioida, että kaikki käsitellyt haasteet ja ratkaisuehdotukset eivät välttämättä sovellu sellaisenaan käytännön erilaisiin suosittelujärjestelmiin. Tämä voi johtua toimialakohtaisista eroista, ratkaisuehdotusten käytäntöön soveltumattomuudesta tai käsiteltyjen asioiden liiasta abstraktiudesta. Karpati ym. (2020) panivat merkille tämän abstraktiuden yleisissä tekoälyyn pohjautuvien teknologioiden eettisten haasteiden kartoituksissa: liian yleistettyjä kartoituksia ei voi hyödyntää sellaisenaan kehitystyössä, sillä ne eivät ole riittävän yksityiskohtaisia. He kuitenkin näkivät esitetyt yleiset haasteet tärkeänä ensiaskeleena kohti perinpohjaisemmin määriteltyjä toimialakohtaisia käytäntöjä, jotka mahdollistaisivat suuremman tutkimustyön tuloksien annin niiden ollessa varsinaisesti sovellettavissa käytäntöön (Karpati ym., 2020). Esimerkkejä käytännönläheisimmistä tutkimuksista ovat Karpatin ym. (2020) artikkeli elintarvikkeiden suosittelujärjestelmien etiikasta ja Poloniolin (2020) artikkeli tieteellisten julkaisujen suosittelujärjestelmien etiikasta.

Käsitellyillä haasteilla ja ratkaisuehdotuksilla on paljon yhteisiä aspektoja. Monet haasteista ovat yhteydessä toisiinsa ja niillä saattaa olla limittyviä ratkaisuehdotuksia, joista olisi hyötyä useammalle kuin yhdelle käsitellylle haasteelle. Eräs monelle eettiselle haasteelle merkillepantava yhteinen tekijä on useassa artikkelissa korostettu kokonaisvaltaisempi lähestymistapa suosittelujärjestelmien eettiseen tarkasteluun, jolla pyritään ottamaan huomioon kaikki sille oleelliset sidosryhmät aikaisemman käyttäjäkeskeisen lähestymistavan sijaan (Karpati ym., 2020; Milano ym., 2021). Ratkaisuehdotusten yhteisenä piirteenä tuli ilmi tarve tehdä jonkinlainen kompromissi erilaisten toimintatapojen, toimien ja tavoiteltavien vaikutusten välillä. Esimerkkejä näistä ovat tasapainottelu käyttäjän oman toimijuuden ja suosittelujärjestelmän käyttäjän puolesta tekemän päätöksenteon välillä (Harambam ym., 2019) sekä tasapainottelu suosittelujen perustelujen syvällisyyden ja niiden tarkkuuden välillä (Shulman & Wolf, 2020). Usein siis tulee tarpeeseen määritellä, että mikä on milloinkin eettisesti hyväksyttävä kompromissi sen lisäksi, että on olemassa ratkaisutapa, joka takaa kyseisen määritetyn kompromissin. Tämä osoittaa omalta osaltaan suosittelujärjestelmien etiikan tutkimuksen oleellisuutta.

Harambam ym. (2018) esittivät käyttäjille jo olemassa olevia ja suosittelujärjestelmästä riippumattomia tapoja vastata useampaan käsitelyyn eettiseen haasteeseen. Ne eivät ole optimaalisia ratkaisuja, eli suosittelujärjestelmään sisäänrakennettuja ja siten useammalle helpommin hyödynnettävissä, mutta niitä voidaan käyttää tarvittaessa. Heidän esittelemänsä neljä tapaa mahdollistavat käyttäjien oman äänen ilmaisun ja vaikuttamisen suosittelujärjestelmän käyttäjälle välittämään informaatioon riippumatta siitä, missä määrin suosittelujärjestelmien ylläpitäjät ovat pyrkineet vastata mahdollisiin eettisiin haasteisiin. Näitä tapoja ovat vuorottelu, tiedostaminen, mukauttaminen ja hämääminen. Ensimmäisessä tavassa, eli vuorottelussa, käytetään esimerkiksi eri palveluntarjoajien uutis-suosittelujärjestelmiä vuorotellen, jotta vastaanotetun tiedon monipuolisuutta ei aseteta vain yhden algoritmin varaan. (Harambam ym., 2018.)



Tiedostaessaan, eli esitetyistä tavoista toisessa, käyttäjällä on perusymmärrys hänen käyttämiensä suosittelujärjestelmien toimintatavoista, mikä mahdollistaa paremman niihin liittyvän päätöksenteon. Kolmannessa tavassa, eli mukauttamisessa, käyttäjä osaa hyödyntää niin suosittelujärjestelmän mahdollisia sisäisiä kuin ulkoisiakin työkaluja vastaanottamaansa informaation hallintaan. (Harambam ym., 2018.) Neljäs tapa on yksityisyyttään vaaliville käyttäjille ja hyödyllinen etenkin silloin, kun järjestelmä ei tarjoa siihen sisäänrakennettuja keinoja. Tämä on hämäys, eli käyttäjä mahdollisuuksien mukaan jättää kertomatta todellisia tietoja itsestään ja tarpeen vaatiessa valehtelee, jotta hänestä ei saada muodostettua todenmukaista käyttäjäprofiilia. On myös mahdollista käyttää erilaisia työkaluja, kuten virtuaalisia erillisverkkoja, jotka salaavat käyttäjän todellisen sijainnin sekä selaimen lisäosia, jotka toteuttavat suosittelujärjestelmän hämäämistä automatisoidusti. (Harambam ym., 2018.) Tällöin käyttäjä ei saa osuvia suositteluita, mutta toimintahistorian ja muiden tietojen ollessa satunnaisia hänen yksityisyytensä on paremmin turvattu.

## 4 YHTEENVETO

Tämä kandidaatintutkielma on pyrkinyt kartoittamaan aiheen tutkimuksessa käsiteltyjä suosittelujärjestelmien tämän päivän eettisiä haasteita ja niiden ratkaisuehdotuksia. Tutkimusmenetelmänä on käytetty narratiivista systemaattista kirjallisuuskatsausta Templierin ja Parén (2015) ohjeita noudattaen.

Tutkielma on pyrkinyt antamaan yleiskuvan suosittelujärjestelmien eettisistä haasteista sekä niiden ratkaisuehdotuksista. Tässä se on mielestäni onnistunut, tarjoten riittävän kattavan aloituskohdan aiheesta kiinnostuneille. Oma kontribuutiona on siis kokoava katsaus suurilta osin hajanaisesta aiheesta sekä pitkälti vastaavien tulosten toistaminen Milanon ym. (2020) tutkimuksen kanssa, joka myös pyrki kartoittamaan suosittelujärjestelmien eettisiä haasteita ja niiden mahdollisia ratkaisuja, joksikin paljon laajemmin. Tämän lisäksi heidän työnsä jaottelee käsiteltyjen haasteiden vaikutukset hyötyyn vaikuttavuuden, oikeuksien loukkaavuuden sekä vaikutusten välittömyyden perusteella (Milano ym., 2020).

Tutkielman rajoitteina on ollut sille osoitetut rajalliset resurssit. Laajemman lähdeaineiston joukon tarkastelulla olisi mahdollisesti löydetty muitakin näkökulmia saatuihin tuloksiin, tai täysin uusia, tässä tutkielmassa käsittelemättömiä haasteita ja niiden ratkaisuehdotuksia. Tästä syystä tutkielman tulokset eivät ole kaikenkattavia, vaan ne antavat yleiskuvauksellisen vastauksen tutkimuskysymyksiin. Jotkin havaitut haasteet myös saivat suhteellisen vähän huomiota valikoituneessa aineistossa, kuten yksityisyys ja vastuuvollisuus.

Eettisistä haasteistaan huolimatta suosittelujärjestelmät ovat hyödyllisiä työkaluja esimerkiksi informaatiotulvaa vastaan, sillä ne auttavat käyttäjiä löytämään heitä kiinnostavia asioita nopeammin ja helpommin. Ne ovat myös erinomaisia järjestelmiä palveluntarjoajien näkökulmasta, jotka esimerkiksi parantavat käyttäjäkokemusta, joka puolestaan sitouttaa käyttäjiä palvelun käyttöön. Tämän tutkielman tarkoituksena ei ollut osoittaa, että suosittelujärjestelmät olisivat haitallisia, vaan havaittuihin haasteisiin puuttumisen myötä niitä voidaan kehittää entistäkin paremmiksi työkaluiksi, jotka palvelevat kaikkia osapuolia mahdollisimman eettisesti.

Tämä tutkielman ja Milanon ym. (2020) katsaukset ovat pääasiallisesti yleisluontoisia, ja ne toimivat ensiaskeleina niissä käsiteltyjen haasteiden ratkaisuun, tuoden aiheen moniulotteisuuden useampien tietoon. Jatkotutkimusaihe ja seuraava askel onkin toimialakohtaiset, eli käytännönläheisemmät, suosittelujärjestelmien eettisten haasteiden ja ratkaisujen tarkastelut, jotka puolestaan näyttäisivät oikean suunnan käytännön ratkaisujen ja suosittelujärjestelmien kehitystyölle.

## LÄHTEET

- Allcott, H. & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211-236. <https://doi.org/10.1257/jep.31.2.211>
- Barocas, S. & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732. <https://doi.org/10.15779/Z38BG31>
- Bradshaw, S., Howard, P. N. (2019). The global disinformation order: 2019 global inventory of organised social media manipulation. *Computational Propaganda Project, Oxford Internet Institute*. Haettu osoitteesta <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>
- Calero Valdez, A., Ziefle, M. & Verbert, K. (2016). HCI for Recommender Systems: The Past, the Present and the Future. *Association for Computing Machinery*. <https://doi.org/10.1145/2959100.2959158>
- Erdélyi, O. & Goldsmith, J. (2018). Regulating Artificial Intelligence: Proposal for a Global Solution. *AIES '20: Proceedings of the 2018 AAA/ACM Conference on AI, Ethics, and Society*, 95-101. <https://doi.org/10.1145/3278721.3278731>
- Friedman A., Knijnenburg B.P., Vanhecke K., Martens L., Berkovsky S. (2015). *Privacy Aspects of Recommender Systems*. Boston, MA: Springer. [https://doi-org.ezproxy.jyu.fi/10.1007/978-1-4899-7637-6\\_19](https://doi-org.ezproxy.jyu.fi/10.1007/978-1-4899-7637-6_19)
- Harambam, J., Bountouridis, D., Makhortykh, M. & van Hoboken, J. (2019). Designing for the better by taking users into account: A qualitative evaluation of user control mechanisms in (news) recommender systems. *RecSys 2019 - 13th ACM Conference on Recommender Systems*, 69-77. <https://doi.org/10.1145/3298689.3347014>
- Harambam, J., Helberger, N. & van Hoboken, J. (2018). Democratizing algorithmic news recommenders: How to materialize voice in a technologically saturated media ecosystem. *Philosophical Transactions of the Royal Society A - Mathematical, Physical and Engineering Sciences*, 376(2133). <https://doi.org/10.1098/rsta.2018.0088>
- Heersmink, R., van Den Hoven, J., van Eck, N. & van Den Berg, J. (2011). Bibliometric mapping of computer and information ethics. *Ethics and Information Technology*, 13(3), 241-249. <https://doi.org/10.1007/s10676-011-9273-7>

- Karpati, D., Najjar, A. & Ambrossio, D. (2020). Ethics of Food Recommender Applications. *AIES '20: Proceedings of the AAA/ACM Conference on AI, Ethics, and Society*. <https://doi.org/10.1145/3375627.3375874>
- Kermary, N. R., Zhao, W., Yang, J., Wu, J., & Pizzato, L. (2020). An Ethical Multi-Stakeholder Recommender System Based on Evolutionary Multi-Objective Optimization. *2020 IEEE International Conference on Services Computing (SCC)*, 478-480. <https://doi.org/10.1109/SCC49832.2020.00074>
- Lonjarret, C., Robardet, C., Plantevit, M., Auburtin, R. & Atzmueller, M. (2020). Why Should I Trust This Item? Explaining the Recommendations of any Model. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 526-535. <https://doi.org/10.1109/DSAA49011.2020.00067>
- Milano, S., Taddeo, M. & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & Society*, 35(4), 957-967. <https://doi.org/10.1007/s00146-020-00950-y>
- Milano, S., Taddeo, M. & Floridi, L. (2021). Ethical aspects of multi-stakeholder recommendation systems. *The Information society*, 37(1), 35-45. <https://doi.org/10.1080/01972243.2020.1832636>
- Mittelstadt, B. (2016). Auditing for transparency in content personalization systems. *International journal of communication*, 10, 4991-5002. Haettu osoitteesta <http://search.ebscohost.com.ezproxy.jyu.fi/login.aspx?direct=true&db=ufh&AN=127361909&site=ehost-live>
- Obar, J. A. (2015). Big Data and The Phantom Public: Walter Lippmann and the fallacy of data privacy self-management. *Big data & society*, 2(2). <https://doi.org/10.1177/2053951715608876>
- Paraschakis, D. (2017). Towards an ethical recommendation framework. *2017 11<sup>th</sup> International Conference on Research Challenges in Information Science (RCIS)*. <https://doi.org/10.1109/RCIS.2017.7956539>
- Polonioli, A. (2020). The ethics of scientific recommender systems. *Scientometrics*, 126(2), 1841-1848. <https://doi.org/10.1007/s11192-020-03766-1>
- Ricci, F., Rokach, L. & Shapira, B. (2011). *Introduction to Recommender Systems Handbook*. Springer. [https://doi.org/10.1007/978-0-387-85820-3\\_1](https://doi.org/10.1007/978-0-387-85820-3_1)
- Sacharidis, D. (2020). Building User Trust in Recommendations via Fairness and Explanations. *UMAP '20: 28<sup>th</sup> ACM Conference on User Modeling, Adaptation and Personalization, Association for Computing Machinery*, 313-314. <https://doi.org/10.1145/3386392.3399995>

- Seaver, N. (2019). Captivating algorithms: Recommender systems as traps. *Journal of material culture*, 24(4), 421-436.  
<https://doi.org/10.1177/1359183518820366>
- Sharma, S. R. R. (2016). Explanations in recommender systems: An overview. *International Journal of Business Information Systems*, 23(2), 248.  
<https://doi.org/10.1504/IJBIS.2016.078909>
- Shrestha, Y. R. & Yang, Y. (2019). Fairness in Algorithmic Decision-Making: Applications in Multi-Winner Voting, Machine Learning, and Recommender Systems. *Algorithms*, 12(9), 199.  
<https://doi.org/10.3390/a12090199>
- Shulman, E. & Wolf, L. (2020). Meta Decision Trees for Explainable Recommendation Systems. *Association for Computing Machinery*, 365-371.  
<https://doi.org/10.1145/3375627.3375876>
- Souali, K., El Afia, A. & Faizi, R. (2011). An automatic ethical-based recommender system for e-commerce. *2011 International Conference on Multimedia Computing Systems*.  
<https://doi.org/10.1109/ICMCS.2011.5945631>
- Tang, T. Y. & Winoto, P. (2016). I should not recommend it to you even if you will like it: The ethics of recommender systems. *The new review of hypermedia and multimedia*, 22(1-2), 111-138.  
<https://doi.org/10.1080/13614568.2015.1052099>
- Templier, M. & Paré, G. (2015). A Framework for Guiding and Evaluating Literature Reviews. *Communications of the Association for Information Systems*, 37, 112. <https://doi.org/10.17705/1CAIS.03706>
- Tintarev, N. & Masthoff, J. (2007). A Survey of Explanations in Recommender Systems. *2007 IEEE 23<sup>rd</sup> International Conference on Data Engineering Workshop*. <https://doi.org/10.1109/ICDEW.2007.4401070>