

## Artikkeli



VERTAISARVIOITU  
KOLLEGIALT GRANSKAD  
PEER-REVIEWED  
www.tsv.fi/tunnus

# Dataa näkyvissä

## Läpinäkyvyys algoritmien ja datan journalistisessa hyödyntämisessä

Läpinäkyvyyden käsite on noussut keskeiseksi viestinnän, median ja politiikan tutkimuksessa sekä viestinnän ja politiikan käytäntöjen eettisessä arvioinnissa. Läpinäkyvyyttä on esitetty keinoksi ylläpitää ja kasvattaa luottamusta journalistista työtä ja sen tuotteita kohtaan interaktiivisessa media- ja viestintäympäristössä. Tiedontuotannon ja päätöksenteon perustuessa yhä enemmän dataan ja sen prosessointiin läpinäkyvyys on noussut keskeiseksi käsitteeksi myös algoritmien päätöksenteon kohdalla. Tässä artikkelissa tarkastelemme läpinäkyvyyden roolia dataa ja algoritmeja sekä journalismia koskevassa keskustelussa ja esitämme mahdollisuuksia dataan ja algoritmeihin tukeutuvan journalismin läpinäkyvyyskäytännöiksi. Dataa hyödynnetään digitaalisessa journalismissa sekä informaation lähteenä manuaalisissa ja automatisoiduissa tuotantoprosesseissa että sisällön jakelussa ja kohdentamisessa yleisöille. Miten journalismi voisi tuoda läpinäkyväksi datan käyttöä, dataa koskevia valintoja ja niihin vaikuttavia tarpeita, tarkoituksia ja arvoja?

**AVAINSANAT:** läpinäkyvyys, journalismi, data, algoritmit, automaatio, luottamus, päätöksenteko, vastuu

Läpinäkyvyyden käsite on noussut keskeiseksi viestinnän, median ja politiikan tutkimuksessa sekä viestinnän ja politiikan käytäntöjen eettisessä arvioinnissa. Taustalla on monia tekijöitä. Läpinäkyvyyttä on ensinnäkin pitkään pidetty tapana estää korruptiota ja edistää esimerkiksi julkisten varojen tarkoituksenmukaista käyttöä (Lindstedt ja Naurin 2010). Toiseksi poliittisen filosofian osallistuvaa tai deliberatiivista demokratiaa ajavat näkemykset ovat viime vuosikymmenten aikana herättäneet paljon keskustelua (Rydenfelt 2019a; 2019b). Deliberatiivisen demokratian keskeisiin ihanteisiin kuuluvat läpinäkyvyys informaatiosta ja informaatiolähteiden taustoista sekä intresseistä päätöksenteon onnistumisen ja legitimitietin takaamiseksi (Lindstedt ja Naurin 2010). Läpinäkyvyyttä on pidetty yhtenä viestinnän eettisistä arvoista (Rydenfelt ja Juholin 2016). Tätäkin laajemmin läpinäkyvyys on usein esitetty yleisenä edellytyksenä vastuullisuuden toteutumiseen (Gorwa ja Garton Ash 2020). Läpinäkyvyyden on ajateltu mahdollistavan toisten toiminnan ja näkemysten arvioinnin ja kritiikin sekä valistuneiden eettisten, poliittisten ja yhteiskunnallisten johtopäätösten vetämisen.

Läpinäkyvyyden käsitteen voi katsoa levinneen journalismia ja mediaa käsittelevään tutkimuskirjallisuuteen muutama vuosikymmen sitten. Bill Kovachin ja Tom Rosenstielin vaikutusvaltainen journalismin käsikirja *Elements of Journalism* esitti läpinäkyvyyden käsitteen vastakohtana tietojen pimittämiselle ja valehtelulle (Kovach ja Rosenstiel 2001, 80–81). Journalismia koskevassa tutkimuksessa ja keskustelussa läpinäkyvyyden toteuttamisen tavoiksi on esitetty toimittajien taustojen ja lähtökohtien esittelyä, toimituksellisten prosessien avaamista yleisölle sekä yleisön osallistamista uutissisältöjen tuottamiseen. Läpinäkyvyyden tavoitteeksi on usein esitetty journalistisen työn ja sen tuotteiden aseman säilyttäminen ja parantaminen. Se on nähty keskeiseksi keinoksi palauttaa, ylläpitää ja kasvattaa luottamusta journalistista työtä ja sen tuotteita kohtaan interaktiivisessa media- ja viestintäympäristössä (Allen 2008; Philips 2010).

Tiedontuotannon ja päätöksenteon perustuessa yhä enemmän dataan ja sen prosessointiin läpinäkyvyys on noussut keskeiseksi käsitteeksi myös algoritmisen päätöksenteon kohdalla. Datalla tarkoitetaan yleisesti raaka-ainetta, jota syntyy, kun maailmaa abstrahoidaan mittareiden tai kategorioiden avulla. Data on ikään kuin rakennuspalikoita informaation ja tiedon tuottamiseksi. Algoritmi puolestaan on proseduuri, jonka perusteella suoritetaan toimintoja, eräänlainen koodattu resepti tai matemaattisten sääntöjen joukko. Suppeasti teknisestä näkökulmasta tarkasteltuna algoritmi on tekninen väline, joka toteuttaa jonkin tehtävän, esimerkiksi järjestää dataa annetun kriteerin mukaisesti (esim. Gillespie 2016). Algoritmilla kuitenkin usein viitataan monimutkaisempaan kokonaisuuteen, algoritmiseen järjestelmään, joka käsittää paitsi ohjelmakoodin ja sen toteuttamat säännöt, myös koodin käyttämän datan sekä taustalla olevat ihmisen tekemät päätökset ja valinnat siitä, mikä on huomionarvoista, tärkeää tai arvokasta. Algoritmien järjestelmä on ihmisten toimien, koodin ja datan muodostama kokonaisuus, joka tekee päätöksiä käyttäen ja tuottaen muodollista, vertailtavaa ja mitattavaa informaatiota (Gillespie 2016; Seaver 2019).

Algoritmien järjestelmien monitahoisten sisäisten ja ulkoisten suhteiden tarkastelu muuttuu helposti vaikeaksi. Keskeiseksi metaforaksi algoritmeja koskevassa keskustelussa on muodostunut *musta laatikko* (Pasquale 2015). Algoritmille syötetään dataa, ja se tuottaa tietyn lopputuloksen, kuten päätöksen tai ratkaisuehdotuksen, kertomatta tarkemmin, miten data valikoitiin ja millä tavoin lopputulokseen päädyttiin. Saadaksemme selville, mitä valintoja algoritmiset prosessit tekevät sekä mitä päämääriä, tarpeita ja arvoja ne heijastelevat, meidän on kyettävä kurkistamaan mustan laatikon sisään. Läpinäkyvyys näyttäytyy näin välttämättömänä ehtona vastuulliselle datan ja algoritmien käytölle (Pasquale 2015; Gorwa ja Garton Ash 2020).

Digitaalinen media on kasvanut keskeiseksi toimijaksi datan hyödyntämisessä ja algoritmien käytössä, ja toimittajien suhtautuminen data-analytiikkaan on muuttunut viimeisen kymmenen vuoden aikana epäilevästä innostuneeksi (Cherubini ja Nielsen 2016). Dataa hyödynnetään digitaalisissa journalismissa sekä manuaalisissa ja automatisoiduissa tuotantoprosesseissa että sisällön jakelussa ja kohdentamisessa yleisöille. Miten journalismi voisi tuoda läpinäkyväksi datan käyttöä, dataa koskevia valintoja ja niihin vaikuttavia tarpeita, tarkoituksia ja arvoja?

Tässä artikkelissa tarkastelemme läpinäkyvyyden roolia dataa ja algoritmeja sekä journalismia koskevassa keskustelussa. Aloitamme kuvaamalla datan ja algoritmien vaikutuksia sosiaaliseen maailmaan. Tämä yleinen datafikaatiokehitys tarjoaa taustan datan ja

algoritmien journalististen käyttötapojen erittelylle, joka pohjautuu toteuttamiimme kyselyihin ja haastatteluihin kotimaisten ja ulkomaisten media-alan toimijoiden kanssa (Haapanen 2020b; Rydenfelt 2021a). Kiinnitämme huomiota siihen, että journalistinen datan ja algoritmien käyttö pohjautuu usein toimitusten ja mediatalojen ulkopuolisiin algoritmiin järjestelmiin ja joutuu niihin monissa yhteyksissä myös mukautumaan. Tämän jälkeen siirrymme tarkastelemaan keskustelua läpinäkyvyydestä sekä sen roolista dataistuvassa ja algoritmien muovaamassa maailmassa. Esittelemme ajatuksen läpinäkyvyydestä keinona ylläpitää ja lisätä journalismin luotettavuutta ja erittelemme sekä jo vakiintuneita että uusia tapoja tehdä journalistisia prosesseja läpinäkyvämmiksi. Käymme läpi myös kritiikkiä, joka kohdistuu läpinäkyvyyteen ihanteena sekä sen teknisiin ja muodollisiin toteutustapoihin. Lopuksi tuomme yhteen nämä keskustelut tarkastellaksemme mahdollisuuksia ja ehdotuksia dataa ja algoritmeja hyödyntävän journalismin läpinäkyvyyskäytännöiksi. Pohdimme, miten journalistiset läpinäkyvyyskäytännöt soveltuvat datan ja algoritmien käyttötapoihin ja arvioimme niiden mahdollisuuksia datan ja sen hyödyntämiseen liittyvien valintojen tekemisessä läpinäkyväksi.

## Datafikaatio ja algoritmiset järjestelmät

Informaatioteknologian sosiaalisia ja yhteiskunnallisia vaikutuksia voidaan hahmottaa kahden toisiinsa liittyvän mutta käsitteellisesti erillisen kehityskulun, digitalisaation ja datafikaation, avulla (esim. Flyverbom, Deibert ja Matten 2019). Digitalisaatio viittaa prosessien ja informaation siirtymiseen informaatioteknologialla käsiteltävään muotoon. Se on tausta-vaikuttajana esimerkiksi kehityksessä teollisesta yhteiskunnasta informaatioyhteiskuntaan. Siinä missä digitalisaation merkitys on ymmärretty jo vuosikymmeniä, datafikaatio (van Dijck 2014) on käsitteenä vasta joitain vuosia vanha. Datafikaatio viittaa luonteeltaan laadullisten, ihmisiin ja sosiaaliseen maailmaan liittyvien ilmiöiden muuttamiseen mitattavaan muotoon. Teknisessä mielessä datafikaatiossa on kyse informaatioteknologian ominaisuudesta: kun informaatioteknologian avulla automatisoidaan asioita, voidaan samalla tuottaa dataa automaation kohteesta (Zuboff 1988). Esimerkiksi uutisisältöjen jakelu informaatioteknologian avulla ei tarkoita ainoastaan sisältöjen ja niiden jakelukanavien digitalisointia vaan myös niiden datafikaatiota: datan luomista esimerkiksi sisältöjen jakelusta ja vastaanottamisesta, kuten klikkauksista, artikkelin näyttökerroista, lukemiseen käytetystä ajasta ja laitteesta, samalla kertaa luetuista muista artikkeleista sekä lukijan muista tekemisistä.

Datafikaation yhteiskunnallisten seurausten kannalta eräs dataa koskeva keskeinen kysymys on episteeminen: millä tavalla data kuvaa ja voi kuvata maailmaa? Voisiko raaka, käsittelemätön data kertoa maailmasta neutraalilla ja objektiivisella tavalla? Jos näin olisi, data voisi olla vapaata arvoalinoista ja ainoastaan sen käyttöön liittyisi ideologisia elementtejä ja poliittisia ulottuvuuksia (Kitchin 2014). Yhteiskuntatieteellisen datafikaation tutkimuksen ytimessä on havainto siitä, että ajatus neutraalista ja arvovapaasta datasta ei pidä paikkaansa. Lisa Gitelmanin ja Virginia Jacksonin (2013) usein siteeratun huomion mukaan data ei ole koskaan ”raakaa” vaan aina jollain tavalla ”keitettyä”: datan valikointiin, käsittelyyn ja analyysiin liittyy arvovalintoja. Data on väistämättä sitä tuottavan sosioteknisen järjestelmän tulos, ja järjestelmään liittyvät valinnat ovat olleet läsnä jo ennen kuin data on luotu ja

tallennettu. Data ei valikoidu tai prosessoidu itsekseen, vaan sekä dataa että sen käsittelyä koskevat ratkaisut sisältävät arvovalintoja, jotka perustuvat kriteereihin paremmuudesta ja sopivuudesta suhteessa käsillä oleviin käyttötarkoituksiin. Esimerkiksi inhimillisen kokemuksen muuttaminen numeeriseksi ja mitattavaksi on valinta, joka johtaa tietynlaisen tiedon tuottamiseen (Berry 2011).

Vaikka datafikaation taustalla on informaatioteknologian ominaisuus, sitä on hyödyllistä ajatella pikemminkin yhteiskunnallisena kuin teknisenä ilmiönä. Kulttuuristen, poliittisten, institutionaalisten ja taloudellisten tekijöiden täytyy kohdata, jotta datafikaatio nykymuodossaan on mahdollista. Datafikaatio tiedontuotannon ja päätöksenteon mahdollistajana voi myös vakiinnuttaa yhteiskunnallisia jakolinjoja ja valtasuhteita (Hintz, Dencik ja Wahl-Jorgensen 2019). Viime vuosien yhteiskuntatieteellistä datatutkimusta (tai kriittistä datatutkimusta, ks. Kennedy 2018) voikin ajatella juuri datafikaation reunaehtojen ja datafikaatiota mahdollistavan sosiokulttuurisen kontekstin kartoittamisena. Kuka tuottaa, kerää, käyttää ja hyödyntää tietoa, ja miten tästä päätetään? Mitä valintoja datan hyödyntämisessä tehdään tai on lupa tehdä? Kirjallisuudessa on pyritty identifioimaan datan käytön nykykäytäntöjen taustalla vaikuttavia ideologisia tekijöitä (esim. dataismi, van Dijck 2014) ja taloudellisia logiikoita (esim. valvontakapitalismi, Zuboff 2015, sekä alustakapitalismi, Srnicek 2017). Lisäksi tutkimuskirjallisuudessa on pohdittu datan käyttämisen ja algoritmisen päätöksenteon keskeisiä sosiaalisia ja eettisiä ongelmia, kuten syrjintää sekä eettisen ja laillisen vastuunkannon ongelmia (esim. Pasquale 2015; O'Neil 2016; Zarsky 2016). Datan käytön ja algoritmisen päätöksenteon eettistä arviointia, oikeudenmukaisuutta, syrjimättömyyttä ja vastuullisuutta on pyritty takaamaan erilaisilla teknisillä ratkaisuilla (Lepri ym. 2018). Viime vuosina keskusteluun ovat nousseet myös datan hallinnan ja kontrolloinnin mallit, joiden tavoitteena on saattaa ihmiset osallisiksi dataa ja sen käyttöä koskeviin päätöksiin (esim. Lehtiniemi 2020; Micheli ym. 2020; Prainsack 2019).

Mitä vaikutuksia datalla ja algoritmeilla on sosiaaliseen maailmaan – mitä data tekee, tai tarkemmin sanoen, mitä datalla tehdään? Angèle Christin (2020) tarjoaa datan yleisimpiä käyttötapoja ja vaikutuksia selventävän typologian erittelemällä viisi mekanismia, joilla data vaikuttaa ympäröivään sosiaaliseen maailmaan. Nämä vaikutusmekanismit toteutuvat tyypillisesti eri toimintojen suorittamiseen tarkoitettujen algoritmisten järjestelmien kautta. Christinin typologiassa data seuraa, samankaltaistaa, järjestää, vaikuttaa käyttäytymiseen ja kasvattaa arvoa. *Seuranta* datan avulla on mahdollista, kun ihmisten toimia ja käyttäytymistä voidaan tallentaa datan muodossa. Seuranta voi koskea esimerkiksi sijaintia, klikkauksia nettipalveluissa, ostoksia luottokortilla, sovelluksen avaamista ja sulkemista tai vaikkapa sensorin mittaamaa fysiologista suuretta. Datafikaation poliittis-taloudellisia yhteyksiä sekä datan avulla tapahtuvaa valvontaa koskeva kriittinen tutkimus käsittelee usein juuri ihmisten toimien seurantaa ja sen muotoutumista datafikaation myötä (esim. van Dijck 2014; Andrejevic ja Gates 2014; Zuboff 2015).

Christinin erottelussa *samankaltaistaminen* taas juontaa juurensa datafikaation peruluonteesta, monimutkaisten ja luonteeltaan laadullisten ilmiöiden muuttamisesta mitattavaan muotoon. Esimerkiksi sosiaalisten ilmiöiden tai ihmisten luonteenpiirteiden kuvaaminen datan avulla häivyttää niiden yksityiskohtia ja tekee ilmiöistä enemmän toistensa kaltaisia. Samankaltaistaminen on datan käytön kannalta oleellista, sillä se mahdollistaa datafioitujen ilmiöiden vertailemisen. *Järjestäminen* viittaa tällaiseen asioiden laittamiseen

arvojärjestykseen datan avulla. Järjestäminen on esimerkiksi luokittelua, joka tapahtuu jonkin seurauksen ennakoitun todennäköisyyden tai kiireellisyyden perusteella, kuten luottokelpoisuutta arvioiva luokitus (esim. Fourcade ja Healy 2017).

*Käyttäytymiseen vaikuttaminen* datan avulla viittaa siihen, miten dataa käytetään tavoitellun lopputuloksen aikaansaamisessa. Suoraviivainen esimerkki voisi olla metriikoiden käyttäminen halutun toiminnan kannustimena työpaikalla erilaisten mittareiden ja tavoitetasojen kautta. Monet käyttäytymiseen vaikuttavat datankäytön tavat eivät ole näin ilmeisiä. Esimerkiksi työympäristöä voidaan muokata tietynlaisten valintojen suuntaan pelkästään työvälineisiin toteutettujen käyttöliittymien ominaisuuksien avulla (esim. Yeung 2016; Gray ym. 2018). *Arvon kasvattaminen* viittaa puolestaan tapoihin, joilla data kasvattaa niiden asioiden arvoa ja legitimitteettiä, joita data koskee. Yksinkertaisimmillaan pelkkä datan tuottaminen ilmiöstä tekee siitä mitattavan ja näkyvän, ja näin painoarvoltaan potentiaalisesti merkittävämmän: datan avulla mitattavat asiat voivat nousta onnistumisen mittareiksi ja päätöksenteon kriteereiksi.

## Datan ja automaation käyttö journalismissa

Journalismissa datan käyttötavat ovat samanlaisia kuin yhteiskunnassa ylipäätään, ja datan roolia journalistisessa tuotantoprosessissa voidaan tarkastella edellä esitetyn Christinin (2020) typologian avulla. Data samankaltaistaa journalistisia sisältöjä: datan avulla erilaiset sisällöt tulevat yhteismitallisiksi ja siten vertailukelpoisiksi. Tämä yhteismitallisuus mahdollistaa myös järjestämisen ja luokittelun. Vaihtoehtoisten artikkeliotsikoiden kokeilu (ns. A/B-testaus) on esimerkki otsikkotyypin järjestämisestä sen perusteella, miten ne palvelevat tavoitteita. Dataa käytetään myös yleisöjen seurantaan. Seuranta tekee mahdolliseksi sisältöjen – journalististen sisältöjen sekä mainosten – kohdentamisen yleisöille heidän aiempien valintojensa perusteella. Tässä journalistiset mediat seuraavat digitaalisen maailman vakiintuneita toiminta- ja ansaintatapoja (West 2019). Artikkeleiden ohessa suositellut muut artikkelit, suosituimpien artikkeleiden listaukset sekä etusivun sisältöjen datapohjaiset järjestämisen tavat ovat esimerkkejä lukijoiden käyttäytymiseen vaikuttamisesta datan avulla. Samoin seurannalla on vaikutuksia journalistiseen päätöksentekoon. Esimerkiksi yleisömetriikoilla voi olla vaikutusta journalistien käsityksiin siitä, millaisilla jutuilla on sekä uutisarvoa että huomioarvoa yleisön silmissä (Anderson 2011a; 2011b; Lamot ja Van Aelst 2019). Yleisömetriikat vaikuttavat näin siihen, mitä pidetään arvokkaana. Missään näistä tilanteista datan käyttö ei ole vaikutuksiltaan neutraalia. Jos mittauksen kohteena ovat sivuvierailut eli klikkaukset, voi houkuttelevan otsikoinnin painoarvo kasvaa ja johtaa hyvin tunnettuun ilmiöön, ”klikkauskuultuuriin” (Anderson 2011a). Klikkaamiselle vaihtoehtoisia mittauksen kohteita on esitetty: artikkelin parissa käytetty aika ja artikkelin sivun loppuun asti vierittäminen voivat näyttäytyä klikkausta osuvampina mittareina, mikäli pyrkimys on datafioida nimenomaan artikkelin lukemista (Haapoja ja Lampinen 2018). Yleisömetriikkana tällaiset mittarit voivat painottaa esimerkiksi pidempien tai loppuun asti sivua vierittämään houkuttelevien juttujen arvoa.

Journalismissa datan hyödyntämisen keskeisenä osana ovat algoritmiset prosessit, joissa käytetty data toimii raaka-aineena. Journalismin tutkimuksessa näihin prosesseihin viita-

taan yleisesti *automaationa* (Caswell ja Dörr 2017; Sirén-Heikel ym. 2019). Joissakin tapauksissa prosessi on luotu yhtä käyttötarkoitusta, kuten yksittäistä uutisjuttua varten. Yleistymässä ovat kuitenkin algoritmiset järjestelmät, joiden tarkoitus on hyödyntää dataa jatkuvasti. Tällä hetkellä journalistisessa mediassa pääasiassa käytetyt algoritmiset järjestelmät voidaan haastattelututkimusten perusteella jakaa karkeasti kolmeen eri tyyppiin (Rydenfelt 2021a). Yksi on varsinaiseen sisällöntuotantoon tarkoitettu automaatio, kuten uutissisältöjä tuottavat uutisbotit. Toinen on automaation käyttö sisällön jakelussa, esimerkiksi uutissisällön suosittelu aiemman käytön perusteella. Kolmanneksi dataa käytetään automatisoiduissa prosesseissa, jotka toimivat journalistisen työn ja muun toimituksellisen työn apuvälineinä. Ne toteuttavat rutiininomaisia työtehtäviä kuten äänen litterointia, oikolukua, kuvanmuokkausta tai tekstin ja kuvan ajastamista yhteen. Lisäksi ne toimivat journalistisen työn ohjauksessa tarjoamalla työvälineitä ja mittareita. Siinä missä ensimmäinen ja toinen automaation käyttötapa ovat ainakin jossain määrin näkyviä myös yleisölle tuotostensa muodossa, kolmas käyttötapa on usein nähtävissä ainoastaan toimituksen sisäpuolella.

Aiemmin toteuttamamme haastattelu- ja kyselytutkimukset tarjoavat käsityksen datan sekä algoritmisen automaation käytöstä ja kehitysaskeleista sekä kansainvälisesti että suomalaisissa journalistisissa medioissa. Haapanen (2020b) toteutti vuosina 2019–2020 kyselyn eri Euroopan maiden medianeuvostojen edustajille, joilta saatiin 22 vastausta, sekä heidän nimeämilleen automaation edelläkävijäasemassa olevien median toimijoille, joilta saatiin 16 vastausta. Lisäksi Haapanen oli mukana toteuttamassa automaation käyttöönotosta ja kehityksestä kotimaisille medioille suunnattua kyselyä, johon vastasi 18 journalistisen median tai mediatalon edustajaa (Grundström, Haapanen ja Ilkka 2019). Rydenfelt (2021a) puolestaan haastatteli tutkimuksessaan kahden suuren suomalaisen median (Yleisradio ja Helsingin Sanomat) automaation kehityksestä sekä etiikasta vastaavien toimijoita (6 haastateltavaa) sekä näiden lisäksi julkisen sanan neuvoston puheenjohtajaa vuonna 2019. Näiden kysely- ja haastattelututkimusten tulosten perusteella nykyiset automaattisen sisällöntuotannon sovellukset niin Suomessa kuin kansainvälisestikin ovat ennemmin yksittäisiä esittelytapauksia, *showcaseja*, kuin vakiintunut tai merkittävä osa minkään mediatoimijan sisällöstä. Lisäksi automatisoitu sisällöntuotanto on varsin kapeaa. Algoritmit eivät valitse itsenäisesti juttuaihetta, kerää aineistoa eri lähteistä ja koosta ja julkaise tämän pohjalta juttua. Sisältöä tuottavan automaation rooli rajoittuu tällä hetkellä raportinomaisiin, esimerkiksi säätä ja ilmanlaatua, urheilu- ja vaalituloksia sekä esimerkiksi rikostilastoja, yritysten tulostietoja ja kiinteistöjen hintoja summaaviin uutisjuttuihin. Näitä uutisaiheita yhdistää se, että niistä on saatavissa systemaattisesti strukturoitua ja pääasiassa numeerista tai kategorista dataa (Haapanen ja Leppänen 2020). Uutistapahtumina automatisoituja juttuaiheita puolestaan yhdistää tapahtumien kulun ennalta-arvattavuus: esimerkiksi jalkapallo-ottelusta tyypillisesti kerrotaan tehdyt maalit ja maalintekijät. Useilla medioilla myös Suomessa on jo käytössä tai suunnitteluasteella erilaisia sisällön kohdentamisen ja suosittelun algoritmisia toteutuksia (Grundström, Haapanen ja Ilkka 2019; Haapanen 2020b). Lisäksi erilaisia automatisoituja järjestelmiä journalistisen työn tukena on ollut käytössä jo pitkään. On odotettavissa, että sekä automaatio että datan muunlainen käyttö lisääntyvät ja monipuolistuvat tulevaisuudessa. Tulosten perusteella uutta on koko ajan kehitteillä, ja innovaatioita myös valuu journalismiin toisenlaisista liiketoimintamalleista kuten algoritmisesta markkinoinnista.

Journalismissa käytössä oleva data ja algoritminen automaatio ovat peräisin sekä toimitusten sisä- että ulkopuolelta (Taulukko 1). Datan käyttö on yhtäältä toimituksen sisäistä toimintaa. Toimittajat käyttävät data-aineistoja ja algoritmeja informaation keräämiseen ja prosessointiin sekä tiedon tuottamiseen. Datatoimitukset laativat automaatioita, jotka muun muassa tuottavat sisältöjä ja toimivat niiden jakelussa yleisöille. Automaation kehitys ja datan käyttö ei kuitenkaan tapahdu yksinomaan toimituksen sisäpuolella. Toimitukset nojaavat usein algoritmeihin, jotka on suunniteltu sen ulkopuolella. Toimittajat käyttävät tiedonhankinnassa ja informaation prosessoinnissa toimituksen ulkopuolisia algoritmisia työkaluja, ja toimitusten käyttämä data on usein sekä peräisin ulkopuolisista lähteistä. Edelleen sisällön jakelussa toimitusten on mukauduttava esimerkiksi sosiaalisen median alustojen algoritmeihin saadakseen sisällöille näkyvyyttä. Mediatalojen institutionaaliseen kontekstiin lukeutuvat myös ulkopuoliset algoritmiset järjestelmät, joiden tekemien valintojen merkitys voi jäädä toimituksiltakin tiedostamatta ja ymmärtämättä.

**Taulukko 1.** Esimerkkejä datan ja algoritmien journalistisista käyttötavoista.

	<b>Informaation kerääminen ja prosessointi</b>	<b>Sisällön jakelu ja kohdentaminen</b>	<b>Prosessien ohjaus ja apuvälineet</b>
<b>Toimituksen sisäinen (usein tiedostettu)</b>	Uutisbotit, datankeruujärjestelmät, datajournalistiset projektit	Suosittelun, kohdentamisen ja personoinnin järjestelmät	Omat toiminnan-ohjausjärjestelmät, metriikat, rutiineja automatisoivat työkalut
<b>Toimituksen ulkopuolinen (joskus huomaamaton)</b>	Hakukoneet, data-lähteet, sosiaalinen media, ulkopuolinen data-analytiikka	Sosiaalisen median alustat ja vastaavat kanavat	Ulkopuoliset ohjelmistot ja metriikat

Sisäpuolisen ja ulkopuolisen raja voi hämärtyä, esimerkiksi kun journalistisessa kontekstissa käytetään ulkopuolisen palveluntarjoajan tuottamia järjestelmiä, jotka kuitenkin on muovattu toimituksen tarpeita varten. Automaatiikka voi sisältää kerrostumia, joista osa on peräisin toimituksen ulkopuolelta ja osa on toimituksen tietoisesti suunnittelun tulosta. Sama voi koskea dataa, joka voidaan valikoida osin ulkopuolisista lähteistä, osin esimerkiksi aiemmin tehdyn toimituksellisen työn tuloksena. Erottelu on kuitenkin erityisen merkityksellinen, kun pohditaan ulkopuolisen datan ja automatisoitujen järjestelmien mahdollisia ongelmia. Ulkopuolinen data voi olla valmiiksi valikoitua tavalla, joka ei lopulta palvele journalistisia tarkoituksia (Diakopoulos 2019; Lindén, Tuulonen ja työryhmä 2019). Haastattelututkimusten perusteella dataan perustuvan journalismin kehittäjien keskeinen huolenaihe onkin ulkopuolisen datan mahdollisesti sisältämät vinoumat, jotka jäävät havaitsematta (Rydenfelt 2021a). Samat ongelmat voivat olla käsillä myös silloin, kun toimituksen sisäpuolella kehitetyt automatisoidut järjestelmät tai datan valikoimisen tavat eivät (enää) vastaa alkuperäistä tarkoitustaan.

Erään esimerkin algoritmisen kontekstin ja ulkopuolisen automaation vaikutuksesta journalistiseen mediaan tarjoaa sosiaalisen median uutisvirta ja sen järjestämiseen liittyvät algoritmiset prosessit. Robyn Caplan ja danah boyd (2018) käyvät läpi Facebookin uutisvirran priorisointijärjestyksen muutoksia 2010-luvulla. Alkuvaiheessa uutisvirran kärkipäähän nousivat jutut, joilla oli paljon tykkäyksiä ja kommentteja. Tämän jälkeen Facebook otti käyttöön erilaisia uutisartikkeleiden laatumittareita, joihin vaikutti muun muassa se aika, jonka käyttäjä vietti Facebookin ulkopuolella artikkelin klikkauksen jälkeen. Seuraavaksi Facebook pyrki pitämään käyttäjät omalla alustallaan korostamalla alustalle suoraan ladattua sisältöä ja tarjoamalla valikoitujen yhdysvaltalaisen mediatoimijoiden uutisartikkeleita luettavaksi Facebookin sisällä. Myöhemmin Facebook alkoi priorisoimaan käyttäjän kaveriverkoston tuottamaa sisältöä mediatalojen uutisten sijaan. Facebookin uutisvirrassa menestyäkseen mediatoimijoiden on näin täytyntä vastaavasti muuttaa omia toimintatapojaan. Soveltamalla Paul DiMaggion ja Walter Powellin (1983) ajatusta institutionaalisten kenttien samankaltaistumisesta Caplan ja boyd (2018) esittävät, että sosiaalisen median uutisvirtoihin liittyvät algoritmiset järjestelmät aiheuttavat samankaltaistumista mediakentällä. Mitä riippuvaisempia toimijat mediakentällä ovat yhdestä organisaatiosta, esimerkiksi Facebookista ja sen uutisvirran tuomasta näkyvyydestä ja verkkoliikenteestä, sitä enemmän mediatoimijoiden on mukautettava omaa toimintaansa tämän organisaation luomiin käytäntöihin ja algoritmeihin.

## Datan ja algoritmisten järjestelmien läpinäkyvyys

Läpinäkyvyydestä on tullut keskeinen käsite dataa ja algoritmisia järjestelmiä koskevassa yhteiskuntatieteellisessä tutkimuksessa ja julkisessa keskustelussa. Datasta ja algoritmeista on näin tullut uusi kohde vaatimuksille tiedon julkisuudesta ja avoimuudesta, joita on esitetty antiikin ajan klassisista ajattelijoista alkaen. Modernin ajan läpinäkyvyyden käsitteen ja sitä koskevan keskustelun taustalla on siirtyminen absoluuttisesta monarkiasta liberaaliin ja porvarilliseen hallintoon valistusaikana (McStay 2014). Varhaisena esimerkkinä tästä kehityksestä mainitaan usein Anders Chydeniuksen Ruotsin valtiopäivillä läpi ajama vuoden 1766 painovapauslaki, joka sisälsi ajatuksen julkisen päätöksenteon avoimuudesta (ks. Erkkilä 2012, 48–50). Kansainvälisesti keskeinen hahmo läpinäkyvyyden käsitteen kehityksessä ja sen merkityksen tarkastelussa lienee kuitenkin Chydeniusta pari vuosikymmentä nuorempi aikalainen, brittifilosofi Jeremy Bentham, joka puolusti päätöksenteon julkisuutta (*publicity*), josta hän käytti myös sanaa *läpinäkyvyys* (*transparency*) (ks. Baume ja Papadopoulos 2015, 3). Utilitaristisen etiikan kannattajana Bentham keskittyi läpinäkyvyyden puolesta argumentoidessaan sen tuottamiin hyötyihin tai etuihin sekä toisaalta haittoihin. Bentham (1843) tarjoaakin läpinäkyvyyden eduista varsin kattavan erittelyn. Hänen mukaansa läpinäkyvyys pitää päättäjät kiinni velvollisuuksissaan, takaa kansan luottamuksen päätöksentekoon sekä auttaa hallinnan alla olevien ihmisten tai kansalaisten toiveiden välittymisessä päättäjille. Näitä kolmea Bentham kutsui merkittäviksi näkökohdiksi läpinäkyvyyden puolesta. Niiden lisäksi läpinäkyvyys tuottaa kolme vähemmän merkittävää seurausta: kansa voi tehdä informoituja valintoja päättäjiensä suhteen (esimerkiksi vaaleissa), kansalla olevaa informaatiota voidaan hyödyntää päätöksenteossa, ja läpinäkyvyys tarjoaa myös huvitusta, jonka senkään merkitystä ei sovi kokonaan unohtaa.



Benthamin erittely on niin kattava, että nykyisessä keskustelussa esitetyt läpinäkyvyyden puolustuspuheenvuorot pitkälti yhdistävät siinä esitettyjä näkökohtia. Läpinäkyvyyttä on ajettu tapana edistää päätöksentekijöiden vastuullista toimintaa: se esimerkiksi estää korruptiota ja haitallista oman edun tavoittelua (Lindstedt ja Naurin 2010). Viime vuosikymmeninä voimistuneet osallistuvat tai deliberatiiviset näkemykset demokratiasta painottavat kansalaisten informoitujen valintojen merkitystä sekä toisaalta kansalaisten informaation tuomista päätöksenteon käyttöön (ks. Rydenfelt 2019b; Rydenfelt 2021b). Deliberatiivisen demokratian näkökulmasta läpinäkyvyys onkin yksi edellytys päättäjien ja kansalaisten väliselle vuorovaikutukselle ja päätöksenteolle. Tätä yleisemmin ja laajemmin viestinnän ja median etiikan piirissä on kehittynyt ajatus yleisöjen oikeudesta saada tietoa näille merkityksellisistä asioista ja tapahtumista; tiedonsaannin mahdollistavaa avoimuutta ja läpinäkyvyyttä on puolestaan pidetty yhtenä viestinnän eettisistä arvoista (ks. Rydenfelt ja Juholin 2016; Rydenfelt 2016). Läpinäkyvyyttä koskeva keskustelu ja siihen liittyvät vaatimukset ovat varsinkin viime vuosikymmeninä ulottuneet paljon poliittisen ja yhteiskunnallisen päätöksenteon ydintä laajemmaksi kaikkeen viranomaisten toimintaan, mutta myös yritysten ja yhteisöjen käytäntöihin sekä ammatilliseen itsesääntelyyn.

Viime vuosina paljon keskustelua on herättänyt datan ja algoritmien läpinäkyvyys, jota koskevien vaatimusten taustalla keskeinen perustelu näyttäisi olevan ensimmäinen Benthamin esittämä näkökohta läpinäkyvyyden puolesta: velvollisuuksissa pysyminen. Laajassa mielessä läpinäkyvyys on tutkimuskirjallisuudessa käydyssä keskustelussa käsitetty dataa ja algoritmeja koskevan informaation saatavuutena, joka on puolestaan käsitetty *tilivelvollisuuden* tai *vastuullisuuden* (accountability) edellytykseksi. Läpinäkyvyys on mielletty oleelliseksi osaksi algoritmisten prosessien tekemistä ymmärrettäviksi ja hallittaviksi sekä eettisen arvioinnin kohteiksi. Muuten algoritmien ”musta laatikko” (Pasquale 2015) imee sisään dataa ja sylkää ulos lopputuloksen, mutta sen tekemät valinnat ja prosessit jäävät tuntemattomiksi. Mustan laatikon prosessin tunteminen ja arvioiminen edellyttävät tietoa siitä, miten data valikoitiin ja millä tavoin lopputulokseen päädyttiin. Tämän vuoksi läpinäkyvyys on nähty välttämättömänä, joskaan ei automaattisesti riittävänä, ehtona vastuulliselle datan ja algoritmien käytölle (Pasquale 2015; Diakopoulos 2016; Berger ja Owetschkin 2019; Gorwa ja Garton Ash 2020). Vaatimukset mustan laatikon avaamiseksi ovat vaikuttaneet myös käytäntöön: läpinäkyvyys on yksi keskeisistä periaatteista esimerkiksi datan käyttöä sääntelevässä EU:n tietosuojasetuksessa (Felzmann ym. 2019).

Toisaalta datan ja algoritmisten järjestelmien läpinäkyvyyttä ratkaisukeinona sekä sen toteuttamisen tapoja on analysoitu myös kriittisesti (ks. Gorwa ja Garton Ash 2020). Läpinäkyvyyden käsite ei ole selvä tai yksiselitteinen, ja on epävarmaa, miten dataa koskevan sääntelyn vaatima läpinäkyvyys käytännössä toteutuu tai toteutettaisiin tiedontuotannon ja päätöksenteon järjestelmien ominaisuutena (Felzmann ym. 2019). Ongelma kytkeytyy laajempaan kysymykseen läpinäkyvyyden roolista eettisen ja yhteiskunnallisen vastuun toteutumisessa. Läpinäkyvyyttä koskevassa tutkimuksessa on usein esitetty, että läpinäkyvyys sellaisenaan ei riitä (Lindstedt ja Naurin 2010). Informaation on tavoitettava kontekstissaan relevantti yleisö, jotta sitä voidaan arvioida. Lisäksi yleisöllä on oltava tapoja vaatia tai saattaa informaation perusteella virheellisesti tai väärin toimineita toimijoita vastuuseen. Algoritmisten järjestelmien tutkimuksessa onkin kritisoitu läpinäkyvyyden toteuttamista tavalla, joka ei todellisuudessa tavoita relevanttia yleisöä. Mike Ananny ja Kate Crawford

(2018) ovat käsitelleet algoritmisten järjestelmien läpinäkyvyyden problematiikkaa asettaen vastakkain pelkän ”näkemisen” ja ”tietämisen ja ymmärtämisen”, joista ensimmäinen voi heidän mukaansa toteutua ilman jälkimmäisiä. Voimme nähdä saamatta tietoa – algoritmisen mustan laatikon sisälle voi kurkistaa ymmärtämättä mitään. Läpinäkyvyyden toteutumista ja toteuttamista monimutkaistaa edelleen se, että algoritmiset järjestelmät ovat tyypillisesti osa toimijoiden verkostoa: algoritmiset järjestelmät eivät ole pelkästään algoritmeja ja dataa, vaan ne pitäisi pikemminkin ymmärtää tilannesidonnaisina kokoonpanoina, jotka koostuvat ihmisistä, koodista ja datasta (Ananny 2016). Ananny ja Crawford (2018) ehdottavatkin järjestelmän sisälle katsomisen tilalle tapauskohtaista näkemystä läpinäkyvyydestä ja sen toteuttamisesta. Kontekstin hahmottaminen on tarpeen, jotta käsitämme, mitä järjestelmän toiminnan ymmärtäminen, kontrollointi ja eettisen vastuullisuuden toteutuminen edellyttävät.

Tämä läpinäkyvyyden kritiikki osoittaa, että algoritmien ja automatisoidun päätöksenteon kohdalla läpinäkyvyyden käsite on joskus liitetty pelkkiin teknisten yksityiskohtien esittelyyn tai saattamiseen muodollisesti tarjolle – esimerkiksi datan, algoritmin tai lähdekoodin tuomiseen saataville jollekin yleisölle. Kritiikki kohdistuuikin ennen kaikkea läpinäkyvyyden toteuttamiseen jonkinlaisena muodollis-teknisenä avoimuutena. Kritiikin mukaan läpinäkyvyyttä ei tulisi käsittää pelkästään teknisenä ja muodollisena ominaisuutena, joka toteutuu ennalta määrätyt ja muuttumattomat ehdot täyttämällä, vaan se tulisi pikemminkin ymmärtää pyrkimyksenä antaa yleisölle ymmärrettävää informaatiota algoritmien toiminnasta, dataa koskevista valinnoista sekä järjestelmien kokonaisuuksista. Sen toteutuminen on tapauskohtaista ja riippuu sekä tilanteesta että yleisöstä. Tietyissä yhteyksissä algoritmisen läpinäkyvyyden toteutumiseen voi riittää esimerkiksi lähdekoodin tai datan avaaminen tai teknisten yksityiskohtien esitys. Monissa yhteyksissä tällaisen esillepanon muuttaminen tarpeelliseksi informaatioksi vaatii kuitenkin taustatietoja ja ajallisia ja tiedollisia resursseja, joita tarkoitettulla yleisöllä ei ole.

Edellä esitetyllä tavalla tutkimuskirjallisuudessa läpinäkyvyyttä on pidetty keinona kasvattaa datan käytön vastuullisuutta. Käytännön toimijoiden keskuudessa tämä ei kuitenkaan välttämättä ole läpinäkyvyyttä tavoittelevien käytäntöjen ainoa tai edes keskeisin päämäärä. Kun läpinäkyvyyden käsite valjastetaan ammatillisen itsesääntelyn tai liiketoiminnasta viestimisen käyttöön, taustalla voivat olla myös pyrkimykset välttää esimerkiksi datan käyttöä koskeva tiukempi sääntely lainsäädännön ja viranomaisvalvonnan muodossa. Asetelma on vastaava kuin toisen datan käyttöön keskeisesti liittyvän käsitteen, informaation yksityisyyden kohdalla. Tutkimuksessa on huomautettu, että nykyinen käsityksemme yksityisyydestä datan käytön ja datafikaation sääntelyn keskeisenä oikeutena tai arvona on samalla digitaalisen talouden mahdollistaja (Coll 2014). Lainsäätäjä on jättänyt vastuuta päätöksestä paljolti ihmisille: tyypillisesti digitaalisia palveluita käyttäessämme annamme palveluntarjoajalle luvan hyödyntää yksityisyyden piiriin kuuluvaa dataa. Tämä sääntelyä koskeva ratkaisu – yksilökohtaiseen valintaan perustuva näkemys yksityisyydestä – voi olla dataa hyödyntäville toimijoille hyödyllinen tai suosiollisempi kuin mahdollisesti ankarampi sääntely. Edelleen käsitys yksityisyydestä datafikaation keskeisenä arvona vaikuttaa tapaamme hahmottaa datan käyttöön liittyviä eettisiä ja oikeudellisia ongelmia, joita tällä hetkellä pyritään ratkaisemaan antamalla käyttäjille lisää mahdollisuuksia vaikuttaa dataa koskeviin ratkaisuihin ja erityisesti kouluttamalla ihmisiä pitämään datastaan parempaa huolta (ks. emt.). Vaatimusten kohteeksi eivät tällöin nouse perinpohjaisemmat muutokset dataa hyödyntävien toimijoiden toiminta-

malleihin. Sama voi koskea myös läpinäkyvyyttä, johon keskittyminen voi olla dataa hyödyntäville toimijoille edullisempaa kuin tiukempi sääntely. Läpinäkyvyyteen keskittyminen voi myös kaventaa näkemyksiä siitä, millä tavoin datan käyttöä koskevia ongelmia tulisi havaita ja pyrkiä korjaamaan. Seuraavassa tarkastelemme journalistisen datankäytön ja algoritmien läpinäkyvyyden toteuttamiseen liittyviä pyrkimyksiä ja mahdollisuuksia näiden näkökohtien valossa.

## Journalistiset läpinäkyvyyskäytännöt

Jo ennen datan ja algoritmien nykyisten käyttötapojen kehitystä on esitetty vaatimuksia journalismia ja sen toimintatapoja koskevan läpinäkyvyyden lisäämisestä. Läpinäkyvyyttä on ehdotettu journalismin uudeksi eettiseksi normiksi tai jopa ”uudeksi objektiivisuudeksi” (Weinberger 2009) korvaamaan journalismin tutkimuksessa jo vuosikymmenten ajan kritisoituja ja kyseenalaistettuja objektiivisuuden ja puolueettomuuden arvoja (esim. Durham 1998; Gauthier 1993). Nämä vaatimukset heijastelevat erityisesti Benthamin toista näkökohtaa läpinäkyvyyden puolesta: sen kykyä kasvattaa kansan luottamusta päätöksentekoon. Journalistisilla tiedotusvälineillä on perinteisesti ollut keskeinen ja omaehtoinen rooli yhteiskunnallisesti tärkeistä aiheista käytävän julkisen keskustelun ylläpitämisessä. Viime aikoina yleisön käsitys median objektiivisuudesta ja puolueettomuudesta on kuitenkin murentunut, ja myös luottamus journalistiseen mediaan on rapautumassa (Edelman 2019). Kun läpinäkyvyyttä ja avoimuutta on esitetty yleisesti reseptiksi luottamuksen lisäämiseksi eri instituutioita kohtaan (esim. Berger ja Owetschkin 2019), ratkaisuksi luottamuksen rapautumisen kierteen pysäyttämiseen ja yleisön luottamuksen palauttamiseen monet journalismin tutkijat ja toimitustyön ammattilaiset ovat esittäneet läpinäkyvyyttä: journalistisen median tulisi avata työprosessejaan yleisölle nykyistä laajemmin (Deuze 2005; Allen 2008; Philips 2010; Karlsson 2010; 2011; McBride ja Rosenstiel 2014; Dörr ja Hollnbuchner 2016; Zamith 2019). Osana journalistista itsesääntelyä läpinäkyvyyden sekä luottamuksen tavoittelu voivat edellä esitetyn mukaisesti myös liittyä yleiseen pyrkimykseen välttää mediaa koskevaa sääntelyä lainsäädännön ja viranomaisvalvonnan muodossa.

On kuitenkin huomattava, että empiirinen tutkimus ei ole pystynyt osoittamaan yhteyttä läpinäkyvyyden ja journalismia kohtaan lisääntyneen luottamuksen välillä. Tähänastisten tulosten perusteella on vain vähän (Chen, Curry ja Whipple 2019; Curry ja Stroud 2019; Karlsson, Clerwall ja Nord 2014; 2017; Meier ja Reimer 2011) tai käytännössä ei lainkaan (Koliska 2015; Roberts 2007) merkkejä siitä, että läpinäkyvyys tukisi luottamuksen säilymistä tai sen kasvattamista. Vaikutus voi olla jopa päinvastainen (esim. Smolkin 2006), jos läpinäkyvyyden nimissä esiin tuodut journalistiset käytännöt eivät vastaakaan yleisön odotuksia (Manninen 2020; Tandoc ja Thomas 2017). Kaiken kaikkiaan tähänastisesta tutkimuksesta ei voida kuitenkaan tehdä pitkälle meneviä tai jyrkkiä johtopäätöksiä. Sikäli kuin luottamuksen rapautuminen liittyy laajempiin yhteiskunnallisiin kehityskulkuihin, journalististen työprosessien avaaminen ei välttämättä johda luottamuksen lisääntymiseen. Lisäksi läpinäkyvyyden mahdolliset positiiviset seuraukset voivat peittyä muiden muuttujien vaikutuksiin, eivätkä tähänastiset tulokset näin ollen välttämättä täsmällisesti heijastele nimenomaan läpinäkyvyyden tarjoamia mahdollisuuksia (Haapanen 2020a).

Journalismin tutkimuksessa läpinäkyvyyden tuottamisen ja toteuttamisen keinoina on tarkasteltu media-alan tuotantoprosesseihinsa ja tuotoksiinsa kehittämiä käytäntöjä, jotka kertovat yleisölle toimittajista ja toimitustyöstä ja mahdollistavat yleisön osallistumisen keskusteluun (Heikkilä ym. 2012; Karlsson 2010; 2011; Koliska 2015; Meier ja Reimer 2011). Tutkimuksessa nämä keinot on jaettu Taulukon 2 mukaisesti kolmeen päätyyppiin (Haapanen 2020a; ks. myös Groenhard ja Bardoel 2012). Ensiksi ovat keinot, jotka pyrkivät tarjoamaan tietoa tuottajista eli toimittajista ja toimituksista. Toinen tyyppi ovat tuotannon prosesseja eli toimitustyötä valottavat ja selittävät keinot. Kolmannet keinot pyrkivät lisäämään läpinäkyvyyttä tarjoamalla yleisölle mahdollisuuden osallistua jutuntekoon tai valmiiden juttujen kommentointiin ja keskusteluun.

**Taulukko 2.** Keskeisiä läpinäkyvyyden toteuttamisen keinoja journalismissa (Haapanen 2020a, 4).

Päämäärä	Esimerkkejä toteutuksesta
<b>Tuottajiin liittyvä läpinäkyvyys</b> – antaa tietoa kirjoittajasta ja organisaatiosta	Toimittajan nimi, esittely, kuva ja yhteystiedot; linkki julkaisun noudattamiin eettisiin ohjeistuksiin; maininta eturistiriidoista.
<b>Tuotantoon liittyvä läpinäkyvyys</b> – selittää jutunteko- ja julkaisemisprosesseja	Jutun yhteydessä tai muualla julkaistut toimituksen kommentit; ulkoiset linkit lähdeaineistoihin ja sisäiset linkit muihin samantapaisiin juttuihin kyseisessä julkaisussa; muunlainen taustatieto kuten lähteenä käytetyn haastattelun kokonaistaltiointi; aikaleimat sekä maininnat ja selitykset jälkikäteen tehdyistä muutoksista ja korjauksista.
<b>Osallistumisen luoma läpinäkyvyys</b> – yleisön osallistaminen vuorovaikutukseen	Foorumi yleisön kommentille joko jutun yhteydessä tai sosiaalisen median kanavilla; yleisön osallistaminen uutistuotantoon yhteistyössä toimituksen kanssa.

Edellä mainituissa tutkimuksissa esitettyjen arvioiden perusteella läpinäkyvyyttä tavoittelevien keinojen toimivuus riippuu monista tekijöistä. Ensinnäkin näiden keinojen kykyyn lisätä läpinäkyvyyttä vaikuttaa se, miten ne otetaan käyttöön toimituksissa ja miten ne onnistutaan vakiinnuttamaan osaksi journalismin pysyviä käytäntöjä ja omaksumaan osaksi toimitusten ammatillista kulttuuria. Vaikka toimittajat suhtautuvat läpinäkyvyyden periaatteisiin myönteisesti (ks. Fengler ym. 2014), heidän halunsa toteuttaa läpinäkyvyyden ideoita käytännössä on ollut laimeampi (ks. Eberwein, Fengler ja Karmasin 2019). Tutkimuksessa on myös arvioitu, että monet läpinäkyvyyden lisäämiseksi tarkoitetut keinot luovat lähinnä illuusiota läpinäkyvyydestä sen sijaan, että ne aidosti kuvaisivat tai pystyisivät kuvaamaan toimittajan ja toimituksen työskentelyä (Koliska 2015; Haapanen 2020a). Läpinäkyvyyden ihanne muuttuu ongelmalliseksi, jos journalistit ja toimitukset ovat tosiasiasa valmiita avaamaan vain ideaalisia ja sosiaalisesti hyväksyttäviä puolia toiminnastaan ja työstään (Koliska 2015). Journalistiseen työhön sisältyy myös seikkoja, joita toimittajat eivät voi ylei-

sölleen läpinäkyvästi avata. Nämä voivat liittyä esimerkiksi lähdesuojaan eli lain suomaan mahdollisuuteen siitä, että toimittajalla on oikeus pitää lähteen henkilöllisyys salassa. Toimittaja voi saada jutuntekoproessin kuluessa haltuunsa muitakin tietoja, jotka eivät ole julkaistavissa. (Haapanen 2020a, 5–6.)

Toiseksi voidaan hyvin kuvitella, että teknisesti ja rutiininomaisesti toteutettu läpinäkyvyys voi kääntyä tarkoitustaan vastaan. Jos nimittäin työprosessista pyritään kuvaamaan kaikki sen yksityiskohdat, saattaa tällainen kuvaus sekä kuormittaa liiaksi toimittajaa että aiheuttaa sen, ettei yleisö enää erota olennaista sisältöä vähemmän tärkeästä (Allen 2008; Vos ja Craft 2017). Myös vähemmän rutiininomaisen läpinäkyvyyden hyödyt voivat osoittautua vähäisiksi. Minkälaisia hyödyllisiä johtopäätöksiä yleisö voi tehdä esimerkiksi siitä, että se saa tietää toimittajan puoluekannan tai keskeiset kiinnostuksen kohteet? Onko toimittaja mahdollisesti tai väistämättä puolueellinen esimerkiksi raportoidessaan kannattamastaan puolueesta, vai tulisiko hänet pikemminkin nähdä kyseisen puolueen erityisenä asiantuntijana? Näiden tekijöiden vuoksi myös journalismin kohdalla läpinäkyvyyden tarve sekä sen toteutuksen onnistuminen ovat vääjäämättä tapauskohtaisia. Pelkkien teknisten ja rutiininomaisten merkintöjen toteuttaminen ei ole vielä todellista läpinäkyvyyttä, ja rutiininomaiset merkinnät voivat muuttua jopa tarkoituksensa vastaiseksi. Tiedonvälityksen ammattilaisina journalistit voivat kuitenkin olla lähtökohtaisesti hyvässä asemassa harkitsemaan tapauskohtaisesti, millä tavoin toteutettu läpinäkyvyys on paikallaan ymmärryksen ja tiedon lisäämiseksi juuri heidän omalle yleisölleen.

## **Dataan liittyvät läpinäkyvyyden käytännöt journalismissa**

Dataa hyödyntävinä toimijoina journalistisiin instituutioihin kohdistunee jatkossa entistä enemmän vaatimuksia datan käyttöä koskevasta läpinäkyvyydestä. Nämä läpinäkyvyyden kysymykset kytkeytyvät kuitenkin voimakkaasti edellä esiteltyyn akateemiseen ja julkiseen keskusteluun, jossa läpinäkyvyyttä on esitetty ratkaisuna ennen kaikkea journalismia kohtaan koetun luottamuksen ylläpitämiseen ja kasvattamiseen. Vaikka edellä läpikäymämme tutkimuksen valossa ei toistaiseksi ole selvää, missä määrin ja millä tavoin läpinäkyvyys tosiasiaa lisää luottamusta, juuri journalismin kohdalla datan ja sen käytön läpinäkyvyyden lisäämiselle voi olla muitakin hyviä perusteita. Päätöksenteon ja tiedontuotannon pohjautuessa entistä enemmän algoritmeihin myös datan ja automaation käytön valaiseminen voi muuttua keskeiseksi osaksi sitä tehtävää, joka journalismilla maailman kuvaajana ja selittäjänä on – ei vain tavaksi lisätä journalismia kohtaan koettua luottamusta.

Journalismissa datan, automaation ja algoritmien käytön ilmoittamisesta ei ole vakiintuneita käytäntöjä, saati kansainvälistä tai usein edes maakohtaista lainsäädännön asetta-  
maa ohjeistusta tai itsesääntelyllistä linjaa (Haapanen 2020b).<sup>1</sup> Millä tavoin journalismi voisi tehdä dataa ja algoritmeja koskevia ratkaisujaan läpinäkyviksi? Esimerkiksi automatisoidun uutistuotannon lähdekoodien tai sen käyttämien datajoukkojen esitleminen tai ”näyttämisen” yleisölle ei riitä todellisen, ymmärrystä ja tietämystä lisäävän läpinäkyvyyden toteutumiseen. Lähestymme kysymystä tarkastelemalla, miten journalismin edellä kuvatut läpinäkyvyyskäytännöt voisivat toteutua niin, että ne valaisisivat edellä eriteltyjä datan ja algoritmien journalistisia käyttötapoja yleisölle ymmärrettävällä tavalla (Taulukko 3). Kiinni-

tämme huomiota siihen, että näiden käytäntöjen kautta tuotaisiin läpinäkyviksi paitsi toimituksen sisäiset tekijät ja valinnat, myös toimituksen ulkopuolisten algoritmisten toimijoiden vaikutus käytettyyn dataan, tehtyihin valintoihin ja journalistisiin sisältöihin. Ulkopuolisen automaation vaikutuksen ja merkityksen tuominen näkyväksi voikin olla tärkeää. Toimituksen käytössä olevat ulkopuoliset automaattiset järjestelmät on usein suunniteltu muuhun (tai ainakin yleisempään) tarkoitukseen kuin siihen, mihin niitä journalismin kontekstissa käytetään. Kärkevä esimerkki on sosiaalinen media, jolla ei tyypillisesti ole varsinaisia journalistisia tarkoituseriä. Kun Facebookin tapa priorisoida uutisvirtaa vaikuttaa siihen, miten uutisjuttujen tai kokonaisten mediatoimijoiden menestystä mitataan, Facebookin algoritmiseen järjestelmään upotetut arvot ja tavoitteet ja sen edustamat menestyksen metriikat tulevat osaksi myös mediatoimijoiden omia arvoja, tavoitteita ja metriikoita. Mediatoimijoihin nähden ulkoinen ja alati muuttuva algoritmisen järjestelmä vaikuttaa näin siihen, millainen artikkeli on arvokas ja miten arvoa pyritään mittaamaan. Helposti datalla mitattava ja algoritmisesti huomioon otettava paremmuuden mittari voi samalla kasvattaa merkitystään suhteessa perinteisiin journalistisiin arvoihin ja tavoitteisiin.

**Taulukko 3.** Journalistiset läpinäkyvyyskäytännöt sekä datan ja algoritmien käyttötavat.

Päämäärä	Informaatio	Esimerkkejä toteutuksesta
<b>Automaattisia toimijoita koskeva läpinäkyvyys</b>	Tiedot automaation käytöstä Tiedot automaation tuottajista ja tekijöistä sekä vastuussa olevista toimijoista Tiedot automaation perusideasta ja tavoitteista Ulkoisen automaation tarjoajan (liike) toimintamalli	Automaatiota koskeva ”tekijäesittely” Ulkoisen palvelun tuote-esittely
<b>Datan prosessointia ja valintoja koskeva läpinäkyvyys</b>	Sisäiset ja ulkoiset datalähteet Ulkoisten hakujärjestelmien käyttö Datan valinnan perusteet Prosessin tarkoitus ja toimintaperiaate Mukautuminen ulkoisiin automaattisiin järjestelmiin Prosessin luomisen vaiheet ja sen korjaukset Virhelähteiden ja epävarmuuksien esittely	Datalähteen tuominen saataville esimerkiksi linkin kautta Lähdekoodin tuottaminen saataville Mahdollisuus vaihtaa prosessin eri vaiheiden tai versioiden välillä Yleisön mahdollisuudet vaikuttaa prosesseihin tai lopputuloksiin esimerkiksi vaihtamalla prosessin asetuksia tai parametreja
<b>Yleisösuhteeseen liittyvä läpinäkyvyys</b>	Tiedot yleisöä koskevan datan keräämisestä ja säilyttämisestä Datan kerääminen ulkoisilta jakelukanavilta (kuten sosiaalisesta mediasta) Datan käytön tarkoitukset ja periaatteet	Kohdennetun tai suositellun sisällön valikointiperusteen esittäminen sisällön yhteydessä Mahdollisuus vaikuttaa sisällön kohdentamiseen omien valintojen tai kohdentamisen pois päältä laittamisen avulla

Journalismin vakiintuneissa läpinäkyvyyskäytännöissä ovat ensinnäkin keskeisiä tekijään liittyvät seikat. Datan ja automaation kohdalla ne liittyvät erityisesti automaattiseen sisällöntuotantoon. Annettu informaatio koskee tällöin sisällöntuotannon osana olevaa automaatiota. Läpinäkyvyyttä lisää sen tuominen esiin, että sisältö on (joiltain osin) tuotettu algoritmisesti tai automaation avulla. Lisäksi esille voidaan tuoda tarkemmin se, millaisesta ja mitä käyttötarkoituksia varten suunnitellusta automaatiosta on kysymys ja kuka automaation on suunnitellut ja tuottanut. Usein sisältöä tuottavat myös toimituksen ulkopuoliset algoritmit, joista voidaan vastaavasti antaa samat tiedot. Toimijoita koskeva läpinäkyvyys ulottuu sisällön tuotannon ohella myös muihin datan ja automaation käyttötapoihin. Myös sisällön kohdentamisessa tai personoinnissa on mukana automaattisia toimijoita, jotka nekin voivat olla toimituksen sisäisiä – esimerkiksi toimituksen itse kehittämiä – tai ulkopuolelta hankittuja palveluja.

Toinen journalismin läpinäkyvyyskäytäntö koskee journalistisen sisällöntuotannon ja julkaisun prosesseja. Läpinäkyväksi voidaan tehdä ensinnäkin valinnat, jotka koskevat datan lähteitä. Mistä (sisäisistä ja ulkoisista) lähteistä data on peräisin, miten se on valikoitu tai tuotettu ja miksi tai miten se sopii käsillä olevaan tarkoitukseen? Varsinkin datajournalismissa läpinäkyvyys on usein samastettu juuri datan lähteiden ja niitä koskevien huomioiden esittämiseen (Zamith 2019). Läpinäkyvyys voidaan kuitenkin ulottaa datan käsittelyyn ja sen algoritmeihin laajemmin kuvaamalla tavat, joilla dataa prosessoidaan, sekä prosesseille suunnitellut päämäärät ja lähtökohdat. Nicholas Diakopoulos ja Michael Koliska (2017) ovat esittäneet, että algoritmisen läpinäkyvyys journalismissa voisi käsittää dataa koskevien valintojen lisäksi myös mallintamista sekä päättelyä koskevat valinnat. Tapoja toteuttaa läpinäkyvyyttä ovat muodollis-tekninen avoimuus kuten lähdekoodin sekä sen kehityshistorian esittäminen, toimituksellisten valintojen esiintuominen sekä informaation tarjoaminen esimerkiksi virheiden havaitsemisen ja korjaamisen standardeista ja tehtävien päätelmien epävarmuudesta (Diakopoulos ja Koliska 2017, 9–12). Tällainen informaatio keskittyy pitkälti ihmisten toimituksen sisäpuolella tekemien ratkaisujen tekemiseen läpinäkyviksi. Proessin ja siinä tehtyjen valintojen valottaminen voidaan kuitenkin ulottaa yhtä lailla myös ulkoisten tekijöiden vaikutukseen: millä tavoin tehdyt valinnat ovat mukautuneet ulkoisten automaattisten järjestelmien ja algoritmien toimintatapaan? Esimerkiksi mediatalot voivat esittää tapoja, joilla sosiaalisen median alustojen vaatimukset näkyvät niiden tuottamien otsikoiden ja juttujen rakenteessa. Tällainen informaatio voisi myös tarkentaa ja korjata yleisölle syntyneitä ”kansanteorioita” (Eslami ym. 2016) sosiaalisen median uutisvirtojen toiminnasta.

Journalismin tyyppillisissä läpinäkyvyyskäytännöissä kolmas on yleisön osallistuminen kommentoimalla ja vuorovaikuttamalla journalististen päätöksentekijöiden kanssa. Vuorovaikutus on mahdollista myös datankäyttöön perustuvien tuotteiden ja sisältöjen kohdalla, ja datankäyttö journalismissa tarjoaa sille vielä lisää mahdollisuuksia. Diakopoulos ja Koliska (2017, 11–13) käsittelevät tällaisia mahdollisuuksia lisätä läpinäkyvyyttä yleisösuhteen kautta käyttöliittymän avulla. Yleisölle voidaan esimerkiksi tarjota mahdollisuus tarkastella julkaisun käyttämiä datajoukkoja ja algoritmeja sekä muuttaa taustalla vaikuttavien mallien ja päättelyiden oletuksia. Yleisön on usein mahdollista vaikuttaa varsinaisten datajournalististen juttujen sisältöön muuttamalla niiden parametreja ja asetuksia. Datajournalistisen jutun avulla yleisö voi esimerkiksi saada laskelman siitä, millä tavoin veronkorotukset vai-

kuttavat juuri hänen menoihinsa syöttämällä laskuriin tietoja kulutustottumuksistaan.<sup>2</sup> Kyse on aktiivisesta ja usein eksplisiittisestä vuorovaikutuksesta, joka itsessään tuo journalismin toimintaa läpinäkyväksi.

Yleisön toimintaa koskevan datan keräämisen ja hyödyntämisen kautta yleisön vaikutus saa kuitenkin myös hyvin toisenlaisen ja huomaamattomamman muodon. Yleisön toiminnan tuottaman datan käyttö ei tule niinkään tavaksi lisätä läpinäkyvyyttä vaan kohteeksi, jota itseään on tarpeen valottaa. Läpinäkyvyys koskee sitä, millä tavoin yleisön toiminnan tuottamaa dataa kerätään ja käytetään sekä mitä sen käytöllä tavoitellaan tai miten datan pohjalta tuotettu informaatio vaikuttaa sisältöjen kohdentamiseen. Läpinäkyvyyttä voi lisätä esittämällä kohdennetun tai suosittelun sisällön valikointiperusteiden itse sisällön yhteydessä sekä antamalla yleisölle mahdollisuuden tehdä omia eksplisiittisiä valintoja kiinnostavista aiheista. Yleisölle voi tarjota myös mahdollisuuden tarkastella sisältöjä ilman personointia ja suosituksia (esim. Diakopoulos ja Koliska 2017, 12).<sup>3</sup> Personointi ja kohdentaminen sekä niiden tavoitteet ovat tästä keskeisiä esimerkkejä, mutta yleisön toiminnasta kerätty data vaikuttaa myös muihin prosesseihin, erityisesti journalistisen sisällön suunnitteluun esimerkiksi kiinnostuksen perusteella. Tällaisen vaikutuksen tekeminen läpinäkyväksi voisi toteutua avaamalla esimerkiksi aihevalinnan perusteita ja suhdetta yleisöä koskevaan dataan itse jutussa.

Näiden näkökohtien valossa voimme tarkastella Julkisen sanan neuvoston (JSN) syksyllä 2019 julkaisemaa lausumaa *uutisautomaatiikan ja personoinnin merkitsemisestä*.<sup>4</sup> Lausuma on kansainvälisesti ensimmäinen journalistisen itsesääntelyelimen tuottama kannanotto automaation ja datan käytön läpinäkyvyyttä koskeviin käytäntöihin (Grundström, Haapanen ja Ilkka 2019). JSN:n lausuman keskeinen lähtökohta on periaatteellinen. Lausumassa kuvataan sen tarkoituksen olevan ”määrittellä algoritmisten apuvälineiden käyttö osaksi journalistista työtä ja antaa yleisölle varmuus siitä, että tiedotusvälineet toimivat algoritmeja käyttäessään avoimesti ja vastuullisesti”. Automaattinen sisällöntuotanto sekä sisällön kohdentaminen tai personointi pyritään lausumalla tuomaan journalistisen päätöksenteon sekä sitä koskevan vastuun ja itsesääntelyn piiriin. JSN:n lausumasta on eriteltävissä kolme konkreettista suositusta, joiden tarkoitus on lisätä läpinäkyvyyttä tiedotusvälineiden toiminnasta. Jos journalistinen sisältö on ”olennaiselta osin tuotettu ja julkaistu automaattisesti”, tämä tulee kertoa yleisölle. Toiseksi JSN suosittelee julkaistujen tietojen lähteen kertomista. Kolmanneksi JSN toteaa, että sisällön personoinnista tai kohdentamisesta sekä siihen käytettävän datan hankkimisesta on kerrottava yleisölle avoimesti.

JSN:n esittämät kolme suositusta heijastelevat edellä eriteltyjä kolmea läpinäkyvyyskäytäntöä (Taulukko 3). Ensinnäkin automaattisesta sisällöntuotannosta kertominen on läpinäkyvyyttä *tekijöistä*. Toiseksi tietojen tai datan lähteiden kertominen on läpinäkyvyyttä *datan prosessoinnista*. Kolmanneksi sisällön personointia koskeva suositus pyrkii lisäämään läpinäkyvyyttä yleisön toiminnan tuottaman datan keräämisestä ja sen käytöstä. Tiedon tai datan lähteen merkitseminen tai siihen linkittäminen ei kuitenkaan vielä kerro, miksi tietolähde on valittu ja mitä muita dataa ja sen prosessointia koskevia valintoja on tehty. Samoin sisällön personoinnin tai automaattisen tuotannon mainitseminen esimerkiksi uutissivustolla ei kerro sitä, millä perusteella tiettyä sisältöä on suositeltu, mihin kaikkeen personoinnilla vaikutetaan ja mihin sillä pyritään. Pelkästään näiden suositusten noudattaminen muuttuu helposti tekniseksi suoritukseksi, joka ei sellaisenaan lisää tietoa ja ymmärrystä datan ja



automaation käytöstä. Kuten lausuman tekstissä todetaankin, sen sisältämät suositukset ovat vasta ensimmäinen avaus uusien läpinäkyvyyskäytäntöjen vakiinnuttamiseksi.

## Johtopäätökset

Journalismista käydyssä akateemisessa ja julkisessa keskustelussa läpinäkyvyyttä on esitetty keinona ennen kaikkea luottamuksen ylläpitämiseen ja kasvattamiseen, joskin epäilyjä on esitetty siitä, missä määrin se todella pystyy näiden päämäärien saavuttamisessa auttamaan. Datan ja sen käytön läpinäkyvyyden lisäämiselle on esitetty muitakin hyviä perusteita. Se auttaa ihmisiä tekemään informoituja päätöksiä (Gorwa ja Garton Ash 2020) sekä ylläpitämään ja lisäämään vastuun toteutumista ja vastuullisuutta (Lindstedt ja Naurin 2010). Yhteiskunnallisen päätöksenteon ja tiedontuotannon pohjautuessa entistä enemmän algoritmeihin myös datan ja automaation käytön valaiseminen voi muuttua keskeiseksi osaksi sitä tehtävää, joka journalismilla maailman kuvaajana ja selittäjänä on. Edelleen mikäli journalismi kykenee valottamaan hyödyllisellä tavalla omaa datankäyttöään, se kertoo samalla datafioituneesta maailmasta – siitä, miten datan käyttö ja algoritmit tällä hetkellä muovaavat ja ohjaavat maailmaa.

Näitä tarkoituksia palvellakseen läpinäkyvyyden toteuttamisen tapojen on kuitenkin lisättävä todellista ymmärrystä ja tarjottava aitoa informaatiota (Ananny ja Crawford 2018). Pelkkien teknisten ja rutiininomaisten merkintöjen toteuttaminen ei vielä ole todellista läpinäkyvyyttä. Esimerkiksi automatisoidun uutistuotannon lähdekoodien tai sen käyttämien datajoukkojen esitleminen tai ”näyttäminen” yleisölle ei riitä läpinäkyvyyden toteuttamiseksi. Ymmärryksen kannalta on myös olennaista, että data ja algoritmit ymmärretään sosioteknisen kokonaisuuden osina. Journalistisen läpinäkyvyyden sisältöinä on usein toimituksen sisäiset tekijät, kuten toimittajat ja näiden suunnittelemat ja toteuttamat prosessit. Dataa koskevan läpinäkyvyyden kohdalla on kuitenkin syytä tuoda esiin ulkoisetkin vaikutukset: datan käytön algoritminen konteksti, joka ohjaa datan ja algoritmien suhteen tehtyjä valintoja. Näistä syistä sekä läpinäkyvyyden tarve että toteuttamisen tapa on arvioitava tapauskohtaisesti. Tehtävä vaikuttaa vaativalta, mutta journalistit tiedonvälityksen ammattilaisina ovat lähtökohtaisesti hyvässä asemassa tapauskohtaisesti harkitsemaan, pohtimaan ja toteuttamaan ymmärryksen ja tiedon lisäämistä juuri omalle yleisölleen.

Läpinäkyvyyden tavoitteena on ymmärrys ja tietämys. Edes aidosti informaatiota tarjoava ja ymmärrystä lisäävä läpinäkyvyys ei kasvata vastuullisuutta, ellei tarkoitettulla yleisöllä ole tapaa puuttua ongelmalliseksi katsomaansa toimintaan (Lindstedt ja Naurin 2010). Tämä on totta myös datan ja algoritmien kohdalla. Datan käytön tunteminen ei riitä, ellei yleisöillä ole tapaa vaikuttaa siihen, millä tavoin ja millä tarkoituksella dataa kerätään ja prosessoidaan. Tilanne on tuttu esimerkiksi sosiaalisen median suurista palveluista, joiden algoritmeihin vaikuttaminen yleisön palautteen tai sääntelyn kautta näyttää osoittautuneen vaikeaksi tai jopa mahdottomaksi.

Yksi journalismin tyypillisistä läpinäkyvyyskäytännöistä on yleisön osallistaminen. Dataan perustuvat digitaaliset journalistiset sisällöt tarjoavat monia uusia tapoja, joilla yleisö kykenee vaikuttamaan sisältöihin interaktiivisesti sekä vaikuttamaan algoritmiseen tapaan, jolla sisältöjä kohdennetaan ja suositellaan. Tästä ei kuitenkaan vielä seuraa, että yleisön

valinnoilla on vaikutusta itse päätöksentekoon – tapahtuu se sitten automatisoidusti tai ihmisen tekemänä yksittäisenä ratkaisuna. Yleisön toiminnasta kerättävän datan vaikutus journalistisiin ratkaisuihin puolestaan tapahtuu lähinnä yleisölle läpinäkyvättömällä tavalla esimerkiksi silloin, kun datan avulla mitattu yleisön kiinnostus määrittää tulevia aiheita ja painotuksia. Datan ja algoritmien käyttöä koskevan informaation ja ymmärryksen kautta journalistinen media voisi tarjota yleisölle aitoja tilaisuuksia vaikuttaa niihin ratkaisuihin, joita media tekee. Journalistit ja viime kädessä päätoimittajat ovat vastuussa lopullisista julkaisua koskevista päätöksistä. Tutkimuksen perusteella algoritmien ja automaation tulo journalismiin ei ole tätä näkemystä horjuttanut (Rydenfelt 2021a). Läpinäkyvyys ja sen mahdollistama dialogi saattaisivat kuitenkin tuoda yleisön odotukset ja näkemykset kuuluviin ja entistä keskeisemmäksi osaksi myös journalistista sisältöä.

Viimeisten vuosien tutkimuksessa ja julkisessa keskustelussa läpinäkyvyyttä on ehdotettu vastaukseksi moniin datafikaation tuottamiin haasteisiin ja ongelmiin sen toteuttamiseen liittyvistä haasteista huolimatta. Ajattelutapa voi houkutelaa dataa hyödyntäviä toimijoita, joiden on mahdollista pyrkiä läpinäkyvyyttä edistämällä välttämään kannaltaan vähemmän edullisia ratkaisuja. Keskittyminen läpinäkyvyyteen datan käyttöä ja hyödyntämistä koskevassa arvioinnissa voi johtaa kaikkien datan käytön ongelmien mieltämiseen läpinäkyvyyden ongelmiksi, joiden ratkaisemiseen pääasiallinen keino on lisätä läpinäkyvyyttä. Tällä tavoin läpinäkyvyydestä – kuten mistä tahansa vastaavasta käsitteestä – voi tulla myös ongelmallisten datafikaatiokäytänteiden mahdollistajia.<sup>5</sup>

## Viitteet

- 1 Esimerkiksi kahdessa kattavassa journalistisia eettisiä koodistoja sisältävässä tietokannassa (EthicNet 2018; Accountable Journalism 2019) ei hakusanoilla "automat\*", "algorithm\*", "robot\*" and "personali\*" tule lainkaan osumia (haettu 8.10.2020).
- 2 Tällaisesta jutusta esimerkki on Helsingin Sanomien 17.9.2019 julkaisema verolaskuri: <https://www.hs.fi/politiikka/art-2000006242236.html>.
- 3 Tällaisesta toimii esimerkkinä Helsingin Sanomien keväällä 2021 lanseeraama personoitu etusivu: <https://www.hs.fi/kotimaa/art-2000007775964.html>.
- 4 Lausuma on luettavissa osoitteessa: <https://www.jsn.fi/lausumat/lausuma-utisautomaatiikan-ja-personoinnin-merkitsemisesta-2019/>
- 5 Kiitämme nimettömiä arvioijia hyödyllisestä palautteesta, Kaarina Nikusta teemanumeron toimitamisesta sekä Helsingin Sanomain säätiötä ja C. V. Åkerlundin mediasäätiötä tutkimuksemme tukemisesta.

## Kirjallisuus

- Accountable Journalism. 2019. Monitoring Media Ethics across the Globe. Luettu 8.10.2020. <https://accountablejournalism.org/>.
- Allen, David S. 2008. "The trouble with transparency: The challenge of doing journalism ethics in a surveillance society." *Journalism Studies* 9 (3): 323–40. <https://doi.org/10.1080/14616700801997224>.
- Ananny, Mike ja Kate Crawford. 2018. "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability." *New Media & Society* 20 (3): 973–89. <https://doi.org/10.1177/1461444816676645>.
- Ananny, Mike. 2016. "Toward an ethics of algorithms: Convening, observation, probability, and timeliness." *Science Technology and Human Values* 41 (1): 93–117. <https://doi.org/10.1177/0162243915606523>.

- Anderson, Christopher W. 2011a. "Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms." *Journalism* 12 (5): 550–66. <https://doi.org/10.1177/1464884911402451>.
- Anderson, Christopher W. 2011b. "Deliberative, agonistic, and algorithmic audiences: Journalism's vision of its public in an age of audience transparency." *International Journal of Communication* 5: 529–47. <https://ijoc.org/index.php/ijoc/article/view/884>.
- Andrejevic, Mark ja Kelly Gates. 2014. "Big Data Surveillance: Introduction." *Surveillance and Society* 12 (2): 185–96. [http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/bds\\_ed](http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/bds_ed).
- Baume, Sandrine ja Yannis Papadopoulos. 2015. "Transparency: from Bentham's inventory of virtuous effects to contemporary evidence-based scepticism." *Critical Review of International Social and Political Philosophy* 212: 169–92. <https://doi.org/10.1080/13698230.2015.1092656>.
- Bentham, Jeremy. 1843. "An Essay on Political Tactics, or Inquiries Concerning the Discipline and Mode of Proceeding Proper to be Observed in Political Assemblies: Principally Applied to the Practice of the British Parliament, and to the Constitution and Situation of the National Assembly of France." Teoksessa *The Works of Jeremy Bentham*. 11 vols. Vol. 2, toimittanut John Bowring. Edinburgh: William Tait. Luettu 11.12.2020. <https://oll.libertyfund.org/title/bowring-the-works-of-jeremy-bentham-vol-2>.
- Berger, Stefan ja Dimitrij Owetschkin, toim. 2019. *Contested transparencies, social movements and the public sphere*. Palgrave Studies in the History of Social Movements. London: Palgrave Macmillan.
- Berry, David. 2011. "The computational turn: Thinking about the digital humanities." *Culture Machine* 12: 1–22. <https://cultremachine.net/wp-content/uploads/2019/01/10-Computational-Turn-440-893-1-PB.pdf>.
- Caplan, Robyn ja danah boyd. 2018. "Isomorphism through algorithms: Institutional dependencies in the case of Facebook." *Big Data & Society* 5 (2): 1–12. <https://doi.org/10.1177/2053951718757253>.
- Caswell, David ja Konstantin Dörr. 2017. "Automated Journalism 2.0: Event-driven narratives." *Journalism Practice* 12 (4): 477–96. <https://doi.org/10.1080/17512786.2017.1320773>
- Chen, Gina M., Alex Curry ja Kelsey N. Whipple. 2019. "Building Trust: What Works for News Organizations." Center for Media Engagement. Luettu 8.10.2020. <https://mediaengagement.org/research/building-trust/>.
- Cherubini, Federica ja Rasmus K. Nielsen. 2016. "Editorial Analytics: How News Media are Developing and Using Audience Data and Metrics." Digital News Project 2016. <https://dx.doi.org/10.2139/ssrn.2739328>.
- Christin, Angèle. 2020. "What data can do: A typology of mechanisms." *International Journal of Communication* 14: 1115–34. <https://ijoc.org/index.php/ijoc/article/view/12220>.
- Coll, Sami. 2014. "Power, knowledge, and the subjects of privacy: Understanding privacy as the ally of surveillance." *Information, Communication and Society* 17 (10): 1250–63. <https://doi.org/10.1080/1369118X.2014.918636>.
- Curry, Alexander L. ja Natalie Jomini Stroud. 2019. "The effects of journalistic transparency on credibility assessments and engagement intentions." *Journalism* 22 (4): 901–18. <https://doi.org/10.1177/1464884919850387>.
- Deuze, Mark. 2005. "What is journalism? Professional identity and ideology of journalists reconsidered." *Journalism* 6 (4): 442–64. <https://doi.org/10.1177/1464884905056815>.
- Diakopoulos, Nicholas 2016. "Accountability in algorithmic decision making." *Communications of the ACM* 59 (2): 56–62. <https://doi.org/10.1145/2844110>.
- Diakopoulos, Nicholas. 2019. *Automating the News. How Algorithms Are Rewriting the Media*. Cambridge: Harvard University Press.
- Diakopoulos, Nicholas ja Michael Koliska. 2017. "Algorithmic transparency in the news media." *Digital Journalism* 5 (7): 809–28. <http://dx.doi.org/10.1080/21670811.2016.1208053>.
- Dimaggio, Paul J ja Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48 (2): 147–60. <https://doi.org/10.2307/2095101>.
- Durham, Meenakshi. 1998. "On the relevance of standpoint epistemology to the practice of journalism: The case for 'strong objectivity.'" *Communication Theory* 8 (2): 117–40. <https://doi.org/10.1111/j.1468-2885.1998.tb00213.x>.
- Dörr, Konstantin N. ja Katharina Hollnbuchner. 2016. "Ethical challenges of algorithmic journalism." *Digital Journalism* 5 (4): 404–19. <https://doi.org/10.1080/21670811.2016.1167612>.
- Eberwein, Tobias, Susanne Fengler ja Matthias Karmasin. 2019. *Media Accountability In The Era Of Posttruth Politics: European Challenges and Perspectives* 1st ed. London: Routledge. <https://doi.org/10.4324/9781351115780>.
- Edelman. 2019. "Edelman Trust barometer: Global report." Luettu 8.10.2020. [https://www.edelman.com/sites/g/files/aatusi91/files/2019-02/2019\\_Edelman\\_Trust\\_Barometer\\_Global\\_Report\\_2.pdf](https://www.edelman.com/sites/g/files/aatusi91/files/2019-02/2019_Edelman_Trust_Barometer_Global_Report_2.pdf).

- Erkkilä, Tero. 2012. *Government Transparency: Impacts and Unintended Consequences*. Palgrave Macmillan.
- Eslami, Motahhare, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton ja Alex Kirlik. 2016. "First I "like" it, then I hide it: Folk theories of social feeds." *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2371–82. <https://doi.org/10.1145/2858036.2858494>.
- EthicNet 2018. Collection of Codes of Journalism Ethics in Europe. Luettu 8.10.2020. <https://research.uta.fi/ethicnet>.
- Felzmann, Heike, Eduard Fosch Villaronga, Christoph Lutz ja Aurelia Tamò-Larrieux. 2019. "Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns." *Big Data @ Society* 6 (1): 1–14. <https://doi.org/10.1177/2053951719860542>.
- Fengler, Susanne, Tobias Eberwein, Gianpetro Mazzoleni ja Colin Porlezza, toim. 2014. *Journalists and Media Accountability. An International Study of News People in the Digital Age*. Bern: Peter Lang.
- Fengler, Susanne ja Dominik Speck. 2019. "Journalism and transparency: A mass communications perspective." Teoksessa *Contested Transparencies, Social Movements and the Public Sphere. Palgrave Studies in the History of Social Movements*, toimittaneet Stefan Berger ja Dimitrij Owetschkin, 119–49. London: Palgrave Macmillan.
- Flyverbom, Mikkel, Ronald Deibert ja Dirk Matten. 2019. "The Governance of Digital Technology, Big Data, and the Internet: New Roles and Responsibilities for Business." *Business @ Society* 58 (1): 3–19. <https://doi.org/10.1177/0007650317727540>.
- Fourcade, Marion ja Kieran Healy. 2017. "Seeing like a market." *Socio-Economic Review* 15 (1): 9–29. <https://doi.org/10.1093/ser/mww033>.
- Gauthier, Gilles. 1993. "In defence of a supposedly outdated notion: The range of application of journalistic objectivity." *Canadian Journal of Communication* 18:4: 497–505. <https://doi.org/10.22230/cjc.1993v18n4a778>.
- Gillespie, Tarleton. 2016. "Algorithm." Teoksessa *Digital Keywords: A Vocabulary of Information Society and Culture*, toimittanut Benjamin Peters, 18–30. New Jersey: Princeton University Press.
- Gitelman, Lisa ja Virginia Jackson. 2013. "Introduction." Teoksessa *"Raw data" is an oxymoron*, toimittanut Lisa Gitelman, 1–14. Cambridge, MA: MIT Press.
- Gorwa, Robert ja Timothy Garton Ash. 2020. "Democratic Transparency in the Platform Society." Teoksessa *Social Media and Democracy: The State of the Field*, toimittaneet Nate Persily ja Josh Tucker, 286–312. Cambridge, UK: Cambridge University Press. <https://doi.org/10.31235/osf.io/ehcy2>.
- Gray, Colin, Yubo Kou, Bryan Battles, Joseph Hoggatt ja Austin Toombs. 2018. The dark patterns side of UX design. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* 534: 1–14. <https://doi.org/10.1145/3173574.3174108>.
- Groenhardt, Harmen ja Jo L. H. Bardoel. 2012. "Conceiving the transparency of journalism: Towards a new media accountability currency." *Studies in Communication Sciences* 12 (1): 6–11. <https://doi.org/10.1016/j.scoms.2012.06.003>.
- Grundström, Elina, Lauri Haapanen ja Sakari Ilkka. 2019. "JSN määritteli algoritmit osaksi journalistista työtä." *Media & viestintä* 42 (4): 253–59. <https://journal.fi/mediaviestinta/article/view/88457>.
- Haapanen, Lauri. 2020a. "Problematising the restoration of trust through transparency: focusing on quoting." *Journalism* (June 2020). <https://doi.org/10.1177/1464884920934236>.
- Haapanen, Lauri. 2020b. *Media councils and self-regulation in the emerging era of news automation*. Helsinki: Council for Mass Media in Finland. Luettu 16.5.2021. <https://presscouncils.eu/userfiles/files/Emerging%20era%20of%20news%20automation%20Haapanen.pdf>.
- Haapanen, Lauri ja Leo Leppänen. 2020. "Recycling a genre for news automation: The production of Valtteri the Election Bot." *AILA Review* 33: 67–85. <https://doi.org/10.1075/aila.00030.haa>.
- Haapoja, Jesse ja Airi Lampinen. 2018. "'Datafied' reading: Framing behavioral data and algorithmic news recommendations." *NordiCHI '18: Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, 125–36. <https://doi.org/10.1145/3240167.3240194>.
- Heikkilä, Heikki, David Domingo, Judith Pies, Michal Glowacki, Michal Kus ja Olivier Baisnée. 2012. "Media accountability goes online: A transnational study on emerging practices and innovations." *MediaAct Working Paper 14/2012*. Journalism Research and Development Centre, University of Tampere, Finland.
- Hintz, Arne, Lina Dencik ja Karin Wahl-Jørgensen. 2019. *Digital Citizenship in a Datafied Society*. Cambridge: Polity Press.
- Karlsson, Michael. 2010. "Rituals of transparency: Evaluating online news outlets' uses of transparency rituals in the United States, United Kingdom and Sweden." *Journalism Studies* 11 (4): 535–45. <https://doi.org/10.1080/14616701003638400>.
- Karlsson, Michael. 2011. "The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority." *Journalism* 12 (3): 279–295. <https://doi.org/10.1177/1464884910388223>.

- Karlsson, Michael, Christer Clerwall ja Lars Nord. 2014. "You ain't seen nothing yet: Transparency's lack of effect on source and message credibility." *Journalism Studies* 15 (5): 668–78. <https://doi.org/10.1080/1461670X.2014.886837>.
- Karlsson, Michael, Christer Clerwall ja Lars Nord. 2017. "Do not stand corrected: Transparency and users' attitudes to inaccurate news and corrections in online journalism." *Journalism and Mass Communication Quarterly* 94 (1): 148–67. <https://doi.org/10.1177/1077699016654680>.
- Kennedy, Helen. 2018. "Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication." *Krisis* 1: 18–30. <https://archive.krisis.eu/living-with-data/>.
- Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.
- Koliska, Michael. 2015. *Transparency and trust in journalism: An examination of values, practices and effects*. Väitöskirja, University of Maryland, College Park, USA. <https://drum.lib.umd.edu/handle/1903/17031>.
- Kovach, Bill ja Tom Rosenstiel. 2001. *The elements of journalism: what newspeople should know and the public should expect*. New York: Crown Publishers.
- Lamot, Kenza ja Peter Van Aelst. 2019. "Beaten by Chartbeat? An Experimental Study on the Effect of Real-Time Audience Analytics on Journalists' News Judgment." *Journalism Studies* 21 (4): 477–93. <https://doi.org/10.1080/1461670X.2019.1686411>.
- Lehtiniemi, Tuukka. 2020. *Imagining the Data Economy*. Väitöskirja, Turun yliopisto. <http://urn.fi/URN:ISBN:978-951-29-8002-4>.
- Lepri, Bruno, Nuria Oliver, Emmanuel Letouzé, Alex Pentland ja Patrick Vinck. 2018. "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes: The Premise, the Proposed Solutions, and the Open Challenges." *Philosophy and Technology* 31 (4): 611–27. <https://doi.org/10.1007/s13347-017-0279-x>.
- Lindén, Carl-Gustav, Hanna Tuulonen ja työryhmä (Asta Bäck, Nicholas Diakopoulos, Lauri Haapanen, Leo Leppänen, Magnus Melin, Myriam Munezero, Stefanie Sirén-Heikel, Caj Södergård ja Hannu Toivonen). 2019. *News Automation: The Rewards, Risks and Realities of 'Machine Journalism'*. WAN-IFRA guide to the field. Reports / The World Association of Newspapers and News Publishers WAN-IFRA.
- Lindstedt, Catharina ja Daniel Naurin. 2010. "Transparency is not enough: Making transparency effective in reducing corruption." *International Political Science Review* 31 (3), 301–22. <https://doi.org/10.1177/019251210377602>.
- McBride, Kelly ja Tom Rosenstiel. 2014. *The New Ethics of Journalism: Principles for the 21st Century*. Thousand Oaks, CA: CQPress.
- McStay, Andrew. 2014. *Privacy and Philosophy: New Media and Affective Protocol*. New York, NY: Peter Lang.
- Manninen, Ville J. E. 2020. "If only they knew: Audience expectations and actual sourcing practices in online journalism." *Journalism Practice* 14 (4): 412–28. <https://doi.org/10.1080/17512786.2019.1625719>.
- Meier, Klaus ja Julius Reimer. 2011. "Transparenz im Journalismus: Instrumente, Konfliktpotentiale, Wirkung." *Publizistik* 56 (2): 133–55. <https://doi.org/10.1007/s11616-011-0116-7>.
- Micheli, Marina, Marisa Ponti, Max Craglia ja Anna Berti Suman. 2020. "Emerging Models of Data Governance in the Age of Datafication." *Big Data @ Society* 7 (2): 1–15. <https://doi.org/10.1177/2053951720948087>.
- O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.
- Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge: Harvard University Press.
- Philips, Angela. 2010. "Transparency and the new ethics of journalism." *Journalism Practice* 4 (3): 373–82. <https://doi.org/10.1080/17512781003642972>.
- Prainsack, Barbara. 2019. "Logged out: Ownership, Exclusion and Public Value in the Digital Data and Information Commons." *Big Data @ Society* 6 (1): 1–15. <https://doi.org/10.1177/2053951719829773>.
- Roberts, Mark C. 2007. *Measuring the Relationship between Journalistic Transparency and Credibility*. South Carolina, SC: University of South Carolina Press.
- Rydenfelt, Henrik. 2016. "Blogimediät ja bloggaamisen etiikka." *Media @ viestintä* 39 (3): 209–24. <https://journal.fi/mediaviestinta/article/view/61421>.
- Rydenfelt, Henrik. 2019a. "Pragmatism, Social Inquiry and the Method of Democracy". Teoksessa *Practices of Citizenship in East Africa. Perspectives from Philosophical Pragmatism*, toimittaneet Katariina Holma ja Tiina Kontinen, 29–43. Lontoo: Routledge.
- Rydenfelt, Henrik. 2019b. "Democracy and Moral Inquiry: Problems of the Methodological Argument." *Transactions of the Charles S. Peirce Society* 55 (2): 254–72.
- Rydenfelt, Henrik. 2021a. "Transforming Media Agency? Approaches to Automation in Finnish Legacy Media." *New Media @ Society*. Tulossa.

- Rydenfelt, Henrik. 2021b. "Recent Problems of the Public." Teoksessa *Pragmatism and Social Philosophy*, toimittanut Michael G. Festl, 237–47. Lontoo: Routledge.
- Rydenfelt, Henrik ja Juholin, Elisa. 2016. "Hyvä viestintä, parempi viestintäkulttuuri." *Procomma Academic* 2016: 12–27.
- Seaver, Nick. 2019. "Knowing Algorithms." Teoksessa *Digital STS. A Field Guide for Science @ Technology Studies*, toimittaneet Janet Vertesi ja David Ribes, 412–22. Princeton: Princeton University Press.
- Sirén-Heikel, Stefanie, Leo Leppänen, Carl-Gustav Lindén ja Asta Bäck. 2019. "Unboxing news automation: Exploring imagined affordances of automation in news journalism." *Nordic Journal of Media Studies* 1 (1): 47–66. <https://doi.org/10.2478/njms-2019-0004>.
- Smolkin, Rachel. 2006. "Too transparent?" *American Journalism Review* 28 (2): 17–23.
- Srnicek, Nick. 2017. *Platform Capitalism*. Cambridge, UK: Polity Press.
- Tandoc, Edson C. Jr. ja Ryan J. Thomas. 2017. "Readers value objectivity over transparency." *Newspaper Research Journal* 38 (1): 32–45. <https://doi.org/10.1177/0739532917698446>.
- van Dijck, José. 2014. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance @ Society* 12 (2): 197–208. <http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/datafication>.
- Weinberger, David. 2009. "Transparency is the New Objectivity." Luettu 12.10.2020. <https://www.hyperorg.com/blogger/2009/07/19/transparency-is-the-new-objectivity>.
- West, Sarah Myers. 2019. "Data Capitalism: Redefining the Logics of Surveillance and Privacy." *Business @ Society* 58 (1): 20–41. <https://doi.org/10.1177/0007650317718185>.
- Vos, Tim P. ja Stephanie Craft. 2017. "The discursive construction of journalistic transparency." *Journalism Studies* 18 (12): 1505–22. <https://doi.org/10.1080/1461670X.2015.1135754>.
- Yeung, Karen. 2016. "'Hypernudge': Big Data as a Mode of Regulation by Design." *Information, Communication @ Society* 20 (1): 118–36. <https://doi.org/10.1080/1369118X.2016.1186713>.
- Zamith, Rodrigo. 2019. "Transparency, interactivity, diversity, and information provenance in everyday data journalism." *Digital Journalism* 7 (4): 470–89. <https://doi.org/10.1080/21670811.2018.1554409>.
- Zarsky, Tal. 2016. "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making." *Science Technology and Human Values* 41 (1): 118–32. <https://doi.org/10.1177/0162243915605575>.
- Zuboff, Shoshana. 1988. *In the Age of the Smart Machine: The future of Work and Power*. New York, NY: Basic Books.
- Zuboff, Shoshana. 2015. "Big other: Surveillance capitalism and the prospects of an information civilization." *Journal of Information Technology* 30: 75–89. <https://doi.org/10.1057/jit.2015.5>.