

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Knuutila, Aleksj; Järvenpää, Riia

Title: Sosiaalisen median moderointi ei ratkaise yhteiskunnallisia ongelmia

Year: 2022

Version: Published version

Copyright: © 2022 Valtiotieteellinen yhdistys

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Knuutila, A., & Järvenpää, R. (2022, 16.2.2022). Sosiaalisen median moderointi ei ratkaise yhteiskunnallisia ongelmia. *Politiikasta*. <https://politiikasta.fi/sosiaalisen-median-moderointi-ei-ratkaise-yhteiskunnallisia-ongelmia/>

SOSIAALISEN MEDIAN MODEROINTI EI RATKAISE YHTEISKUNNALLISIA ONGELMIA

16.2.2022 [Politiikasta lehti](#) | [Artikkelit](#) | [Aleksi Knuutila, Riia Järvenpää](#)



Creative christians / Unsplash

Tehokkaampi misinformaation ja muun vahingollisen sisällön poisto sosiaalisesta mediasta nähdään nykyään ratkaisuna moneen yhteiskunnalliseen ongelmaan. Onko moderaatioon nojaava politiikka kuitenkin alustajättien eduksi, heikompien ryhmien kustannuksella?

Facebookin entinen työntekijä, yrityksen sisäisiä asiakirjoja julkisuuteen luovuttanut **Frances Haugen** oli [Britannian parlamentin kuultavana](#) 25. lokakuuta 2021. Haugenin vuotamista tiedoista käy ilmi, että yritys on tietoinen vahingollisen tai lainvastaisen sisällön leviämisestä alustallaan, eikä ole puuttunut siihen. Haugen mainitsee esimerkkeinä vahingollisesta sisällöstä väkivallan lietsomisen Rohingya-vähemmistöä vastaan Myanmarissa sekä rokotteita koskevat harhaanjohtavat sisällöt.

Todistusaineiston lisäksi Haugenilla on ehdotuksia siitä, miten somejättien tarkempi sääntely voisi ratkaista ongelmia. Tiedon näkyvyyttä säätelevien algoritmien rukkaamisen

lisäksi Haugen esittää, että **alustan tulisi ottaa vastuu siellä julkaistavista sisällöistä** ja estää ja poistaa haitalliset sisällöt nykyistä tehokkaammin.

Tilin sulkeminen tapahtui ilman varoitusta ja ilman selkeästi annettua syytä. Päätöksestä oli käytännössä mahdoton valittaa.

Seuraavana päivän Googlen omistama YouTube-videopalvelu **sulki ilman varoitusta** Novara Media -nimisen, pienen vasemmistolaisen toimituksen YouTube-kanavan. Novara Media on verkkoon keskittyvä pieni joukko toimittajia, jotka kuitenkin lukeutuvat Englannin merkittävimpiin sosiaalidemokratian äänenkannattajiin.

Ilman näkyvyyttä verkossa Novaran toimittajat menettäisivät pian työnsä. Novaran kanavan sulkeminen noudatti samaa kaavaa kuin monet muut YouTube-moderaation kohteeksi joutuneet: päätös tapahtui ilman varoitusta ja ilman selkeästi annettua syytä. Päätöksestä oli käytännössä mahdoton valittaa.

Novara sai pian kanavansa takaisin, mutta vain luomalla alustaa kohtaan painetta omien katsojiensa kautta. **YouTube ilmoitti tehneensä virheen.** Kaikilla ei ole samanlaista kykyä vastata heihin kohdistuviin moderointipäätöksiin.

Sisältöjen valvonta nähdään ratkaisuksi yhteiskunnan ongelmiin

Yhä useammin koetaan, että sosiaalisessa mediassa tartuntataudin tavoin leviävän **terveysdisinformaation** ja **väkivaltaisten ääriilikkeiden** tukahduttaminen tai jopa **“yhteisen todellisuuden pelastaminen”** vaatii yhä tehokkaampaa sosiaalisen median sisältöjen valvontaa.

Tällainen tapa ajatella yhteiskunnallisia ongelmia tekee niistä helposti hallittavia, jopa teknisiä, etenkin kun alustayritykset lupaavat entistä tarkempia tekoälyratkaisuja sisällön seulomiseen. Kirjailija **Evgeny Morozov** on kutsunut tällaista ongelmien yksinkertaistamista **“ratkaisuismiksi”** (*engl. solutionism*), jossa monimutkaiset yhteiskunnalliset tilanteet nähdään läpinäkyvinä ja optimoitavina prosesseina.

Liian yksinkertaistava tapa ajatella yhteiskunnallisia ongelmia tekee niistä helposti hallittavia, jopa teknisiä, etenkin kun alustayritykset lupaavat entistä tarkempia tekoälyratkaisuja sisällön seulomiseen.

Kanavien sulkeminen tai sisällön poisto ei kuitenkaan aina saa aikaan toivottuja muutoksia. Yhdysvaltojen entinen presidentti **Donald Trump** sai porttikiellon Facebookiin ja Twitteriin kansallisen turvallisuuden vaarantamisen vuoksi, mutta hän **ylläpitää otettaan republikaanipuolueesta.** Salaliittoryhmittymä QAnon **on edelleen voimissaan,** vaikka somejätit ovat kertoneet siivonneensa sen alustoiltaan Senaattiin vuoden 2021 alussa tapahtuneen hyökkäyksen jälkeen.

Tiukkoja rajoituksia vaativien pitäisi ymmärtää somessa leviävien ilmiöiden taustalla olevia tekijöitä monipuolisemmin, sen sijaan että ajattelisivat niiden suosion – ja koko ongelman – johtuvan sosiaalisesta mediasta.

Alustojen matka neutraalista välittäjistä vastuunkantajiksi

Vuonna 2005 Twitterin yhteisösäännöissä ensimmäisissä lauseissa ilmoitettiin, että käyttäjät itse ovat vastuussa kirjoittamastaan eikä Twitter seuraa tai sensuroi sisältöä, ja vielä vuonna 2018 Facebookin omistaja **Mark Zuckerberg** kertoi, että **sananvapauden rajoittaminen ei kuulu yrityksen tehtäviin**. Alustayritykset katsoivat pitkään edukseen asemoida itsensä neutraaleiksi viestintäkanaviksi, jotka ottavat mahdollisimman vähän velvollisuuksia välittämästään sisällöstä. Oikeusjärjestys onkin useimmissa maissa **tarjonnut mediakanaville vastuuvapautta** (*liability privilege*), eli vaatinut niitä rajoittamaan viestintää vain yritysten tietäessä sen sisältävän jotain laitonta.

Viimeisen 10 vuoden aikana mediatilan säätely on kulkenut “**vastuuttamisen tietä**”, jossa vastuuvapautta on tulkittu kapeammin ja siihen on kytketty velvollisuuksia. Esimerkiksi saksalainen **NetzDG-laki** vaatii alustayrityksiltä entistä nopeampaa reagoitua: selvästi laitton vihapuhe vähemmistöjä vastaan on poistettava alustoilta päivässä. EU on koventanut painetta sisältöjen moderointiin, ja **uusiuu** **tiukennuksia** on todennäköisesti luvassa.

Sosiaalisen median alustojen voi olla kannattavinta poistaa mahdollisimman paljon sisältöjä varmuuden vuoksi.

Tahot, jotka uskovat sisältöjen entistä tiukemman valvonnan olevan ratkaisu ongelmiin, kannattavat usein myös sosiaalisen median alustojen vastuun laajentamista. Tämän vastuun kantaminen voi olla alustoille riski ja näkyä heidän palveluidensa tuottamisen kustannuksissa. Päätöksenteon ja rajoitusten toteuttamisen siirtäminen alustojen vastuulle voi kuitenkin myös lisätä näiden valtaa, ja niiden voi olla kannattavinta poistaa mahdollisimman paljon sisältöjä varmuuden vuoksi.

Lakieteilijä **Frank Pasquale** on kuvannut alustojen valtaa “**toiminnallisen suvereniteetin**” (*functional sovereignty*) käsitteen kautta. Pasqualen mukaan alustat usein ottavat itselleen sellaisia rooleja, jotka aikaisemmin ovat kuuluneet julkishallinnolle, esimerkiksi Amazonin luodessa vaatimuksia myyjille omille verkkomarkkinoilleen.

Yhtä lailla NetzDG:n kaltaiset lait siirtävät julkishallinnon asettamien kielletyn viestinnän sääntöjen täytäntöönpanon alustayrityksille. Yritykset eivät kuitenkaan tarjoa asiakkailleen samoja oikeuksia kuin julkishallinto tarjoaa esimerkiksi sen tekemien päätösten riitauttamiseen. Erityisen ongelmallista tämä on, jos alustojen harteille annetaan **päätäntävalta rajanvedosta misinformaation** kaltaisten, **vaikeasti määriteltävien** käsitteiden kohdalla.

Moderoinnin seuraukset: tahaton hiljentäminen

Sosiaalisen median valvonnan yksi päämäärä on haavoittuvaisten ryhmien olevien suojele, kuten esimerkiksi vähemmistöjen suojele häirinnältä. Viestinnän rajoittamisella voi siis olla selkeitä hyötyjä, mutta samalla se myös usein synnyttää haittoja. Moderoinnin yksi kääntöpuoli johtuu puhtaista virheistä: sisältöä voidaan poistaa silloinkin, kun se ei ole vahingollista tai ei riko yhteisösääntöjä. Tällaiset virheet ovat yleisiä etenkin silloin kun, valvontaa tehdään automatisoidusti ja suurella skaalalla.

Usein tällaisen tahattoman hiljentämisen kohteeksi joutuvat heikommassa asemassa olevat. Esimerkiksi Facebook on toistuvasti **poistanut palestiinalaisten näkökulmaa kuvaavaa tietoa**, joka liittyy heidän konfliktiinsa Israelin kanssa, esimerkiksi vuonna 2021 Gazan pommitusten aikana. Sisältöä seulovien järjestelmien kun on vaikea erottaa väkivallan seurauksien kuvaus sen ihannoinnista.

Virheet ovat yleisiä etenkin silloin kun, valvontaa tehdään automatisoidusti ja suurella skaalalla.

Myös esimerkiksi Twitterin vihapuhetta automaattisesti tunnistavat järjestelmät **poistavat helpommin afroamerikkalaisten kirjoittamia viestejä**. Heikommassa asemassa on myös vaikeampi yrittää korjata tilannetta ja puolustautua tällaisilta päätöksiltä.

Yritysten mukaan automaatio on ainoa tapa valvoa niissä liikkuvaa massiivista tietomäärää. Koneoppivat sovellukset, niin sanotut **tekoälyalgoritmit** hoitavat epäilyttävistä sisällöistä ilmoittamisen esimerkiksi YouTubessa lähes **sataprosenttisesti**. Alustat käyttävät ihmismoderoijia ja faktantarkastajia etenkin arvioimaan tekoälyn epäilyttäväksi liputtamia julkaisuja, ja tekoäly poistaa tai merkitsee nämä misinformaatioksi luokitellut sisällöt jatkossa.

Myös paljon puhuttu kieli, kuten hindi, voi silti jäädä alustojen kehityksessä paitsioon, koska yritykset eivät ole priorisoineet kieltä ymmärtävien ohjelmistojen kehittämiseen.

Jos tekoälyä ei ole koulutettu ymmärtämään alustalla käytettyä kieltä eikä alueen tilannetta tuntevia faktantarkastajia ole palkattu, ei se osaa puuttua oikeisiin ongelmiin. Esimerkiksi väkivaltaisten ääri liikkeiden ja valtioita ulkoa uhkaavien tahojen viestit eivät ole samanlaisia eri maissa.

Tekoälyn voi olla teknisesti vaikeampaa ymmärtää pienten kielikuntien edustajia kuten **esimerkiksi suomea**. Toisaalta myös paljon puhuttu kieli, **kuten hindi**, voi silti jäädä alustojen kehityksessä paitsioon, koska yritykset eivät ole priorisoineet kieltä ymmärtävien ohjelmistojen kehittämiseen.

Hallitukset hiljentämässä kriitikkojaan misinformaation varjolla

Viime vuosina on monissa maissa, niin demokraattisissa kuin autoritäärisissäkin, säädetty uusia lakeja, joiden tavoitteena on misinformaation torjuminen. Yksi tutkimus löysi **174 tällaista lakia**, joista suurin osa on syntynyt vuoden 2015 jälkeen. Esimerkiksi Venäjä säätöi vuonna 2020 uuden lain, joka **kriminalisoi väärän tiedon levittämisen** pandemian kaltaisissa hätätilanteissa.

Pahimmillaan kyse on siitä, että liberaaleissa demokratioissa käytettyjä puhetapoja ja argumentteja – taistelua misinformaatiota vastaan – käytetään sumuverhona omien poliittisten vastustajien hiljentämiseen. Esimerkiksi **Intian hallitus on poistanut** omaa politiikkaansa kritisoivia viestejä Twitteristä ja Facebookista, sillä verukkeella että misinformaatio on vaaraksi terveydelle. Turkissa **satoja ihmisiä pidätettiin vuonna 2020** ”virheellisistä ja provokatiivisista” koronaan liittyvistä postauksista.

Etenkin terveyteen liittyvä tieto on jatkuvan muutoksen kohteena ja auktoriteettkaan eivät välttämättä ole yhteisymmärryksessä, etenkin kansainvälisesti.

Epädemokraattiset hallinnot voivat siis viitata laajasti hyväksytyihin väitteisiin disinformaation vaaroista ja sosiaalisen median säätelyn välttämättömyydestä omien tarkoitusperiensä edistämiseksi.

Pandemia on nostanut misinformaation torjumisen agendalle, mutta myös näyttänyt, että kyse on epävakaasta kategoriasta. Pandemian alussa Facebook pyrki järjestelmällisesti poistamaan kaikki viestit, joissa väitettiin koronaviruksen syntyneen laboratorio-olosuhteissa. Vuoden 2021 toukokuussa se **luopui tästä säännöstä**, koska kyse ei ollutkaan ilmiselvästi väärästä väitteestä.

Vastaavasti alusta **kielsi kasvomaskeihin liittyvät mainokset** silloin kun niiden hyödyt olivat vielä kiistanalaisia. Etenkin terveyteen liittyvä tieto on jatkuvan muutoksen kohteena ja auktoriteettkaan eivät välttämättä ole yhteisymmärryksessä, etenkin kansainvälisesti.

Vaihtoehtoja sisällön poistolle

Amerikkalainen lakitieteilijä Frank Pasquale on kartoittanut erilaisia poliittisia vaihtoehtoja **alustojen vallan** suitsimiseksi jakamalla ne kahteen joukkoon, jeffersonilaisiin sekä hamiltonilaisiin Yhdysvaltojen ”perustajaisien” **Thomas Jeffersonin** ja **Alexander Hamiltonin**, ja näiden poliittisten näkemuserojen, mukaan. Näistä ensimmäisten mielestä ongelma on pohjimmiltaan alustojen koko, ja alustat tulisi pilkkoa pienemmiksi palveluiksi. Hamiltonilaiset puolestaan uskovat, että alustojen arvo syntyy nimenomaan niiden suurista käyttäjämääristä – ne ovat luonnollisia monopoleja, joiden toimintaa valtiovallan pitää tarkkaan säädellä.

Myös moderoinnin tulevaisuus saattaa seurata jompaa kumpaa näistä suuntaviivoista: joko sosiaalinen media sirpaloituu pienempiin tiloihin, jotka itse päättävät omista ohjenuoristaan. Tai sitten niin sanotut alustajätit säilyttävät asemansa, mutta julkinen valta määrittelee entistä tarkemmin missä raja sallitun ja kielletyn viestinnän välillä kulkee.

Oli moderointi järjestettynä millä tavalla tahansa, se todennäköisesti säilyy konfliktien aiheena ja synnyttää monenlaisia tahattomia vahinkoja. Internet-kulttuurin tutkijat **Whitney Phillips** ja **Ryan Milner** ovat kuvanneet, miten monia digitaalisen median piirteitä, kuten trollausta tai meemejä, ei voi nähdä yksinomaan positiivisina tai negatiivisina, vaan **pohjimmiltaan ristiriitaisina, ambivalentteina ilmiöinä.**

Joko sosiaalinen media sirpaloituu pienempiin tiloihin, jotka itse päättävät omista ohjenuoristaan. Tai sitten niin sanotut alustajätit säilyttävät asemansa, mutta julkinen valta määrittelee entistä tarkemmin missä raja sallitun ja kielletyn viestinnän välillä kulkee.

Moderointiin liittyy samanlaista ristiriitaisuutta: rajoitukset voivat olla perusteltuja tai jopa välttämättömiä. Silti meillä ei ole olemassa niitä instituutioita ja rakenteita, joiden avulla rajoitusten toteuttaminen tapahtuisi luotettavasti tai reilusti. Emme voi olettaa, että alustayritykset – yhtään sen enempää kuin valtiovaltakaan – olisivat neutraaleja tuomareita, jotka asettavat ja seuraavat rajoituksia yleisen edun mukaisesti.

Keskustelussa sisältörajoituksista myös siihen liittyvistä riskeistä ja mahdollisesta vahingosta tulisi puhua rehellisesti. Moderointi tulisi aina nähdä kompromissina, johon liittyy mahdollisia myönteisiä mutta myös kielteisiä vaikutuksia. Vahinkoihin kuuluvat usein heikommassa asemassa olevien ryhmien hiljentäminen sekä esimerkiksi misinformaation **torjunnan nimissä** tehty hallituksia arvostelevien äänen sensurointi. Tilanteessa, joissa alustoilta vaaditaan entistä hanakampaa väliintuloa viestintään, pitää myös tämän vallan väärinkäyttöä valvoa.

Moderointi tulisi aina nähdä kompromissina, johon liittyy mahdollisia myönteisiä mutta myös kielteisiä vaikutuksia.

Misinformaation ja vihapuheen kaltaisiin ongelmiin tulisi siis hakea **myös muita ratkaisuja** kuin suuryritysten ja tekoälyn käsiin annettu entistä tehokkaampi sisältöjen moderointi. Nykytilanteessa yhden yrityksen tekninen **ongelma voi lamaannuttaa** ison osan viestintäkenttää, ja toisaalta **Yhdysvalloissa suunniteltu** sisällön valvonta ei välttämättä torju esimerkiksi Suomeen kohdistuvia turvallisuusuhkia.

Sosiaalisen median sääntelyä pitäisi siis täydentää **tukemalla vaihtoehtoisten viestintäalustojen kasvua.** Jos mainostuloja ohjautuisi nykyistä enemmän paikallismedioille, mahdollistaisi tämä nykyistä paremmat resurssit suomalaiselle journalismille ja paikallisten verkkoyhteisöjen kehittämiseksi. Verkkokeskustelun laatua voi parantaa muutenkin kuin **pelkällä kepillä.**

Aleksi Knuutila on tutkijatohtori Jyväskylän yliopistossa ja tutkii poliittista viestintää sekä poliittisten yhteisöjen toimintaa verkkotilassa.

Riia Järvenpää on väitöskirjatutkija Helsingin yliopiston Social Data Centressä. Hän tutkii viestinnän vaikuttavuutta ja terveyteen liittyviä somedebatteja ja kuuluu vihreiden tietopolitiikka-työryhmään.