Tekoälykäs informaatio-vaikuttaminen

Jussi Toivanen, Jukka Niittymaa ja Vilma Luoma-aho

ProCom – Viestinnän ammattilaiset ry
ISBN 978-952-65488-6-9

Tekoälyteknologioiden nopea kehitys on herättänyt paljon keskustelua niiden vaikutuksista yhteiskuntiin, organisaatioiden toimintaan ja ihmisten arkeen. Tekoälykkään viestinnän aikakausi tuo myös mukanaan uudenlaisia informaatioympäristössä nousevia uhkia. Tässä artikkelissa keskitytään tarkastelemaan, miten tekoäly muuttaa informaatioympäristöämme ja siellä nousevia uhkia. Artikkelissa tarkastellaan myös pimeän suunnittelun ja informaatiovaikuttamisen uudenlaisia keinoja. Tarkoitus on auttaa tunnistamaan, miten tekoälyä voidaan hyödyntää vihamielisessä informaatiovaikuttamisessa. Informaatiovaikuttamisen keinot ovat muuttuneet entistä kohdennetummiksi, henkilökohtaisemmiksi ja sitä kautta tehokkaammiksi. Lopuksi perehdytään siihen, miten viestinnän ammattilaisten ja organisaatioiden tulisi varautua tekoälykkään viestinnän aikakauden uhkiin.

Informaatiovaikuttaminen, disinformaatio, misinformaatio, generatiivinen tekoäly, tekoäly, deepfake, dark design, vaikutusyritykset, tekoälykäs disinformaatio

Johdanto

Informaatiovaikuttaminen seuraa aina aikaansa. Esimerkiksi yhteiskunnalliset muutokset, maailmapoliittinen tilanne ja teknologinen kehitys vaikuttavat siihen, millaisia keinoja, tekniikoita ja teknologioita vaikuttamistoiminnassa kulloinkin käytetään. Myös vaikuttamistoiminnan kohteissa tapahtuvat muutokset heijastelevat tätä edellä kuvattua muutosta. Avoimeen demokraattiseen yhteiskuntaan kuuluu vapaa ja avoin kansalaiskeskustelu. Nykyaikainen digitalisoitunut yhteiskunta, sen instituutiot ja organisaatiot nojaavat toiminnassaan pitkälti kansalaisten luottamukseen (World Economic Forum 2022). Tätä luottamusta on pidetty demokraattisen yhteiskunnan arvokkaimpana voimavarana, mutta sitä on erityisen helppo rapauttaa generatiivisten tekoälyjen mahdollistamilla uusilla keinoilla. Viestintätoimisto Edelmanin 2024 Luottamusbarometri summaakin infosodan ja hakkeroinnin nouseviksi yhteiskunnan globaaleiksi uhkiksi.

World Economic Forum (2022) onkin määritellyt disinformaation kansainvälisesti yhteiskuntia lyhyellä aikavälillä eniten uhkaavaksi tekijäksi. Generatiivisen tekoälyn kehittyminen muuttaa laajasti viestinnän kautta yhteiskunnan, ihmisten, instituutioiden ja organisaatioiden välistä vuorovaikutusta. Uudet algoritmit ja sovellukset tarjoavat mahdollisuuksia myös toimijoille, jotka haluavat valjastaa tekoälyn informaatiovaikuttamisen, kuten disinformaation, tuottamisen ja välittämisen välineeksi. Tekoälyn avulla tuotettujen deepfake (”syvävääreennös”) -teosten muodostama uhka yhteiskunnille on herättänyt viimeisten vuosien aikana paljon huolta ja kysymyksiä mm. siitä, mihin ylipäätään voimme enää uskoa ja luottaa. Suurin uusi ongelma on disinformaation entistä tehokkaampi ja vaikuttavampi levitys, sillä räätälöidysti ja kohdennetusti tuotettu väärä tieto uppoaa yksilöihin ja yleisöihin entistä paremmin (Meikle 2023, 145; Wardle & AbdAllah 2023; Aïmeur ym. 2023; Anderson ym. 2017).

Uusien uhkien tunnistaminen edellyttää organisaatioilta päivitettyä osaamista ja teknologian käyttöön ottamista sekä päivittäisessä toiminnassa että myös informaatioympäristön seurannassa ja analysoinnissa. Nykyaikaisen informaatioympäristön hahmottaminen ja ymmärtäminen edellyttävät perinteisen mediaseurannan rinnalle työkaluja, jotka kykenevät analysoimaan laajasti informaatioympäristössä nopeasti syntyviä teemoja ja verkostoja.

Pärjätäkseen tekoälykkään viestinnän aikakaudella organisaatioiden on ymmärrettävä laajasti omaa toimialaansa ja koko yhteiskuntaan vaikuttavia kehityskulkuja. Valtionhallinnon tehostetun viestinnän ohjeessa (Valtioneuvoston kanslia 2019b) tuodaan hyvin esiin, mitä onnistunut toiminta kaikissa tilanteissa edellyttää. Nimittäin sitä, että johtamisen, viestinnän ja tilannekuvan väliset yhteydet ovat selvät ja toimivat. Tämä on hyvä periaate, joka koskee kaikkia organisaatiota, myös valtionhallinnon ulkopuolella. Tämän päivän maailmassa laaja-alainen tilannekuva, joka sisältää myös informaatioympäristön tilannekuvan, on perusta kaikelle toiminnalle. Liian suppea tilanneymmärrys tai -tilannekuva tekee organisaatioista haavoittuvia erilaisille uhkille. Tulkittaessa informaatioympäristössä esiin nousevia ilmiöitä tai jopa yksittäisiä viestinnällisiä toimenpiteitä on aina muistettava tarkastella kulloistakin viestinnällistä kontekstia: Mihin tämä liittyy yhteiskunnan tasolla? Kuka tästä hyötyy?

Viime vuosina viestinnän ammattilaiset ovat entistä useammin kohdanneet ja joutuneet miettimään toimintatapoja, joilla uudenlaisiin nopeasti kehittyviin uhkiin, kuten kyber- ja informaatiovaikuttamiseen, varaudutaan ja vastataan. Esimerkkeinä näistä voidaan mainita arkipäiväistyneet palvelunestohyökkäykset, tietojenkalastelu ja disinformaation levittäminen. Teknologinen kehitys on myös tuonut mukaan uuden haastavan sidosryhmän: feikkijoukot (fakeholders, katso esim. Luomaaho 2015). Feikkijoukoilla viitataan teknologioihin ja tahoihin, jotka piiloutuvat tai tekeytyvät joksikin muuksi kuin ovat vaikuttaakseen muihin haluamallaan tavalla. Haastavan tästä joukosta tekee se, että tekoälyn kehittyessä feikkijoukkojen tunnistaminen käy entistä vaikeammaksi. Samalla teknologisen kehityksen myötä niiden vaikuttamiskeinot muuttuvat entistä tehokkaammiksi ja sitä myötä vaarallisemmiksi: feikkijoukoista muokataan meille uskottavia lähteitä, esimerkiksi jäljittelemällä meidän oman lähipiirimme jäseniä.

Kaikki ihmiset ja organisaatiot voivat käytännössä joutua vihamielisen informaatiovaikuttamisen kohteeksi (Olivieri ym. 2023). Syynä kohteeksi joutumiseen voivat olla esimerkiksi organisaation tekemät päätökset ja toimenpiteet tai laajemmin yhteiskunnallinen ja turvallisuuspoliittinen tilanne. Syy voi olla myös organisaation ulkopuolella, kuten esimerkiksi alihankkijana toimivan yrityksen omissa toimenpiteissä. Hetki ja konteksti vaikuttavat entistä useammin siihen, joutuuko

organisaatio vaikuttamistoiminnan kohteeksi. Organisaatio ei voi enää ajatella, että "emme me ketään kiinnosta" tai "ettei informaatiovaikuttaminen kosketa meitä".

Teknologisen kehityksen myötä, ja etenkin tekoälyaikakaudella, toden ja valheen, faktan ja fiktion tunnistaminen verkossa ja sähköisessä mediassa muuttuu entistä haastavammaksi. Valheet rakennetaan erilaisten teknologioiden avulla niin, että ne muistuttavat mahdollisimman paljon todellisuutta. Näin niiden tunnistaminen valheeksi on entistä vaikeampaa (Wardle & AbdAllah 2023; Aïmeur ym. 2023; Andersson ym. 2017).

Tässä artikkelissa tarkastellaan tekoälykästä informaatiovaikuttamista, ja kysytään, miten informaatiovaikuttaminen muuttuu tekoälyn kehittymisen myötä. Artikkelissa vedetään yhteen informaatiovaikuttamista ja siihen liittyvää keskustelua tekoälystä. Tarkastelussa on myös se, miten kehittyviä teknologioita hyödynnetään vaikuttamistoiminnassa, ja miten organisaatioiden ja viestijöiden tulisi varautua tekoälykkään viestinnän aikakauden uusiin informaatioympäristösisä nouseviin uhkiin.

Tämän artikkelin kirjoittamiseen on käytetty testimuotoisesti laajaa kielimallia GPT-4 ChatGPT Team -lisenssin kautta aihepiirin tutkimusten kartoittamiseen ja Taulukon 1 muotoiluun. Lisäksi oletettavasti lukuisat eri tekoälyalgoritmit ovat vaikuttaneet tekstin sisältöön muun muassa hakukoneiden, pilvipalvelujen ja uutissivustojen kautta monilla tavoilla, joista sen paremmin niiden käyttäjät kuin tämän artikkelin kirjoittajatkaan eivät ole enää pitkään aikaan olleet tietoisia.

Oikean tiedon saatavuuden merkitys korostuu

Tekoäly tarvitsee toimiakseen dataa, ja erilaiset sovellutukset hyödyntävät avoimista lähteistä saatavilla olevaa tietoa. On hyvä huomioida, että jos omasta organisaatiosta saatavilla oleva tieto on vanhaa tai virheellistä, esimerkiksi generatiiviset tekoälytekstisovellukset eivät pelkästään jaa virheellistä tietoa sellaisenaan eteenpäin, vaan ne yhdistävät sen muihin asiakokonaisuuksiin, jolloin misinformaation (tahattoman väärän tiedon) leviämisen vaikuttavuus ja laajuus kertautuvat ja kasvavat. Paras valhe sekoittaa aina vähän totuutta mukaan. Myös omien tietojen tarpeeton salaaminen tekoälyltä internetissä, esim. kieltämällä tekoälyn ryömijän

(engl. crawler) pääsy verkkosivustolle robots.txt-tiedostoon kirjattavalla kiellolla, voi lisätä mahdollisuuksia misinformaation synnylle ja leviämiselle. Jos ihmiset lisääntyvissä määrin kysyvät tietoja tekoälyltä, eivätkä enää ”googlaa” tai tarkista tiedon alkuperää, voivat ainakin nykyiset tekoälyt (datan puuttuessa) paikata aukkoja tiedoissaan hallusinaatiolla eli tarkistamattomalla tiedolla (ks. hallusinaatiosta esim. Fui-Hoon Nah ym. 2023).

2020-luvulla tekoälyn ja siihen liittyvien sovellusten ja palveluiden kehittymisen yhteydessä on syytä palata kysymykseen digitaalisesta jalanjäljestä. Hyvässä tarkoituksessa verkossa jaetut kuvat ja videot luovat digitaalisia jälkiä. Nämä tiedot ovat käytännössä kenen tahansa kaapattavissa ja hyödynnettävissä myös rikollisiin tarkoituksiin. Tiedoilla voidaan luoda esimerkiksi deepfake-videoita tai -äänitallenteita tai väärennettyjä kuvia. Kaikkea saatavilla olevaa dataa voidaan hyödyntää muun muassa identiteettivarkauksissa tai huijauspuheluissa. Organisaation ja sen työntekijöiden datan suojelusta onkin tulossa entistä tärkeämpi tehtävä myös organisaatioille.

Informaatiovaikuttaminen

Mistä informaatiovaikuttamisessa sitten on kysymys? Tässä artikkelissa informaatiovaikuttamista käsitellään ja tarkastellaan pahantahtoisen tai vihamielisen vaikuttamistoiminnan (malign information influece) näkökulmasta (Yadav ym. 2023). Yhteistä aiheen eri määritelmille on se, että (informaatio)vaikuttamistoiminta on tarkoituksellista ja/tai strategista toimintaa, jonka tavoitteena on informaatiota tai kohteen käsityksiä ja tulkintoja manipuloimalla saada kohde toimimaan vaikuttajan haluamalla tavalla. Yleisenä vihamielisen vaikuttamisen tavoitteena voi olla pyrkimys esimerkiksi rapauttaa ihmisten luottamusta yhteiskuntaan, päättäjiin tai demokraattisiin instituutioihin (Ingram 2020; Pamment ym. 2018).

Informaatiovaikuttamisessa voidaan hyödyntää useita tekniikoita, teknologioita ja menetelmiä, kuten esimerkiksi disinformaation levittämistä, väärennettyjä henkilöllisyyksiä ja verkostojen tai ajatusten hakkerointia. Usein näitä erilaisia keinoja yhdistetään ja käytetään samanaikaisesti (Pamment ym. 2018; Valtioneuvoston kanslia 2019a, 35;

Myndigheten för Samhällskydd och Beredskap 2018, 26; Thomas ym. 2020). Käytännössä toiminta on näkynyt esimerkiksi erilaisina valeprofiileina somealustoilla, automatisoitujen tilien (bottien) hyödyntämisenä sisältöjen levittämisessä, palvelunestohyökkäyksinä, maalittamisena ja häirintänä sekä valeasiakirjoina ja -kuvina. Näiden tarkoituksena on ollut esimerkiksi vaikuttaa ihmisten käsityksiin todellisuudesta tai pyrkiä muuttamaan heidän mielipiteitään tai asenteitaan.

Toimijoina informaatiovaikuttamisessa nähdään joko pelkästään valtiolliset toimijat (Myndigheten för Samhällskydd och Beredskap 2018, 9), mutta myös laajemminkin näitä keinovalikoimaan kuuluvia toimintatapoja ja menetelmiä käyttävät erilaiset yhteenliittymät ja järjestöt (esimerkiksi terroristijärjestö ISIS) tai jopa yksittäiset kansalaiset (Thomas ym. 2020; Culloty ym. 2021, 3–4; Valtioneuvoston kanslia 2019a, 11 ja 16).

Informaatiovaikuttamisen yhteydessä keskustellaan usein dis-, mis- ja malinformaatiosta, joita usein sekoitetaan toisiinsa, vaikka termeillä onkin painotuseroja (Wardle 2020; Northeastern University 2024; Santos D-Amorim & Mirand 2021):

- **Misinformaatio** on virheellistä harhaanjohtavaa tietoa, jota sen levittäjä luulee todeksi. Hän ei siis tiedä, että tieto on virheellistä, vaan esimerkiksi jakaa sitä eteenpäin vilpittömästi sen todenmukaisuuteen uskoen. Esimerkiksi organisaatiosta saatetaan levittää vahingossa väärää puhelinnumeroa asiakaspalveluun.
- **Disinformaatio** eroaa misformaatiosta tarkoitusperän ja motiivin suhteen. Disinformaatio on tarkoituksellisesti tuotettua valheellista tai vääristeltyä tietoa, jota levittämällä toimija pyrkii vaikuttamaan tiedon vastaanottajiin ja kuluttajiin. Esimerkiksi organisaation palveluista saatetaan levittää keksittyjä huonoja kokemuksia sosiaalisessa mediassa, joissa varotetaan ostamasta sen palveluita.
- **Malinformaatio** on sinänsä todenmukaista tietoa, jota käytetään ja levitetään vahingoittamisen tarkoituksessa. Kohteita ovat esimerkiksi valtio, yksilöt, yhteiskunnan instituutiot tai yhteisöt. Esimerkiksi organisaatiota koskevaa asiayhteydestään irroitettua tietoa levitetään tarkoituksena rapauttaa ihmisten luottamusta kyseiseen organisaatioon.

Misinformaatio voidaan nähdä myös eräänlaisena yläkäsitteenä, jonka alakäsitteenä puolestaan on disinformaatio. Näin määriteltynä disinformaatiolla tarkoitetaan ”intentionaalista misinformaatiota” (Brown 2020).

Vaikuttamistoiminta muuttuu

Pyrittäessä ymmärtämään informaatiovaikuttamista tekoälykkäällä aikakaudella on tärkeä tiedostaa kyber- ja informaatiovaikuttamisen yhtäläisyydet ja merkitys toisilleen. Tämän päivän pitkälle digitalisoituneessa yhteiskunnassa tieto ja meidän jokaisen lähettämät viestit liikkuvat kyberavaruudessa. Käytämme päivittäin erilaisia palveluita (mukaan lukien sosiaalisen median alustat), jotka ovat osa kyberavaruutta. Kyse on aina myös ihmisten ja teknologioiden välisestä suhteesta. (Bronk 2024) Toiminnan tavoite ratkaisee sen, onko kyse pelkästä kiusanteosta vai voidaanko sen katsoa olevan informaatiovaikuttamista, pyritäänkö toiminnalla esimerkiksi rapauttamaan ihmisten luottamusta johonkin palveluun, henkilöön, yritykseen tai laajemmin digitaaliseen yhteiskuntaan.

Informaatiovaikuttamista tapahtuu reaalimaailman lisäksi myös kyberavaruudessa. Kyberhyökkäyksissä käytettävät tekniikat ja menetelmät hyödyntävät informaatiovaikuttamisessa käytettyjä keinoja. Uhreihin pyritään vaikuttamaan manipuloimalla, uhkailemalla ja muokkaamalla heidän käsityksiään todellisuudesta. Myös informaatiovaikuttamisen keinoihin kuuluvat kyberhyökkäykset, kuten palvelunestohyökkäykset tai Hack and Leak -tyyppinen toiminta, jossa tietoverkkoihin murtautumalla hankittuja tietoa käytetään informaatiovaikuttamisessa.

Esimerkiksi palvelunestohyökkäyksä, joita on nähty kasvavissa määrin viime vuosien aikana, on pidetty muun muassa verkkoaktivismina eli haktivismina. Mielipiteen osoittamisena tai protestina vastauksena esimerkiksi jonkin valtion tai organisaatioiden tekemiin päätöksiin. (CISA 2022; Dahan & Pasha 2023; Burgess 2022)

Tekoälykkäällä aikakaudella esimerkiksi deepfaket ovat työkalu sekä informaatio- että kybervaikuttamisessa. Toiminnassa, jolla kohde yritetään saada tekemään haluttuja tekoja, avaamaan haittaohjelman sisältävän liitteen sähköpostissa tai uskomaan huijauksiin. Tällaiseen

kyberavaruuden ja informaatioympäristön yhdistävään toimintaan ja manipulointiin tekoälyteknologiat tarjoavat aivan uudenlaisia sekä entistä tehokkaampia ja vaikuttavampia keinoja. (Aksela ym. 2022; Bronk 2024)

Miten vaikuttamistoiminnan muodot muuttuvat tekoälyaikakaudella ja millaisia haasteita ne muodostavat yhteiskunnille, organisaatioille ja yksilöille? Tässä yhteydessä kannattaa tarkastella sitä, miten tekoälyä on hyödynnetty kyberhyökkäyksissä ja mihin suuntaan niiden nähdään kyberympäristössä kehittyvän.

Tekoälyn hyödyntämistä kyberrikoksissa tutkineet Kalifornian yliopiston tutkijat nostavat analyysissaan deepfaket tekoälypohjaisen kyberrikollisuuden vakavimmaksi uhkaksi. Selvityksessään tutkijat kartoittivat tapoja, joilla tekoälyä voidaan käyttää rikoksissa. Sen avulla voidaan tunnistaa kohteiden heikkouksia, luoda kohteiden huijaamiseen valesisältöjä, kiristää heitä tai pyrkiä rapauttamaan heidän luotettavuuttaan. Tekoäly voi myös suorittaa toimenpiteitä, joita rikoksentekijät eivät itse voi tai halua tehdä. Vaikka metodit näissä ovat uusia, rikokset, joihin niissä pyritään ovat tutkijoiden mukaan hyvin perinteisiä. Toimenpiteiden kohteena voivat olla yksilöt, instituutiot, yritykset tai niiden asiakkaat, hallinto, julkinen diskurssi tai sosiaalinen koheesio. Rikosten motiiveja voivat lla taloudellisen hyödyn saavuttaminen, epäjärjestyksen lietsominen, politiikkaan vaikuttaminen, kosto tai maineiden tai suhteiden vahingoittaminen. Tekojen taustalla voi olla myös nihilistinen pyrkimys tuhoon, väkivaltaan tai vandalismiin. (Caldwell ym. 2020)

Kuten kybervaikuttamisessa myös informaatiovaikuttamisessa voidaan hyödyntää tekoälyä monin eri tavoin. Tarkasteltaessa tekoälyn vaikutuksia kyberturvallisuuteen tekoälykkään informaatiovaikuttamisen muutoksia on summattu alla olevaan Taulukkoon 1. Se esittää tiiviisti erilaisia tekoälyn tuomia mahdollisuuksia hyökkäykseen, puolustukseen ja varautumiseen tai kumpaankin. Mahdollisuudet taulukossa ovat listattuina aakkosjärjestyksessä, eikä niiden järjestyksestä voi tehdä päätelmiä vaikutuksen suuruudesta. Taulukko 1 perustuu (Aksela ym. 2022) esittämiin mahdollisuuksiin, ja sitä on rikastettu tämän artikkelin kirjoittajien aihepiirin asiantuntemuksella.

Taulukko 1.
Tekoälyn vaikutuksia informaatiovaikuttamiseen

TEKOÄLYN MAHDOLLISUUDET	KUVAUS JA VAIKUTUS	HYÖKKÄYS TAI PUOLUSTUS JA VARAUTUMINEN VAI MOLEMMAT?
1. Ajallisten rajoitteiden poistaminen	Mahdollistaa, nopeuttaa ja helpottaa toimintaa ympäri vuorokauden.	**Molemmat:** edistää hyökkäyksen nopeutta ja joustavuutta, mutta mahdollistaa myös jatkuvan puolustuksen ja varautumisen.
2. Hyökkäyksen kohdistaminen useisiin kohteisiin	Mahdollistaa resurssien hajauttamisen ja kohdeorganisaation kuormituksen lisäämisen.	**Hyökkäys:** Lisää hyökkäyksen tehokkuutta hajauttamalla puolustuksen voimavaroja.
3. Hyökkäysten laajuuden, kattavuuden ja nopeuden lisääminen	Lisää hyökkäysten tehokkuutta, mutta vaatii myös puolustukselta vastaavaa nopeutta ja kattavuutta.	**Molemmat:** Parantaa hyökkäyksen mahdollisuuksia mutta vaatii puolustukselta nopeita ja laaja-alaisia toimenpiteitä vastaamiseksi.
4. Kohdeorganisaation tilannekuvan hämärtäminen	Pyrkii haittaamaan organisaation kykyä johtaa toimintaansa ja viestiä tehokkaasti.	**Hyökkäys:** Hyökkäystaktiikka, jonka ymmärtäminen on kuitenkin olennaista puolustuksen ja varautumisen kannalta.
5. Kyber- ja informaatiovaikuttamisen integraatio	Yhdistää kyber- ja informaatiovaikuttamisen toiminnot, mikä hyödyttää sekä hyökkäystä että puolustusta.	**Molemmat:** Edellyttää sekä hyökkääjien että puolustajien ymmärrystä toistensa strategioista ja motiiveista tehostaen molempien toimintaa.
6. Pidempikestoinen vaikuttamistoiminta	Tekoäly voidaan kohdentaa toimimaan halutun aihepiirin parissa pidempiaikaisesti joko pienellä tai olemattomalla valvonnalla.	**Molemmat:** Mahdollistaa tehostetun ja pitkäjänteisen toiminnan sekä hyökkäyksen että puolustuksen näkökulmasta tehostaen molempien strategioita.
7. Sopivien kohteiden tunnistaminen	Auttaa tunnistamaan haavoittuvat kohteet sekä hyökkäys- että puolustusstrategioissa.	**Molemmat:** Auttaa hyökkääjiä kohdentamaan toimintansa, samalla kun puolustajat voivat tunnistaa ja suojata potentiaalisia heikkouksia.
8. Tekijän jälkien peittäminen	Vaikeuttaa toiminnan alkuperän jäljittämistä.	**Hyökkäys:** Suosii hyökkääjiä piilottamalla toiminnan lähteen, mutta opettaa puolustusta kehittämään parempia tunnistusmenetelmiä.
9. Tulevien vastatoimien ennakointi	Tukee puolustuksen ja varautumisen suunnittelua ymmärtämällä mahdollisia tulevia hyökkäyksiä.	**Puolustus ja varautuminen:** Keskittyy ennakointiin ja valmiuden parantamiseen, mutta antaa myös hyökkääjille tietoa puolustuksen mahdollisista vastatoimista.
10. Yleisön käsitysten muokkaaminen	Käyttää informaatiovaikuttamista mielipiteiden ja asenteiden ohjaamiseen.	**Molemmat:** Toimii sekä hyökkäyksen välineenä että puolustuksen strategiana korostaen tarvetta informaation hallintaan.

Edellä esitetystä taulukosta voidaan todeta yhteenvetona, että tekoälyä voidaan käyttää laaja-alaisesti sekä hyökkäykseen että puolustukseen ja varautumiseen, eikä tekoälyn mahdollisuuksia kannata sivuuttaa millään toiminnan tasolla. Kaikesta tekoälyhypetyksestä ja uhkakuvien maalailusta huolimatta on tärkeä muistaa, että tekoälyllä ja sen käytöllä on vielä rajoitteita. Lisäksi on hyvä muistaa, että tekoäly ei korvaa ”perinteisen” viestinnän tarvetta, vaan valmius siihen on tärkeää säilyttää myös tekoälyttömiä skenaarioita ja tilanteita varten. Näitä voivat olla esimerkiksi häiriötilanteet, joissa tekoälyt eivät ole kyber- tai muun hyökkäyksen takia käytössä, kriittisissä järjestelmissä tai teknologioissa esiintyy muu käytön estävä syy (kuten sähkö- tai tietoliikennekatko, tulipalo, järjestelmärikko tai avainhenkilöt eivät ole syystä tai toisesta tavoitettavissa) tai käsitellään tietoa, jota ei voida tai haluta tekoälyllä käsitellä.

Deepfake esimerkkinä tekoälykkään vaikuttamistoiminnan työkaluista

Sekä kyber- että informaatiovaikuttamisessa hyödynnetään deepfake-teknologioita. Deepfakeista puhuttaessa on tärkeä huomata, että termillä viitataan useisiin tekniikoihin. Näitä, lähes poikkeuksetta englanniksi käytettyjä termejä ovat esimerkiksi face-swap, lip-sync, puppet-master, face synthesis, attribution manipulation, audio-deepfaket (Masood ym. 2023). Face-swap-tekniikassa kohteen kasvot korvataan toisilla, jolloin videolla näkyvä hahmo tekee asioita, joita hän ei ole koskaan tehnyt. Tekniikkaa voidaan käyttää esimerkiksi henkilön maineen vahingoittamiseksi, kun saadaan tilanne näyttämään siltä, että henkilö tekee jotain, mitä hän ei todellisuudessa tee. Lip-sync-tekniikassa yhdistetään videolla näkyvän henkilön huulien liikkeet äänitallenteeseen. Näin videolla näkyvä hahmo puhuu asioita, joita hän ei ole koskaan sanonut. Kolmannessa ns. puppet-master-tekniikassa deepfake jäljittelee kohteen kasvoilla näkyviä ilmeitä ja pään liikkeitä. Tekniikan tavoitteena on kaapata henkilön ilmeet tai vartalo kuvitteelliseen videoon. Neljännessä ja viidennessä tekniikassa luodaan realistisia valokuvia kasvoista. Näitä tekniikoita käytetään esimerkiksi valeprofiilien luomiseen sosiaalisessa mediassa. Viimeisessä tekniikassa eli audio-deepfakeissa luodaan henkilön äänellä tuotettuja valheellisia sisältöjä.

Huonosti ja huonolla kielellä kirjoitetuista huijaussähköposteista on jo nyt siirrytty huimasti eteenpäin. Helmikuun alussa 2024 Suomessakin uutisoitiin tapauksesta, jossa hongkongilaiselta pankilta oli huijattu 23 miljoonaa euroa deepfake-videon avulla (Chen & Magramo 2024). Pankin työntekijä oli uskonut keskustelevansa videopuhelussa esihenkilöidensä kanssa ja tehnyt heidän pyynnöstään suuret rahasiirrot. Myös Suomessa on nähty huijausyritys, jossa on käytetty deepfake-äänitallennetta. Suomalaisen yrityksen työntekijöille Whatsappissa lähetetyllä viestillä toimitusjohtaja pyysi työntekijöitä tekemään ison rahasiirron. Viestin taustalla olevat rikolliset olivat luoneet tekoälyn avulla yrityksen johtajalta kuulostavan ääniviestin. (MTV-STT 2024)

Deepfaket eivät ole vain tallenteita, vaan niitä voidaan tehdä myös reaaliaikaisesti, jolloin esimerkiksi videoneuvottelussa aidon kuuloista keskustelua käyvä henkilö on tekoälyllä luotu virtuaalinen hahmo. Vuonna 2021 useat europarlamentaarikot saivat videoneuvottelukutsun venäläisen oppositiopoliitikko Aleksein Navalnyin avustajilta. Videopuhelussa näkyvät ”henkilöt” olivat tekoälyfiltterillä luotuja ”varsin aidonnäköisiä” kopioita todellisista henkilöistä (Roth 2021).

Deepfaken tunnistaminen

Deepfake-videoiden tunnistamiseen on kehitetty ja kehitteillä useita teknologioita. Teknologiaa voidaan käyttää myös käänteisesti selvitettäessä videon alkuperää ja taustoja. Ohjelmistoja kehittävät tutkimuslaitosten ja viranomaisten lisäksi myös esimerkiksi yritykset. Teknologiat keskittyvät tunnistamaan videoilta epäjohdonmukaisuuksia, kuten varjojen ja heijastusten puutteita, puheen, ilmeiden ja eleiden epäsymmetriaa (Somers 2020). Esimerkiksi kansainvälinen teknologiayritys INTEL on kehittänyt ohjelmiston, joka analysoi videolla näkyvän henkilön kasvoista verenkierron aiheuttamia värimuutoksia (INTEL 2023).

Tässä yhteydessä on syytä muistuttaa, että deepfake-teknologioita kehittävien ja niiden tunnistamiseen ohjelmistoja kehittävien toimijoiden välillä on käynnissä jatkuva kilpajuoksu. Kun tunnistus- ja torjuntamenetelmiä kehitetään, vastapuolella pyritään jatkuvasti kehittämään menetelmiä ja tekniikoita, jotka tekevät olemassa olevat tunnistusteknologiat hyödyttömiksi (Dickinson 2019).

Deepfake-videoiden tunnistamiseen pätevät samat periaatteet kuin informaatiovaikuttamisen yleisempäänkin tunnistamiseen. Tekniikoiden ja teknologioiden lisäksi on syytä aina muistaa lähdekritiikki ja tiedostaa, että kaikelle viestinnälle on aina tarkoitus, tarve ja syy. Kanavat ja välineet ovat vain keino välittää viestiä eteenpäin. Kääntämällä katse peiliin ja tarkastelemalla omia asenteita ja ennakkoluuloja sekä suhtautumalla rakentavalla tavalla kriittisesti kohtaamaamme viestintään pääsee jo pitkälle. Jos sosiaalisessa mediassa esiin tulevassa videossa aurinko paistaa siniseltä taivaalta, mutta kuvassa ei näy varjoja luonnollisissa paikoissaan, on kyse hyvin todennäköisesti valheellisesta sisällöstä.

Pimeä suunnittelu ja viestinnän kaappaaminen

Graham Meiklen mukaan ”luottamus on keskeistä viestinnälle” (2023, 118), ja hän vertaakin synteettisen median sisältöjä (deepfakeja) ”likaiseen pommiin”, joka haastavat perustavanlaatuiset käsitykset muihin luottamisesta. Hän myös muistuttaa sosiaalisen median palveluiden vastuusta alustoillaan leviävän disinformaation (mukaan lukien deepfaket) estämisessä ja nostaa esiin alustojen liiketoimintamalliin liittyvät ongelmat. (Emt.)

Suunnittelun ja kognitiotieteiden maailmassa puhutaan pimeän suunnittelun malleista (dark design patterns, Kollmer & Eckhardt 2023; Widdicks ym. 2022), joiden tarkoituksena on koukuttaa ihminen toimimaan verkossa tavalla, joka voi olla hänelle itselleen haitallista. Tällaiset koukuttavat toiminnot voivat olla esimerkiksi manipuloivia ohjeita ja puoliksi palvelussa jatkamaan pakottamista, joita usein pidetään melko normaaleina toimintatapoina esimerkiksi tietokonepeleissä. Tällaiseen suunnitteluun eivät sorru ainoastaan totalitaariset valtiot, vaan myös brändejä ja organisaatioita on syytetty sellaisten verkkopalveluiden tietoisesta suunnittelusta, jotka esimerkiksi koukuttavat käyttäjiä mukaan epäeettisellä tavalla (Widdicks ym. 2022) osittain niiden tehokkuuden takia. Facebookin Meta-yhtiötä on useana vuonna syytetty juuri tällaisesta manipuloivasta suunnittelusta, vaikka ylikäytön riskit ovat tiedossa (Murphy 2023). Samoin myös voimakkaasti kasvanut kiinalainen lyhytvideosome TikTok on usein kriittisen tarkastelun kohteena, sillä sen läpinäkymätön toiminta ja datan keräämisen motiivit nostavat toistuvasti epäilyksiä mediassa.

Pimeän suunnittelun malleissa uhreja on monia: käyttäjien lisäksi ne rakentavat uutta normaalia, jolloin hyväksymme tietomme väärinkäytön jo lähtökohtaisesti saadaksemme esimerkiksi räätälöidympää palvelua. Manipulaatio ja väärään suuntaan ohjaus ovat pimeiden mallien suunnitellussa keskeisiä, kun käyttäjää ohjataan esimerkiksi tekemään nopeita päätöksiä ilman harkintaa tai väärä valinta saadaan näyttämään hyvältä (Kollmer & Eckhardt 2023).

Digitaalisen jalanjäljen ja datan merkityksen yhteydessä on hyvä nostaa esiin viestinnän tutkimuksessa puhuttu ”viestinnän kaappaaminen”, jolloin jokin ulkopuolinen taho kaappaa esimerkiksi jonkin organisaation kampanjan sloganin, visuaalisen ilmeen tai sisällöt kääntäen niiden merkityksen päinvastaiseksi alkuperäisen kampanjaan nähden (Hautala ym. 2024). Perinteisesti tämä on tapahtunut esimerkiksi irrottamalla kampanjan logo kontekstistaan ja alkuperäisestä yhteydestään tai tekemällä kampanjan verkkosivua muistuttava verkkosivu päinvastaisella sisällöllä. Lisäksi kampanjan ilmettä on hyödynnetty oman sisällön jakamisessa. Lisäksi useista videoista on saatettu irrottaa pätkiä ja yhdistää ne kokonaan uudeksi videoksi, jossa lopullinen video ei ole ollut totta (MTV-STT 2019; Vaarala 2020).

Tekoälyaikakaudella viestinnän kaappaamiseen tarjoutuu uusia mahdollisuuksia. Esimerkiksi poliittiseen kampanjaan osallistuvasta julkisuuden henkilöstä voidaan tehdä deepfake-video, jossa hän arvostelee kampanjaa, tai hänestä tehdään kampanjaan kuuluvaksi väitetty video, jota ei oikeasti ole tehty. Lisäksi hänestä voidaan tehdä hänen mainettaan tai luottamusta romuttamaan pyrkivä deepfake-video tai äänitallenne. Käytännön esimerkkinä 29.2.2024 uutisoitu ääninäyttelijä Reidar Waseniuksen äänen kaappaus aikuisviihdemainokseen ilman lupaa (ks. Jonsson 2024). Tämän tyyppiseen toimintaan tekoäly tarjoaa uudenlaisia mahdollisuuksia erityisesti visuaaliseen viestintään ja audiovisuaalisiin sisältöihin.

Yhteenveto ja johtopäätökset

Helposti käytettävät ja käyttöönotettavat teknologiat voivat laskea uusien toimijoiden kynnystä lähteä vaikuttamistoimintaan mukaan. Niin sanotuille vanhoille tekijöille teknologinen kehitys ja uudet teknologiat

tarjoavat mahdollisuuksia muuttaa ja kehittää toimintaansa. Tekoälyaikakauden viestinnällisiin haasteisiin varautuminen edellyttää vahvaa ymmärrystä oman organisaation toimialasta, toimintaympäristöstä sekä erilaisista vuorovaikutussuhteista ja verkostoista, jotka liittyvät ja voivat vaikuttaa oman organisaation toimintaan.

Vaikka demokratiaan kuuluu avoin keskustelu, se voidaan kääntää myös demokraattista yhteiskuntaa vastaan (Pamment ym. 2018). Tekoälykkään viestijän onkin syytä olla tietoinen häneen mahdollisesti kohdistuvasta informaatiovaikuttamisesta, sen keinoista ja misinformaation riskeistä (generatiivisia) tekoälyjä käyttäessä. Esimerkiksi tekoälyn tietojen tarkistaminen tekoälyllä aiheuttaa silmukan, jossa mahdollisesti hallusinoiva algoritmi toimii tiedon vartijana löytämättä omia ongelmiaan (ks. Niittymaa & Luoma-aho tässä julkaisussa). Tämä voi johtaa pahimmillaan datan pilaantumiseen, kun tulevat tekoälyt ja viestijät ottavat väärän tiedon oikeana. Onkin ehdotettu, että viestinnän ammattilaisista tulee tekoälyn aikakaudella entistä vahvemmin ”boxturnereita” eli sen pienellä printatun lisätiedon selvittäjiä, jota tuotteisiin ja palveluihin liittyy mutta jota harvoin julkisesti esitetään (Badham & Luoma-aho, 2023). Tulevaisuuden viestinnän ammattilaisten työ vaatii enemmän kriittistä pohdintaa ja taustoitusta, kun selvitettävänä on sidosryhmien vaikutusyritysten aitous, datan analysointi ja organisaation omien suunnitelmien kestävyys ja eettisyys.

Onnistunut vastaaminen vihamieliseen informaatiovaikuttamiseen edellyttää varautumista, suunnitelmallisuutta ja harjoittelua. Tämä pitää sisällään myös harjaantumisen kulloinkin yleisesti käytössä olevien viestintäteknologioiden käyttöön. Jos teknologioita ja niiden toimintalogiikkaa ja mahdollisuuksia sekä hyvään että pahaan ei ymmärrä, on vaikea myöskään varautua ennalta arvaamattomaan. Viestijän työ myös tällä rintamalla teknistyy, ja viestinnän ammattilaisten seurantaan tulee laajempi kuva kuin vain oman brändin tai organisaation kehitys.

Lisäksi viestinnän ammattilaisista tulee organisaatioidensa ”omatunto”: tekoälykkään viestinnän myötä nousee kysymys, käytetäänkö organisaatiossa dataa ja tekoälyä kestävällä ja läpinäkyvällä tavalla, joka hyödyttää yhteiskuntaa. Jos ei, mitä organisaatio voisi tehdä muuttaakseen tätä parempaan suuntaan? Vastuu sovellusten käytöstä ja niiden tuottaman sisällön oikeellisuuden tarkistamisesta ja jakamisesta on aina viime kädessä sen käyttäjällä. Mikäli organisaatio käyttää

tekoälyä toiminnassaan ja viestinnässään, kannattaa siitä kertoa mahdollisimman avoimesti välttääkseen arvon tuhoutumista (ks. Niittymaa & Luoma-aho tässä julkaisussa).

Tekoäly ei poista tarvetta tarkastella muunlaisia kriisejä tai uhkia ja varautua niihin. Tekoälyn varaan ei voi rakentaa oman organisaationsa valmiutta vastata informaatioympäristössä ja sen ulkopuolella nouseviin uhkiin ja kriiseihin. On todennäköistä, että tulevaisuudessa tarvitaan jatkossakin sekä ilman tekoälyä että myös generatiivisen tekoälyn kanssa tehtyä viestintää.

Näistä voit aloittaa

1) **Tietoturva nyt! Kun jokainen päivä voi olla aprillipäivä – Mistä deepfakeissa on kysymys? Traficomin Kyberturvallisuuskeskuksen tietopaketti deepfakeista ja niiden tunnistamisesta. https://www.kyberturvallisuuskeskus.fi/fi/ajankohtaista/kun-jokainen-paiva-voi-olla-aprillipaiva-mista-deepfakeissa-kysymys**
2) **AI in Support of StratCom Capabilities. Naton strategisen viestinnän osaamiskeskuksen (NATO StratCom CoE) tuore raportti keinoista, joilla tekoäly voi tukea strategista viestintää. https://stratcomcoe.org/publications/ai-in-support-of-stratcom-capabilities/296**
3) **Assessing the risks and opportunities posed by AI-enhanced influence operations on social media: Lundin yliopiston julkaisema artikkeli tekoälyn vaikutuksista informaatiovaikuttamiseen sosiaalisessa mediassa. https://portal.research.lu.se/sv/publications/assessing-the-risks-and-opportunities-posed-by-ai-enhanced-influe**
4) **EUvsDisinfo-verkkosivusto. EU:n ulkosuhdehallinnon East StratCom Task Forcen ylläpitämä informaatiovaikuttamisen tunnistamiseen keskittyvä sivusto. Sivusto on erittäin hyödyllinen tietopankki kaikille informaatiovaikuttamisesta kiinnostuneille. https://euvsdisinfo.eu/**
5) **The WAYBACK Machine: Tietojen tarkastuksessa hyödyllinen sivusto, josta voi tarkastaa muuttuneita tietoja nettisivuilla. https://help.archive.org/help/wayback-machine-general-information/**

Jussi Toivasella on pitkä kokemus valtionhallinnon viestinnästä. Hän on työskennellyt viestinnän eri tehtävissä mm. Pääesikunnassa, Poliisihallituksessa, sisäministeriössä ja valtioneuvoston kansliassa. Jussi aloitti Traficomin Kyberturvallisuuskeskuksen viestintäpäällikkönä huhtikuussa 2022. Työskennellessään valtioneuvoston kansliassa kymmenen vuoden ajan hän vastasi muun muassa kriisiviestintään ja informaatiovaikuttamiseen liittyvistä asiakokonaisuuksista. Nykyisessään tehtävässään Kyberturvallisuuskeskuksessa Toivasen tehtäviin kuuluu johtaa ja kehittää keskuksen viestintää. Koulutukseltaan Toivanen on valtiotieteiden maisteri ja filosofian maisteri Turun yliopistosta.

Jukka Niittymaa on väitöskirjatutkija Jyväskylän yliopiston kauppakorkeakoulun viestintäjohtamisen oppiaineessa ja tekoäly- ja innovaatiojohtaja luova toimisto Sherpassa. Jukalla on laaja-alaista käytännön kokemusta eri GenAI-sovellusten käytöstä viestinnän ja markkinoinnin asiakasprojekteissa, ja työssään Jukka myös valmentaa suomalaisia yrityksiä tekoälykkääseen aikaan. Jukan väitöstutkimus käsittelee generatiivisten tekoälyjen vaikutusta viestintään ja markkinointiin.

Vilma Luoma-aho on viestinnän johtamisen professori ja koulutuksesta vastaava varadekaani Jyväskylän yliopiston kauppakorkeakoulussa. Hänen päätutkimusalueitaan ovat teknologiavälitteinen viestintä, haastavat sidosryhmät ja informaatiovaikuttaminen. Hän johtaa parhaillaan Maanpuolustuksen kannatussäätiön rahoittamaa tutkimusprojektia viestinnän kaappauksista sekä kirjoittaa uutta oppikirjaa digitaalisen viestinnän johtamisesta. Vilma on ProComin hallituksen entinen puheenjohtaja ja kunniajäsen.

Kirjallisuus

Aïmeur, Esma; Amri, Sabine & Brassard, Gilles (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining.* 13, 30. https://doi.org/10.1007/s13278-023-01028-5

Aksela, Matti; Marchal, Samuel; Patel, Andrew & Rosenstedt, Lina (2022). *Tekoälyn mahdollistamat kyberhyökkäykset.* Traficomin julkaisuja 30/2022. https://www.traficom.fi/sites/default/files/media/publication/TRAFICOM_Teko%C3%A4lyn_mahdollistamat_kyberhy%C3%B6kk%C3%A4ykset%202022-12-12_web.pdf

Anderson, Janna & Rainie, Lee (2017). *The Future of Truth and Misinformation Online*. Pew Research Center. https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/

Badham, Mark; Luoma-aho, Vilma & Valentini, Chiara (2023). "A revised digital media–arena framework guiding strategic communication in digital environments", *Journal of Communication Management.* https://doi.org/10.1108/JCOM-03-2023-0031

Bronk, Chris (2024). New Problems in Hybrid Warfare: Cyber Meets Cognition, 11.4.2024. *The Defence Horizon Journal.* https://tdhj.org/blog/post/hybrid-warfare-cyber-cognition/

Brown, Étienne (2020). Regulating the Spread of Online Misinformation. Teoksessa: Hannon Michael. & de Ridder, Jeroen. (toim.) *The Routledge Handbook of Political Epistemology*. Routledge, New York

Burgess, Matt (27.12.2022). Hacktivism Is Back and Messier Than Ever. *The Wired.* https://www.wired.com/story/hacktivism-russia-ukraine-ddos/

Caldwell, Matthew; Andrews, Jerone T.A.; Tanay, Thomas & Griffin, Lewis (2020). AI-enabled future crime. *Crime Sci* 9, 14-. https://doi.org/10.1186/s40163-020-00123-8

Chen, Heather. & Magramo, Kathleen (4.2.2024). Finance worker pays out $25 million after video call with deepfake 'chief financial officer. *CNN.* https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

CISA (9.5.2022). *Russian State-Sponsored and Criminal Cyber Threats to Critical Infrastructure.* https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-110a

Culloty, Eileen. & Suiter, Jane (2021). *Disinformation and Manipulation in Digital Media*. Routledge.

Dahan, Amir & Pasha, Syed (17.3.2023). *KillNet and affiliate hacktivist groups targeting healthcare with DDoS attacks.* Microsoft. https://www.microsoft.com/en-us/security/blog/2023/03/17/killnet-and-affiliate-hacktivist-groups-targeting-healthcare-with-ddos-attacks/

Dickinson, Ben (2.9.2019). AI Tools Can Detect Deepfakes, But for How Long? *PCMag* https://uk.pcmag.com/news/122383/ai-tools-can-detect-deepfakes-but-for-how-long

Fui-Hoon Nah, Fiona; Zheng, Ruilin; Cai, Jingyuan; Siau, Keng & Chen, Langtao (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*. 25. 1-28. 10.1080/15228053.2023.2233814.

Hautala Miriam; Luoma-aho, Vilma & Brown, Jason (2024). Communication Hijacked? New Vulnerabilities in the Digital Media Arenas. Teoksessa: Bowen, Shannon & Erzikova, Elina (toim.) *Handbook of Innovations in Strategic Communication*, Edward Elgar Publishing (tuleva)

Ingram, Haroro J. (2020). The Strategic Logic of State and Non-State Malign 'Influence Activities', *The RUSI Journal,* 165:1, 12–24. https://doi.org/10.1080/03071847.2020.1727156

Intel (14.11.2022) Intel Introduces Real-Time Deepfake Detector https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html

Jonsson, Sören (2024). Erotisk dejtingsajt stal Reidar Wasenius röst. *HBL* 29.2.2024. Saatavilla: https://www.hbl.fi/artikel/dbdec147-109a-5cf7-8692-3edeeade7ccf (luettu 1.3.2024).

Kollmer, Tim & Eckhardt, Andreas. (2023). Dark Patterns. *Business Information Systems Engineering* 65, 201–208. https://doi.org/10.1007/s12599-022-00783-7

Lievonen, Matias; Bowden, Jana & Luoma-aho, Vilma (2023). Towards a typology of negative engagement behavior in social media. *Service Industries Journal*, *43*:3-4, 238–259. https://doi.org/10.1080/02642069.2022.2121961

Luoma-aho, Vilma (2015). Understanding Stakeholder Engagement: Faith-holders, Hateholders & Fakeholders. *RJ-IPR : Research Journal of the Insitute for Public Relations*, *2*(1). http://www.instituteforpr.org/understanding-stakeholder-engagement-faith-holders-hateholders-fakeholders/

Masood, Momina; Nawaz, Mariam; Malik, Javed; Irtaza, Aun & Hafiz, Malik (2023). Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. *Appl Intell* 53, 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

Meikle, Graham (2023). *Deepfakes*. Polity Press.

MSB, Myndigheten för Samhällsskydd och Beredskap (2018). *Att Möta informationspåverkan - Handbok för kommunikatörer.* https://www.mpf.se/assets/uploads/2022/06/Att-mota-informationspaverkan-handbok-for-kommunikatorer.pdf

MTV-STT (23.11.2019). Jussi Halla-ahon ja Maria Ohisalon eduskuntakeskustelusta levisi väärennetty video – varapuhemies Haatainen: ”Ilmiölle on pantava stoppi”. Saatavilla: https://www.mtvuutiset.fi/artikkeli/jussi-halla-ahon-ja-maria-ohisalon-eduskuntakeskustelusta-levisi-vaarennetty-video-varapuhemies-haatainen-ilmiolle-on-pantava-stoppi/7635302#gs.54kdrc (luettu 20.4.2024)

MTV & STT (25.2.2024). Häijy tekoälyllä tehty huijausyritys paljastui Suomessa – ”Hyvin aidon kuuloinen” *MTV.* Saatavilla: https://www.mtvuutiset.fi/artikkeli/haijy-tekoalylla-tehty-huijausyritys-paljastui-suomessa-hyvin-aidon-kuuloinen/8885622#gs.5jj7ix (luettu 20.4.2024)

Murphy, Hannah (24.1.2023). US states accuse Meta of ‘manipulative’ practices towards young users. *Financial Times.* https://www.ft.com/content/78c46a01-30b2-40e1-a8c9-a920b7d7d5d1

Niittymaa, Jukka & Luoma-aho, Vilma (2024). Tekoälykäs viestintä. Teoksessa: Niittymaa, Jukka & Luoma-aho, Vilma (toim.). *ProComma Academic 2024: Tekoälykäs viestintä*. Helsinki: ProCom.

Northeastern University (30.1.2024). Fake News/Misinformation/Disinformation: What is Fake News? https://subjectguides.lib.neu.edu/fakenews

Olivieri, Mirko; Mäkelä, Rosa-Maria; Romenti, Stefania & Luoma-aho, Vilma (2023). Digital corporate communication and disinformation. Teoksessa: Vilma. Luoma-aho & Mark Badham (toim.). *Handbook on Digital Corporate Communication* (s. 426–438). Edward Elgar. https://doi.org/10.4337/9781802201963.00042

Pamment, James; Nothaft, Howard; Agardh-Twetman, Henrik & Fjällhed, Alicia (2018). Countering Hostile Influence - the State of the Art. Lund University. https://rib.msb.se/filer/pdf/28697.pdf

Rafique, Rimsha, Gantassi, Ramsha; Amin, Rashid; Frnda, Jaroslav; Mustapha, Aida & Alsheri, Asma (2023). Deep fake detection and classification using error-level analysis and deep learning. *Sci Rep* 13, 7422 https://doi.org/10.1038/s41598-023-34629-3

Roth, Andrew (22.4.2021). European MPs targeted by deepfake video calls imitating Russian opposition. *Guardian.* https://www.theguardian.com/world/2021/apr/22/european-mps-targeted-by-deepfake-video-calls-imitating-russian-opposition

Santos-d’Amorim, Karen. & Miranda, Májory Fernandes de Oliveira (2021). Misinformation, disinformation, and malinformation: clarifying the definitions and examples in disinfodemic times. *Encontros Bibli Revista Eletrônica de Biblioteconomia e Ciência da Informação*. DOI:10.5007/1518-2924.2021.e76900

Somers, Meredith (21.6.2020). Deepfakes, explained. MIT Sloan School of Management. https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

Thomas, Elise; Thompson, Natalie & Wanless, Alicia (2020). The Challenges of Countering Influence Operations. Carnegie Endowment for International Peace. https://carnegieendowment.org/2020/06/10/challenges-of-countering-influence-operations-pub-82031

Vaarala, Viljami. (24.11.2020). Somevaikuttajat iskivät THL:n koronakampanjaan levittämällä väärää tietoa, muu some riensi apuun. *Yleisradio.fi.* https://yle.fi/a/3-11662185

Valtioneuvoston kanslia, VNK (2019a). Informaatiovaikuttamiseen vastaaminen – Opas viestijöille. Valtioneuvoston kanslian julkaisuja 11/2019. https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161512/VNK_11_2019_Informaatiovaikuttamisen%20vastaaminen_web.pdf

Valtioneuvoston kanslia, VNK (2019b). *Valtionhallinnon tehostetun viestinnän ohje – viestintä normaalioloissa ja häiriötilanteissa.* Valtioneuvoston kanslian julkaisuja 23/2019. http://urn.fi/URN:ISBN:978-952-287-815-1

Wardle, Claire & AbdAllah, AbdelHalim (2023). The Information Environment and Its Influence on Misinformation Effects. Teoksessa: Purnat, Tina; Nguyen, Tim., Briand, Sylvie (toim.). *Managing Infodemics in the 21st Century.* Springer, Cham. https://doi.org/10.1007/978-3-031-27789-4_4

Wardle, Claire (22.9.2020). Understanding Information disorder. https://firstdraftnews.org/long-form-article/understanding-information-disorder/

Widdicks, Kelly; Remy, Christian; Bates, Oliver; Friday, Adrian & Hazas, Mike (2022). Escaping unsustainable digital interactions: Toward "more meaningful" and "moderate" online experiences, *International Journal of Human-Computer Studies*, Volume 165. 102853, ISSN 1071-5819, https://doi.org/10.1016/j.ijhcs.2022.102853.

World Economic Forum (2022). Earning Digital Trust: Decision-Making for Trustworthy Technologies. *Insight Report* 11/2022. https://www3.weforum.org/docs/WEF_Earning_Digital_Trust_2022.pdf

Yadav, Kamya; Riedl, Martin; Wanless, Alicia & Woolley Samuel (2023). What Makes an Influence Operation Malign?. *Carnegie Endowment for International Peace.* Saatavilla: https://carnegieendowment.org/2023/08/07/what-makes-influence-operation-malign-pub-90323