



Maanpuolustuskorkeakoulu

Sotatekniikan laitos

Julkaisusarja 2: Tutkimuselosteita nro 4

Katsaus tekoälyyn

Tekoäly ja sen sovellukset

Arto Mutanen, Antti Rissanen, Kalle Saastamoinen (toim.)

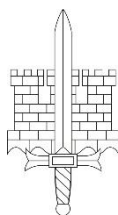
MAANPUOLUSTUSKORKEAKOULU
SOTATEKNIIKAN LAITOS
JULKAISUSARJA 2: TUTKIMUSSELOSTEITA NRO 4

NATIONAL DEFENCE UNIVERSITY
DEPARTMENT OF MILITARY TECHNOLOGY
SERIES 2: RESEARCH REPORTS NO. 4

KATSAUS TEKOÄLYYN

TEKOÄLY JA SEN SOVELLUTUKSET

ARTO MUTANEN, ANTTI RISSANEN
& KALLE SAASTAMOINEN
(TOIM.)



MAANPUOLUSTUSKORKEAKOULU
SOTATEKNIIKAN LAITOS
HELSINKI 2023

Arto Mutanen, Antti Rissanen & Kalle Saastamoinen: *Katsaus Tekoälyyn – Tekoäly ja sen sovellutukset*

Maanpuolustuskorkeakoulu

Sotatekniikan laitos

Julkaisusarja 2: Tutkimusselosteita nro 4

National Defence University

Department of Military Technology

Series 2: Research Reports No. 4

Uusimmat julkaisut pdf-muodossa: <http://www.doria.fi/handle/10024/73990>

© Tekijät & Maanpuolustuskorkeakoulu

ISBN 978-951-25-3332-9 (nid.)

ISBN 978-951-25-3333-6 (pdf)

ISSN 2737-0615 (verkkojulkaisu)

Maanpuolustuskorkeakoulu – Sotatekniikan laitos

National Defence University – Department of Military Technology



Tämä teos on lisensoitu Creative Commons BY-NC 4.0 -käyttöluvalla. Tarkastele käyttöluvaa osoitteessa <https://creativecommons.org/licenses/by-nc/4.0/deed.fi>

PunaMusta Oy
Joensuu 2023



KIITOKSET

Tätä kirjaa edeltänyt seminaarisarja tukeutui professori Jouko Vankan johtaman koneoppimiseen keskittyneen ryhmän tekemään tutkimukseen. Kirja perustuu laajemmin tutkimusryhmän ympärillä toimivan toimijaverkoston seminaarisarjassa pitämiin esityksiin. Olemme kiitollisia seminaarisarjan virittämästä avoimesta intellektuaalisesta ilmapiiristä. Ansio kirjasta kuuluu kirjoittajille, ilman heidän suurta panostustansa, ei kirjaa olisi voinut syntyä: kiitokset kaikille kirjoittajille arvokkaasta työstä kirjamme hyväksi. Hankkeen mahdollisti Maanpuolustuskorkeakoulun Sotatekniikan laitos, lisäksi Merisotakoulu tuki osaltaan hankkeen toteutusta, mistä suuri kiitos. Artikkel ”Using Simulation to Evaluate SARS-CoV-2 Transmission in Finnish Defence Forces’ Leave Arrangements” esitettiin seminaarissa, ja on julkaistu esityksen jälkeen ISMS Conference 2020:n artikkelikokoelmassa kaksipalstaisena tekstinä (<https://www.doria.fi/handle/10024/178092>).

Kiitos mahdollisuudesta julkaista artikkeli uudestaan taitettuna myös tässä kirjassa.

JOHDANTO

Arto Mutanen, Antti Rissanen & Kalle Saastamoinen

Tekoälyn arkipäiväistyminen on muuttanut niin työelämää, harrasteita kuin normaalia kotiarkea. Tekoälyyn liittyy paljon toiveita sen tuomasta inhimillisen elämän helpotuksista ja toisaalta pelkoja, joita tekoälyn kehitys voi lopulta tuottaa. Teknologia mahdollistaa yhä paremman yhteydenpidon ihmisten välillä ja tekoälysovellutukset voivat parantaa niin hoivaa kuin opetusta. Samalla pelkona on, että koneiden tullessa ihmisten väliin, aidot inhimilliset kohtaamiset vähenevät. Työelämässä tekoälyn toivotaan tekevät tylsät ”mekanisoitavissa olevat” työt ja siten lisäävän inhimillistä vapautta ja luovuutta. Ongelmana tällaisessa ajatuksessa on, että luovuus ei ole vapaata ideointia, vaan perusasioiden parissa työskentelyä, johon luovuus lomituu.

Digitalisaatioon ja tekoälyyn kytkeytyvät toiveet eivät tietenkään ole perusteettomia. Ne muistuttavat Baconin optimismia, jossa ajatus oli, että systemaattinen (induktiivinen) tiede tuottaa uutta todenmukaista kausaalitietoa, joka mahdollistaa ihmisten tuottavan tiettyjä toivottuja lopputuloksia tai estävän ei-toivottujen tulosten toteutumisen, mikä lopulta johtaa inhimillisen elämän helpottumiseen ja paranemiseen. Tämä optimismi elähdytti uuden ajan tieteen ja tiedepohjaisen tekniikan kehitystä. Kuitenkin myös kritiikkiä esiintyi. Erityisesti Frankfurtin koulun toimesta kritiikki toi esiin teettellis-tekniisen elämänmuodon perustana olevan rationaliteetin ongelmallisuuden. Tämä kritiikki on saanut uutta ajankohtaisuutta nykyisten pirullisten ongelmien (wicked problems) aikakaudella. Joudumme tietenkin puntaroimaan, toistaako tekoälyyn liittyvä optimismi baconilaisen optimismin ongelmallista sokeutta tiedon valoon kytkeytyviä varjoja kohtaan. Tekoälyn rinnalla tarvitsemme inhimillistä viisautta.

Onnistumisen kannalta keskeinen kysymys on tekoälyn perustana oleva älykkyys. Puhuttaessa tekoälystä, ei ole lainkaan selvää, mitä tällöin älykkyydellä tarkoitetaan. Inhimillinen älykkyys ei ole redusoitavissa kalkyloivaan järkeen, vaan siihen kuuluu asioiden harkintaa ja puntarointia, joita on erittäin vaikeaa täsmällisesti luonnehtia. Siten tekoälytutkimukseen liittyy olennaisesti älykkyiden ja järkevyyden perusteiden tutkimus. Tämä pitää sisällään niin tiedonhankinnan logiikan analyysin kuin myös arvojen ja tunteiden roolin systematisointia osana älykkyyttä.

Jo nyt useat tekoälyperustaiset järjestelmät tehostavat ja parantavat työprosesseja. Samalla tietenkin ihmisten asema ja rooli muuntuu, usein tavalla, jota ei ole kyetty ennakoidaan. Ei ole selvää, säilyykö ihmiset ohjaajan ja johtajan roolissa vai ottaako tulevaisuuden tehokkaat ja inhimillisen järkevyyden ylittävät koneet (singulariteetti) ihmiset hallintaansa. Vaikka tällainen ongelma saattaa vaikuttaa science fiction -kirjallisuuden ongelmalta, niin ei tule väheksyä ongelmaa. Yhtäältä tällaisten koneiden mahdollisuus on todellinen ja toisaalta tämän ongelman pohtiminen auttaa jäsentämään tiettyjä keskeisiä tekoälyyn liittyviä käsitteellisiä ongelmia systemaattisesti.

Tekoälyn vaikutus ihmisiin ja ihmisyhteisöihin tulee esiin jo nykyisten tekoälyjärjestelmien toteutuksissa. Onkin hyödyllistä tutustua nykyhetken tekoälyteknologian omi-

naisuuksiin ja mahdollisuuksiin. Valtaosa tekoälytutkimuksesta painottuu arkipäiväiseen tekoälytekniikkaan, jota tapaamme niin kodinkoneissa, autoissa kuin sotilaallisissa laitteissa. Digitalisoituvassa, datakeskeisessä ympäristössä on hyödyllistä, että tekoäly tuodaan osaksi lääketieteellistä diagnostiikkaa tai sodan edellyttämän informaation hankintaa. Kuitenkaan ei ole selvää, miten iso-osa potilaan hoidon tai sodanjohdon päätöksenteosta voidaan osoittaa tekoälyn vastuulle. Ongelmia tulee esiin jo nykyisten tekoälyjärjestelmien puitteissa.

Digitalisaation ansiosta tekoäly voi muuntaa teknisistä infrastruktuuria radikaalisti. Useat valtiot panostavat tekoälyn tutkimuksen lisäksi sovellutusten kehitykseen. On selvää, ettei kehityksen tahdissa voida pysyä ilman systemaattista satsausta. Suomessa julkishallinto on linjannut, että meidän tulee olla tekoälyn soveltamisen kärkimaiden joukossa. Tavoite on mielekäs. Mutta täytäntöönpano edellyttää systemaattista panostusta niin tutkimukseen, koulutukseen kuin teollisuutteen. Tämä edellyttää sekä oman tiedon että taidon tuotantoa, mutta myös laajempaa verkottumista kansainvälisiin osajaverkostoihin. Myös puolustushallinnossa on linjattu, että puolustushallinto kehittää tekoälyosaamistaan tutkimuksen, koulutuksen ja kansainvälisen verkostoitumisen kautta (Puolustusministeriö 2020; <https://julkaisut.valtioneuvosto.fi/handle/10024/162370>)

Maanpuolustuskorkeakoulun sotatekniikan laitoksessa toimii professori Jouko Vanhan johtamana tekoälyyn keskittynyt tutkimusryhmä. Keväällä 2019 syntyi ajatus, että olisi hyvä kutsua asia-alueesta kiinnostuneita keskustelemaan tekoälytutkimuksen ajankohtaisista teemoista. Vapaamuotoinen logiikkaan ja tekoälyyn keskittynyt seminaarisarja aloitettiin keväällä 2020, mutta koronapandemia keskeytti seminaarisarjan, jota jatkettiin syksyllä 2020 hybridimuotoisena. Kirjan artikkelit perustuvat seminaarisarjassa pidettyihin esityksiin.

Tekoäly ei ole mikään uusi asia. Ihmisen kiinnostus niin robotiikkaan kuin tietokoneisiin on vanhaa perua. Tietenkään tällainen havainto ei sumenna nykyisen tietotekniikan ja tekoälytutkimuksen laadullisesti uusia piirteitä. Samalla on muistettava, että matematiikkaan ja logiikkaan liittyvä algoritmisen ajattelu, jonka juuret ulottuvat aina antiikkiin (ja kauemminkin) on edelleenkin tekoälyn perustana. Algoritmisen lähestymistavan valta-asemaa ovat horjuttaneet sumeaan logiikkaan perustuvat järjestelmät ja konnektionistiset järjestelmät, joiden avulla on toteutettu erittäin kiinnostavaa teoreettista työtä. Niihin perustuvat sovellutukset ovat myös käytännössä osoittautuneet toimiviksi.

Algoritmisen ajattelun perusteet ovat kehittyneet matemaattisen ajattelun kehityksen mukana. Algoritmisuus on liittynyt matemaatikkoja kiehtoneeseen matemaattisen ajattelun mekanisointiin. Samalla logiikan ja matematiikan historiaan liittyy myös algoritmisen ja mekanisoitavissa olevan ajattelun kehittyminen. Tästä esimerkkinä riittää mainita seuraavat: Boolean kehittämä, hänen nimeään kantava algebran järjestelmä, toimii edelleen tietokoneiden logiikan perustana. Vastaavasti Fregenin kehittämä predikaattilogiikan järjestelmä loi mahdollisuuden inhimillisen tiedon systemaattisen loogisen esittämisen. Predikaattilogiikka on edelleen monien tekoälyjärjestelmien perustana oleva logiikka. Lisäksi on mainittava Turing, joka muotoili tietokoneen teoreettisen

mallin (Turingin koneet), joka on edelleen teoreettisesti kiinnostava. Kirjoituksissaan Turing pohti näiden koneiden roolia kehitellessään oppivia koneita ja tekoälyn perusteita.

Tekoäly ei ole vain älyllistä (matemaattista ja loogista) ajattelua, vaan myös toimintaa, mikä kytkeytyy robotiikkaan. Toimintaa on myös mahdollista jäsentää loogis-käsitteellisesti selkeällä tavalla, jota tehdään filosofisessa teon teoriassa. Tiivistetysti voidaan sanoa, että teon teoria mahdollistaa toiminnan algoritmisoinnin. Siten teon teoria voidaan nähdä keinotekoisien älykkään toiminnan teoreettisena perustana.

Ei ole syytä ajatella, että algoritmisen ajattelu olisi ainoa mahdollinen lähestymistapa teoreettiseen tai toiminnalliseen älykkyyteen. On hyvä ymmärtää, ettei älykkyys pelkisty informaation loogiseen prosessointiin ja ettei älykäs toiminta ole vain todellisuuden algoritmista muuntelua. Älykkyyteen liittyy oivaltaminen ja oppiminen, joihin liittyy yllätyksellisyyttä ja intuitiivisuutta, joiden täsmällinen luonnehdinta on vaikeaa ellei mahdotonta.

Tekoälyjärjestelmät ovat osoittautuneet hyödyllisiksi ja tehokkaiksi moniin tarkoituksiin. Useat tekoälyn liitetystä toiveista ovat osoittautuneet todenmukaisiksi, mutta samalla osa peloista on saanut uutta uskottavuutta. Tekoäly on alkanut muistuttaa inhimillistä älykkyyttä, mutta ei ole lainkaan selvää saavutetaanko lopulta aitoa ja ehkä inhimillistä älykkyyttä parempaa älykkyyttä (vahva tekoäly) vai ihmisille hyödyllisiä sovellutussensitiivisiä tekoälyjärjestelmiä (heikko tekoäly). On kuitenkin tärkeää avoimin mielin tutkia tekoälyn mahdollisuuksia niin konkreettisten sovellutuksia kehittämällä kuin teoreettisesti tutkimalla. Tekoäly on tutkimuksellisesti kiinnostava, nopeasti kehittyvä ala.

Aloittaessamme seminaarisarjaa ja sittemmin kootessamme siihen kytkeytyvää kirjaa oli taustalla ajatus alan kiinnostavuuden tuominen yliopistomme ihmisten tietoisuuteen. Siksi toivomme, että kirja tuo esiin innostavia ja kiinnostavia kysymyksenasetteluja, joihin tulevaisuudessa on mahdollisuus tarttua ja tutkia ja opiskella asiaa yhä syvemmälle. Toivomme, että kirja tarjoaa tutkijoille, opettajille ja opiskelijoille innostavia näkökulmia, joita on mahdollista lähteä kehittelemään ja syventämään. Tutkimus ja opiskelu on aina yhteistoimintaa ja kannustammekin kaikkia lukijoita innostumaan niin tutkimaan kuin toimimaan yhteistyössä alan muiden toimijoiden kanssa.

RESUME

Postgraduate seminar series with a title artificial intelligence and logic held at the Department of Military Technology of the National Defence University in 2020 and 2021. This book is a collection of some of talks that were presented in the seminar. The papers address ethics in autonomous military technology, algorithmic thinking, application of fuzzy logic in the Armed Forces, logic and skill, use of artificial intelligence technologies in information and psychological warfare, cyber security risks of artificial intelligence applications and algorithms, information acquisition strategies and artificial intelligence to support learning. This set of papers tries to give some insight to current issues of the artificial intelligence and logic.

Seminars at the Department of Military Technology have a tradition to make publication based on the presentations, but only as an internal publication of the Finnish Defence Forces and in such a way has not hindered publication of the papers e.g. in international conferences. Publication of these papers in peer reviewed conferences has indeed been always the goal of the seminar, since it teaches writing conference level papers. We still hope that an internal publication in the department series is useful to the Finnish Defence Forces by offering an easy access to these papers.

Professor Jouko Vankka

SISÄLLYS

Kiitokset	i
Johdanto	ii
Resume	v
Algoritminen ajattelu <i>Kalle Saastamoinen</i>	1
Taidon logiikka <i>Eero Kallio</i>	8
Episteeminen logiikka <i>Arto Mutanen</i>	21
Tiedonhankinnan strategioista <i>Ilpo Halonen</i>	34
Turing ja tekoäly <i>Arto Mutanen</i>	40
Ethics in autonomous military technology – a call for ethical competency <i>Mireille Farzan</i>	56
Using Simulation to Evaluate SARS-CoV-2 Transmission in Finnish Defence Forces' Leave Arrangements <i>Ville Rissanen et al.</i>	69
Sumeaa logiikkaa sotaväessä <i>Riitta Penttinen</i>	76
Tekoälysovellusten ja algoritmien kyberturvallisuusriskit <i>Janne Mutanen</i>	88
Tekoäly oppimisen tukena – kehityksen vaiheita ja näkökulmia <i>Antti Rissanen</i>	98
Tekoälyavustettu Tilannejohtaminen <i>Lauri Kilpeläinen</i>	107
Tekoäly ja psykologinen vaikuttaminen <i>Saara Jantunen</i>	118
Kirjoittajat	127

ALGORITMINEN AJATTELU

Kalle Saastamoinen

“Algoritmi ei ole vain vaiheittainen ongelmanratkaisumenettely, tietokonekoodirivien sarja tai mekaaninen kielellinen ilmaisu, vaan se on myös ontologinen rakennelma, jolla on syvällisiä filosofisia, sosiaalisia, suunnittelu- ja taiteellisia vaikutuksia.”

Terzidis, K. (2006). *Algorithmic architecture*. Routledge.

Jobdanto

Algoritmi on äärellinen sarja selkeästi määriteltyjä toimintoja jonkin tehtävän suorittamiseksi, tavoitteen saavuttamiseksi tai ongelman ratkaisemiseksi (Webster, 2022). Tällainen toiminta on tyypillistä ihmisille ja toistuu jatkuvasti meidän jokapäiväisessä elämässä osana arki- ja työrutiinejamme. Suunnitelmallinen ajattelu on yleensä algoritmista, koska kaikki suunnitelmallinen toiminta sisältää vaiheita ja loogisia rakenteita. Puolustusvoimissa algoritmisen ajattelu esiintyy erityisesti osana kaikkea operatiivista toimintaa.

Algoritmia käyttivät jo babylonialaiset matemaatikot n. 2500 ekr. jakolaskuun (Barbin et al, 2012). Myöhemmin kreikkalaiset matemaatikot etsivät alkulukuja Eratostheneen seulalla ja kahden luvun suurinta yhteistä tekijää Eukleideen algoritmilla 240 ekr (Cooke, 2011). Arabialaiset matemaatikot (mm. Al-Kindi) käyttivät salakirjoitus algoritmeja, jotka perustuivat taajuusanalyysiin jo 900-luvulla (Dooley, 2013). Sanaa algoritmi käytettiin 900-luvulla kuvaamaan Persialaisen matemaatikon Muhammad ibn Mūsā al-Khwārizmīn kirjassaan *Al-jabr* esittämiä sääntöjä ja tekniikoita yhtälöiden ratkaisemiseksi. Myöhemmin algoritmi yleistyi tarkoittamaan mitä hyvänsä joukkoa sääntöjä tai tekniikoita ongelmien ratkaisemiseksi (Barbin et al, 2012).

Algoritmeja käytetään laajalti toimintojen automatisoimiseen prosesseissa, tuotannossa, suunnittelussa, ohjelmoinnissa, päättelyssä ja tiedon käsittelyssä. Esimerkiksi koneoppiminen ja sitä kautta tekoäly perustuu algoritmien käyttöön. Koneoppimisalgoritmit korvaavat jo nykyisellään runsaasti ihmisten tekemää työtä. Niiden etuihin kuuluu, että opittuaan jonkin toiminnan ne kykenevät aina tekemään saman edullisesti, väsymättä ja vähintään opitulla tarkkuudella.

Algoritmisen Ajattelu

Algoritmiselle ajattelulle on tyypillistä tavoitteellisuus, mikä näkyy yleensä jonkin tehtävän suorittamisena, toiminnon toteuttamisena tai ongelman ratkaisemisena. Algoritmisessa ajattelussa päädytään ongelman ratkaisuun vaiheiden kautta. Tästä syystä algoritmi voidaan esittää sarjana askeleita käyttäen vuokaaviota tai ns. pseudokoodia, joka esittää algoritmin perusrakenteen, molemmat voidaan helposti kääntää mille tahansa ohjelmointikielelle. Samankaltaisia ongelmia pystytään ratkaisemaan jo käytetyin algoritmein ja metodein. Tyypillisesti algoritmista ajattelua kuvaamaan käytetään aritmeettisia ja loogisia operaattoreita, sekä silmukoita ja iteraatioita toistamaan tehtä-

vää riittävän monta kertaa. Algoritmisessa ajattelussa tyypillisesti esiintyviä teemoja ovat (Mannila et al, 2014):

Data: kerääminen, manipulointi, analysointi, esittäminen, tulkinta, ennustaminen ja päättely.

Prosessointi: ongelman osittaminen, algoritmisointi, abstraktointi ja rinnakkaistaminen.

Automaatio ja simulointi: toistuvat tehtävät tehdään koneella ja simuloimalla luodaan uutta tietoa.

Algoritmi on määritelmänsä mukaisesti äärellinen joukko peräkkäisiä toimintoja. Hyvä algoritmi suorittaa sille annetun tehtävä tai ratkaisee ongelman riittävällä tarkkuudella ja nopeudella. Professori Donald Ervin Knuthin mukaan tietokone algoritmilla on aina seuraavat perustavanlaatuiset ominaisuudet (Knuth, 1968):

Algoritmin täytyy aina päättyä äärellisen askelmäärän jälkeen.

Jokaisen algoritmin askeleen täytyy olla täsmällisesti määritetty.

Algoritmilla on nolla tai useampia syötteitä.

Algoritmilla on yksi tai useampia ulostuloja: suureita, joilla on määritetty suhde syötteisiin.

Algoritmin oletetaan olevan tehokas, operaatioiden täytyy olla riittävän yksinkertaisia, jotta ne voidaan tehdä täsmällisesti ja äärellisessä ajassa.

Kohta 3 Knuthin määrittelyssä voi johtaa jopa filosofisluonteisiin pohdiskeluihin aiheena: ”Millainen on algoritmi, jolla on nolla syötettä?”.

Algoritmista ajattelua esiintyy kaikkialla, myös taiteissa. Käsitteellisessä taiteessahan määritelmällisesti idea tai käsite, jonka taiteilija esittelee, on tärkeä, ei teos itsessään. Mikäli ajatellaan ideat luokiksi, joista ammennetaan uusia teoksia, päästään käsitteeseen algoritminen taide. Algoritmisessa taiteessa luodaan esimerkiksi kuva lukuisien kuvien luokasta, jolloin luovuuden ytimessä on algoritmi, joka kuvan tekee. Taidekasvatuksella on tärkeä rooli yhteyden luomisessa taiteellisen ja teknologisen ajattelun välillä (Nake, 2017).

Luovuuden yhteydessä puhutaan laskennallisesta luovuudesta “computational creativity”, jonka leikkauskohdasta löytyy tekoälyä, kognitiivista psykologiaa ja taiteita. Tämän tavoitteina on tehdä ohjelmia, jotka jäljittelevät ihmisen luovuutta, ymmärtää ihmisten luovuutta algoritmiselta kannalta ja luoda ohjelmia, jotka vahvistavat ihmisten luovuutta olematta välttämättä itse luovia. Kolmannessa kohdassa voidaan puhua algoritmeista tukiluovuutena. Algoritmeille erityisen sopiva konsepti on ainakin evoluutionaalinen luovuus (Shneiderman, 1999), joka määritelmällisesti koostuu seuraavista osista:

Kokoa (collect), opi jo tehdyistä teoksista ja tiedoista, esim. Google.

Ota huomioon (relate), keskustele vertaisten ja asiantuntijoiden kanssa prosessin kaikissa vaiheissa, esim. asiantuntijajärjestelmät, AHP.

Luo (create), tutki, yhdistä ja vertaile mahdollisia lopputulemia, esim. vahvistusoppiminen.

Lahjoita (donate), eristä hyödylliset tulokset ja lisää ne tietokantaan, esim. vahvistusoppiminen.

Toinen helposti algoritmisoitavissa oleva käsitteen määrittely on Margaret Bodenin (Boden, 2010) määrittely yhdistelevästä luovuudesta (combinatorial creativity):

Tutun esineen sijoittaminen tuntemattomaan ympäristöön.

Kahden keinotekoisesti erilaisen kohteen tai genren yhdistäminen.

Tutun esineen vertaaminen keinotekoisesti toisiinsa liittymättömään ja semanttisesti etäiseen käsitteeseen.

Uuden ja odottamattoman ominaisuuden lisääminen olemassa olevaan konseptiin.

Ikonisen kuvan käyttäminen yhdestä verkkotunnuksesta verkkotunnuksessa liittymättömään tai yhteen sopimattomaan ideaan tai tuotteeseen.

Algoritmisella ajattelulla on luonnollisesti hyvin keskeinen rooli matematiikan oppimisessa. Peruslaskutoimitusten algoritmien opetus etenee siten, että opetus aloitetaan peruskoulun alaluokilla ja niiden käyttöä varmennetaan ja sovellusta lisätään myöhemmillä vuosiluokilla. Ala-asteen loputtua oppilaalle pitäisi olla ainakin kehittynyt lukukäsitys, lukujono ja kyky laskea yksinkertaisia laskutoimituksia. (Ilveskoski et al, 2004)

Kasvatustieteissä tieto jaetaan käsitteelliseen ja menetelmälliseen tietoon. Käsitteellisellä tiedolla tarkoitetaan ymmärrystä matemaattisten käsitteiden välisistä suhteista. Se rakentuu uuden tiedon linkittyessä jo aikaisemmin opittuun tietoon tai jo opittujen asioiden linkittyessä toisiinsa. Käsitteellisellä tiedolla on kaksi tasoa. Primaarisella tasolla (primary level) henkilö osaa yhdistää saman kontekstiin kuuluvia asioita toisiinsa. Tällöin hän osaa esimerkiksi soveltaa yhteenlaskualgoritmin muistinvientisääntöä erilaisiin yhteenlaskutehtäviin, mutta ei hallitse muistinvientiä kertolaskun yhteenlaskuvälivaiheessa. Kun yhteys pystytään muodostamaan myös eri konteksteihin kuuluvien käsitteiden välillä, konseptuaalinen tieto on reflektiivisellä tasolla (reflective level). (Hiebert & Lefevre, 1986, 4–5) Reflektiiviselle tasolle on päästy esimerkiksi silloin, kun henkilö ymmärtää, että lukujen asettelu paikkajärjestelmän mukaisesti yhteenlaskualgoritmissa on eräs esimerkki yleisestä ideasta siitä, että allekkain laskettaessa luvut järjestetään niiden paikka-arvon mukaisesti. Menetelmällinen tieto puolestaan viittaa tietoon matemaattisesta symbolijärjestelmästä sekä säännöistä, algoritmeista ja muista menetelmistä, joita käytetään matemaattisten ongelmien ratkaisemisessa. (Hiebert & Lefevre 1986, 3–6). Menetelmällinen tieto voi näkyä toiminnon vaiheiden suorit-

tamisena automatisoituneesti, kun taas käsitteellinen tieto vaatii aina ajattelua. (Haapasalo 2003, 3–4.) Algoritmien pinnallinen oppiminen johtaa taidon nopeaan unohtamiseen, ellei laskualgoritmia jatkuvasti harjoitella. Suorituksessa tämä näkyy virheiden lisääntymisenä ja algoritmien sääntöjen sekoittumisena (Lehtinen, 1993).

Entisen Neuvostoliiton koulujärjestelmän perustui hyvin pitkälti algoritmisen ajattelun kehittämiseen. Neuvostoliittolainen kasvatuksen pioneeri Lev Landan kuvasi ihmisen oppimista loogisten rakenteiden avulla ja tarjosi teoreettisen perustan erityisten opetustietokoneiden kehittämiseksi. (Babintseva, 2021)

Vahvistusoppiminen

Vahvistusoppimisessa systeemi oppii jatkuvasti parantamaan toimintaansa. Vahvistusoppiminen on trial-and-error menetelmällä tapahtuvaa oppimista. Tässä menetelmässä saadaan tieto, kuinka hyvä tai oikein on algoritmilla saavutettu lopputulos. Toisin sanottuna algoritmi tietää milloin se on päätenyt haluttuun lopputulemaan, ja näin se hiljalleen oppii positiivisen ja negatiivisen palautteen kautta toimimaan toivotulla tavalla myös uusien syötteiden kanssa. Alun perin idea oli yksinkertaisesti ajatus oppimisjärjestelmästä, joka haluaa jotain ja joka mukauttaa käyttäytymistään maksimoidakseen positiivisen palautteen ympäristöstään. Tämä oli "hedonistista" oppimista tai, kuten nyt sanotaan, vahvistusoppimista. (Sutton ja Barto, 2018)

Mikäli systeemin pitää toimia ympäristössä, joka jatkuvasti muuttuu, tai se on hyvin vaikeasti määriteltävissä, vahvistusoppiminen on luonnollinen vaihtoehto. Reaaliaikaisessa datavirrasta oppimisessa kierretään keskeiset syvien neuroverkkojen ongelmat eli vaatimukset valtavasta opetusdatan määrästä, huono virheensietokyky, ongelman rakenteen ymmärtämisen vaikeudet ja neuroverkossa tapahtuvan oppimisen läpinäkyvättömyys (black-box).

Tavoitteena on kehittää kone, joka pärjää ihmistä paremmin tehtävissään. Vahvistusoppimisen tunnetuimpia menestystarinoita ovat pelit, kuten Shakki ja Go, muita menestyksekkäitä alueita ovat nykyään säätötekniikka ja robotiikka. Kaikissa näissä on olennaista, että päätöksenteko tehdään muuttuvissa ympäristöissä. Päätöksenteko on usein monimutkainen prosessi, jonka alle voidaan sijoittaa kaikki muut tietotekniikkaan liitettävät asiat, kuten: tiedon hankinta ja sen säilyttäminen, kommunikointi, toiminnot ja niiden toimeenpano ja koneoppiminen.

Päätöksenteko voidaan armeijassa ja muuallakin jakaa kolmeen kategoriaan strategiseen, taktiseen ja reaaliaikaiseen. Näissä kaikissa on paitsi erilainen ajallinen vaikutusjakso myös eri tason johtajat tekevät näitä päätöksiä. Strategisessa päätöksenteossa meidän tulee kyetä simuloimaan päätöksiä tulevaisuudessa, jotka eivät ole riippuvaisia nykytilanteesta. Taktisessa päätöksenteossa on kyettävä simuloimaan päätöksiä, jotka ovat riippuvaisia nykytilanteesta. Reaaliaikaisessa päätöksenteossa meidän tulee kyetä simuloimaan nyt tehtävien päätöksien vaikutuksia tulevaisuudessa. Vaikka kaikkia näitä simuloitteja mallinnetaan eri tavoin johtuen niiden erilaisesta aikajaksotuksesta ja monimutkaisuudesta niitä kaikkia sitoo päätöksien ja tapahtumien sitoutuminen toisiinsa ajallisesti. Informaation kasvaessa tilanteiden ja päätöksien suhteesta kyetään

simulointimallien tarkkuutta kasvattamaan. Olennaista tässä on, että kaikki päätöksenteko on sarjana etenevää, eli ne voidaan luokitella kuuluviksi alueeseen nimeltä Sequential Decision Analysis (SDA), joka on Warren B. Powellin määritelmän mukaisesti (Powell, 2019) hyvin laaja alue.

Jobtopäätökset

Algoritminen ajattelu on keskeinen osa ihmisyyttä ja kaikki tekevät sitä ajatellessaan johdonmukaisesti, miettiessään toimintojensa vaikutuksia ja kuitenkin yleensä täysin tiedostamattaan.

Algoritmin toiminnan syvälinenkin pohdinta on tarpeen, kun rakennetaan automaattisesti toimivia systeemeitä ja erityisesti pyrkimyksessämme rakentaa ns. älykkäitä systeemeitä, jotka tukevat päätöksiä tai jopa tekevät päätöksiä puolestamme.

Rakennettaessa päätöksentekoon kykeneviä systeemeitä ei pidä hyödyntää pelkästään ns. koneoppimien menetelmiä vaan käyttää kaikkia tunnettuja ja toimiviksi havaittuja menetelmiä esimerkkeinä mainittakoon stokastinen päätöksenteko, optimointi, simulointi ja peliteoria. Tärkeää on hyödyntää myös ihmisten tietämys rakennettaessa ja arvioitaessa näiden systeemeiden tuottamia päätöksiä.

Puolustusvoimien kannattaa rakentaa itselleen avoin verkko, joka kerää turvallisuuden liittyvää dataa Internetistä käyttäen verkossa toimivia hakurobotteja. Lisäksi tarvitaan suljettu keskitetty tietovaranto, jonne kerätään operatiivisesti hyödyllistä dataa.

Puolustusvoimien kannattaa panostaa upseereiden lisäkoulutukseen erityisesti operatiivista päätöksentekoa tukevien tieteenalojen osalta. Lisäksi kannattaa panostaa tietotekniikan päätösprosessien käytännön ja teorian osaamisen hyödyntämiseen yliopistoista ja teollisuudesta lisäämällä yhteistyötä ja palkkaamalla.

Tekoälyssä tulevaisuudessa tullaan tutkimaan entistä enemmän vahvistusoppimista (SDA) ja sen liittämistä jo olemassa oleviin systeemeihin. Tekoäly systeemit tulevat entistä älykkäämmiksi, kun systeemit alkavat oikeasti oppia, eikä niitä rakenneta vain luottaen suuriin datamääriin.

Tietokoneiden käytön ja ohjelmoinnin lisääntyessä harrastuksena kaikilla aloilla ihmisten tietoisuus omasta algoritmisesta ajattelusta kasvaa.

Lähteet

Babintseva, E. (2021) “Overtake and Surpass”: Soviet Algorithmic Thinking as a Re-invention of Western Theories during the Cold War’, in M. Solovey and C. Dayé (eds) *Cold War Social Science: Transnational Entanglements*. Cham: Springer International Publishing, pp. 45–72. Available at: https://doi.org/10.1007/978-3-030-70246-5_2.

Barbin, E. et al. (2012) *A history of algorithms: from the pebble to the microchip*. Springer Science & Business Media.

Boden, M. (2010) *Can computer models help us to understand human creativity?* « *On the Human*. Available at: <https://nationalhumanitiescenter.org/on-the-human/2010/05/can-computer-models-help-us-to-understand-human-creativity/> (Accessed: 21 November 2022).

Cooke, R.L. (2011) *The history of mathematics: A brief course*. John Wiley & Sons.

Definition of ALGORITHM (no date). Available at: <https://www.merriam-webster.com/dictionary/algorithm> (Accessed: 20 September 2022).

Dooley, J.F. (2013) *A Brief History of Cryptology and Cryptographic Algorithms*. Cham: Springer International Publishing (SpringerBriefs in Computer Science). Available at: <https://doi.org/10.1007/978-3-319-01628-3>.

Hiebert, J. and Lefevre, P. (1986) 'Conceptual and procedural knowledge in mathematics: An introductory analysis', in *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, pp. 1–27.

Ilveskoski, M. and Suvilehto, T. (2004) 'Peruslaskutoimitusten algoritmien hallinta toisen asteen ammatillisten opintojen alussa teknisillä aloilla'. Available at: <https://jyx.jyu.fi/handle/123456789/7888> (Accessed: 20 September 2022).

Knuth, D.E. (1973) *The art of computer programming. 1: Fundamental algorithms*. 2nd ed. Reading, Mass: Addison-Wesley.

Lehtinen, E. and Kinnunen, R. (1993) 'Matemaattisista oppimisvaikeuksista', in M. Vauras (ed.) *Oppimisvaikeudet ja opetuksen kehittäminen. Katsaus Turun yliopiston Oppimistutkimuksen Keskuksen toimintaan ja tutkimukseen*. Rauma: Oy West Point.

Mannila, L. *et al.* (2014) 'Computational Thinking in K-9 Education', in *Proceedings of the Working Group Reports of the 2014 on Innovation & Technology in Computer Science Education Conference*. New York, NY, USA: Association for Computing Machinery (ITiCSE-WGR '14), pp. 1–29. Available at: <https://doi.org/10.1145/2713609.2713610>.

Nake, F. and Grabowski, S. (2017) 'Think the image, don't make it! On algorithmic thinking, art education, and re-coding', *Journal of Science and Technology of the Arts*, 9(3), pp. 21–31.

Powell, W.B. (2021) 'From reinforcement learning to optimal control: A unified framework for sequential decisions', in *Handbook of Reinforcement Learning and Control*. Springer, pp. 29–74.

Shneiderman, B. (1999) 'Creating creativity for everyone: User interfaces for supporting innovation. University of Maryland', *Dep [t of Computer Science. Technical Report*.

Sutton, R.S. and Barto, A. (2018) *Reinforcement learning: an introduction*. Second edition. Cambridge, Massachusetts London, England: The MIT Press (Adaptive computation and machine learning).

Terzidis, K. (2006) *Algorithmic architecture*. Routledge.

‘The conflict between conceptual and procedural knowledge: Should we need to understand in order to be able to do, or vice versa?’ (2003) in Haapasalo, L., *Towards meaningful mathematics and science education: proceedings on the IXX Symposium of the Finnish Mathematics and Science Education Research Association*. Joensuu: Joensuu Yliopiston (Kasvatustieteiden tiedekunnan selosteita / Joensuu Yliopisto, 86).

TAIDON LOGIIKKA

Eero Kallio

Jobdanto

Käytännön elämässä taitoa arvioidaan usein sen käytännöllisyyden (toimivuus, vaikuttavuus) kannalta. Taitavuus kohdistuu tällöin tavoitteen saavuttamiseen. Taitoa voidaan arvioida myös esteettisyyden kannalta, jolloin arvioinnin kohteena ovat toiminnan näkyviin piirteisiin liittyvät kriteerit. Taidolla on tyypillisesti joko fyysinen tai mentaalinen yleisolemus. Vaikka erottelu on perusteltu, tulee huomioida, että näiden luokkien välillä on useita kytkentöjä.

Ruumiilliset taidot tulevat aina esiin, kun toiminta liittyy agentin kehon toimimiseen. Mentaaliset taidot korostuvat, kun älylliset toiminnot, kuten strategiset ja taktiset taidot, alkavat vaikuttaa toiminnan tuloksellisuuteen. Toiminnan kontekstilla on suuri merkitys sille, millaiset taidot ovat oleellisia. Teknologinen selitys sitoo yhteen rakenteellisen (fyysisen) ja toiminnallisen (mentaalisen) selityksen.

Agentin ja teknologian suhdetta voi tässä esitetyn valossa käsitellä tavoitteeseen suunnattuna intentionaalisisena toimintana ja tämän rinnalla tapahtuvina fyysisinä tekoina. Tavoitteet ohjaavat tekoja, joita voidaan saavuttaa välineiden ja niiden taitavan käytön avulla. Ihmisen mahdollisuuksia rajoittavat tässä mielessä vain loogiset mahdollisuudet, joita voidaan toteuttaa fysikaalisten mahdollisuuksien rajoissa.

On ilmeistä, että kaikki taito on jossain määrin tietopitoista; mutta samalla on hyvä tiedostaa, että kaikki tieto on taitopitoista. Tarkoitukseni on tarkastella taidon ja tiedon kontekstuaalisuutta eli kytkentää tiedon ja taidon esiintymistilanteeseen. Teknologia ilmenee kaikkialla ympärillämme nykyisessä tieteellis-teknisessä elämässämme. Teknologisoitunutta ympäristöämme voidaan kutsua teknosysteemiksi, joka luo puitteet toiminnallemme. Ihmisen toiminta tässä ympäristössä luo uusia näkökulmia tiedon ja taidon rooliin yksilön toiminnoissa.

Tieto ja taito

Niiniluoto (1989, 52) kuvaa Aristoteleen määrittelemää taitoa (tekhne) oikeana järkipäisenä tekemisvalmiutena, johon liittyy käsitys siitä, miten valmistettava tulos syntyy. Tämän perusteella taidon takana on sen harjoittamiseen liittyvää käsittämistä tai tietoa. Tällaista taidon harjoittamiseen liittyvää tietoa on uuden ajan alusta lähtien alettu kutsua teknologiaksi – tekhnin logokseksi eli opiksi teknologiasta. Vaikka teknologia sanana viittaakin teollisuuden keinoihin tai teknisiin välineisiin, on sana laajasti tulkittavissa juuri *tekhnen* (taito) *logoksesi* (tieto, oppi, perusta) – opiksi taidosta.

Oppi taidosta kytkeytyy Gilbert Rylen tekemään erotteluun kahden erilaisen tietämisen lajin välille. Hän kuvaa taitoon liittyvää tietämistä propositionaalisisena *tietää että* (know that) ja tekemiseen liittyvänä *tietää kuinka* (know how). Rylen kritiikki kohdistuu väittämään, että know how olisi aina käännettävissä takaisin väitelauseeksi. Niini-

luoto erottaa tässä vaiheessa pelkän osaamisen ja know how'n. Hän käyttää taitamiiseen liittyvästä tiedosta termiä taitotieto, jolla hän tarkoittaa kielellisesti ilmaistavissa olevaa tietoa jonkin taidon harjoittamisen tehokkaimmasta tavasta – taitotieto ilmaisee keinojen ja tavoitteiden suhteita koskevaa välineellistä tietoa. (Niiniluoto 1989, 52-53; Niiniluoto 2002, 137; Ryle 2009, 18-19.)

Tieto on perinteisesti jaettu kahteen toisiinsa sidoksissa olevaan lajiin: teoreettiseen (theoretical reason) ja käytännölliseen tietoon (practical reason). Erottelusta huolimatta nämä kaksi ovat erottamattomasti kietoutuneet toisiinsa. Hintikka (1974, 81) käsittelee tekijän tiedon (makers knowledge) perinnettä ainoana todellisena inhimillisenä tietona – meillä voi olla teoreettista tietoa vain asioista, joita olemme itse tuottaneet, olemme tuottamassa tai voimme tuottaa. Vaikka eri yksilöt olisivat valmistamassa tai tekemässä samaa asiaa, on heidän käsityksensä niin kohteesta kuin siihen pääsemisestä erilaisia. Tekijän tiedon takana on ajatus, jonka mukaan vain monimutkaisen mekanismin tekijä voi tuntea sen, sillä vain hän tuntee sen tarkoituksen. Pelkästään mekanismin toimintaa seuraamalla ei voi ennustaa sen tulosta.

Tekijän tieto (makers knowledge) on hyvin läheisesti yhteydessä käytännölliseen tietoon tai järkeen (practical reason tai practical knowledge), jota kutsutaan toisinaan myös intentionaaliseksi tiedoksi. Käytännöllinen tieto on aikaansaamisen, toisiinsa liittyvien toimenpiteiden konkreettisten suhteiden ymmärtämistä. Käytännöllisen tiedon tai tekijän tiedon luonteen ymmärtää hyvin, kun näitä vertaa spekulatiiviseen tietoon, joka on johdettu päätelemällä tunnetuista asioista ilman kokemusten ja havaintojen vaikutusta. Tekijän tieto on enemmän kuin yksin päättelyllä tai yksin havainnoilla saavutettu tieto – se pitää sisällään havaittavissa olevia konkreettisia tekoja ja niiden taustalla olevan intentionaalisuuden. (Hintikka 1974, 84-85; Niiniluoto 1989, 51.)

Tekijän tietoa tai käytännöllistä tietoa käsiteltäessä on tärkeää olla unohtamatta kiinteää yhteyttä yksilön teoreettisen tiedon ja käytännöllisen tiedon välillä. Yksilön kausaalisten riippuvuuksien tietämyksen kasvaessa myös hänen kykynsä kehittää käytännöllistä tietämystä paranee (Hintikka 1974, 86). Teoreettinen järki on irti yksilön käytännöllistä järkeä rajoittavasta havaitusta maailmasta. Teoreettisesti hahmotettu todellisuus laajentaa mahdollisuuksia harjoittaa käytännön toimia järkeilemällä ennalta tulevia tekoja ja niiden vaikutuksia. Käsitteellinen syiden ja seurausten rakenne tuottaa yksilölle tietoisuutta (awareness tai knowledge) mahdollisuuksista päästä tekijän tiedon taustalla olevaan intentionaaliseen tavoitteeseen. Kokemustiedolla on tässä mielessä käsitteellisiä edellytyksiä (Niiniluoto 2002, 144).

Taitoa käsitellään tyypillisesti ajatellen joko sen fyysistä tai mentaalista yleisolemusta. Vaikka erottelu on perusteltu, tulee huomioida, että näiden luokkien välillä on useita kytkentöjä. Fyysiset taidot tulevat aina esiin, kun toiminta liittyy kehon toimimiseen. Mentaaliset taidot korostuvat, kun älylliset toiminnot, kuten strategiset ja taktiset taidot, alkavat vaikuttaa toiminnan tuloksellisuuteen. Toiminnan kontekstilla on suuri merkitys sille, millaiset taidot ovat oleellisia (Siitonen 2008, 50, 52.)

Klassisen tiedon määritelmän mukaan tiedolla on kolme ehtoa: totuus-, oikeus- ja uskomusehto. Arto Siitonen on rakentanut klassisen tiedon määritelmän mukaisella tavalla analogisen taidon määritelmän. Se pitää sisällään kolme ehtoa:

1. tehtävyysehdon,
2. perusteltavuusehdon ja
3. uskomusehdon.

Tässä tehtävyysehto vastaa tiedon määritelmän totuusehtoa, eli taidon tuottaman teon on oltava tehtävissä. Perusteltavuusehdolla tarkoitetaan taitajan harjaantumista, joka oikeuttaa uskomisen siihen, että tekijä taitaa oikean teon. Harjaantuminen erottaa taidon hyvästä onnesta, perustelee uskomuksen. (Siitonen 2007; Mutanen 2015, 67-68.)

Taito ja toiminta

Tiedon ja taidon yhtymäpisteen, sulauma-alueen, etsiminen lienee mielekästä aloittaa yksilön toiminnan tarkastelusta. Tämä tarkastelu on perusteltua kytkeä inhimillisen toiminnan keskeiseen kohtaan, tekoon; agentin toteuttamaan perusteltuihin uskomuksiin ankkuroituun intentionaaliseen toimintaan: uskomukseen siitä, että hänen aloittamansa teko johtaa haluttuun lopputulokseen. Kokonaisuus yhdistää syy-seuraussuhteen teoreettisen tiedon (tietää että) ja käytännöllisen tiedon (tietää kuinka) itse totuudenmukaisesti päätteleväksi valmiudeksi tehdä (Mutanen 2015, 67).

Jarmo Toiskallio (2009, 51) yhdistää taidot keskeiseksi osaksi toimintaa. Toimintakyvyn teoria lähtee liikkeelle ajatuksesta, ettei elämää ja kykyä toimia voida milloinkaan muotoilla valmiiksi tuotteeksi. Toimintakyky ei ole normatiivista vaan yksilöllisesti ilmenevää kykyä olla todellisuudessa. Toimintakyky ei kuitenkaan ole pelkästään yksilöllisesti muodostuvaa. Arkiajattelu usein mieltää mielen ja ruumiin toisistaan erillisiksi. Toimintakyky sen sijaan hahmottaa ihmisen kokonaisuudeksi, jonka kaikki ominaisuudet, fyysinen, psyykinen, sosiaalinen ja eettinen, vaikuttavat toisiinsa. (Toiskallio 2009, 48, 50-51.)

Ero toimimisen ja käyttäytymisen välillä on merkittävä. Tietämisen ja taitamisen sisäistäminen mahdollistaa toiminnan. Pintaoppimisen perusteella tietoja ja taitoa kytetään toistamaan sellaisina kuin ne on opetettu. Syväoppimisen tuloksena sisäistetyt tiedot ja taidot mahdollistavat soveltamisen uusissa erilaisissa tilanteissa. Kyky hahmottaa opitut tiedot ja taidot osana kokonaisuuksia mahdollistaa niiden soveltamisen – kyseessä ei ole opitun tempun toistaminen vaan harkintaan perustuva teko. (PVKK 1998, 16, 36.)

Toiskallio käsittelee toimintakykyä suhteessa käytännölliseen viisauteen eli fronesikseen. Kyseessä on teoreettisen ja objektiivisen tiedon sijaan subjektiivinen suhde valitseviin tilanteisiin ja toisiin ihmisiin. Käytännöllisen viisauden kannalta ihmisen mieli ja ruumis ovat yksi kokonaisuutena – tiedot ja taidot, teoriat ja käytännöt, ovat kiinteässä yhteydessä toisiinsa. Niitä ei voida siirtää yksilöltä toiselle. Opettaja ei voi antaa

tietoa ja taitoa opiskelijalle, vaan oppijan on itse rakennettava oma tietoisuus ja taitavuus. Oma tietoisuus ja taitavuus ovat yksilön suhde todellisuuteen. Suhde ei ole staatinen vaan se on jatkuvassa muutoksessa. (Toiskallio 2009, 55, 60; PVKK 1998, 22.)

Toimiminen ja taitaminen liittyvät käytännölliseen viisauteen. Tämän ydin on toimimisen luonteen ymmärtämisessä – se ei ole pelkkää suorittamista vaan tarkoituksenmukaista tavoiteorientoitunutta toimintaa. Toimintakyvystä puhuttaessa keskitytään ihmiseen. Sotilaspedagogisesta näkökulmasta toimiminen ja taitaminen ovat ihmiseen liittyviä yksilöllisiä tekijöitä, siinä missä osaamista voidaan käyttää myös koneista ja tekoälystä puhuttaessa. Osaaminen on tässä mielessä vain teknistä suorittamista, käyttäytymistä, siinä missä taitaminen ja toimiminen pitävät sisällään teon vastuullisuuden. Toiminta edellyttää itsenäistä, omaan ajatteluun perustuvaa, päätöksentekoa (Toiskallio 2016, 83).

Bill Pollard (2006, 2) pyrkii esittämään tavanmukaisen toiminnan ja muun käyttäytymisen välisen eron asettamalla vastakkain tavat toiminnan uskottavana selittäjänä ja muun käyttäytymisen jonain, mitä ihmiset kohtaavat ja kokevat. Hän kritisoi teon filosofian ajattelutapaa, jossa psykologiset tekijät näyttäytyvät aina selittävänä tekijänä toiminnan taustalla. Pollard pyrkii osoittamaan, että tapojen selittäminen (habit explanation) on kykenevä selittämään tekoja. Tapojen selittämisen huomioiminen mahdollistaa tässä suhteessa laajemman ihmisagentin ymmärryksen – psykologian ohella käyttäytyminen voidaan selittää myös tavanmukaisena toimintana.

Breivik (2014, 156) erottelee automaattiset tavat ja rutiinit sekä ns. habituaation, jonka ansiosta ammattimaiset urheilijat kykenevät dynaamisesti yhdistämään ajattelua ja automaattista toimintaa suorituksensa aikana. Pollard (2006, 10) käyttää samaa termiä kuvaamaan kertauksen kautta saavutettua kyvykkyyttä, joka mahdollistaa toimimisen ilman ennakkosuunnittelua. Ajattelun ja automaation dynaamisuus ilmeneekin reagoinnin ja tilanteenarvioinnin samanaikaisuutena. Tässä mielessä toimijuuden taso vaihtelee taitavuuden ja dynaamisuuden mukaan.

Tahdonteon vaikutus tapaan muodostaa mielenkiintoisen kytkennän kykyihin ja taitoihin. Agentin mahdollisuus kehittää tapansa muistuttaa taitojen kehittämistä. Breivik (2014, 156-157) esittää taitavan pianistin sormiautomaation ja rutiininomaisen työmatka-autoilun eroa: vaikka ammattimaisen pianistin sormityöskentely toimii kuin automaatio, sen taustalla oleva soittaminen ei ole rutini vaan ennemminkin erittäin keskittyntä tavoitteellista toimintaa, joka kehittää itseään toteutuessaan. Pacherie (2007, 12-13) käsittelee toiminnan tietoisuuden ja toimijuuden kokemisen suhdetta samankaltaisella esimerkillä, jossa automaattinen käsien liikuttaminen voidaan kokea niin kontrollin ulkopuolisena tapahtumana kuin sisäisen ohjauksen tuloksena. Pollard (2006, 7) esittää tämänkaltaisen taidon hankinnan ja taitavuuden toteuttamisen kytkeytyvän vahvasti agentin olemukseen. Soran Reader (2007, 585) tuo esiin tähän liittyvän agentin roolin: agentin kyky vaikuttaa halutessaan asioiden kulkuun korostaa valinnan mahdollisuuden, vapauden merkitystä toimijuuden osana.

Toinen keskeinen tekijä liittyy tapojen hankinnan prosessiin. Kyse on tällöin tapojen kehittämisestä taitavuuden näkökulmasta aloittelijasta ammattilaiseksi. Tässä näkö-

kulmassa taidon oppiminen ja kehittäminen edellyttää keskittyntä ja tavoitteellista harjoittelua. Automaation taso saavutetaan, kun taidon toteuttaminen muuttuu helpommaksi kuin taidon toteuttamatta jättäminen. Taidosta tulee osa agentin olemusta. (Pollard 2006, 7)

Sookermanny (2011) käsittelee sotilastaitoa määritellesään perinteistä jakoa yleisten taitojen ja erityistaitojen välillä. Hän esittää taidot sääntökeskeisinä, universalistisina, tai olosuhteet huomioivina, kontekstuaalisina. Asetelmassa sääntökeskeinen taitaminen tulee ymmärtää käskyjä vastaanottavan varusmieskoulutetun sotilaan tietoa vastaanottavana roolina ja vastaavasti kontekstuaalinen näkökulma upseerin tai kouluttajan tietoa hallitsevana tilana.

Pollard (2006, 8) esittääkin, että intervention mahdollisuus ja taidon hankinnan perustavanlaatuinen yhteys toimijuuteen luovat perusteen pitää tapojen toteuttamista toimintana. Toiskallion (2009, 53) mukaan toimimisen pitää lähteä yksilöstä itsestään, jotta kyseessä olisi todellinen teko eikä pelkkää normien mukaista käyttäytymistä. Tässä mielessä toimijuuden taso ja taitavuus eivät kulje käsi kädessä. Toistamalla opittu taitavuus ilman taidon hyödyntämisen päämäärää ei täytä toimijuuden kriteerejä. Taitavuus esiintyy ennemminkin toiminnan mahdollistavana elementtinä.

Taito ja tekniikka

Käytännölliseen tietoon liittyvä harkinta, käytännöllinen järkeily, kohdistuu Aristoteleen mukaan keinoihin, ei lopputuloksiin. Koska tekeminen ylipäätään on halutun lopputuloksen saavuttamista, ei lopputuloksen harkinta ole tässä mielessä oleellista. Haluttu lopputulos perustuu oletukseen, kuten potilaan parantamiseen tai toimivan lainsäädännön luomiseen. Harkinnan kohteeksi muodostuu se, miten haluttuun lopputulokseen päästään. Useista vaihtoehdoista valitaan harkinnan perusteella paras. Harkinta on askeleittain etenevä ketju, jolla saavutetaan ymmärrys päämäärään pääsemisestä. (Hintikka 1974, 89-90.)

Syiden ja seurausten suhdetta on pohdittu läpi aikojen pääsemättä asiasta yksimielisyyteen. Yksinkertaistettuna tästä yhteydestä on riittävää todeta: kun ilmiö on toisen ilmiön syy, täytyy aina ensimmäisen tapahtuessa toisen seurata sitä ollen ensimmäisen seuraus. Syy-yhteyksien käsittäminen ja löytäminen mahdollistaa ennustamisen – ensimmäisen ilmiön tapahtuessa sitä seuraavan ilmiön voi ennustaa. Luonnontieteissä tehtävät kokeet ovat näiden syiden ja seurausten toteutumisten ehtojen tahallaan aiheuttamista. Tekniikalla tarkoitammekin tässä yhteydessä syy-yhteyksiin liittyvän tietämyksemme soveltamista käytäntöön. (von Wright 1961, 177-178.)

G. H. von Wrightin muotoilema tekninen normi kuvaa tietyn keinon ja päämäärän välisen suhteen. Se kytkee yhteen päämäärän ja siihen johtavan keinon. Kyseessä on preskriptiivinen eli normatiivinen lause muotoa ”*Jos olet tilanteessa S ja haluat A, sinun tulee tehdä T.*” Praktisena syllogismina esitettyinä kyseessä on tilanne, jossa tahdon tilanteessa S päästä lopputulokseen A, jolloin minun on tehtävä T:

Haluan tehdä majastani asuttavan (A).

Jos en lämmitä (T) majaa, majasta ei tule asuttavaa (A).

Minun on lämmitettävä maja.

Ensimmäinen premissi ilmaisee intentiota – sitä, mitä haluan saada aikaiseksi. Toinen premissi ilmentää syy-yhteyttä, jolla tavoite saadaan toteutumaan. Johtopäätös (3) on eräänlainen julistus, joka ei välttämättä vielä ole teko, mutta antaa teolle syyn – järkevä agentti ryhtyy välittömästi toimimaan. Tekniseen normiin liittyvä syy-yhteys on kuitenkin mahdollista kuvata deskriptiivisesti, jolloin se on mahdollista testata tieteellisin menetelmin. (Hughes&Kroes&Zwart 2007, 208; Mutanen 2006, 142; Mutanen 2016a, 86, 88.)

Toisen premissin ilmentämä syy-yhteys kytkee tavoitteen (ends) asuttavasta majasta välineisiin (means), jolla tavoite on saavutettavissa. Keino, majan lämmittäminen, voidaan toteuttaa kontekstiin kytkeytyvillä välineillä (means), jotka kytkevät intentionaalisuuden instrumentaaliseen järkeen. Praktinen syllogismi toimii rationaalisuuskriteerinä, joka lopulta kytkee toisiinsa inhimillisen tavoitteen, kontekstuaalisen syy-yhteyden ja teknisen ratkaisun tavoitteen saavuttamiseen.

Syy-yhteyksien tunnistaminen tavoitekeskeisestä toiminnasta on erittäin monimutkaista. Havainnoimalla kohteen toimintaa tavoitetaan ainoastaan sen havaittavissa olevia toimintoja. Se, miksi nämä toiminnot tapahtuvat, jää havaitisijalta näkemättä. Toiminnalla ja toimijan käsityksillä siitä, miten tavoite on saavutettavissa, on kiinteä yhteys. Tämä ei kuitenkaan todenna syy-seuraussuhdetta vaan toimijan käsitystä siitä, miten hän pääsee haluttuun tavoitteeseen. Toiminta perustuu käsitykseen kausaalisista suhteista, mutta tämä käsitys jättää toiminnan ja tavoitteen ulkopuoliset kausaalisuhteet huomioimatta. (Hintikka 1974, 85.)

Inhimillisten toimijoiden mahdollisuus harkintaan erottaa niiden tekojen syy-yhteydet fyysisten objektien välisistä suhteista. Inhimillinen harkinta mahdollistaa erilaiset valinnat näennäisesti samanlaisissa tilanteissa. Inhimillinen syy-yhteys kertoo siitä, miksi joku on tehnyt jotain. Tämä kuitenkin ei ole tae siitä, että näin tapahtuisi varmasti uudelleen. Fyysisten objektien välinen vuorovaikutus käydään puhtaasti kausaalisin yhteyksin, kun taas ihmisten käyttäytymiseen liittyy merkittävässä määrin intentionaalista harkintaa (Kroes 2002, 293).

Herbert A. Simon (1996, 4) määrittelee keinotekoisien (artificial) luonnollisen vastakohtana: *"Produced by art rather than by nature; not genuine or natural; affected; not pertaining to the essence of the matter."* Se on ihmisen taidon, ei luonnon, valmistamaa. Se on keinotekoisia ja epäaitoa – jo kielellisesti ilmaistuna hieman negatiivista ja epäluotettavaa. Kuitenkin keskeisenä osana keinotekoisien olemusta on sen kiinteä kytkös tekijään tai käyttäjään – keinotekoisella on merkitystä vain, kun sen käyttötarkoitus, syy, ymmärretään.

Simon (1996, 6) käsittelee artefaktia sen sisäisten ja ulkoisten ympäristöjen piirteiden kytkentäkohtana. Sen sisäinen ympäristö kuvaa artefaktia substanssina – fyysisenä

järjestelmänä. Sen ulkoinen ympäristö keskittyy toiminnalliseen kontekstiin, jossa artefakti toteuttaa tarkoitustaan. Artefaktin luonteen kannalta keskeistä on ulottuvuuksien erottamattomuus – kello, joka ei näytä aikaa oikein, ei voi toteuttaa kellon tehtävää, mutta voi kelvata koristeeksi.

Tekninen artefakti on samaan aikaan ihmisen tuottama fyysinen objekti ja objekti, jolla on tekninen toiminnallinen tarkoitus – se ei ole puhtaasti fyysisen maailman eikä inhimillisen maailman tuote. Sen toiminta voidaan selittää fyysisen maailman kausaalisina prosesseina, mutta jos siltä poistetaan sen intentionaalinen toiminnallisuus, se menettää artefaktin luonteensa. (Kroes 2002, 294.)

Teknologisessa selityksessä tulee huomioida teknologisen objektin dualistinen luonne: se on toisaalta fyysinen objekti, jonka rakenne noudattaa fysiikan lakeja, ja toisaalta toiminnallinen väline, joka saa merkityksensä inhimillisen toiminnan välikappaleena – lopputulokseen pääsemisen välineenä. Tämän objektin luonne on riippuvainen sen molemmista piirteistä: fyysisestä rakenteesta ja toiminnallisesta kontekstista. Suunnittelussa ei ole kyse pelkästään siitä, miten tämä objekti rakennetaan. Se pitää sisällään myös, vähintään implisiittisen, kuvauksen siitä, miten tämä fyysinen järjestelmä toteuttaa siltä vaaditun toiminnallisuuden – teknologisen selityksen objektin luonteesta. (Kroes 1998; 2002, 291.)

Teknologisen selityksen rooli on sitoa yhteen teknologisen objektin rakenteellisen ja toiminnallisen selityksen kuulu. Toiminta, joka on kuvattu intentionaalisella kielellä, selitetään rakenteen termein, joka kuvataan ei-intentionaalisella kielellä. Teknologisen välineen toimintaperiaatteita ei ole valmiiksi olemassa luonnonlaeissa – ne hyödyntävät luonnonlakeja, mutta niiden toiminnallisuus on kiinni objektin käyttötarkoituksessa. (Luonnon)tieteellinen tieto ei voi teknologian tapaan selittää objektin toiminnallisuutta ennen kuin toiminnallisuus on olemassa. (Kroes 1998.)

Teknisen artefaktin duaaliluonne liittyy kahteen tapaan, jolla käsitteellistämme maailmasta saamamme havainnot. Käsitteellistämme toisaalta maailman fyysisinä asioina, jotka liittyvät toisiinsa kausaalisin kytkennöin, ja toisaalta ihmisagenttien intentionaalisina käsityksinä maailmasta ja intentionaalisina toimina sen suhteen. Ensimmäisessä käsityksessä teemme rakenteellisen tai fyysisen käsitteellistämisen, jossa kausaatiolla on keskeinen rooli. Jälkimmäisessä keskitymme kausaation sijaan päättelyyn (reasoning), joka selittää inhimillisen agentin intentionaalista toimintaa. (Kroes 2002, 293.)

Kroes (2002, 294) kuvaa näiden kahden eri käsitteellistämistavan eroja esimerkillä, jossa käden nostaminen äänestämistilanteessa voidaan selittää molempien tapojen kautta: sille on toisaalta fysiologinen ja toisaalta järkeen perustuva lähtökohta. Teknistä artefaktia ajatellen käsitteellistäminen on samankaltaista. Kyseessä on esine, jolla on fyysinen rakenne ja tekninen toiminnallisuus, jotka ihminen on tuottanut ymmärtäen sen käyttötarkoituksen. Esine voidaan käsitteellistää sen fyysisen rakenteen perusteella, jolloin sen toiminta voidaan selittää kausaalisten kytkentöjen perusteella. Olakseen tekninen artefakti sen tulee kuitenkin pitää sisällään myös intentionaalisen käsitteellistämisen kuvaama toiminnallisuus. Kumpikaan käsitteellistämistapa ei yksin

riitä täyttämään teknisen artefaktin kokonaisuutta – sillä on sekä fyysinen että intentionaalinen luonne.

Agentin ja teknologian suhdetta voi tässä esitetyn valossa käsitellä tavoitteeseen suunnattuna intentionaalisenä toimintana ja tämän rinnalla tapahtuvina fyysisinä tekoina. Tavoitteet ohjaavat tekoja, joita voidaan saavuttaa välineiden ja niiden taitavan käytön avulla. Ihmisen mahdollisuuksia rajoittavat tässä mielessä vain loogiset mahdollisuudet, joita voidaan toteuttaa fyysikaalisten mahdollisuuksien rajoissa (Niiniluoto 2020, 136).

Teknologisen objektin fyysistä olemusta ja sen tuottamaa toimintaa voidaan selittää kausaalisiin kytkennöin. Esimerkiksi männän liike sylinterissä ja liikkeen tuottaman energian siirtäminen voidaan kuvata ilmiöinä ja niiden ominaisuuksina. Näiden ilmiöiden perusteella ei voida kuitenkaan deduktiivisesti päätellä sitä, millaisen toiminnallisuuden objekti tuottaa – se on kontekstiriippuvaista. Ilmiöillä voidaan tuottaa eri tarkoituksissa erilaisia lopputuloksia. Vastaavasti tulemme tulokseen, jossa käyttötarvoksesta ei voida suoraan johtaa objektin rakenteellisia kuvauksia. Suunnittelussa on tässä mielessä kyse tarvittavien toiminnallisuuden kääntämisessä rakenteelliselle kielelle. (Kroes 1998.)

Tekniikka ja taidon malli

Malliteorian avulla voidaan määritellä loogisen totuuden ja loogisen seurauksen käsitteet (Niiniluoto 1980, 96). Malliteoriassa tarkastellaan kielen semantiikkaa eli merkitystä malleissa, jotka logiikassa määritellään täsmällisesti joukko-opillisiksi struktuureiksi. Loogisella totuudella tarkoitetaan totuutta kaikissa mahdollisissa maailmoissa. Siten loogiset totuudet ovat tosia kielen malleissa. Sanomme, että lause A on lauseen B looginen seuraus, jos B:n totuus takaa A:n totuuden – A on siis tosi jokaisessa mallissa, missä B on tosi. Näin nähdään helposti, että deduktion on totuuden säilyttävää päättelyä. Tämän tyyppisessä lauselogiikassa totuusarvoja on vain kaksi, tosi ja epätosi. Moniarvologiikassa totuusarvoja voi olla enemmän kuin kaksi (Niiniluoto 1980, 96-98).

Hughes&Kroes&Zwart (2007, 208) pyrkivät esittämään praktisen syllogismin syy-seuraussuhdetta malliteorian keinoin. Tavoitteena on hyödyntää formalismia selkeyttämään ja yksilöimään analyysin kannalta relevantit tekijät. Tämän kaltainen matemaattisten mallien käsittely edellyttää tiettyä idealisointia ja poikkeamaa luonnollisesta kielestä. Lisäksi tulee huomioida, ettei agentilla ole tiedossaan kaikkia häntä ympäröiviä tosiasioita, vaan hänen tulee järkeillä syiden ja seurausten suhteet oikeiksi uskomiansa perusteiden kautta (Hughes&Kroes&Zwart 2007, 210).

Jos olen rangaistuslaukauskilpailussa K ja haluan voittaa V, minun on tehtävä maali M:

- 1) Voittoa edellyttää maalin syntymistä. $M \leftrightarrow V$
- 2) Onnistunutta laukausta seuraa maali. $L_o \rightarrow M$

3) Maali voi syntyä, jos torjunta epäonnistuu. $\neg T \rightarrow M$

Tavoitteeseen pääsemisen kannalta keskeistä on voittamisen olosuhteiden muodostuminen. Tämä edellyttää minimissään pallon liiketekijöitä, jotka mahdollistavat pallon maaliin päätyksen. Rangaistupotkun olosuhdetta voidaan tarkastella ajan ja etäisyyden suhteen. Potkaisijan kannalta edullinen kohde muodostuu tilaan, joka on maalin rajojen sisällä ja mahdollisimman kaukana maalivahdin lähtöpisteestä. Pallon, kohteen ja maalivahdin sijainnit muodostavat suorakulmaisen kolmion, jossa kulma A on pallon lähtöpiste, kulma B on maalipiste ja kulma C on maalivahti. Kateetti a on maalivahdin etäisyys maalipisteestä B ja hypotenuusa c pallon etäisyys maalipisteestä. Laukauksen onnistuessa pallo kulkee hypotenuusan c ennen kuin maalivahti ehtii reagoida kateetin a suuntaisesti ja saavuttaa kulman B.

Laukaukselta edellytetään minimissään:

- 1) Laukaus toteutuu, jos pallon suunta ja nopeus ovat riittävät osumaan maaliin.
- 2) Laukaus on onnistunut, jos pallon suunta ja nopeus ovat riittäviä maaliin osumiseen ja torjunnan epäonnistumiseen.

Laukauksen toteutumista ja onnistumista voidaan ajatella osumatarkkuuden todennäköisyyden kannalta. Mitä onnistuneempi laukaus on, sitä pienempi on ympyrän muotoisen hajontakuvion säde. Vastaavasti hajonnan kasvaessa hajontakuvion säde kasvaa. Torjunnan onnistumista kuvaa maalin keskipisteessä olevasta maalivahdista määritettävä torjuntasäde, joka voidaan määrittää maalivahdin ulottuvuuden (etäisyys ja nopeus) suhteen. Lopputuloksena tapahtumaa voidaan tarkastella suorakulmiona, jonka keskelle muodostuu osan maalista peittävä ympyrän muotoinen torjunta-ala, ja keskipisteestä etäisimpiin reunoihin ja kulmiin maalit mahdollistava tyhjä alue.

Käytännön järjestyksen kannalta on löydettävissä ainakin kolme keskeistä tekijää (Hughes&Kroes&Zwart (2007, 214):

- 1) heikko riittävä keino: A tekeminen voi aiheuttaa B
 - Maali Φ voi syntyä jossain olosuhteessa, jossa laukaus α toteutuu.
 - Laukaus osuu alueelle, joilla sekä torjunta-alue että laukauksen hajontakuvio ovat päällekkäin.
- 2) (vahva) riittävä keino: A tekeminen aiheuttaa B
 - Maali Φ syntyy aina kun laukaus α onnistuu.
 - Laukaus osuu alueelle, jolle maalivahdin torjunta-alue ei ulotu.
- 3) välttämätön keino: B ei aiheudu, ellei agentti tee A
 - Maali Φ ei voi syntyä, ellei laukaus α toteudu.
 - Laukaus jää täysin maalivahdin torjunta-alueelle tai laukaus menee maalin ulkopuolelle.

Maali Φ syntyy aina kun laukaus α onnistuu: $[\alpha]\Phi$. Laukaus osuu alueelle, jolle maali-
vahdin torjunta-alue ei ulotu. Maali Φ voi syntyä jossain olosuhteessa, jossa laukaus α
toteutuu: $\langle\alpha\rangle\Phi$. Laukaus osuu alueelle, joilla sekä torjunta-alue että laukauksen ha-
jontakuvio ovat päällekkäin. Laukauksen onnistuminen riippuu vaiheesta, jossa pallo
saa liikesuunnan ja -voiman.

On kuitenkin huomioitava, että jos agentti ei osaa tai kykene tekemään A, mikään
keino ei johda B. Välttämättömät keinot vaikuttavat olevan luonteeltaan hankalampia
kuin riittävät keinot. A on riittävä niissä tilanteissa, joissa sillä saavuttaa halutun pää-
määrän. Välttämättömät keinot ovat sen sijaan monimutkaisia (Hughes&Kroes
&Zwart (2007, 217):

- 1) Välttämättömän keinon ei tarvitse olla riittävä, sillä sitä ei ilmaista heikon tai
vahvan riittävän keinon termein.
- 2) Välttämättömän keinon ei tarvitse olla välittömästi välttämätön – agentti voi
tehdä muita asioita ennen kuin tekee A.
- 3) Peräkkäisiä välttämättömiä keinoja ei välttämättä tarvitse tehdä kerralla saa-
vuttaakseen tavoitteen.

Vaikka rangaistulaukauskilpailun voittaminen edellyttää maalin tekemistä, jokaisen yri-
tyksen ei tarvitse päätyä maaliin. Voittaminen edellyttää maalia, jonka edellytys on
laukaus. Jotta vastustaja ei-voittaisi, tulee vastustajan olla tekemättä maalia. Voittami-
nen ja onnistunut laukaus vaikuttavat toteutuvan jos

- 1) pelaajan osumatarkkuus palloon on riittävä
- 2) pelaajan osuman voima ja suunta ovat riittävät.

Rangaistuspotkun taito voidaan tällaisessa mallissa kiteyttää pelaajan (agentin) kykyyn
toistaa onnistunut potku (osumatarkkuus ja voimavektori) mahdollisimman optimaal-
lisena. Taidon kehittäminen keskittyy tässä tilanteessa potkun osumapisteen hajonnan
pienentämiseen sekä osumavoiman suunnan vakioimiseen ja osumaenergian vaki-
oimiseen. Tässä suhteessa potkaisijan liikeradat voitaisiin esittää teknisenä kuvauksena
ihmisruumiin voimantuotannosta. Kuvaus olisi tällöin vastaavan kaltainen kuin höy-
rykoneen tekninen kuvaus ilman funktiota.

Lopuksi

Lokaalinen keinojen ja päämäärien välinen semantiikka mahdollistaa yksinkertaisten
kausaaliyhteyksien käsittelyn. Käsittely on varsin kapea, mutta sen etuna on yksinker-
taisuus. On kuitenkin huomioitava, että käsittelytapa on monin tavoin vajavainen ver-
rattuna luonnollisen kielen syy-seuraussuhteeseen (Hughes&Kroes&Zwart 2007,
224):

- 1) Tapa ei ole käyttökelpoinen rajoittuneen agentin järkeilyn ymmärtämiseen – jos agentti ei ole tietoinen ympäristön faktoista, hän ei ole kykenevä arvioimaan niiden vaikutusta.
- 2) Tapa ilmaisee kausaalisuhteen vain yhdessä erityisessä olotilassa.
- 3) Tapa jättää huomioimatta välitavoitteiden merkityksen.

Rajatun taitotilanteen käsittely mahdollistaa tarkastelun viemisen teknisen ja inhimillisen selittämisen rajapinnalle. Tekninen kuvaus tai intentionaalinen toiminta eivät yksin kykene selittämään tapahtumaa yhtä kattavasti – fyysinen ja intentionaalinen luonne yhdessä muodostavat kokonaisuuden. Tiukka rajausta tuo mukanaan sovellettavuuden haasteellisuuden. Rangaistuspotkuesimerkki osoittaa tarkastelun suppeuden. Kyseessä on optimoidun suorituksen tehokkuuden maksimoiminen, eikä tapahtumassa huomioida esimerkiksi perinteistä intention salaamista eli harhauttamisen vaikutusta toiseen toimijaan.

Esimerkkiä voi luonteeltaan ajatella suhteessa Turingin koneeseen. Esimerkissä käsitellään yhtä mallia, jonka tavoite on ratkaista yhtä asiaa optimaalisella tavalla. On toki huomioitava, ettei yksinkertaisintakaan tekoa voida formalisoida laskennan tapaan. Jalkapallo on maailman suurin ja seuratuin urheilulaji, johon liittyy valtavasti intohimoa ja taituruuden tunnustamista. Tämän lajin yhden huipentuman, rangaistuspotkun, yksinkertaistaminen kaavan tasolle tuntuu miltei rikolliselta. On hyvä kuitenkin kysyä Turingin (1950) ideaa vapaasti soveltaen, voimmeko erottaa ihmisen ja koneen suorittaman rangaistuspotkun toisistaan, jos emme näe, kumpi potkun suorittaa?

Rajoitteistaan huolimatta yksinkertaistaminen mahdollistaa tietyn rajatun ilmiön tarkastelun viemisen lähemmäs fyysisen ja intentionaalisen kokonaisuuden liityntää. On vaikea kuvitella, että optimaalinen vartalonliike pallon suhteen on riittävä kaava rangaistuspotkutaidon kehittämiseen. Yhtä lailla on kuviteltavissa se, että hyvän ratkaisun kaavalla voidaan tukea taidon kehittämistä kyseisellä alalla. Taidon tekninen kuvaus ilman kontekstia lienee käsitteen tasolla mahdotonta. Vastaavasti mahdottomalta kuulostaa se, ettei taitavuutta voida rajatuissa tilanteissa muotoilla yksinkertaisiksi säännöiksi. Taitavuus ilmenee, kun ihminen operoi maailmassa luoden siellä jotain kuvitteellista. Taitavuuden luonne on tältä osin hyvin samankaltaista kuin artefaktin olemus; sillä on inhimillinen tavoite ja toteutuakseen sen on saatava fyysinen rakenne.

Lähteet

Breivik, G. (2014) 'Sporting Knowledge and the Problem of Knowing How', *Journal of the Philosophy of Sport*, 41(2), pp. 143–162. Available at: <https://doi.org/10.1080/00948705.2014.911102>.

'Computer Machinery and Intelligence' (1950) in Turing, A. M. and Copeland, B. J., (2004) *The essential Turing: seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life, plus the secrets of Enigma*. Oxford New York: Clarendon Press Oxford University Press, pp. 441–464.

- Hintikka, J. (1974) *Knowledge and the Known*. Dordrecht-Holland/Boston-USA: D. Reidel Publishing Company. Available at: <https://doi.org/10.1007/978-94-010-2217-0>.
- Hughes, J., Kroes, P. and Zwart, S. (2007) 'A Semantics for Means-end Relations', *Synthese*, 158(2), pp. 207–231. Available at: <https://doi.org/10.1007/s11229-006-9036-x>.
- Kroes, P. (1998) 'Technological Explanations: The Relation Between Structure and Function of Technological Objects', *Techné: Research in Philosophy and Technology*, 3(3), pp. 124–134. Available at: <https://doi.org/techne19983325>.
- Mutanen, A. (2006) 'Sotilaan taidosta ja Tiedosta', *Tiede ja Ase*, 2006(64).
- Mutanen, A. (2015) 'Tiedosta ja taidosta', *Ammattikasvatuksen aikakauskirja*, 17(1), pp. 58–73.
- Mutanen, A. (2016) 'About Practical Problem Solving', *Problemos*, 89, pp. 85–94. Available at: <https://doi.org/10.15388/Problemos.2016.89.9889>.
- Niiniluoto, I. (1980) *Johdatus tieteenfilosofiaan: käsitteen- ja teorianmuodostus*. Helsinki: Otava.
- Niiniluoto, I. (1989) *Informaatio, tieto ja yhteiskunta: filosofinen käsitteanalyysi*. Helsinki: Valtion painatuskeskus.
- Niiniluoto, I. (2002) *Johdatus tieteenfilosofiaan: käsitteen- ja teorianmuodostus*. Keuruu: Otavan Kirjapanimo Oy.
- Niiniluoto, I. (2020) *Tekniikan filosofia*. Tallinna: Gaudeamus.
- Pacherie, E. (2007) 'The Sense of Control and the Sense of Agency', *PSYCHE: An Interdisciplinary Journal of Research On Consciousness*, 13(1), pp. 1–30.
- Pollard, B. (2006) 'Explaining Actions with Habits', *American Philosophical Quarterly*, 43(1), pp. 57–69.
- Puolustusvoimien Koulutuksen Kehittämiskeskus (1998) *Sotilaspedagogiikan perusteet*. Hämeenlinna: Karisto Oy.
- Reader, S. (2007) 'The Other Side of Agency', *Philosophy*, 82(322), pp. 579–604.
- Ryle, G. (2009) *The Concept of Mind*. London and New York: Routledge.
- Siitonen, A. (2007) 'Taito ja tieto', in H. Kotila et al. (eds) *Taidon tieto*. Helsinki: Edita Prima Oy.

Siitonen, A. (2008) 'On soldiers' skills', in A. Mutanen (ed.) *The Many Faces of Military Studies: a search for Fundamental Questions*. Helsinki: Edita Prima Oy (Publications of the Finnish Naval Academy, 1/2008), pp. 50–54.

Simon, H.A. (1996) *The sciences of the artificial*. Cambridge, Mass., London, England: MIT Press.

Sookermany, A.M. (2012) 'What Is a Skillful Soldier? An Epistemological Foundation for Understanding Military Skill Acquisition in (Post) Modernized Armed Forces', *Armed Forces & Society*, 38(4), pp. 582–603. Available at: <https://doi.org/10.1177/0095327X11418320>.

Toiskallio, J. (2009) 'Toimintakyky sotilaspedagogiikan käsitteenä', in Mäkinen, J. and Toiskallio, J., *Sotilaspedagogiikka : sotiluuden ja toimintakyvyn teoriaa ja käytäntöä*. Helsinki: Maanpuolustuskorkeakoulu; Johtamisen ja sotilaspedagogiikan laitos, pp. 48–73. Available at: <http://www.doria.fi/handle/10024/74178> (Accessed: 21 September 2022).

Toiskallio, J. (2016) 'Eettisyys ja toimintakyky', in Jantunen, T., *Reijo Wilenius: henkisen kasvun ja toiminnan filosofi*. Edited by M. Niinivirta and M. Wilenius. Kotka: Kopio Niini Oy (Snellman-korkeakoulun tutkimusosaston julkaisu, 1/2016).

Von Wright, G.H. (1974) *Ajatus ja julistus*. Porvoo: Werner Söderström Osakeyhtiön Laakapaino.

EPISTEEMINEN LOGIIKKA

Arto Mutanen

Johdanto

Tietojenkäsittelytieteessä määritelmällisesti tarkastellaan tietokoneissa tai tietokoneiden avulla tapahtuvaa tiedon käsittelyä. Tietokoneet ovat elektronisia ”laskimia” (computer), jotka käsittelevät 0-1 -merkkijonoja. Miten tällaiset voivat ylipäättään käsitellä tietoa? Tarkoitetaanko tiedolla tällöin samaa kuin arkikielessä tai tieto-opissa? Puhuttaessa tekoälystä, joudumme yhä syvemmin pureutumaan yksityiskohtaisiin käsitteellisiin ongelmiin. Siten tiedon yksityiskohtainen tarkastelu on perustavaa laatua oleva tehtävä niin tietojenkäsittelytieteelle kuin tekoälytutkimukselle.

Filosofinen tieto-oppi keskittyy tutkimaan tiedon luonnetta, sen mahdollisuutta ja rajoja. Tieto-opilliset tarkastelut lähtevät liikkeelle klassisesta tiedonkäsityksestä, jonka muotoilut palautuvat aina Platonin dialogiin *Theaitetos*. Klassinen tiedonkäsitys sanoo, että tieto on hyvin perusteltu tosi uskomus. Gettier julkaisi vuonna 1963 erittäin tärkeän artikkelin. Artikkelin oli erittäin lyhyt, mutta siinä Gettier osoitti muutamilla hyvin laadituilla esimerkeillä, että klassisessa tiedonkäsityksessä on jotain perustavaa laatua olevaa vikaa. Tämä on rikastuttanut tieto-opillista tutkimusta; nykyinen tieto-opillinen keskustelu on rikasta ja monipuolista. Gettierin artikkelin merkitystä kuvaa hyvin se, että erittäin usein tieto-opin tarkasteluissa osoitetaan, että ehdotettu muotoilu selvittää Gettier-ongelmat. (Lammenranta 2022.)

Kuitenkin tietojenkäsittelytieteessä ja tekoälytutkimuksessa yleinen tieto-opillinen keskustelu ei ole riittävää, vaan tarvitaan keinoja muotoilla tieto-opilliset kysymykset formaalisti tai loogisesti. Filosofisessa logiikassa on pitkä modaalilogiikan traditio, jonka juuret ulottuvat aina Aristoteleen ja Platonin filosofiaan. Perustana on ns. aleettiset modaaliteetit (mahdollisuus ja välttämättömyys), joiden tulkinta on osoittautunut erittäin kiinnostavaksi. Aristoteelinen temporaalinen tulkinta (White 1980) on nykyään valtaosin vaihtunut tulkintaan, jossa voidaan tarkastella (loogisesti) samanaikaisia mahdollisuuksia. Tällöin modaaliteettien semantiikka perustuu ns. mahdollisten maailmojen semantiikalle (Menzel 2021).

Von Wright toi esiin kirjassaan *An Essay in Modal Logic* (1951) aleettisten, episteemisten, deonttisten ja eksistentiaalisten modaaliteettien välisen kiinnostavan loogisen samankaltaisuuden. Tällaiset havainnot tuovat esiin tiettyjen loogisten periaatteiden heidelmällisyyden modaaliteettien tarkastelussa. (Kts. myös von Wright 1977.) Modaalilogiikka on edistynyt voimakkaasti 1950-luvulta lähtien (Copeland 2002; Garson 2021). Episteeminen logiikka on – nimensä mukaisesti – tiedon loogiseen tarkasteluun keskittynyt modaalilogiikan ala, jonka eräs perustavimmista lähteistä on Jaakko Hintikan vuonna 1962 julkaisema teos *Knowledge and Belief. An Introduction to the Logic of the Two Notions*, joka on ollut innoituksen lähteenä niin filosofeille, loogikoille kuin tekoälytutkijoille. Tiedon logiikkaan liittyvästä monipuolisesta tutkimuksesta saa kattavan kuvan kirjasta Arló-Costa, Hendricks & van Benthem (2016).

Nykyajan logiikassa on esillä valtava määrä erilaisia logiikan systeemejä (Susan Haack 1974; 1978; Halonen 1995). Usein erotellaan ns. traditionaalinen logiikka, jonka malliesimerkki on aristoteelinen syllogistiikka, jolla on tietysti suuri historiallinen arvo. Vaikka syllogistiikalla on filosofisesti kiinnostavia piirteitä, niin varsinaisesti logiikan tutkimuksen kannalta tämä on jäänyt sivuun. Klassisella logiikalla viitataan lähinnä lauselogiikkaan ja predikaattilogiikkaan, joihin logiikan perusopetus usein keskittyy. Esimerkiksi tekoälyn perusteita tarkastelevassa yleisteoksessa Russell & Norvig (2003) tekoälytutkimuksen keskeisiä ideoita tarkastellaan nimenomaan predikaattilogiikan kehityksessä. Klassisen logiikan rinnalla esiintyy sen kanssa kilpailevia systeemejä, kuten poikkeavat logiikat, moniarvologiat, parakonsistenttilogiat ja intuitionistinen logiikka. Toisaalta on myös klassisen logiikan laajennuksia, joita ovat erilaiset modaali-logiikat, kuten aleettisten modaaliteettien logiikka, episteeminen logiikka, deonttinen logiikka ja aikalogiikka. (Susan Haack 1978; Halonen 1995).

Logiikalla voidaan tarkoittaa loogis-matemaattista systeemiä (kalkyyli), jonka eräs merkkiteos on Boolean vuonna 1847 julkaistu teos *The Mathematical Analysis of Logic, being an Essay towards a Calculus of Deductive Reasoning*. Toisaalta logiikalla voidaan tarkoittaa loogis-käsitteelliseen tarkkuuteen pyrkivää tutkimusta, jonka eräs merkkiteos on Fregen vuonna 1879 julkaistu teos *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Tämä teos tuo esiin logiikan mahdollisuuden filosofisen analyysin metodisena keinona, jota myöhemmin ovat toteuttaneet esimerkiksi Russell, jonka essee *On Denoting* (1905) on erinomainen esimerkki tästä lähestymistavasta. Näiden lähestymistapojen välillä ei suinkaan ole ristiriitaa, vaan ne tukevat toinen toisiaan. Kuitenkin niillä on selkeä painotusero. (von Wright 1964; Juti 2013.)

Olemme edellä viitanneet mahdollisten maailmojen semantiikkaan, joka on osoittautunut teoreettisesti erittäin tärkeäksi. Esimerkiksi, normatiivisten lauseiden totuuden luonnehdinta on luontevaa mahdollisten maailmojen semantiikassa, joten niiden logiikka saa täsmällisen merkityksen. Yleisesti mahdollisten maailmojen semantiikka antaa teoreettisesti eheän viitekehyksen modaalikäsitteiden tarkasteluun. Lisäksi mahdollisten maailmojen semantiikka tuo konkreettisesti esiin yhteyden modaalilogiikan ja ensimmäisen kertaluvun logiikan välillä. (Hintikka 1969.)

Logiikasta

Logiikka on formaali oppiaine, jossa tarkastellaan lauseiden ja argumenttien muotoa, ei niiden sisältöä. Tämä tulee esiin hyvin tarkastelemalla mitä tahansa logiikan perusoppikirjaa, kuten esimerkiksi Miettinen (1995), Salminen & Väänänen (1992) ja Väänänen (1987). Näistä kirjoista saa hyvän kuvan siitä, mitä logiikassa opetetaan ja miten sitä opetetaan. Usein logiikan perusopetuksessa opetetaan ensin lauselogiikka, jonka varaan opetetaan predikaattilogiikan perusominaisuudet. Lauselogiikka tuo esiin logiikan perusajatukset (merkinnät, todistukset ja tulokset). Predikaattilogiikka mahdollistaa syvällisemmän tutustumisen logiikan tuloksiin ja sovellutuksiin. Predikaattilogiikka on varsin rikas looginen järjestelmä, johon tutustuminen on sinällään kiinnostavaa, mutta sillä on myös paljon käytännöllisesti tärkeitä sovellutuksia. Monien sovellutusten kannalta predikaattilogiikka tarjoaa riittävän rikkaan loogisen struktuurin perustavien ideoiden muotoiluun, mikä tulee hyvin esiin esimerkiksi kirjasta Russell & Norvig

(2003). Lauselogiikka, yksinkertaisuudesta huolimatta, on osoittautunut teoreettisesti hedelmälliseksi esimerkiksi koneoppimisen kannalta (Anthony & Biggs 1997).

Logiikassa tehty erottelu syntaksin ja semantiikan välillä on erittäin merkittävä. Syntaksilla tarkoitetaan logiikan kielioppia eli logiikan symbolijärjestelmää, lauseenmuodostussääntöjä ja päättelysääntöjä, jotka määrittelevät kielen formaalina merkkijärjestelmänä, jolla ei tarvitse olla mitään merkitystä. Olennaista on tarkastella mitä annettusta lausejoukosta voidaan päättelysääntöjen avulla johtaa tai todistaa. Olkoon Ω lausejoukko. Sanomme, että lause S on johdettavissa lausejoukosta Ω , jos ja vain jos on olemassa äärellinen jono ϕ_1, \dots, ϕ_n siten, että $\phi_n = S$ ja että kaikille $i, i \in \{1, \dots, n\}$, ϕ_i on joko (i) aksiomi, (ii) kuuluu joukkoon Ω tai (iii) on päätelty päättelysäännöllä jonon aikaisemmista lauseista. Edellä mainittu lausejono on lauseen S deduktiojoukosta Ω . Jos lause on johdettavissa tyhjästä joukosta, niin sanomme lausetta todistuvaksi ja kyseistä jonoa sen deduktioksi. Tällainen todistusteoreettinen tarkastelu on syntaktista. Semantiikalla tarkoitetaan kielen merkitysoppia, joka logiikassa tunnetaan malliteorianana. Kielen malli on mikä tahansa joukko-opillinen struktuuri, joka antaa kielelle tulkinnan. Kullakin kielellä voi olla useita erilaisia tulkintoja. Tulkinnan määrittää ns. tulkintafunktio, joka määrittellään induktiivisesti. Malliteoriassa keskeinen käsite totuus mallissa tai toteutuvuus mallissa. Sanomme lausetta validiksi (päteväksi), jos se on tosi jokaisessa kielen mallissa. Malliteoria on osoittautunut rikkaaksi lähestymistavaksi logiikkaan. Gödelin täydellisyyslause tuo tärkeän yhteyden todistusteoreettisten ja malliteoreettisten menetelmien välille, sen mukaan (ensimmäisen kertaluvun) predikaattilogiikassa pätee seuraava: kielen lause S on todistuva, jos ja vain jos se on validi. (Addison, Henkin & Tarski 1965; Hodges 1993; Miettinen 1995.)

Ensimmäisen kertaluvun logiikka ja sen malliteoria antavat lujan perustan niin loogiselle kuin käsitteelliselle tekoälytutkimukselle. Modaalilogiikat saadaan ensimmäisen kertaluvun logiikasta liittämällä siihen sopivat modaalioperaattorit. Esimerkiksi aleettisten modaliteettien kohdalla operaattorit ' \square ' (välttämättä) ja ' \diamond ' (mahdollisesti) ja episteemisessä logiikassa ' K_a ' (a tietää, että) ja doksastisessa logiikassa ' B_a ' (a uskoo, että) ja deonttisessa logiikassa ' P ' (sallittu) ja ' O ' (pakollinen). Näin saadaan logiikka, jossa voidaan ilmaista p :n välttämättömyys kaavalla ' $\square p$ '. Episteemisessä logiikassa voidaan esittää agentin tietoa ilmaisevia lauseita: $K_a p$ (a tietää, että p) tai $K_a(p \& q)$ (a tietää, että p ja q). Modaalikontekstissa kvanttorit ovat olleet laajan keskustelun kohteena. On tärkeää pystyä ymmärtämään vivahteikkaita käsitteellisiä erotteluja, joita tulee esiin esimerkiksi tarkastelemalla lauseita " $(\exists x)K_a \phi(x)$ " ja " $K_a(\exists x)\phi(x)$ ", joissa olennaista on tarkastella kvanttorien ja episteemisen operaattorin välisiä suhteita: edellisessä eksistenssi kvanttori ei ole episteemisen operaattorin alassa, kun taas jälkimmäisessä se on. Tällaiset eivät ole vain hienostelevaa saivartelua, vaan näillä on olennainen merkitys.

Myös modaalilogiikassa niin todistusteoreettinen kuin malliteoreettinen lähestymistapa ovat molemmat mahdollisia. Todistusteoreettisessa lähestymistavassa painotus on aksiomatisoinnissa ja formaaleissa todistuksissa. Aksiomatisoinneissa lähtökohtana on lauselogiikan tai predikaattilogiikan aksiomatisoinnit, joihin lisätään relevantit operaattorit, jolloin voidaan tarkastella esimerkiksi millä edellytyksellä voidaan saada tietopillisesti luonteva tulkinta, että tiedon tulee pitää paikkaansa eli $K_a p \rightarrow p$. Toisaalta

looginen muotoilu tuo esiin, tai ehkä aiheuttaa, tiettyjä ongelmia, kuten loogisen kaikkietävyuden ongelma: Jos $K_a p$ ja $p \rightarrow q$, niin seuraako myös, että $K_a q$? Jos katsomme, että em. seuraussuhde pätee, niin henkilö tietää kunkin yksittäisen tietonsa kaikki loogiset seuraukset ja erikoistapauksena, kun lähdemme tyhjästä tiedosta, niin henkilö tietää kaikki loogiset totuudet (looginen kaikkietävyys). Tämä olisi ikävä asia ainakin logiikan opettajille, jotka jäisivät työttömiksi saman tien. Jos taas emme hyväksy seuraussuhdetta päteväksi, niin tulee rajaamisen ongelmaan antaa ratkaisu, sillä ainakin jotkut triviaalit seuraukset ovat järkevää hyväksyä. Tällöin voimme luonnehtia erilaisia tiedon lajeja, jotka ovat sekä loogisesti että filosofisesti kiinnostavia (Hintikka 1962; 1984).

Mahdollisten maailmojen semantiikka antaa rikkaan ja luontevan tulkinnan modaalikäsitteille. Tällöin tulkinta annetaan tietystä malliluokassa, jonka mallien välillä vallitsee ns. saavutettavuusrelaatio (Copeland 2002). Esimerkiksi tiedon tulkinta: ” $K_a p$ ” on tosi mallissa w , jos ja vain jos p on tosi kaikissa malleissa w' , joille pätee wRw' , missä R on saavutettavuusrelaatio. Intuitiivisesti tämä tarkoittaa, että p on tosi kaikissa niissä maailmoissa w' , jotka ovat yhteensopivia a :n tiedon (maailmassa w) kanssa.

Mahdollisten maailmojen semantiikka tuo hyvin esiin, miksi kvantifiointi on ongelmallinen modaalikontekstissa. Siirryttäessä maailmasta toiseen ei ole mitenkään ilmeistä, miten yksilöiden identifioiminen tapahtuu. Tosiasiassa tämä voi tapahtua eri tavoin. Joka tapauksessa modaalikontekstissa yksilön identifikaatio tulee kiinnostavaksi sekä loogis-käsitteellisesti että filosofisesti. (Hintikka 1969.)

Episteemisestä logiikasta

Episteemisen logiikan taustalla on havainto, että episteeminen ilmauksia muotoa ” a tietää, että p ” voidaan tutkia täsmällisin loogisin keinoin. Episteemisessä logiikassa otetaan logiikkaan episteeminen operaattori ” K_a ”, jossa K on tieto-operaattori ja a ilmaisee tiedon agentin. Näin ” $K_a p$ ” ilmaisee, että ” a tietää, että p ”. Episteemisen logiikan semantiikka voidaan muotoilla, modaalilogiikan tavoin, mahdollisten maailmojen semantiikan avulla. Siten edellä olevan nojalla modaalioperaattori K toimii universaalikvanttorin tavoin: jokaisessa a :n episteemisessä maailmassa pätee, että p . Klassinen tiedon määritelmä sanoo, että tiedon tulee olla totta, joten saavutettavuusrelaation tulee olla reflektiivinen. (Hintikka 1962; Hendricks & Symons 2006.)

Se, että tiedon tulee olla totta, on erittäin vahva oletus. Kuitenkin tämä on pitkälle myös ns. arki ajattelun taustalla. On erittäin vaikea ymmärtää, jos joku väittää tietävänsä jonkin asian ja samalla sanovat, ettei se ole totta (Hintikka 1962). Tällaisten episteemisesti ongelmallisten ilmauksien tarkastelu on tuonut paljon lisätietoa tiedon logiikasta.

Uskomusta ilmaisevalla ilmauksilla ” a uskoo, että p ” on kiinnostavia samankaltaisuuksia kuin episteemisillä ilmauksilla, mutta myös tiettyjä eroja, kuten se, ettei uskomuksesta seuraa totuutta. Siten edellä mainitut saavutettavuusrelaatiot episteemisessä ja uskomisen logiikassa poikkeavat toisistaan; episteemisessä saavutettavuusrelaation

tulee olla refleksiivinen, mutta uskomisen logiikassa ei. (Hintikka 1962; Hendricks & Symons 2006.)

Episteeminen logiikka on kiinnostava modaalilogiikan erityisala (Oppy 1998), jolla on kiinnostavia yhteyksiä edellä mainittuun uskomisen logiikkaan, mutta myös havainnon logiikkaan (Niiniluoto 1982), muistin logiikkaan (Schumacher 1992), kuvittelun logiikkaan (Niiniluoto 1985) ja toivon logiikkaan (Day 1970). Toisaalta tieto liittyy toimintaan, joten tiedon logiikka kytkeytyy myös toiminnan logiikkaan (von Wright 1963; Kallion artikkeli tässä kirjassa). Episteemisellä logiikalla on sovellutuksia ja relevanssia filosofiassa, tietojenkäsittelytieteessä, tekoälytutkimuksessa ja päätöksentekoteoriassa.

Episteeminen logiikka on ollut luonteeltaan staattista; siinä on tarkasteltu tiettyjen tietolauseiden ominaisuuksia ja episteemisten päättelyiden pätevyyttä. Tämä sinällään tärkeä aihepiiri ei kuitenkaan kuvaa sitä dynaamisuutta, joka vallitsee ihmisten tiedollisessa todellisuudessa. Ihminen ei vain omaa tietoa, vaan hankkii ja käyttää sitä. Siten on tarve kehittää malleja, joilla on dynaamisia ominaisuuksia, jotka voivat kuvata paremmin ihmisten tiedollista toimintaa. (Hintikka 2007; Baltag & Bryan 2016; Gierasimczuk, Hendricks & de Jongh 2014; van Ditmarsch, van der Hoek & Kooi 2007.)

Jaakko Hintikka kehitti jo 1970-luvulla tiedon etsimisen logiikkaa, joka on osoittautunut rikkaaksi kokonaisvaltaiseksi malliksi, jota yleisesti kutsutaan tutkimuksen kyselymalliksi (kts. tarkemmin Halosen artikkeli tässä kirjassa). Kannaltamme olennaista Hintikan mallissa on sen yksityiskohtainen tiedon rakentumisen looginen analyysi. Jo mallin logiikka on episteemisesti orientoitunut, mutta mallin kehittyneissä versioissa mukana on lisäksi episteemiset operaattorit, jotka konkreettistavat mallin nimenomaan episteemisenä logiikkana (Hintikka, Halonen & Mutanen 2002). Lähtökohtainen ajatus Hintikalla oli yksittäisen ihmisen tiedonhankinta, mutta malli kehittyi samalla tieteellisen tiedon hankinnan malliksi, joten tiedon agentiksi tuli yksittäisen ihmisen ohella myös tiedeyhteisö. Hintikan mallin kiinnostava piirre tekoälytutkimuksen kannalta liittyy loogis-käsitteellisen tarkasteluun. Siten ei ole mitään syytä rajoittaa tiedonetsintää vain ihmisiin, vaan aivan yhtä hyvin tiedon agentti voi olla tietokone.

Tällaiset dynaamiset lähestymistavat ovat yleistyneet 1900-luvun puolella. Nykyisin on useita erilaisia dynaamisia logiikoita jopa siinä määrin, että puhutaan logiikan dynaamisesta käännteestä (Gochet 2002; Gierasimczuk, Hendricks & de Jongh 2014). Van Bentham (2003) esittää, että dynaaminen käänne tuo huomion keskiöön päättelyprosessin, uskomusten muutoksen ja argumentaation. Näin huomioon tuleva päättelyn pätevyyden (staattisen) arvioinnin sijaan päättelyn strategiset piirteet ja toiminnot (episteemiset teot). Nämä ovat keskeisiä niin episteemisessä logiikassa kuin oppimisen teoriassa. (Baltag & Renne 2016; van Ditmarsch, van der Hoek & Kooi 2007; Gierasimczuk, Hendricks & de Jongh 2014; Hintikka 2007.)

Eräs kiinnostava episteemisen logiikan piirre liittyy kollektiivisen tiedon problematiikkaan. Vaikka Hintikan malli mahdollistaa agentin olevan periaatteessa myös ryhmän, niin ryhmän tietämisen problematiikka on oma ongelma-alueensa, jota tarkastellaan

artikkelissa Hendricks & Rendsvig (2014). Tiedon logiikan kollektiivisista päätteistä kts. tarkemmin van Ditmarsch, van der Hoek & Kooi (2007).

Kollektiivinen tieto¹

Episteemisessä logiikassa, kuten myös yleisessä tieto-opissa, tarkastellaan lähtökohtaisesti yhden agentin tietoa. Kuitenkaan ei sinällään ole rajoitusta, millainen tiedon agentti on. On teoreettisesti kiinnostavaa tarkastella, mitä tarkoittaa, että jokin ryhmä tietää jotain. Ryhmä voi olla periaatteessa millainen tahansa. Tietenkin puhuttaessa jonkin organisaation tiedosta, merkitys voi poiketa seuraavassa esitetystä. Organisaatio voidaan ymmärtää itsenäisenä agenttina. Organisaatiot ovat legaalisia agentteja. On tietysti kiinnostavaa tarkastella voisiko organisaatiot olla episteemisiä agentteja (kts. kiinalaisen huoneen argumentti; Searle 1989, 31-34) tai moraalisia agentteja (Lovett & Riedener 2021). Tällainen tarkastelu on tärkeää ja kiinnostavaa, mutta seuraavassa tarkastelemme, miten voimme analysoida kollektiivisen tiedon aiemmin esitetyn yksilötiedon avulla.

Tarkastellaan ryhmän $G = \{a_1, \dots, a_n\}$ tietoa. Sanottaessa, että ”ryhmä G tietää, että p ” voimme tarkoittaa monia eri asioita. Seuraavat kaksi tulkintaa ovat luonteeltaan vaihtoehtoja:

Sanomme, että ”ryhmä G tietää, että p ” (E_{GP}), jos ja vain jos jokainen ryhmän G jäsen tietää, että p . Tämä voidaan siten merkitä seuraavalla tavalla

$$E_{GP} = K_{a_1}p \wedge K_{a_2}p \wedge \dots \wedge K_{a_n}p,$$

mikä voidaan ilmaista myös seuraavalla tavalla

$$E_{GP} =_{df} (\forall a_i)_{i \in \{1, \dots, n\}} K_{a_i}p.$$

Sanomme, että ”ryhmä G tietää, että p ” (S_{GP}), jos ja vain jos joku ryhmän G jäsenistä tietää, että p . Tämä voidaan siten merkitä seuraavalla tavalla

$$S_{GP} = K_{a_1}p \vee K_{a_2}p \vee \dots \vee K_{a_n}p,$$

mikä voidaan ilmaista myös seuraavalla tavalla

$$S_{GP} =_{df} (\exists a_i)_{i \in \{1, \dots, n\}} K_{a_i}p.$$

On selvää, että E_G on loogisesti vahvempi kuin S_G eli $E_{GP} \vdash S_{GP}$.

Nämä molemmat tulkinnat ovat kiinnostavia. Jälkimmäinen tulkinta on siinä merkityksessä kiinnostava, ettei ole useinkaan syytä olettaa kaikkien tietävän. Esimerkiksi tieteellinen tieto ei edellytä, että jokainen alan tutkija tietää puheena olevan asian, riittää, että se on jonkun tutkijan tiedossa siten, että hän tietää sen. Toisaalta vaikka edellytys, että kaikki tietävät on liian vahva, niin jonkun ryhmän jäsenen tieto on selvästi

¹ Tässä kappaleessa seurataan tarkasti artikkelia Hendricks & Rendsvig 2014.

liian heikko. Näiden kahden väliin jää ryhmän jaettu tieto (distributed knowledge). Ryhmän G jaetulla tiedolla tarkoitetaan, että ryhmä G jakaa tietonsa siten, että jokaisen ryhmän jäsenen on mahdollista saada tietää toisten tietämä asia. Tämä voidaan tulkita eri tavoin, eräs tulkinta, joka edellä olevasta muotoilusta voi nousta esiin voidaan ilmaista seuraavalla tavalla:

$$D_{Gp} =_{df} (\forall a_i)_{i \in \{1, \dots, n\}} \diamond K_{a_i} p,$$

Tässä ilmaistaan, että jokaiselle ryhmän jäsenelle on mahdollista, että hän tietää, että p , mikä on liian heikko ehto. Siten tämä ei ole juuri se tulkinta, jota tavoitellaan.

Esimerkiksi tieteellinen tieto on jaettua tietoa siinä merkityksessä, että tutkijat jakavat (julkaisemalla, luennoimalla, jne.) tietonsa tiedeyhteisön yhteiseen käyttöön. Siten kuka tahansa voi, ainakin periaatteessa, saada tietoonsa minkä tahansa tieteellisen tuloksen. Tutkimuksista voidaan nähdä, miten tieto on rakentunut. Siten toisen tutkijan on mahdollista omaksua tieto, kritikoida, korjata, kehittää tai soveltaa sitä. (Hendricks & Rendsvig 2014; Hintikka 2007.)

Kollektiivinen tieto ei ole vain yksiolotteista ensimmäisen kertaluvun tietoa, joka esittää jonkun agentin tiedosta. Formaalisti tällainen ensimmäisen kertaluvun tiedon esityksessä esiintyy ainoastaan yksi tieto-operaattori. Kollektiiviseen tietoon liittyy myös tieto toisten tiedosta, jonka formaalissa esityksessä esiintyy kaksi tieto-operaattoria, tällaista sanotaan toisen kertaluvun tiedoksi. On selvää, että näin voidaan generoida mielivaltaisen monennen kertaluvun tieto. Teknisesti tämä voidaan ilmaista seuraavalla tavalla:

Ensimmäisen kertaluvun tieto (K^1)

$K_{a_1} p; K_{a_2} p, \dots$

Toisen kertaluvun tieto (K^2)

$K_{a_1} K_{a_1} p; K_{a_2} K_{a_1} p, K_{a_1} K_{a_2} p, \dots$

Siten on mahdollista määrittellä yleisesti korkean kertaluvun tieto induktiivisesti. Määrittelemällä ensimmäisen kertaluvun tieto kuten edellä ja $n:n, n > 1$, kertaluvun tieto on KK^{n-1} . Tässä määritelmässä ei tarvitse merkitä agenttia indeksiin.

Voimme korkeamman kertaluvun tiedon avulla määrittellä yhteisen tiedon (common knowledge). Määritelmää voidaan havainnollistaa seuraavalla esimerkillä. Olkoon aikomuksena tehdä hyökkäys kahden joukon toimesta. Joukot sijaitsevat lähekkäisillä kukkuloilla, joiden välillä ei ole suoraa yhteyttä. Hyökkäys tulee tehdä molemmista kukkuloista samanaikaisesti. Ensimmäisen kukkulan komentaja päättää, että hyvä aika olisi aamulla kello 6. Tietoa välittämään lähetetään lähetti toiselle kukkulalle. Nyt tieto, onko lähetin tieto saavuttanut toisen kukkulan, ei voida suoraan varmistaa, joten toisen kukkulan komentaja laittaa lähetin, joka vie tiedon, että viesti on saapunut. Nyt kuitenkin toisen kukkulan komentaja ei saa tietoa, onko lähetti vienyt varmistustiedon, joten jälleen joudutaan lähettämään lähetti. On helppoa havaita, ettei mikään määrä

lähettejä riittää varmistamaan yhteistä tietoa. Tämä osoittaa, ettei ryhmän tieto merkityksessä E_{Gp} riittää varmistamaan yhteistä tietoa. Tämä voidaan yleistää yhteisen tiedon määritelmäksi. (Hendricks & Rendvig 2014.)

On kiinnostavaa tarkastella, miten ryhmän tiedosta merkityksessä E_{Gp} voitaisiin rakentaa yhteinen tieto C_{Gp} . Yhteinen tieto, että p , C_{Gp} , edellyttää, että jokainen ryhmän jäsen tietää, että p ja jokainen tietää, että jokainen tietää, että p , ja Siten yhteinen tieto on ryhmän tiedon merkityksessä E_{Gp} päättymätön iteraatio. Näin ollen yhteinen tieto voidaan määritellä määrittelemällä operaattorille E_{Gp} iteraatiot induktiivisesti. Olkoon merkityksessä $E^1_{Gp} = E_{Gp}$ ja $E^n_{Gp} = E_{Gp}E^{n-1}_{Gp}$, kun $n > 1$. Tällöin yhteinen tieto $C_{Gp} =_{df} \bigcup_{n=1}^{\infty} E^n_{Gp}$. Tällainen tieto on edellytys koordinoitulle toiminnalle. Luonnollisesti, kuten edellä oleva esimerkki jo osoittaa, että joudumme toimimaan usein epävarmuuden vallitessa.

Yhteisen tiedon vaikea saavutettavuus ei tee järkevää toimintaamme mahdottomaksi. Usein onnistumme varsin hyvin luomaan riittävän yhteisen tiedon tai ymmärryksen niin asioiden tilasta, tavoitteista kuin keinoista. On kuitenkin hyvä huomata, että yhteisen tiedon puute aiheuttaa ihmisille monissa tilanteissa hankaluuksia. Monissa tilanteissa arvelemme muiden tavoitteet tai intentiot väärin ja sillä tavoin muokkaamme omaan toimintaamme. Tällöin toimintamme ei perustu yhteiseen tietoon, vaan yleiseen tietämättömyyteen (pluralistic ignorance). (Hendricks & Rendvig 2014.)

Mekanismista

Episteemisen logiikan relevanssia tekoälytutkimukselle voidaan perustellusti epäillä. Kuitenkin episteemisen logiikan perusajatus on selkeyttää episteemisten käsitteiden loogiset ominaisuudet. Siten yhteys tekoälytutkimukseen on nähtävissä. Haaparanta (1995, 41) tuo tämän yhteyden selkeästi esiin seuraavalla tavalla:

”Nk. klassinen tekoäly lähtee liikkeelle Boolean kalkyylin ja Turingin koneen ideoista. Boolean kalkyyllissa sanasto, lauseenmuodostussäännöt ja transformaatio säännöt on annettu, ja Boole itse toteaa jopa, että jokaista hänen kalkyyllinsa piirrettä vastaa jokin ihmismielen piirre. Tämän mallin mukaan looginen ajattelu on tietyn sanaston ja annettujen sääntöjen avulla tapahtuvaa laskemista.” (Haaparanta 1995, 41)

Siten tarkastellessamme episteemistä logiikkaa tulee muistaa, että looginen päättely ajatellaan usein olevan mekaanista päättelysääntöjen noudattamista. Mekaanisuus, johon viitataan, on täsmällisesti määritetty; Turingin koneet tai rekursiivisuus antavat täsmällisen muotoilun mekaanisuudelle. Siten mekaanisuus samaistuu komputationaalisuuteen (laskettavuuteen). Tämä ei tietenkään ole uusi ajatus, vaan sen edeltäjiä ovat esimerkiksi Descartes, Hobbes ja La Mettrie. (Haaparanta 1995; Airaksinen 2006.)

Ei ole lainkaan vieras ajatus kognitiivisessä, että komputationaalisuus on ihmisajattelua rajoittava ominaisuus. Kuitenkin komputationalismi on kohdannut laajaa kritiikkiä. Lucas (1961) argumentoi, että Gödelin epätäydellisyyslausekset osoittavat komputationalismin rajoittuneisuuden ihmisajattelun tarkastelussa. Scheutz (2002) tarjoaa ajankohtaisemman esityksen komputationalismin kritiikistä. Ei ole kovin yleisesti

hyväksytyä olettaa, että aivot olisivat ”laskukone”. Kuitenkin komputationalismin hyväksyntä neurotieteen tasolla tuo esiin komputaation indeterminismin ongelman (Fresco, Copeland & Wolf 2021).

Tekoäly aktualisoituu tietokoneissa, joiden ohjelmistot ovat, ainakin periaatteessa, komputationaalisia. Seuraako tästä, että tekoäly tulee sidotuksi komputationalismiin? Tällöin, mikäli kognitiotieteen komputationaalisuuskritiikki olisi osuvaa, seuraisi, että vahva tekoäly perustuisi virheellisiin oletuksiin (Lucas 1961). Tällainen johtopäätös olisi kuitenkin liian nopea, sillä on mahdollista laajentaa mekanismin käsitettä siten, että voimme antaa täsmällisen määritelmän hyperkomputaatiolle, joka on laajempi kuin tavanomainen komputaatio (Copeland 2002; Putnam 1961). Copeland (2002) määrittelee suppean ja laajan mekanismin, jotka luonnehtivat tavanomaisen komputaation ja laajemman hyperkomputaation.

Se, miten ymmärrämme mekaanisuuden, on historiallisesti muuntunut paljon. Edellä viitattu Copelandin erottelu suppean ja laajan mekaanisuuden välillä on erittäin tärkeä. Usein mekaanisuudesta puhuttaessa, taustalla on ajatus suppeasta mekaanisuudesta. Ns. klassisessa mekanistisessa ajattelussa, kuten Descartesin käsityksestä ihmisestä koneena, taustalla on tietty funktionaalinen ajattelu. Mekanistiseen ajatteluun vaikutti materialistinen filosofia kuten Hobbesilla tai La Mettrieilla, joka ajatteli, Descartesista poiketen, ajatteli, että ei vain ruumis vaan koko ihminen on kone. (Copeland 2002.) Tämä funktionaalinen ajattelu tulee erottaa instrumentaalisesta ajattelusta. Funktionaalinen ajattelu ilmenee edelleen mm. Heideggerin tekniikan filosofiassa (Airaksinen 2005).

1900-luvun mekanistinen ajattelu erityisesti tekoälyyn liittyen liittyy logiikan kehitykseen. Turingin, Gödelin, Postin, Kleenen ja Churchin työt loivat selkeän käsityksen mekaanisen päättelyn luonteesta. Tietojenkäsittelytieteessä edelleenkin hyvin tunnettu Boolean (1847) kehittämä Boolean algebra antaa hyvän kuvan mekaanisen päättelyn luonteesta. Looginen ajattelu on valtaosin kytkeytynyt suppeaan mekanismiin. Tähän loogiseen mekaanisuuteen kytkeytyy esimerkiksi Newellin materiaalien symbolijärjestelmien, kiinalaisen huoneen ajatuskoe ja Blockin homunculus-head -argumentti. (Block 2002.)

Kuitenkaan, kuten Turing (1938) osoitti, ei ole mahdollista systeemiä laajentamalla paeta Gödelin epätäydellisyystuloksia. Laajemmalla komputaatiolla on kuitenkin piirteitä, joita pidämme inhimilliselle tiedonhankinnalle olennaisina, kuten virheistä oppiminen tai lopullisen varmuuden puuttuminen. (Copeland 2002.) Onkin tärkeää huomata, että tekoäly ei mitenkään ole sidottu suppeaan komputationalismiin. Erityisesti on hyvä huomata, ettei Turing myöskään sitoutunut siihen, vaikka hänen mielipiteensä on osin vaikea tämentää (Copeland 2002; 2004; Mutanen & Halonen 2019.)

Episteeminen logiikka ei ole sidottu ensimmäisen kertaluvun predikaattilogiikkaan, vaan on olemassa episteemisen logiikan laajennuksia tai yleistyksiä, joissa vapaudutaan joistakin ”standardin” episteemisen logiikan rajoitteista. Logiikassa yleisemminkin on kiinnitetty huomiota muutokseen (dynamic turn). Episteemiseen logiikkaan on tämä dynaaminen käänne tullut eri tavoin. Baltag & Renne (2016) on yleisesitys keskeisistä

dynaamisen episteemisen logiikan lähestymistavoista. Hintikka (2007) on kehitellyt episteemisen logiikan dynaamisia puolia artikkelissaan ”Second Generation Epistemic Logic”, jossa systematisoidaan tiedon etsinnän logiikkaa (kts. Halosen artikkeli tässä kirjassa). Tiedon etsintä on oikeastaan malliesimerkki siitä, miten rationaalinen tiedonetsintäprosessi etenee. Tiedonetsinnän dynaamisia piirteitä on lähestytty myös nojautumalla metodisesti laskettavuuden teoriaan (Hendricks 2006). Tämä on ollut kiinnostava traditio, jossa on painotettu nimenomaan tiedon etsinnän strategisia puolia, minkä valittu metodinen orientaatio tekee luontevaksi. Tämä traditio alkaa laskettavuuden teorian laajennuksista ja näiden laajennusten soveltamisesta oppimisen analyysiin (Gold 1967; Putnam 1961). Tämä osoittautui hedelmälliseksi lähestymistavaksi, jota oli mahdollista laajentaa oppimisen teoriaan (Osherson, Stob & Weinstein 1986), tiettenfilosofiaan (Kelly 1996; Martin & Osherson 1998). Hendricks (2006) laajentaa lähestymistavan monimodaaliseksi teoriaksi. Edelleen laskettavuuden teorian laajennukset tuovat kiinnostavaa teoreettista ja käsitteellistä ymmärrystä laskettavuuteen (Hyperkomputaatio; Copeland 2002; Syropoulos 2019) ja tämän mielenfilosofisiin sovellutuksiin (Bringsjord & Zenzen 2003)

Loogisesta päättelystä

Episteemisen logiikan taustalla on ajatus, että tiedon logiikkaa voidaan tarkastella tavallisen (ensimmäisen kertaluvun) logiikan viitekehyksessä. Siten logiikan taustalla oleva ajatus loogisesta päättelystä symbolimanipulaationa on tullut osaksi myös episteemistä logiikkaa. Kuitenkaan tällainen ajatus ei oikeastaan vastaa sitä, mitä tiedämme ihmisten tiedonhankinnasta ja -käsittelystä. Oikeastaan symbolimanipulaatio ei kerro myöskään loogisesta päättelystä kovin paljoa. Tietenkin logiikan systeemit ovat muotoiltu formaaleina järjestelminä, joiden puitteissa looginen päättely voidaan esittää mekaanisena symbolimanipulaationa. Tämä ajatus logiikasta on ollut mukana myös episteemisen logiikan, mutta myös yleisemminkin modaalilogiikan, kehittämisessä. Edellä mainittu dynaaminen käänne oli pyrkimys vastata tällaisen lähestymistavan tiettyihin ongelmiin.

Formaalin lähestymistavan taustalla vaikuttaa olevan myös syvempi ongelma, mikä voidaan havaita tarkasteltaessa esimerkiksi Fregen (1879) loogista systeemiä. Fregen kehittelemä systeemi, vaikka monessa suhteessa vaikeasti jäsenyvä, niin tietyssä selkeässä mielessä erittäin intuitiivinen: katsomalla voi nähdä päättelyn rakenteen. Frege ”vetoaa hahmontunnistuskyykyymme sekä lauseanalyysissään että tavassaan esittää päättelyjä” (Haaparanta 1995). Tämä monessa mielessä erittäin luonteva tapa ajatella on hiljalleen hävinnyt formaaleista järjestelmistä. Oikeastaan ns. puolalainen notaatio on tässä suunnassa yksi kaikkein pisimmälle viety notaatiotapa. (Haaparanta 2018.)

Inhimillinen päättely ei välttämättä noudata merkintätapoja. Onkin kiinnostavaa nähdä, että kuvallinen ajattelu on ollut mukana kaiken aikaa. Fregen ohella esimerkiksi Peirce painotti paljon ajattelumme kuvallisuutta, hänelle matemaattinen symbolijärjestelmä oli olennaisesti kuvallista (Haaparanta 1995; 2018; Pietarinen 2012). Tämä saattaa monelle olla vaikea myöntää, mutta Peirce vetoaa näkemyksessään Kantin näkemykseen intuitiivisuudesta. Tätä teemaa on Hintikka kehitellyt matematiikan ja logiikan filosofiassaan sekä tiettenfilosofiassaan. (Hintikka 1969; 2007.) Yleisemminkin

kuvallinen päättely on tullut laajan kiinnostuksen kohteeksi (Zimmermann & Cunningham 1991).

Päättelyssä ei siten ole kyse formaalista symbolien manipulaatiosta, vaan tietyn loogisen rakenteen tunnistamisesta. Siis päättely ei etene merkityksettömien kaavojen kuljettamana, vaan tiettyjen hahmojen tunnistamisen kautta. Tällöin päättelyä ei tarvitse kielellistää ja sitten symbolisesti manipuloida, vaan päättely tapahtuu suoraan tiettyjen hahmojen tai kuvioiden kautta. (Haaparanta 1995; de Toffoli 2017). Tekoälytutkimuksessa tämä sama ajatus on otettu lähtökohdaksi ns. konnektionistisissa malleissa, joissa tiedollinen päättely tapahtuu verkostomaisesti ilman, että tieto formuloidaan symbolisesti tai kielellisesti. Tämän teeman analyysi ei kuulu tämän esseen piiriin. Asiasta tarkemmin kts. Penttinen tässä kirjassa.

Lähteet

Baltag, A. and Renne, B. (2016) ‘Dynamic Epistemic Logic’, in E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Winter 2016. Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/win2016/entries/dynamic-epistemic/>.

van Benthem, J. (2003) ‘Logic and the dynamics of information’, *Minds and Machines*, 13(4), pp. 503–519. Available at: <https://doi.org/10.1023/a:1026296600182>.

Block, N. (2002) ‘Troubles with Functionalism (Excerpt)’, in D.J. Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings*. Oup Usa.

Bringsjord, S., Zenzen, M.J. and Zenzen, M. (2003) *Superminds: People Harness Hypercomputation, and More*. Springer Netherlands (Studies in Cognitive Systems). Available at: <https://books.google.fi/books?id=LBVoDSnEVTIC>.

De Toffoli, S. (2017) ‘Chasing’the diagram—The use of visualizations in algebraic reasoning’, *The Review of Symbolic Logic*, 10(1), pp. 158–186.

van Ditmarsch, H., van der Hoek, W. and Kooi, B. (2007) *Dynamic Epistemic Logic*. Springer.

Floridi, L. and Sanders, J.W. (2004) ‘On the morality of artificial agents’, *Minds and machines*, 14(3), pp. 349–379.

Fresco, N., Copeland, B.J. and Wolf, M.J. (2021) ‘The indeterminacy of computation’, *Synthese*, 199(5), pp. 12753–12775.

Garson, J. (2021) ‘Modal Logic’, in E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Summer 2021. Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/sum2021/entries/logic-modal/>.

Gierasimczuk, N., Hendricks, V.F. and de Jongh, D. (2014) ‘Logic and Learning’, in A. Baltag and S. Smets (eds) *Johan van Benthem on Logic and Information Dynamics*. Cham:

Springer International Publishing (Outstanding Contributions to Logic), pp. 267–288. Available at: https://doi.org/10.1007/978-3-319-06025-5_10.

Hendricks, V.F. and Rendsvig, R.K. (2016) ‘The philosophy of distributed information’, in L. Floridi (ed.) *The Routledge Handbook of Philosophy of information*, Routledge, London, pp. 120–136.

Hendricks, V.F. and Roy, O. (eds) (2010) *Epistemic Logic: 5 Questions*. Automatic Press.

Hintikka, J. (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press.

Hintikka, J. (1969) *Models for Modalities*. Dordrecht, Holland: D. Reidel Publishing Company. Available at: https://doi.org/10.1007/978-94-010-1711-4_6.

Hintikka, J. (1984) ‘Rules, utilities, and strategies in dialogical games’, in J. Hintikka and L. Vaina (eds) *Cognitive constraints on communication*. Dordrecht: D. Reidel Publishing Company, pp. 277–294. Available at: https://doi.org/10.1007/978-94-010-9188-6_16.

Hintikka, J., Halonen, I. and Mutanen, A. (2002) ‘Interrogative logic as a general theory of reasoning’, in D.M. Gabbay et al. (eds) *Studies in Logic and Practical Reasoning*. Amsterdam: Elsevier Science B.V. (Handbook of the Logic of Argument and Inference), pp. 295–337. Available at: [https://doi.org/10.1016/S1570-2464\(02\)80009-X](https://doi.org/10.1016/S1570-2464(02)80009-X).

Lammenranta, M. (2022) *Johdatus tieto-oppiin*. Gaudeamus.

Lovett, A. and Riedener, S. (2021) ‘Group agents and moral status: what can we owe to organizations?’, *Canadian Journal of Philosophy*, 51(3), pp. 221–238. Available at: <https://doi.org/doi:10.1017/can.2021.8>.

MALABOU, C. and SHREAD, C. (2019) *Morphing Intelligence: From IQ Measurement to Artificial Brains*. Columbia University Press. Available at: <http://www.jstor.org/stable/10.7312/mala18736> (Accessed: 21 November 2022).

Menzel, C. (2021) ‘Possible Worlds’, in E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Fall 2021. Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/fall2021/entries/possible-worlds/>.

Miettinen, S.K. (1995) *Logiikan peruskurssi*. Gaudeamus.

Niiniluoto, I. (1985) ‘Imagination and Fiction’, *Journal of Semantics*, 4(3), pp. 209–222. Available at: <https://doi.org/10.1093/jos/4.3.209>.

Oppy, G. (1998) ‘Propositional attitudes’. Available at: <https://doi.org/10.4324/9780415249126-V028-1>.

Pietarinen, Ahti-Veikko (2012) 'Peirce and the logic of image', 2012(192), pp. 251–261. Available at: <https://doi.org/10.1515/sem-2012-0083>.

Rendsvig, R. and Symons, J. (2021) 'Epistemic Logic', in E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Summer 2021. Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/sum2021/entries/logic-epistemic/>.

Russell, S.J. and Norvig, P. (2003) *Artificial intelligence - a modern approach*. 2nd edn. Upper Saddle River, New Jersey: Prentice Hall (Prentice Hall series in artificial intelligence).

Salminen, H. and Väänänen, J. (1992) *Jobdatus logiikkaan*. Gaudeamus.

Schumacher, C. (1992) 'A logic for memory', in D. Pearce and H. Wansing (eds) *Non-classical Logics and Information Processing*. Berlin/Heidelberg: Springer-Verlag (Lecture Notes in Computer Science), pp. 23–45. Available at: <https://doi.org/10.1007/BFb0031921>.

Searle, J.R. (1989) *Minds, brains and science*. Penguin Books.

Syropoulos, A. (2019) 'On TAE machines and their computational power', *Logica Universalis*, 13(2), pp. 165–170. Available at: <https://doi.org/10.1007/s11787-018-0196-5>.

Väänänen, J. (1987) *Matemaattinen logiikka*. Gaudeamus.

White, M.J. (1980) 'Necessity and unactualized possibilities in Aristotle', *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 38(3), pp. 287–298.

Zimmermann, W. and Cunningham, S. (1991) *Visualization in teaching and learning mathematics*. Mathematical Association of America.

TIEDONHANKINNAN STRATEGIOISTA

Ilpo Halonen

Jobdanto

Ateenassa vaikuttanutta Sokratesta (470–399 eaa.) sanotaan usein yhdeksi kaikkien aikojen merkittävimmäksi filosofiksi, ja häntä kutsutaan koko länsimaisen filosofian isäksi. Yksi tärkeä seikka Sokrateen opeissa oli, että hän asetti totuuden pelkän vakuuttavuuden ja menestyksen yläpuolelle. Sokraattiseen asenteseen kuuluu myös kriittisyys. Ainakin Sokrateesta ja hänen oppilaastaan Platonista asti tiede pyrittiin määrittelemään siten, että sen kohteena on muuttumaton aito tieto. Tästä seurasi näkemys tiedon varmuudesta. Ihanteeksi tällaisessa tiedekäsityksessä muodostui ideaalisia geometrisia olioita koskeva todistuva tieto. Geometrinen tiedon muodostus on mahdollista nähdä yleisemminkin kokeellisen tieteen metodologisena ihanteena. (Hintikka 1969.)

Sokrateen asenne heijastuu filosofiassa sokraattisessa menetelmässä, jolla yleensä tarkoitetaan oman ajattelun harjaannuttamista vuoropuhelussa toisten kanssa. Vuonna 2007 ilmestyneessä kirjassaan *Socratic Epistemology. Explorations of Knowledge-Seeking by Questioning* suomalainen filosofi Jaakko Hintikka (1929-2015) pyrki esittämään tiedonhankinnan kyselyprosessina – eräänlaisena sokraattisen menetelmän päivitettyä versiona.

Platonin tunnetuin oppilas Aristoteles yleistti opettajiensa menetelmää edelleen osoittamalla, kuinka dialektiikan, vuoropuhelun, argumentaation säännöt perustuivat niitä yleisempään todistamisen menetelmään, ja että tätä menetelmää voitiin tutkia myös käytännön väittelytilanteista riippumatta. Välttämättömyyden ja todistamisen käsitteet olivat keskeisiä antiikin luonnontieteiden ja matematiikan kehityksessä. Aristoteles kehitti teorian väitelauseiden välttämättömille seuraussuhteille perustuvasta todistamisesta, ja tämä on lajissaan ensimmäinen. Riku Jutin (2013, 60-61) mukaan teoriaa kehittäessään ”Aristoteles tuli perustaneeksi muodollisen logiikan tutkimuksen. Kukaan aikaisempi ajattelija ei ollut yrittänyt mitään vastaavaa, ja kesti yli kaksituhatta vuotta ennen kuin sitä yritettiin samassa mittakaavassa uudestaan.” Näin sokraattista menetelmää voidaan hyvin pitää länsimaisen tieteellisen ajattelun ja päättelyn kantamuotona.

Tiedonhankinnan interrogatiivimalli (kyselymalli)

Professori Jaakko Hintikka toi nykyaikaisen tutkimusprosessin sokraattiset juuret näkyvästi esiin kehittämässään *tiedonhankinnan kyselymallissa*. Jaakko Hintikka oli ”kautta aikojen kansainvälisesti tunnetuimpia suomalaisia tutkijoita, joka teki loisteliaan uran inhimillisen ajattelun välineiden tutkijana Suomen ja Yhdysvaltojen yliopistoissa” (HS 14.8.2015). Hintikka tutki tiedon ja uskomuksen käsitteitä jo vuonna 1962 julkaistussa klassisessa teoksessaan *Knowledge and Belief - An Introduction to the Logic of the Two Notions*. Tuo kirja merkitsi tärkeätä askelta episteemisen logiikan kehityksessä. Vuonna 1976 Hintikka julkaisi kirjan kysymysten ja vastausten logiikasta. Hintikan näkemysten kes-

keisenä osana oli jo pitkään se, että sokraattinen menetelmä ei ole hävinnyt minnekään, vaan se toimii edelleen systemaattisen filosofian osana. Tätä systemaattista sokraattista kyselymenetelmää Hintikka pyrki päivittämään 1980-luvulta alkaen, vaikka osa näistä ajatuksista oli syntynyt jo 1960-luvulla. Näin syntynyt *tiedonhankinnan interrogatiivimalli* eli *kyselymalli* antaa yleisen luonnehdinnan kaikelle tiedonhankinnalle loogisten askelten ja kysymys-vastaus –askelten vuoropuheluna. Hintikan aikaisemmat tulokset toimivat tämän mallin systemaattisina työvälineinä siten, että malli yhdistää kysymysten ja vastausten logiikkaan kehittyneemmän ns. toisen sukupolven episteemisen logiikan ja näin korostaa erityisesti mallin dynaamista luonnetta.

Tiedonhankinnan kyselymalli on luonteeltaan loogis-käsitteellinen tai metodologinen, joten malli ei kuvaa tosiasiallista (empiiristä) inhimillistä päättelyä vaan päättelyn loogis-käsitteellisiä piirteitä. Näin ollen mallin soveltaminen niin yksittäisen tieteen tekijän päättelyn kuin yleisen tieteellisen päättelyn luonnehtimiseen on mahdollista. Kyselymallissa tutkijan voidaan ajatella lähtevän liikkeelle sellaisista väitelauseiden avulla ilmaistavissa olevista tiedoista, jotka hänellä jo on. Työssään hän pyrkii etenemään kohti sellaista uutta tietoa, jota hänellä ei vielä ole. Hän pyrkii kohti totuutta - kohti uutta väitelausein ilmaistavissa olevaa tietoa. Tätä peräkkäisten väitelauseiden muodostamaa jonoa voidaan sanoa tutkijan argumentointiketjuksi. Tähän ketjuun voidaan tuoda uusi lause kahdella tavalla: joko ketjun uuden väitelauseen sisältämä informaatio sisältyy ketjun aikaisempiin väitelauseisiin tai sitten se ei sisälly. Ensimmäisessä tapauksessa ketjuun muodostettu uusi väitelause on aikaisempien lauseiden looginen (tai deduktiivinen) seuraus. Tällaiset argumentaatioketjun askelet ovat logiikan tutkimuskohteita.

Toinen vaihtoehto on se, että uuden askelen sisältämä informaatio ei sisälly aikaisempiin väitelauseisiin vaan on aidosti uutta. Tällaisissa tapauksissa on luonnollista kysyä, mistä informaatio on saatu. Tämä on luonnollista aina, kun olemme tekemisissä rationaalisen argumentoinnin kanssa. Varsinkin tieteellisessä toiminnasta puhuttaessa kaikkien tiedonhankintaan ja päättelyyn kuuluu olennaisena osana perustelun vaatimus. Koska tieteellisessä ongelmanratkaisussa tutkija on aktiivinen toimija, voidaan ajatella, että hän saa uuden informaation tavalla tai toisella vastauksena omaan aktiiviseen toimintaansa. Näin argumentaatioketjun uutta informaatiota voidaan pitää vastauksena kysymykseen, jonka tutkija on esittänyt käyttämälleen informaatiolähteelle. Näin ollen mitä tahansa päättely- tai perusteluketjua voidaan ajatella kysymys-vastaus -prosessina.

Keksimisen tie ja todistamisen tie

Perinteisen luonnehdinnan mukaan loogiset totuudet ovat totuuksia kaikissa mahdollisissa maailmoissa, siis kaikkien mahdollisten asiaintilojen vallitessa. Samoin looginen seuraaminen on “universaalista” seuraamista - seuraamista kaikkien asiaintilojen vallitessa. Tästä seuraa, että puhdas logiikka ei kerro mitään satunnaisista totuuksista tai seurauksista, jotka vallitsevat tai eivät vallitse esimerkiksi aktuaalisessa maailmassa. Logiikka on se kivijalka, jonka päälle koko tieteellinenkin ajattelu rakentuu. Mutta pelkän logiikan varassa ei päästä pitkälle. Kun mestarisalapoliisi Nero Wolfe toteaa Rex Stoutin tarinassa *Ei aivan tarpeeksi kuollut*, että ”lähden siitä oletuksesta, että te joko

murhasitte Ann Amoryn tai ette, mikä tuntuu järkeenkäyvältä”, hän epäilemättä on oikeassa. Mutta salapoliisin varsinainen työ alkaa vasta tästä vaiheesta - olennaista on selvittää, kumpi näistä kahdesta vaihtoehdosta on juuri meidän reaali maailmassamme se oikea. Empiirisessä tutkimuksessa kiinnostus kohdistuu loogisten totuuksien lisäksi kontingentteihin totuuksiin ja epätotuuksiin, lauseisiin, jotka ovat tosia joidenkin mutta epätosia toisten asiointilojen vallitessa.

Uutta informaatiota kuvattiin edellä tutkijan saamana vastauksena esittämäänsä kysymykseen. Informaatio määrittyy sen poissulkemien vaihtoehtojen perusteella. Jonkun asian tietäminen on taitoa eliminoida vaihtoehtoja. Informaatio sallii meidän sulkeistaa tutkimuksestamme pois tiettyjä vaihtoehtoja, joita olisimme voineet muuten pitää mahdollisina. Informaatio tekee tutkijalle mahdolliseksi sulkea pois sellaisia vaihtoehtoja, joiden olisi muuten voinut kuvitella toteutuvan. Usein sanotaankin, että informaatiolla tarkoitetaan epävarmuuden vähentymistä. Tutkijan täytyy kysyessään olla kyseisen aihepiirin asiantuntija. Hänellä täytyy olla tietoja ja taitoja esittää sellaisia kysymyksiä, joihin saatavien vastausten avulla voidaan toteuttaa tiettyjä vaihtoehtoja ja sulkea siten samalla tiettyjä toisia pois. Tämä liittyy tutkijan käyttämiin strategioihin, joihin palaan alla.

Joskus on ajateltu, että edellä esitelty kahden askeltyypin erottelu on ollut liian jyrkkä. Tämä erottelu näkyy erittäin hyvin hypoteettis-deduktiivisessa tiedekäsityksessä, jonka mukaan ”keksimisen tie” ja ”todistamisen tie” on pidettävä jyrkästi erillään. Logiikka kuuluu ainoastaan todistamiseen, hypoteesin deduktiivisten seurausten johtamisen ja testaamisen vaiheeseen. Filosofiasa on vallinnut yleinen yksimielisyys siitä, että emme voi muotoilla määritteleviä sääntöjä keksimiselle, toisin sanoen aidosti uuden tiedon hankkimiselle. Sen sijaan hypoteesit - tiedonhankinnan kohteena olevat ongelmien ratkaisuehdotukset – keksitään jotenkin muulla tavalla. (Ks. tarkemmin esim. Niiniluoto 1983, luku III:3.) Jaakko Hintikan kyselymalli pyrkii osoittamaan, että kaikenlaisen luotettavan tiedon löytämisen vaiheisiin kuuluvat molemmat askeltyypit käsi kädessä.

Jaakko Hintikan kumppaneineen kehittämän mallin perusajatus on siis sellainen, että kaikki argumentaatioprosessit (ja siis erityisesti tieteellisen tutkimuksen ongelmanratkaisuprosessit) voidaan esittää loogisten askelten ja kyselyaskelten vuoropuheluna. Jokaisessa prosessin vaiheessa tutkija voi tehdä valinnan näiden kahden välillä: jos hänellä on tarpeeksi informaatiota, hän valitsee tyypillisesti loogisen päättelyaskeleen hallitsemansa informaation käsittelyä varten. Mutta jos hän tarvitsee uutta informaatiota, hänen täytyy kääntyä kyseisessä tilanteessa parhaan mahdollisen informaatiolähteen puoleen. On tärkeää oivaltaa, että erityisesti tutkijan argumentoinnissa ja tieteellisessä tiedonhankinnassa olennaista on taito osata kääntyä oikealla hetkellä sellaisen informaatiolähteen puoleen, joka on paras mahdollinen kyseisessä tilanteessa.

Mutta olennaista on, että kyselymallissa ei keskitytä pelkästään yksittäiseen kysymykseen saatavaan vastaukseen ja sen mukanaan tuomaan lisäinformaatioon. Olennaista on tutkia kokonaista kysymysten ja vastausten prosessia. Näin ollen mallissa liitetään toisiinsa kysymysten ja vastausten logiikka sekä episteeminen logiikka, joihin tuodaan mukaan dynaaminen painotus. Tutkija - toimiessaan prosessissa kysyjänä - pyrkii

saavuttamaan episteemisen (tiedollisen) tavoitteensa hankkimalla tarvittavan määrän informaatiota.

Määrittelevät ja strategiset säännöt

Jaakko Hintikka tekee käsitteellisen eron kahdentyyppisten sääntöjen välille: ensimmäisiä hän kutsuu määritteleviksi säännöiksi ja toisia strategisiksi säännöiksi. Edellisten sääntöjen avulla määritellään ”pelattava peli” eli argumentoinnin ollessa kysymyksessä lyödään lukkoon se, minkä tyyppisen argumentoinnin kanssa kulloinkin ollaan tekemisissä (tieteellinen päättely, deduktiivinen/induktiivinen päättely, väittely, poliittinen päätöksenteko, arkikeskustelu ...). Määrittelevistä säännöistä mainitaan usein shakkipelin säännöt. Ne kertovat, mitkä siirrot ovat kyseisessä pelissä mahdollisia tai sallittuja. Sen sijaan ne eivät vielä kerro meille, minkälaiset siirrot ovat hyviä tai minikäläiset kenties mestarillisia. Sen tekevät tilannekohtaiset strategiset säännöt – tai paremminkin näiden kokoelmat, strategiat. Pelkkien määrittelevien sääntöjen tunteminen ei tee kenestäkään mestaripelaajaa. Mestaripelaajan on luonnollisesti hallittava pelisäännöt perustaksi, mutta vasta erilaisten strategioiden kehittäminen kutakin erilaista tilannetta varten antaa hänelle avaimet pärjäämiseen.

Tämä liittyy olennaisella tavalla tieteelliseen tiedonhankintaan. Jaakko Hintikka toteaa, että nimenomaan tieteellisten argumentointiprosessien yhteydessä logikan päättelysäännöt ovat määritteleviä sääntöjä. Ne eivät tee kenestäkään mestaria tieteellisessä ongelmanratkaisussa. Logikan säännöt voivat kertoa ainoastaan, miten tutkija voi välttää virheet tutkimustyössään. Mutta menestyäkseen työssään ja löytääkseen uusia kenties merkittäviäkin tieteellisiä tuloksia tutkija tarvitsee toisenlaisia sääntöjä, kullekin eri argumentaatiotyypille ominaista erilaisten strategisten sääntöjen kokoelmaa.

Interrogatiivimalli antaa luontevan metodologisen viitekehyksen tarkastella tutkimusprosessia kokonaisuudessaan. Interrogatiivimallin ideana on, että tutkimus on strategista tiedon etsintää. Väitteen verifioiminen tarkoittaa tutkimusprosessin osoittavan väitteen todeksi ja falsifioitavuus tarkoittaa vastaavasti tutkimusprosessin osoittavan väitteen epätodeksi. Mikäli saatava informaatio on totta, niin interrogatiivimallin lopputulos tosiasiallisesti joko verifioi tai falsifioi väitteen. Luonnollisesti tieteessäkään ei ole mahdollista olettaa kaiken saatavan informaation olevan totta. Siten tulee kehittää metodologisia periaatteita, jotka mahdollistavat saatavan informaation totuudenmukaisuuden systemaattisen arvioinnin tutkimusprosessin kuluessa. Interrogatiivimallissa tämä on toteutettu ns. sulkeistamisoperaation avulla. (Ks. tarkemmin Hintikka, Halonen & Mutanen 2002.)

Näin ollen keksiminen ja todistaminen ovat Hintikan mallin mukaan yhden ja saman tiedonhankintaprosessin eri puolia. Molemmat tehtävät kuuluvat luotettavan tiedon löytämisen strategioihin. Koska tiedonhankintaa koskevien strategioiden arvioiminen voi loppujen lopuksi koskea vain kokonaisia strategioita, todistamisen aluetta ei voida käsitellä erillään keksimisen alueesta. Filosofisen tutkimuksen kohteena ”todistamisen tietä” tulee tutkia aina ”keksimisen tiehen” kytkeytyneenä.

Tekoäly, Turing ja interrogatiivimalli

Mutta onko tällä edellä esitetyllä mitään tekemistä tekoälyn ja tekoälytutkimuksen kanssa? Olemme käsitelleet tätä kysymystä laajemmin yhdessä Arto Mutasen kanssa kirjoittamassamme artikkelissa ”Turingin testi, interrogatiivimalli ja tekoäly” (Mutanen & Halonen 2020), jossa lähestymme teemaa Alan Turingin esittämien ajatusten kautta. Alan Turing (1912–1954) oli lyhyen elämänsä aikana edelläkävijä mm. tietojenkäsittelytieteen ja tekoälyn aloilla. Monet hänen esittämistään ideoista ovat vasta nykyään tulleet ajankohtaisiksi. Etsiessään vastausta kysymykseen ”voiko kone ajatella” Turing muotoili ns. Turingin testin, jota kehittäessään hän päätyi testissään nojaamaan keskustelumenetelmään, jonka tarkoitus oli arvioida koneen (tai muun kohteen) älyllisyyttä. Testillä pyrittiin testaamaan, kykeneekö kyselijä erottamaan ihmisen ja koneen näiden antamien vastausten perusteella.

Kuitenkin tässä kirjoituksessa esiteltyä Jaakko Hintikan interrogatiivimallia tarkastelemalla voidaan väittää, että älyllinen toiminta ei ilmene parhaiten niinkään vastaajan roolissa toimittuna, vaan erilaisten ongelmatilanteiden ratkaisijana ja siten strategisten kysymysten esittäjänä. Turingin logiikka vuonna 1936 kirjoitetussa artikkelissa on juuri tämän ajatuksen mukainen. Turing etsi keinoja muotoilla vastaavaa myös tekoälyn kohdalla.

Turing etsi uransa varrella älykkyyttä ja luovuutta kiinnostavilla ja filosofisesti rikkailla menetelmillä. Turingin ajatusten kehittäminen suuntautui lopulta testin sellaiseen muotoiluun, joka lopulta oli selkeästi mitattavissa. Tällainen testi ei kerro aivan niin paljon, kuin Turing siltä aluksi odotti. Kuitenkin hänen kehittäjänsä on täynnä kiinnostavia luonnehdintoja, joita olisi tärkeää ja mielenkiintoista tutkia edelleen. Yksi suunta, jonka havaitsimme aikaisemmassa artikkelissamme, on se, että Turingin testiä olisi kiinnostavaa lähteä kehittämään nimenomaan interrogatiivimallin pohjalta tiedon etsinnän osoittaessa älykkyyttä. Tällä olisi suora yhteys Turingin omiin ideoihin ja hänen kehittämäänsä metodiseen lähestymistapaan. Testi olisi Turingin ajattelua seuraten nimenomaan metodinen testi, jonka periaatteessa voisivat hyväksytysti suorittaa yhtä hyvin ihmiset kuin koneetkin. Kuitenkin testin läpäisy edellyttäisi toimintaa, jota kutsuisimme älykkääksi. Testi on periaatteessa sellainen, ettei tiedon etsinnän strategiaa ole mahdollista mekanisoida, mikä on testin kannalta hyvä ominaisuus.

Näin olisi hedelmällistä yhdistää kahden aikamme merkittävän ajattelijan - Alan Turingin ja Jaakko Hintikan - ideoita. Tuloksena voisimme ymmärtää entistä paremmin ajattelun välineiden luonnetta myös tekoälytutkimukseen sovellettuna.

Lähteet

Halonen, I. (2004) ‘Argumentaatio ja ammatillinen asiantuntijuus’, in H. Kotila and A. Mutanen (eds) *Tutkiva ja kehittävä ammattikorkeakoulu*. Helsinki: Edita, pp. 255–265.

Halonen, I. (2012) ‘Tieteellinen asiantuntijuus’, in A. Mutanen (ed.) *Näkökulmia asiantuntijaksi kasvamiseen*. Helsinki: Merisotakoulun julkaisuja (Merisotakoulun julkaisuja, A, 2012), pp. 20–26.

Halonen, I. and Mutanen, A. (2006) 'Asiantuntijuus epistiikan, etiikan ja estetiikan kohtaamisena', in O. Lampinen et al. (eds) *Ammattien kutsu: ammattikorkeakoulut ja estetiikka*. Helsinki (Opetusministeriön julkaisuja, 2006: 18), pp. 144–154.

Hintikka, J. (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca (NY). Uusintapainos: King's College Publications, London 2005.

Hintikka, J. (1969) *Tieto on valtaa*. Porvoo, Helsinki: WSOY.

Hintikka, J. (1976) 'The semantics of questions and the questions of semantics', *Acta philosophica Fennica*, 28(4).

Hintikka, J., Halonen, I. and Mutanen, A. (2002) 'Interrogative logic as a general theory of reasoning', in D.M. Gabbay et al. (eds) *Studies in Logic and Practical Reasoning*. Amsterdam: Elsevier Science B.V. (Handbook of the Logic of Argument and Inference), pp. 295–337. Available at: [https://doi.org/10.1016/S1570-2464\(02\)80009-X](https://doi.org/10.1016/S1570-2464(02)80009-X).

Hintikka, J., Hiipakka, J. and Vilkkö, R. (2001) *Filosofian köyhyys ja rikekaus: nykyfilosofian kartoitusta*. Helsinki: Art House.

Hintikka, K.J.J. (2007) *Socratic Epistemology: explorations of knowledge-seeking by questioning*. New York: Cambridge university press.

Juti, R. (2013) *Tiedon filosofia*. Gaudeamus.

Mutanen, A. and Halonen, I. (2020) 'Turingin Testi, Interrogatiivimalli Ja Tekoäly', *Ajatus*, 77, pp. 169–204.

Niiniluoto, I. (1980) *Johdatus tieteenfilosofiaan: käsitteen- ja teorianmuodostus*. Helsinki: Otava.

Niiniluoto, I. (1983) *Tieteellinen päättely ja selittäminen*. Helsingissä: Otava.

Niiniluoto, I. (2003) *Totuuden rakastaminen: tieteenfilosofisia esseitä*. Helsingissä: Otava.

Siitonen, A. and Halonen, I. (1997) *Ajattelu ja argumentointi*. Helsinki: WSOY.

TURING JA TEKOÄLY

Arto Mutanen

Jobdanto

Tekoäly on ollut sekä julkisen että tieteellisen mielenkiinnon kohteena jo pitkään. Tieteiskirjallisuus (science fiction) ja viihdeteollisuus ovat tuoneet esiin kiinnostavia piirteitä asiaan liittyen. On pohdittu koneiden tietoisuuden mahdollisuutta ja luonnetta (I Robot) tai ihmisen ja koneen välistä suhdetta tai kyborgin luonnetta (Robocop), lisäksi on tarkasteltu yhteiskunnallisen valvonnan roolia (Minority Report). Tieteiskirjallisuuden anti mielikuvituksen ruokkijana voi olla hyväkin, mutta usein tieteiskirjallisuuden teosten taustalla olevat perusteet ovat moninaiset ja siten vaikea identifioida. (Niiniluoto 1990; Lammenranta 1987; Vinge 1993.) Vaikka taide voikin herättää kiinnostavia kysymyksenasetteluja, niin tekoälyn perusteiden selkeyttäminen edellyttää systemaattista käsitteellistä analyysia.

Tieteellisen ajattelun kannalta on olennaista, että pyritään mahdollisimman selkään ja systemaattiseen ajatuksenmuotoiluun. Tällainen ei kuitenkaan rajaa tarkasteltavia teemoja, kuten keskustelu posthumanismista ja transhumanismista osoittavat (Nayar 2013; Ross 2020; Pihlström 2021). Koskinen (2019, 56) kuvaa eroa post- ja transhumanismin välillä sanomalla, että ”siinä missä posthumanismi lähtee humanismin ja etenkin valistushumanismin kritiikistä, transhumanismi – jota on kehitelty aivan muualla kuin humanistisilla aloilla – on teknologiaoptimistista valistushumanismia”. Posthumanistille ihminen on osa teknologian varaan rakentuvaa systeemiä ja transhumanistille tulevaisuuden tekninen kehitys auttaa poistamaan ihmisen olennaisen riittämättömyyden. Innoittajina tieteelliseen tutkimukseen voi toimia niin arkielämä, kaukokirjallisuus kuin myyttiset tarinat.

Tietojenkäsittelytieteessä (computer science) on kehitetty oppivia koneita (machine learning), joiden ominaisuudet nojaavat tiettyihin logiikan ja todennäköisyyslaskennan perusajatuksiin (Anthony & Biggs 1992). Koneoppimisessa tavoitteena on tutkia ja kehittää koneita, joilla on kyky oppia kokemuksen perusteella. Koneoppiminen perustuu usein joihinkin heuristiikkoihin, jotka muuntavat koneen ohjelmistoa siten, että koneen toiminta muistuttaa ihmisen toimintaa hänen oppiessaan. Koneoppimisessa usein tutkitaan suhteellisen rajoitettuja oppimistilanteita; tutkimuksessa on paljon tarkastelu pelejä, kuten shakkia. Tämän tradition juuret ulottuvat aina logiikan historian juuriin, Aristoteleen logiikkaan. Maanpuolustuskorkeakoulussa on koneoppimiseen liittyvää tutkimusta tehty professori Jouko Vankan tutkimusryhmässä.

Tekoäly (Artificial Intelligence; AI) on laaja ja vaikeasti jäsentyvä ala. Koneoppiminen kytkeytyy osaltaan tekoälyyn, sillä onhan oppiminen eräs älykkyyden tunnusmerkeistä. Voidaan kysyä, onko älykkyys nimenomaan ihmisille ominainen piirre tai ovatko kaikki ihmiset (samalla tavoin) älykkäitä. Älykkyydellä on eri tieteenaloilla erilaisia merkityksiä ja lisäksi sanan ’älykkyys’ historialliset käyttötavat ja merkitykset ovat vaihdelleet ajan saatossa (Lewis 1967). Se, miten älykkyyttä luonnehditaan, vaikuttaa paljon siihen, millaiseksi keskustelu tekoälystä muotoutuu.

Tutkimusalueen nimenä ”tekoäly” esiintyy ensimmäisen kerran John McCarthyn vuonna 1956 organisoiman konferenssin nimessä ”The Dartmouth Summer Research Project of Artificial Intelligence” (Copeland 1993, 8). Konferenssiin osallistui laaja joukko ansioituneita tutkijoita, joiden työt tunnetaan edelleen, McCarthyn lisäksi mukana olivat esimerkiksi Marvin Minsky, Claude Shannon, Allen Newell ja Herbert Simon. Alkuperäinen ehdotus tilaisuuden ohjelmaksi on julkaistu vuonna 2006 AI Magazine -lehdessä. Ohjelmasta näkee yhtäältä tuolle ajalle ominaisista kysymyksenasettelua, mutta samalla syvällistä pyrkimystä löytää hedelmällisiä kysymyksenasetteluja uudelle tutkimussuunnalle.

Puhuttaessa tekoälystä, ei ole varmaa ajatellaanko tekoälyn olevan ”oikeaa” älyä vai jotain muuta, joka vain muistuttaa ”oikeaa” älyä. Tämä ilmenee esimerkiksi puhuttaessa vahvasta tai heikosta tekoälystä. Vahvaan tekoälyyn liittyy ajatus, että tietokone voi olla sananmukaisessa merkityksessä älykäs. Heikossa tekoälyssä ajatus on, että tekoäly rajoittuu joihinkin täsmällisesti rajattuihin tehtäviin tai aloihin liittyvään ”älykyyteen” ja siten tietokoneen älykkyys on rajattua ja puutteellista älykkyyttä. (Copeland 1993; Hallamaa, Haikarainen, & Kalliokoski 2021.)

Tekoälykeskustelun taustalla on usein ajatus, että kone on älykäs, jos sen toiminta muistuttaa riittävässä määrin ihmisen älykästä toimintaa. Vastaavalla tavalla myös eläinten älykkyyttä usein arvioidaan suhteessa ihmisten älykyyteen. Tällainen lähtökohta edellyttää, että inhimillinen älykkyys on sekä hyvin ymmärretty asia että älykyyden malliesimerkki. Kuitenkaan ei tiedetä, mitä inhimillinen älykkyys viime kädessä on tai onko se tosiasiaa edes älykyyden malliesimerkki. (Hintikka 1990.; Chalmers 2010) Edelleen on kyseenalaista, että olisi jokin tietty ihmisen älykkyys. Tekninen älykkyys, joka ilmenee esimerkiksi työkalujen valmistuksessa ja käytössä, esiintyy laajalti ihmisyyhteisöissä ja myös eläimillä. Ihmisen älykkyys ei kuitenkaan rajaudu ainoastaan tällaiseen tekniseen älykyyteen (instrumentaalinen rationaliteetti), vaan on myös jotain, mitä ehkä voisimme kutsua viisaudeksi. Viisauden luonnehdinta on osoittautunut vaikeammaksi tehtäväksi kuin teknisen älykyyden. (von Wright 1987.)

Tiettyjen rajattujen toimintojen yhteydessä, kuten joidenkin erityisalojen asiantuntijana toimimisessa tai joidenkin pelien, kuten ristinolla tai shakki, pelaamisessa voidaan nähdä älykyyden ilmentymistä. Turing (1945) kehitteli ajatusta shakkia pelaavista koneista. Tekoälytutkimuksessa tällaisten rajoitettujen tilanteiden tarkastelut ovat osoittautuneet hedelmällisiksi. Esimerkiksi tietokoneita on pyritty opettamaan pelaamaan joitakin pelejä. Harjoittelun kautta koneet ovat sekä pystyneet parantamaan pelaamistaan että oppimaan pelistrategioita, jotka parantavat niiden ”pelisilmää”. (Turing 1953; Michie 1961.)

Pelien pelistrategiat voidaan luokitella niiden hedelmällisyyden tai neuvokkuuden perusteella. Hyvät pelistrategiat ilmentävät pelaajan oivaltavuutta ja taituruutta. Siten jos koneelle voidaan opettaa neuvokkaita pelistrategioita, niin tällöin saisimme perusteita ainakin heikon tekoälyn hyväksymiselle. Tietenkin koneen kyetessä itse oppimaan pelistrategioita, koneiden rajatulle älykyydelle saataisiin lisäperusteita. Siten Michien (1961), ehkä vaatimattomilta vaikuttavat, tulokset olivat tärkeitä. Ajatus, että koneet mekaanisina laskentakoneina eivät kykene luovaan ajatteluun, tulisi olla nähtävissä

myös koneiden pelistrategioista. Turing (1948) pyrki kehittämään tähän tilanteeseen testiä, joka osoittaisi, ettei usko koneen pelistrategioiden mekanistisuuteen pitäisi paikkaansa. Hänen kehittämänsä testi oli tunnistamispeli, jossa ihmisen tuli tunnistaa peliä seuraamalla kumpi kahdesta pelaajista on kone.

Tietokoneiden ohjelmat ovat tiettyihin tehtäviin laadittuja ohjelmointikielisiä sääntökokoelmia. Tällaisen järjestelmän oppiminen edellyttää, että kone kykenee muuttamaan omaa ohjelmaansa tavalla tai toisella. Turing (1947) kehitti ajatusta oppivista koneista, joiden oppiminen perustui koneen kykyyn muuttaa omaa ohjelmistoaan. Kuitenkin tällainen järjestelmä on sekä raskas että joustamaton. Ohjelmiston muokkaaminen voi tapahtua sääntöjä muokkaamalla tai vaihtamalla. Toisaalta oppiakseen on mahdollista myös rikastuttaa taustalla olevaa kieltä, jolloin uuden ohjelman laatiminen saattaa osoittautua vaikeaksi tehtäväksi. Oppimisen tutkimuksessa on tarkasteltu oppimisen edellyttämää kielen rikastuttamista (diSessa 2014). Lisäksi ihmisen oppiminen ei – ainakaan aivotasolla – tapahdu tällaisen kielellisen mallin avulla, vaan tiettyjen hermoverkkojen uudenlaisen aktivaatiotasojen muutoksiin perustuen. Tällaisia malleja, joita jo Turing kehitti (Turing 1950, 416-418; Copeland 2004, 360, 403), kutsutaan konnektionistisiksi malleiksi. (Russell & Norvig 2003).

Turing

Alan Turing (1912-1954) teki lyhyen elämänsä aika laaja-alaista ja merkittävää työtä. Hänet tunnetaan edelleen hyvin hänen töistään logiikassa (ratkeamattomuus) ja tietojenkäsittelytieteessä (Turingin kone). Turing työskenteli useilla eri tieteenaloilla, joista mainittakoon seuraavat: filosofia, matematiikka, tietojenkäsittelytiede, kryptoanalyysi, kognitiotiede, tekoäly sekä keinoelämä (Copeland 2012). Turingin merkitys nimenomaan tekoälytutkijana on suuri. Hän kehitti monia ideoita, jotka veivät ajattelua uusiin suuntiin. Näiden erilaisten ajatusten varaan on vaikeaa rakentaa yhtenäistä teoreettista näkemystä. Usein on tärkeää tarkastella keskustelua tietyistä yksityiskohtaisista kysymyksistä, jolloin Turingin ajattelun terävyys tulee esiin (Copeland & Shagrir 2013). Monet Turingin ideoista loivat uusia ja hedelmällisiä teoreettisia näkökulmia ja sellaisina ne olivat lähes vallankumouksellisia. Näin tulee esiin Turingin ajattelun teoreettisen mielikuvituksen rikkaus ja hedelmällisyys; hän kykeni hahmottelemaan monia tulevaisuuden kehityskulkuja omissa töissään.

Turingin tekoälyyn liittyvään tutkimukseen voidaan lähteä tutustumaan tarkastelemalla hänen nimeään kantavaa testiä, Turingin testiä, jonka hän julkaisi vuonna 1950 ilmestyneessä artikkelissaan ”Computing Machinery And Intelligence”. Artikkelissa Turing pohtii kysymystä ”voiko kone ajatella”. Vastausta etsiessään hän muotoilee pelin, jota hän kutsuu jäljittelypeliksi (imitation game), joka tunnetaan edellä mainitulla Turingin testin (Turing Test) nimellä. Turingin muotoilu on seuraava:

”Ongelman uutta muotoa voidaan kuvailla pelillä, jota sanomme ’jäljittelypeliksi’. Sitä pelaavat kolme ihmistä: mies (A), nainen (B) ja kysymysten esittäjä (C), joka voi edustaa kumpaa tahansa sukupuolta. Kysyjä sijoitetaan eri huoneeseen kuin kaksi muuta. Pelin tavoitteena on, että kysymysten esittäjä saa selville, kumpi vastaajista on mies ja kumpi on nainen.” (Laina kirjasta Hodges, 2000, 50–51.)

Ei liene kovin selvää, mitä tällaisella pelillä tavoitellaan. Nykyisen lukijan silmiin pelin asetelma vaikuttaa jopa seksistiseltä ja siten nykyisin tällaista muotoilua tuskin kukaan enää tekisi. Kuitenkin tämä muotoilu kuvaa Turingin tyyliä artikkelissa; hänen argumentaationsa on sekä terävää että provosoivaa. Turing jatkaa pelin kehittelyä seuraavalla tavalla: ”Esitämme nyt kysymyksen: ’Mitä tapahtuu, kun kone omaksuu tässä pelissä osan A?’ Päätyykö kysymysten esittäjä väärään johtopäätökseen yhtä usein näin pelattaessa, kuin jos peliä pelataan miehen ja naisen välillä?” (ibid.). Turingin muotoilu ei varmaankaan ole paras mahdollinen; tyylin provosoivuus vain vahvistuu. Hänen testinsä onkin herättänyt poleemisia yhteiskunnallisia tulkintoja ja ajatuksen teoreettista vähättelyä (kts. esim. Blas & Cardenas 2013; Mackinnon 2017). Kuitenkin testin tietojenkäsittelytieteellinen merkitys on tullut myös laajalti huomioiduksi (Moor 2003).

Turingin kieli ja argumentaatio on poleemisuudesta huolimatta kiehtovaa: artikkelin teoreettinen tavoite on olennaisen tärkeä ja syvälinen. Kysymys ”voivatko koneet ajatella” hukkuu helposti filosofiseen syvämielisyteen. Turing toi testinsä avulla kysymykselle merkityksen, jota oli mahdollista tarkastella sekä filosofisesti että tieteellisesti. Turing katsoi, että hänen testinsä pystyi tuomaan esiin em. suuresta kysymyksestä olennaisia puolia esiin. Artikkelissaan Turing käy läpi, miten hän ajatteli pelin etenevän. Se, että Turing katsoi kyselypelin soveltuvan tehtävään, liittyy tietokoneiden ja ihmisten väliseen ilmeiseen eroavuuteen: älyllisyyttä on mahdollista havaita tai testata keskustelun avulla, joka mahdollistaa testattavan asian kannalta epäolennaisten erojen huomiotta jättämisen. Kysely ja keskustelu tunnetusti ilmentävät ajattelua ja älykkyyttä. Keskustelussa kysymysten alaa ei olla suoranaisesti rajoitettu. Toisaalta niin kysymys- kuin vastausstrategiat voivat olla melkein millaiset tahansa. Tässä tulee mahdollisuus koneen naamioitua ”inhimilliseksi” esimerkiksi hidastamalla toimintaansa ja näin ikään kuin miettimällä asioita ja/tai vastaamalla väärin laskutehtävään ja näin osoittaa inhimillistä erehtyvyyttä. Siten Turingin oletus on, että ihminen ilmentää älykkyyttään keskusteluissa ja että on mahdollista tunnistaa inhimillisen älykkyyden piirteitä keskustelun kuluessa. Näin ollen testi ainakin osin testaa aiottua asiaa. (Copeland 1993; Mutanen & Halonen 2019.)

Keskustelulla on vankka rooli ihmisyhteisöissä. Ihmisten kanssakäymisessä keskustelut ja kyselyt ovat esillä suuressa osassa inhimillistä kanssakäymistämme; olisikin kiinnostavaa tarkastella laadullisen tutkimuksen erilaisten haastattelu- ja keskustelumenetelmien (kts. esim. Hyvärinen, Nikander & Ruusuvuori 2017) ennako-oletuksia tästä näkökulmasta. Kuitenkaan keskustelut eivät mitenkään välttämättä noudata joitakin yleisiä rationaalisuusperiaatteita. Esimerkiksi keskusteluissa esiin tulevia vaikeita kysymyksiä ihmiset kiertävät mitä erilaisimmin keinoin. Asioiden vierestä puhuminen lieenee yleisesti tiedossa oleva tapa. Tällainen strategia on esimerkiksi se, että otetaan edellisestä puheenvuorosta tai kysymyksestä satunnainen sana, jonka ympärille rakennetaan jatkokommentti tai vastaus. Tällaisia strategioita on mahdollista opettaa myös tietokoneohjelmille, esimerkiksi tietokone ohjelmat Parry ja Eliza käyttävät tällaisia ns. triviaaleja strategioita (Copeland 1993).

On perusteltua epäillä Turingin testin kykyä tuottaa vastus Turingin alkuperäiseen kysymykseen. Turing (1950) tiedosti ongelman. Hän tarkasteli systemaattisesti testiinsä esitettäviä potentiaalisia vasta-argumentteja. Ongelmista huolimatta, Turingin testi tuo

olennaisia kysymyksenasetteluja tekoälytutkimukseen. Turingin testin läpäisyn informatiivisuus on ongelmallinen asia. Vardi (2014) katsoi, että Turing ei ehkä itse selvittäisi omaa testiään. Näin ollen testin väärin negatiivisten tulosten mahdollisuus tulisi ottaa huomioon. Toisaalta ukrainalainen 13-vuotias Eugene Goostman läpäisi testin. Hänen strategiansa oli vaikeiden kysymysten ”inhimillinen” välttely. Myöhemmin osoittautui, että Eugene Goostman oli kone. (Aamo 2014.) Eugene Goostmanin strategia muistuttaa edellä mainittuja triviaaleja strategioita, joten testin virheellisten positiivisten tulosten ongelma tulisi myös arvioida.

Älykkyydestä

Turing ajatteli, että hänen testinsä mahdollistaa osapuolten intellektuaalisten piirteiden kriittisen arvioinnin. Sillä jos emme kykene erottamaan ihmistä ja konetta tässä suhteessa, niin meidän tulisi myöntää koneella olevan älykkyyttä (Hodges 2000, 51). Se, että testi on muotoiltu provokatiivisin kielivalinnoin, ei pitäisi estää näkemästä sen taustalla olevaa järkevää ydintä. Tämä lienee vaikeaa, kuten Hodges (2000, 50) tuo esiin: ”Jos Turingin tarkoituksena oli ärsyttää humanisti-intellektuelleja tällä seuraleikkijohdatuksellaan luovan mielen ongelmiin, niin hän kyllä onnistui siinä.”

Turing oli varsin vaikeasti hahmottuva kirjoittaja. Hänen tekstinsä olivat tarkkaa loogista ja matemaattista kieltä, osin hän oli kielellisesti rikas ja käsinkosketeltavan konkreettinen, mutta paikoin myös tietoisesti ärsyttävän provokatiivinen. Tämä ilmenee myös hänen teoreettisissa näkemyksissään. Matematiikan filosofiassaan hän oli lähinnä formalisti, mutta paikoin hänen tekstinsä on luontevaa ja selkeää konkreettisten tilanteiden analyysia. Nämä molemmat puolet tulevat esiin hänen artikkelissaan ”On Computable Numbers, With an Application to the Entscheidungsproblem”, joka julkaistiin vuonna 1936. Tämä artikkeli on logiikan historian yksi merkittävimmistä yksittäisistä julkaisuista. Siinä Turing todistaa logiikan (matematiikan) ratkeamattomaksi. Ratkeamattomuusongelman systemaattinen tarkastelu edellyttää, että ratkaisumenetelmä on hyvin määritelty. Matematiikassa on esitetty valtaisa joukko erilaisia tiettyjen ongelmaluokkien ratkaisumenetelmiä. Kuitenkaan näiden avulla ei voida vastata yleiseen ongelmaan. Muotoillakseen yleisen ratkaisumenetelmän Turing määrittelee lasentakoneet, jotka tunnetaan Turingin koneina. Turingin määritelmä oli nerokas, yhtäältä ne toteuttavat loogisen eleganssin, mutta ovat lisäksi intuitiivisesti oivalliset. Ne määrittelevät yleisen ratkaisumenetelmän käsitteen ja lisäksi ne toimivat tietokoneen teoreettisena mallina. Siten Turingin tulos pätee myös tietokoneiden kohdalla, jolloin se tarkoittaa, ettei ole yleistä keinoa varmistaa, että tietokoneohjelma pysähtyy tai tekee, mitä sen on aiottu tehdä. (Hodges 2000, 14–15.)

Turingin tulos täydentää Gödelin vuonna 1931 julkaisemaa epätäydellisyyslausetta, johon Turing oli tutustunut opiskeluaikanaan kuunnellessaan Newmanin luentoja matematiikan perusteista. Gödelin tulos osoitti, ettei mikään logiikan tai matematiikan järjestelmä, joka sisältää lukuteorian voi olla sekä täydellinen että ristiriidaton. Gödelin epätäydellisyyslause tarkoittaa mm. sitä, että jos matematiikka on ristiriidaton, niin siinä on tosia lauseita, joita ei voida todistaa. (Hodges 2000, 13–14.) Tietenkään Gödelin tulosten jälkeen negatiivinen vastaus ratkeavuusongelmaan ei ollut yllätys.

Kuitenkin jo se, että Turing kykenee luentokurssin pohjalta työskentelemään itsenäisesti tällaisen vaativan ongelman parissa, kertoo hänestä erittäin paljon.

Turing kirjoitti artikkelinsa tuntematta saman ongelman aiempia ratkaisuja, joissa ratkaisumenetelmä poikkesi Turingin muotoilemasta. Muita ratkaisumenetelmiä olivat λ-määriteltävyys (Church; Kleene) ja rekursiivisuus (Gödel; Herbrand). Post muotoilema ratkaisumenetelmä muistutti olennaisesti Turingin muotoilemaa menetelmää. Teoreettisesti erittäin kiinnostava seikka on, että kaikki nämä johtivat samaan laskettavien funktioiden luokkaan. Siten Church (1936) esitti teesin, jossa hän toteaa, että λ-määriteltävyys kiinnittää intuitiivisesti laskettavissa olevien funktioiden luokan. Tietenkään tällaista teesiä ei voi todistaa, mutta sille antaa tiettyä luotettavuutta se, että kaikki yritykset määritellä laskettavien funktioiden luokka on tosiasiaa ollut tämä sama luokka funktioita. (Raatikainen 1995.)

Inhimillinen laskenta

Edellä viitattiin Turingin intuitiivisesti oivalliseen tapaan määrittää esittämäänsä ratkaisumenetelmää. Muotoillakseen etsimänsä menetelmän hän analysoi inhimillistä laskentatapahtumaa. Hän kutsuu inhimillistä laskijaa *computeriksi*. Turingin analyysi on loogis-käsitteellinen analyysi, jonka tavoite on löytää välttämättömät ja riittävät tekijät, jotka määrittävät laskentaa. Näin hän suorittaa ”introspektiivista analyysia”, jonka tuloksena on *mekaaninen ihminen (computer)*, joka pystyy suorittamaan annetut laskutehtävät (Sieg 2000, 2). Tämän pohjalta hän kykenee muotoilemaan yksityiskohtaisesti Turingin koneet, jotka ovat määriteltäviä tehtäväkohtaisesti. Turing pystyi osoittamaan Turingin koneiden yleisyyden, jonka nojalla hän muotoilee ns. Turingin teesin, joka sanoo, että Turingin koneet pystyvät suorittamaan samat tehtävät kuin hänen muotoilemansa mekaaniset ihmiset.

Turingin (1936) tarkasteli esimerkkinä päättymätöntä desimaalilukua, mikä aiheuttaa, että kone laskee desimaali desimaalilta, muttei varsinaisesti koskaan pysähdy. Tällaiset koneet voidaan luontevasti yleistää, jolloin saadaan määritellyksi hyperlaskettavuus (Copeland 2002b; Marcer & Fedorec 1994; Bringsjord & Zenzen 2003; Hintikka & Mutanen 1998; Syropoulos 2019). Hyperlaskettavuutta on mahdollista tulkita filosofisesti, jolloin saadaan esimerkiksi formaali oppimisen teoria (Gold 1967; Oserson, Stob & Weinstein 1986) tai mielenfilosofian tulkintoja (Bringsjord & Zenzen 2003).

Turingin (1936) muotoili ns. universaalikoneen, joka periaatteessa kykenee tekemään minkä tahansa Turingin koneen koneet tehtävän. Universaalikone tuo esiin ohjelmitavuuden mahdollisuuden (Copeland 2004). Samalla näin tulee esiin myös mahdollisuus, että kone kykenee oppimaan (Turing 1948). Näin Turingin (1936) analyysi tuo esiin kiinnostavia yhteyksiä ihmisen ja koneen välillä niiden suorittaessa älyllistä toimintaa (Copeland 2012, 186). Tämän jälkeen on tietysti luontevaa tarkastella koneiden mahdollisuutta oppia, jossa pelit toimivat hyvinä esimerkkeinä (Copeland 1993; Michie 1961). Oppimisessa on tietysti keskeistä kekseliäisyys ja intuitiivisuus, joita Turing (1938, 192–194) tarkastelee. Turingin mukaan kekseliäisyys voidaan korvata lujalla työllä. Intuitiivisuus ei samalla tavoin ole korvattavissa. Kirjeessään Newmanille

hän tuo esiin, että intuitio voi ohjata kekseliäisyyttä erilaisten menetelmien käyttöön (Copeland 2004, 135).

Copeland (1993) esittää Turingin testille erilaisia kriittisiä huomioita. Tässä tarkastelemme niistä vain yhtä, jolla on kannaltamme suuri merkitys. Puhuttaessa tekoälystä joudumme ottamaan kantaa siihen, miten ajattelemme koneen mallintavan (simuloi-
van) älykkyyttä. Voimme ajatella, että se mallintaa olematta varsinaisesti älykäs. Tällöin voimme ajatella, että tekoälyohjelma tuo esiin joitakin piirteitä älykkyydestä, esimerkiksi kyky ratkoa rajattuja ongelmia. Tämä on itse asiassa heikon tekoälyn käsitys. Toisaalta voimme ajatella, että tekoäly mallintaa älykkyyttä olemalla tosiasiallisesti älykäs. Tällöin tekoälyohjelma on vahvan tekoälyn mukaisesti tosiasiallisesti älykäs. Chalmers (2010) tarkastelee systemaattisesti tällaista simulointiargumenttia (emulation argument). Esimerkiksi Lucas (1961) argumentoi Gödelin edellä mainittuun epätäydellisyystulokseen nojaten, ettei ihmisälyä ole mahdollista simuloida, sillä älykkyyden mekaanisuusajatus on virheellinen. Tällainen argumentti edellyttää esimerkiksi sitä, että mekanismi ymmärretään suppeasti (Copeland 2002).

Tämä saattaa vaikuttaa turhalta ja saivartelevalta keskustelulta. Kuitenkin tällainen on olennaisen tärkeää pohdittaessa, mitä tekoäly on ja mihin sillä voidaan pyrkiä tai mihin sillä tulisi pyrkiä. Turingin muotoilu ratkaisuongelmaan on tässä suhteessa erittäin kiinnostava. Turingin analyysi ihmislaskennasta tuottaa lopputulokseksi abstraktin ja formaalinen laskentamenetelmän, jolla on olennaiset ihmislaskennan piirteet. Siten Turingilla on hyvät perusteet käsitykselleen, että hänen määrittelemänsä koneet kiinnittävät ihmislaskennan alan. Näin Turingin koneet tosiasiallisesti laskevat. Kuitenkaan tällainen rajoitettu tosiasiallinen simulaatio ei vielä simuloi älykkyyttä vahvassa merkityksessä. (Copeland 1993.)

Turing kehitteli tekoölyyn liittyviä ajatuksiaan varsin pitkään. Hän etsi erilaisia keinoja toteuttaa älykkyyttä. Hän kehitteli jopa konnektionismia ennakoivia ideoita (Turing 1948; Copeland 2004). Kuten jo edellä on tuotu esiin, pelit olivat yksi kiinnostuksen kohde. Turing (1948) kehitteli testiä, jolla voitaisiin tarkistaa, kyetäänkö kone ja ihminen erottelmaan pelistrategioidensa perusteella. Tämä on erittäin olennainen asia. Yhtäältä meillä ei ole oikeastaan kovin hyvin jäsentynyt käsitys, millainen on hyvä pelistrategia, vaikka monissa yksittäisissä tilanteissa kylläkin pystymme arvioimaan strategian järkevyyttä tai luovuutta. On hyvä käsitys, esimerkiksi millaisin strategioin shakin mestaritason pelaajat pelaavat, mikä voi olla apuna niin pelistrategioiden arvioinnissa kuin koneita kehitettäessä. (Kts. Good 1965.)

Turingin testi tarkastelee erittäin olennaista tekoälyn kysymystä. Kuitenkin testin puutteellisuus on selvää, sillä se on periaatteessa mahdollista läpäistä trivialilla strategialla ja siinä voi tulla hylätyksi tietyn intellektuaalisen jäykkyyden perusteella. Turingilla on varmaan ollut mielessään aiemmin kehittämänsä strategiatesti kehitellessään Turingin testiä. Turingin perusajatus on erittäin oivallinen, yhtäältä strategia lienee älykkyyden kannalta samalla tavoin perustava kuin ratkaisumenetelmä on laskettavuuden kannalta. (Mutanen & Halonen 2019.)

Turing (1936) analysoi inhimillistä laskentaa ja näin kykeni muotoilemaan logiikalle ratkaisumenetelmän, jonka avulla hän kykeni ratkaisemaan ratkeamattomuusongelman. Turingin testin taustalla on samankaltainen ajatus. On luontevaa ajatella, että keskustelussa inhimillinen älykkyys ilmenee. Turing (1950) pohtii tätä asiaa ja antaa valoinnoilleen luontevat perustelut. Kuitenkaan ei mikään tahansa keskustelu ilmennä älykkyyttä. Strateginen keskustelu kylläkin ilmentää älykkyyttä, joten tulisi kehittää keinoja arvioida strategioita. Ehkä Turingilla hänen muotoillessaan testiä oli mielessään hänen omat keskustelunsa tutkimusryhmänsä puitteissa, jotka olivat strategisesti oivallisia eli innovatiivisia ja luovia (Hodges 2000). Ns. arkikeskustelussa vastaavaa strategisuutta ei samassa määrin ilmene.

On kuitenkin keskustelukonteksteja, joissa strategisuus on olennaisen tärkeässä roolissa. Esimerkiksi etsittäessä uutta tietoa tai ratkottaessa käsillä olevaa ongelmaa, strategisuus on ensiarvoisen tärkeää. Näissä yhteyksissä strategisuus on kontekstuaalista tavoitteellisuutta. Tällainen muistuttaa monessa suhteessa Turingin tilannetta hänen kirjoittaessaan vuoden 1936 artikkeliaan. Kuitenkaan tiedonetsintä tai ongelmanratkaisu eivät välttämättä pääty lopulliseen ratkaisuun eli tällaiset prosessit eivät ole rekursiivisia, mikä tekee asian matemaattis-logisesti ongelmalliseksi. Tosin jo edellä viittasimme hyperrekursiivisuuteen, jonka avulla on mahdollista tarttua tällaisiin ongelmiin. Tämä voidaan yleistää myös tieteenfilosofiseen tieteellisen päättelyn analyysiin (Kelly 1996; Martin & Osherson 1998). Näin on mahdollista saada loogisesti täsmällinen analyysi tiedonetsinnän tai ongelmanratkaisun logiikalle. Hintikka on kehittänyt yleisen rationaalisuuden mallin, joka perustuu kysymysten logiikalle. Tässä mallissa ajatus on, että strategisten kysymys-vastaus -ketjujen avulla on mahdollista strategisesti etsiä uutta tietoa (Hintikka, Halonen & Mutanen 2002; Halosen artikkeli tässä kirjassa). Mallia on sovellettu myös opettamisen ja oppimisen teoriaan (Hintikka 1982; Hakkarainen & Sintonen 2002).

Olemme tarkastelleet Turingin keskeisiä ajatuksia tekoälyyn liittyen. Eräs keskeinen huomio on se, että Turingilla oli keskeisenä ideana löytää niitä tekijöitä, joiden avulla on mahdollista luonnehtia älykästä toimintaa. Näiden olennaisten tekijöiden identifioinnissa pelien tutkimuksella on keskeinen rooli; pelien pelaaminen on malliesimerkki strategisesta toiminnasta, jota siis pidetään samalla myös älykkyyttä ilmentävän toiminnan perustana. Systemaattisuus tai strategisuus on olennaisen tärkeässä roolissa ja siten oikeastaan olennaista on tiettyyn ”terveen järjen ajattelun ja päättelyn” (Niiniluoto 1999, 65) ymmärrykseen, eikä niinkään, että lähtökohtaisesti tietäisimme, mitä älykkyys on (Hintikka 1990). Tämä ilmenee epäsuorasti Turingin testistä. Oikeastaan nimenomaan tästä syystä mallia tulee täydentää systemaattisella strategisella orientatiolla, johon Hintikan interrogatiivimalli tarjoaa erinomaisen teoreettisen viitekehyksen. Malli avaa myös moraalifilosofiaan liittyviä kysymyksenasetteluja, joita Turing ei systemaattisesti tarkastellut, mutta tässä yhteydessä on hyvä tehdä lyhyt katsaus tekoälyn etiikan kannalta olennaisiin kysymyksiin (tarkemmin kts. Asaron artikkeli tässä kirjassa).

Puhuttaessa tekoälystä on ajatuksena, että meillä voi, ainakin periaatteessa, olla keino-tekoisia älykkäitä agentteja. On samalla luontevaa kysyä, onko nämä kenotekoiset älykkäät toimijat samalla tavoin vastuullisia kuin ihmiset. Tämä tuo käsiimme kysymyksen,

onko mahdollista, että meillä on myös keinotekoisia moraalisia agenteja (Floridi & Sanders 2004). Luonteva tapa lähestyä asiaa edellä olevan nojalla on tietenkin lähteä etsimään moraaliagentuuria osoittavaa testiä, jolle luonteva nimi olisi moraalinen Turingin testi (Moral Turing Test) (Gerdes & Øhrstrøm 2015). Kuitenkin moraalisuus liittyy toimintaan, joten sitä ei ole samalla tavoin redusoitavissa keskusteluun kuin älykkyys redusoidaan alkuperäisessä Turingin testissä.

Tekniikka, toiminta ja vastuullisuus

Älykkyyteen liittyy myös vastuullisuus, jota tarkastelemme seuraavassa. Yhtäältä keskusteluun liittyvä tiedollinen vastuullisuus ei tule esiin ns. arkikeskustelun kohdalla, sillä arkikeskustelun tavoitteet vaihtelevat tilanteen mukaan. Tietenkin tietty luottamuksellisuus liittyy kaikkeen keskusteluun. Pyrittäessä etsimään tietoa eli asian oikeaa laitaa, on metodisesti keskeistä rehellisyys. Tieteenfilosofiassa tuodaan esiin ns. episteemisiä arvoja, joita tieteellisen tiedonhankinnan katsotaan edellyttävän. Nämä ovat luonteeltaan metodisia arvoja eikä niinkään eettisiä. Seuraavassa tarkastelemme lähemmin moraalitoimijuuden problematiikkaa ja katsomme, miten sen vaikuttaa tai tulisi vaikuttaa tekoälykeskusteluun.

Hakli & Mäkelä (2019) tuo esiin yleisesti hyväksytyn periaatteen, että vastuullisuuteen liittyy tieto ja kontrolli. Kuitenkaan tällainen ehto ei ole kovin sensitiivinen, sillä monissa tilanteissa tämä ei toimi erottavana tekijänä. Moraalitoimijuus ilmenee käytännössä siten, että hyväksymme yleisesti ihmisten olevan vastuullisia teoistaan ja niiden seurauksista. Lähtökohtaisesti kontrolli ja tieto -ehto vaikuttaa toimivalta: jos avaan ikkunan, niin olen siitä ja sen seurauksista vastuussa.

Kullakin teolla on periaatteessa lukematon seurausten luokka, mikä tekee teon seurausten arvioinnin käytännössä mahdottomaksi. Edelleen kunkin tehdyn teon sijaan tekijän on mahdollista tehdä lukuisia muita tekoja, kuten esimerkiksi jättää teko tekemättä. Siten kontrolli ja tieto -ehto eivät päde teon seurauksista tai teolle vaihtoehtoisista teoista puhuttaessa. Kuitenkin yleisesti katsomme, että teon intentionaalisuuteen liittyy mukaan myös vastuullisuus. Se, kuinka pitkälle tekijän vastuullisuus kantaa, on sekä teoreettisesti että käytännöllisesti tärkeä kysymys. (Moore 1905.)

Nykyisin puhutaan paljon myös ns. yhteistoimijuudesta, jossa tietty teko suoritetaan useamman toimijan tekojen yhteisvaikutuksesta, joista luontevia esimerkkejä löytyy sotilaskontekstissa runsaasti, kuten esimerkiksi jonkin rakennuksen valtaus. Useamman toimijan teoissa vastuullisuuskysymys on yksilötekoa vaikeammin jäsentävä. Yhteistoiminnalliset teot voidaan luokitella eri tyyppeihin. Esimerkiksi joissakin yhteistoimissa kunkin osallistujan osuus tulee tehdä samanaikaisesti, kuten esimerkiksi kannettaessa painavaa taakkaa. Tervehtiminen tai keskustelu ovat esimerkkejä yhteistoiminnasta, jossa yhteys on luonteeltaan käsitteellinen. Rakennuksen piirittäminen on selkeä useamman toimijan yhteistoiminto. Joissakin tapauksissa yksittäiset osateot on mahdollista suorittaa toisistaan riippumatta, kuten esimerkiksi lattian imurointi ja mattojen tamppaus, joiden yhteisvaikutuksena huoneiston siivous tulee tehdyksi. (Tuomela 1982; 1983.) Kannettaessa raskasta taakkaa yhdessä prosessi demonstroi yhteisen intention. Tervehtiminen edellyttää toteutuakseen tietyn sosiaalisen käytännön, jonka

puitteissa tervehtiminen tapahtuu. Kollektiivinen teko edellyttää yhteistä intentiota teon tekemiseen ja luottamus siihen, että toinen tai toiset aikovat myös toteuttaa oman osuutensa.

Kollektiiviseen tekoon liittyy myös kollektiivinen vastuu. Henkilöiden kantaessa pöydän paikkaan, jossa se aiheuttaa onnettomuuden, ovat kantajat vastuussa aiheuttamastaan onnettomuudesta. On ongelmallista ajatella, että vastuu ikään kuin jakautuisi osiin kunkin saadessa vain osan moraalivastuusta. Tässä yhteydessä on hyvä huomata, että juridinen ja moraalinen vastuu eroavat toisistaan. (Mäkelä & Hakli 2018.) On kuitenkin hyvä huomata, että ihmiselle voi olla vastuu jostakin kollektiivisesti toteutuneesta asiasta, vaikka hän tai kukaan muukaan ei ole sitä mitenkään intentionaalisesti tehnyt. Tällainen voi olla esimerkiksi ilmastonmuutoksen aiheuttaminen. (Hormio 2020.)

Teknisten välineiden käytön ei pääsääntöisesti katsota alentavan henkilön vastuullisuutta. Ihminen on vastuussa murhasta riippumatta kuristaako vai ampuuko hän uhrin. Gunkel (2020) katsoo, että tämä pätee, olipa kyse miten monimutkaisesta teknologiasta tahansa (instrumentaalinen teknologiakäsitys). On jälleen hyvä havaita, että juridinen vastuullisuus on herkempi tilanteiden yksityiskohdille siten, että tekniikan tai ympäristön rajoittaessa agentin toiminnan vapautta, myös juridinen vastuu voi heiketä. Tulee lisäksi huomioida, ettei moraalivastuu kulje kausaliteetin mukana. (Mäkelä & Hakli 2018.)

Puhuttaessa tekoälystä tulee tarkastella myös tekoälyn vastuullisuutta. Instrumentaalinen teknologiakäsitys katsoo, ettei tekoälyn vastuullisuudesta ole mielekästä puhua. Puhe on tällöin metaforista ja siten se helposti sotkee enemmän kuin selkeyttää kuten Mäkelä & Hakli (2018) argumentoivat jalat maassa ja mieli kirkkaana. Ehkä kuitenkin asia jää auki: kuka on vastuussa autonomisen asejärjestelmän tekemistä iskuista tai robotitaistelijan tekemistä toimista? Tällaiseen epämääräiseen epäilyyn vastataksemme on hyvä ottaa jalat irti maasta ja miettiä ”mitä jos ...” -menetelmällä. Tämä on hyvä luovan ajattelun menetelmä, jota on hyvä harjaantua tekemään (Hintikka & Bachman 1991).

Tekoälystä puhuttaessa helposti ajattelemme tietokoneiden olevan älykkäitä jossain rajatussa merkityksessä. Mielessämme saattavat olla pelejä pelaavat koneet, erilaiset asiantuntijajärjestelmät, itseohjautuvat autot tai asejärjestelmät. Tällöin älykkyys on jonkin tietyn spesifin ongelmatyyppin ”älykäs” ratkominen. Näissä yhteyksissä on luontevaa kysyä, missä mielessä kone simuloi älykkyyttä. Autonomisuus on tällaisissa yhteyksissä rajattua. Tällaisissa tilanteissa oppiminen tarkoittaa ”harjaantumista” koneelle ominaisen ongelmatyyppin ongelmien ratkomiseen. Myöskään tällaisissa rajatuissa tapauksissa koneen ratkaisujen ennakointi ei onnistu niiden suunnittelijoilta, ohjelmoijilta, omistajilta eikä käyttäjiltä. (Mäkelä & Halki 2018.)

Olisi kuitenkin mahdollista ajatella koneita, jotka oppivat kuten ihmiset, mitä jo Turing (1948) pohdi. Good (1965) otti ajatuksen vakavasti hänen tarkastellessaan ultra-älykkäitä koneita, joiden älykkyys ylittää ihmisälyn. Tällaiset koneet eivät ole vain ihmisten älykkäitä orjia, vaan ne kykenevät suunnittelemaan ja toteuttamaan yhä älyk-

käämpiä koneita. Goodin (1965) mukaan hänen tarkastelunsa on puhtaasti spekulatiivista. Vaikka ultraälykkäiden koneiden rakentaminen ei ollut mahdollista, niin hän kuitenkin luonnehtii tällaisten koneiden perusidea tavalla, joka on myöhemmin herättänyt mielenkiintoa (Chalmers 2010).

Teknologiseksi singulariteetiksi kutsutaan sitä pistettä, jossa koneet ohittavat ihmisälykkyyden (Niiniluoto 2020, 66). Chalmers (2010) ottaa Goodin ultraälykkäät koneet ja singulariteetin filosofisen analyysin kohteiksi. Käytännöllisesti singulariteetti olisi merkityksellinen tapahtuma, johon liittyy paljon toivoa, mutta myös pelkoja. Filosofisesti on tietysti kiinnostavaa tarkastella, millaisia tällaiset koneet voisivat olla. Onko niillä tietoisuus? Ovatko ne moraalisisubjekteja? Miten tällaisiin kysymyksiin oikeastaan tulisi vastata? Tällaiset käytännölliset ja filosofiset kysymykset eivät ole erillisiä, vaan ne kytkeytyvät osaksi tekoälytutkimusta, mutta myös osaksi tekoälytodellisuutta, kuten esimerkiksi keskustelu autonomisista asejärjestelmistä tuo esiin (Sparrow 2007; Asaron artikkeli tässä kirjassa).

Singulariteetin ajatuksessa perusidea on, että on mahdollista kehittää tekoälyä, joka se on ”hereillä” (awake) ja ihmisälyä korkeampaa. Siten sen kyky oppia ja kehittyä autonomisesti johtaa siihen, että tällainen tekoäly kykenee kehittämään tekoälyä, joka on sitä itseään älykkäämpää. (Vinge 1993; Chalmers 2010.) Oikeastaan tällainen argumentti vie Turingin (1947; 1948) ajatuksen, että tietokoneen tulisi kyetä oppimaan ihmislapsen tavoin. Tällainen oppimisprosessi johtaa aina vain älykkäämpään tietokoneiden sukupolveen. Tällaisen kehityksen jatkuessa lopulta ympärillämme on ultraälykkäiden koneiden (machine sapiens) järjestelmä. (Vinge 1993.) Chalmers (2010) tarkastelee, mitä tämä tarkoittaa ihmisten kannalta: jäämmekö vain sivusta seuraajiksi tai toisen luokan ”kansalaisiksi” älykkäiden koneiden hallitsemassa yhteiskunnissa?

Tekoälyn kehitys ei ole saavuttanut singulariteetin edellyttämää tasoa. Ihmisillä on kuitenkin vaikeuksia suhtautua tekoälyn sävyttämään todellisuuteen. Koneet jo nykyisin voittavat ihmiset monilla erityisalueilla. Samalla helposti ryhdymme odottamaan ihmisiltä koneiden kaltaista virheettömyyttä, nopeutta ja väsymättömyyttä, mikä aiheuttaa ns. teknostressiä. Eräs keino vastata tähän on painottaa ihmisen muita ominaisuuksia, kuten tunteita tai vapaata tahtoa. (Niiniluoto 1990, 124-125.) Inhimillinen älykkyys ei kuitenkaan ole vain välineellistä tai laskelmoivaa rationaalisuutta, vaan syvällisempää kulttuurista ymmärrystä (von Wright 1985). On selvää, että edellä oleva luonnehdinta ultraälykkäistä koneista ja niihin liittyvästä singulariteetista edellyttää, että koneiden olisi mahdollista saavuttaa älykkyys aidossa merkityksessä (vahva tekoäly). Sekä Vinge (1993) että Chalmers (2010) tarkastelevat tekoälyn kehittymistä asteittaisena prosessina, jossa koneiden älykkyys hiljalleen, sukupolvi sukupolvelta kehittyy.

Itsenäiseen harkintaan kykenevä tekoälyagentti tekee päätöksiä, joiden kohdalla ihmisten sanotaan olevan vastuullisia. Näin eettisyys tulee väistämättä ottaa huomioon tekoälysovellutuksista puhuttaessa. Kuka on vastuussa tekoälyagentin tekemistä päätöksistä. Edellä viitattujen rajoitetun autonomian toimijoiden kohdalla on luontevaa katsoa, että ihminen, joka ohjaa toimintaa, on myös vastuullinen. Itsenäisen tekoälyagentin kohdalla tilanne on olennaisesti erilainen. Voidaanko tekoälyagenttia pitää vastuullisena? Mitä tämä tarkoittaisi? Tietenkin näiden väliin jää suuri joukko toimijoita,

joiden kohdalla ei ole varsinaista ohjaajaa, mutta ei ole kuitenkaan järkevää katsoa tekoälytoimijaa aidosti vastuulliseksi. Tekoäly voi oppia kokemuksesta siten, että ei ohjelmoija tai suunnittelija edes periaatteessa voi tietää, miten tekoälytoimija tulee kuskakin tilanteessa toimimaan. Autojen kohdalla meillä on helposti mielikuva, joka ei vastaa todellista tilannetta. Bartneck, Lütge, Wagner & Welsh (2021, 83-84) esittelee kuusiportaisen luokituksen autojen itseohjautuvuudesta, jossa ylin taso on täydellisen automaation taso, jossa auto suorittaa kaikissa olosuhteissa ajamisfunktiot. Heidän mukaansa tällä hetkellä autot ovat kolmannella tai neljännellä tasolla, jossa parhaimmillaan kuljettaja voi osittain luovuttaa toiminnot autolle. Edes kuvatulla korkeimmalla tasolla ei liene syytä kutsua autoa moraaliagentiksi. Joka tapauksessa on erittäin tärkeää pohtia tekoälyn vastuullisuutta systemaattisesti (Hakli & Mäkelä 2019).

Tekoälyn vastuullisuus on erittäin vaikea kysymys, jota Bartneck, Lütge, Wagner & Welsh (2021, 87) tarkastelevat itseohjautuvien autojen kohdalla. Ei varmaankaan ole mielekästä pitää kuljettajaa vastuullisena niitä toimista, joissa autolla on ohjaus hallinnassa. Tällaisissa tilanteissa Saksan eettinen koodi määrittää auton tai ohjelmiston valmistajayrityksen vastuulliseksi tahoksi. Siten konetta ei katsota kykeneväksi kantaa vastuuta, joten vastuu siirtyy luontevasti yritykselle. Tietenkin tällöin joudutaan pohtimaan miten yritys voi olla eettisesti vastuullinen toimija (Ollila 1993). Yleisemminkin keskustelu yritysten vastuullisuudesta on erittäin laajaa (Garriga & Mele 2004).

Puhe tekoälyn vastuullisuudesta on sen yleisyydestä huolimatta varsin ongelmallista. Lähtökohtaisesti eettinen vastuullisuus on vain ihmisille kuuluva piirre, joka on muilla toimijoilla metaforisesti (Pihlström 2020). Ei ole lainkaan selvää, voiko tekoäly olla missään tilanteessa vastuullinen; ei ole myöskään selvää, mitä tämä oikeastaan tarkoittaisi. Ongelman luonnetta tuo hyvin esiin ihmisyksilöiden kokemat moraalivammat, jotka kytkeytyvät olennaisesti moraalivastuuseen ja siihen kytkeytyviin syyllisyyteen ja häpeään (Cahill, Kinghorn & Dugdale 2022). Siten on hyvä pohtia tarkemmin singulariteettia, onko ultraälykäs tekoäly vastuullinen ja mikä ihmisten rooli olisi eettisesti vastuullisten ultraälykkäiden koneiden rinnalla. (Halki & Mäkelä 2019; Mäkelä & Halki 2018; Gunkel 2020; Sparrow 2007; Niiniluoto 1990.)

Kun itseohjautuva auto tunnistaa hyvin jalankulkijan ja pyöräilijän, mutta ei tunnista pyörää taluttavaa ihmistä ihmiseksi ja ajaa hänen yli, niin kuka on vastuussa? Edellä olevan nojalla ei ainakaan auto tai sitä ohjaava tekoäly. Kuljettaja ei määritelmällisesti ole ohjaaja, miten hän voisi olla vastuussa? Auton valmistaja on ostanut ohjelmiston ulkopuoliselta toimittajalta, joten valmistaja ei ole vastuussa, vaikka ”Ruotsissa Volvo on kertonut ottavansa vastuun itsestään ajavan auton tekemästä kolarista” (Valkonen 2019, 14). Ulkopuolinen ohjelmistoyritys on tilannut ohjelmiston alihankkijalta, joka on teettänyt työn ohjelmoijalla, joka on ohjelmoinut auton oppimaan kokemuksesta, joten hän ei pysty vastaamaan auton toiminnoista. Usein ohjelmoijia on useita, joista kukaan ei oikeastaan hallitse kokonaisuutta. Tietysti asioita tapahtuu, eihän kukaan vastaa tulivuoren purkauksesta ja sen urheiksi joutuneiden kohtalostakaan. Eikö kuitenkin rakentaja tai rakennuttaja vastaa, jos voidaan osoittaa, että talo on rakennettu turvallisuudesta piittaamatta?

Kollektiivisen toiminnan kohdalla vastuu jakautuu kaikkien toimintaan osallistuvien kesken. On luontevaa ajatella tällainen jaettu tai yhteinen vastuullisuus. Kuitenkin on hyvä huomata, ettei eettinen vastuullisuus ole samalla tavoin rajautuva kuin juridinen vastuullisuus. Eettinen vastuullisuus ei jakaudu aritmeettisesti osallistujien kesken. Voisiko vastuu tekoälyn kohdalla olla tällainen jaettu vastuu? Mäkelä & Hakli (2018) esittävät, että eettinen vastuu on jaettavissa vain moraalisubjektien kesken, joten tekoälyn kanssa jaettu vastuullisuus ei näin toimisi lainkaan. Edellä olevan nojalla on syytä hyväksyä tällainen johtopäätös. Oikeastaan samalla tulee esiin ongelma, miten ihminen voisi jakaa vastuun ultraälykkäiden koneiden kanssa, jotka ovat ehkä jossain merkityksessä kykeneviä moraalisubjektuuteen, mutta eivät ole ihmisyhteisöjen jäseniä. Siten edellä mainittu moraalinen Turingin testi lienee erittäin vaikea muotoilla siten, että se antaisi meille relevanttia tietoa moraalivastuusta.

Puhuttaessa esimerkiksi autonomisista asejärjestelmistä, on syytä tarkastella yksityiskohtaisesti, mitä autonomisuudella tarkoitetaan. Monet nykyisin toiminnassa olevista järjestelmistä on siinä määrin autonomista, että ne voivat toimia ilman suoraa ihmiskontrollia, vaikka toiminta edellyttää ihmisten ohjausta. Siten niiden autonomisuus on varsin rajoitettua. Tekoälyn lupaama itsenäiseen harkintaan ja päätöksentekoon kykenevä robottitaistelija lienee vasta suunnitteluasteella. (Sparrow 2007).

Lähteet

Anderson, M. and Anderson, S.L. (eds) (2011) *Machine Ethics*. Cambridge: Cambridge University Press (eBook). Available at: <https://doi.org/10.1017/CBO9780511978036>.

Bartneck, C. *et al.* (2021) *An Introduction to Ethics in Robotics and AI*. Cham: Springer International Publishing (SpringerBriefs in Ethics). Available at: <https://doi.org/10.1007/978-3-030-51110-4>.

Cahill, J.M., Kinghorn, W. and Dugdale, L. (2022) 'Repairing moral injury takes a team: what clinicians can learn from combat veterans', *Journal of Medical Ethics*, 2022. Available at: <https://doi.org/10.1136/medethics-2022-108163>.

Chalmers, D.J. (2010) 'The singularity: A philosophical analysis', *Journal of Consciousness Studies*, 17(9–10), pp. 7–65.

diSessa, A.A. (2014) 'A History of Conceptual Change Research: Threads and Fault Lines', in R.K. Sawyer (ed.) *The Cambridge Handbook of the Learning Sciences*. 2nd edn. Cambridge University Press, pp. 88–108. Available at: <https://doi.org/10.1017/CBO9781139519526.007>.

Floridi, L. and Sanders, J.W. (2004) 'On the morality of artificial agents', *Minds and machines*, 14(3), pp. 349–379.

Gärdenfors, P. (2014) 'Computational complexity and cognitive science: How the body and the world help the mind be efficient', in A. Baltag and S. Smets (eds) *Johan*

van Benthem on Logic and Information Dynamics. Cham: Springer (Outstanding Contributions to Logic, 5), pp. 825–833. Available at: https://doi.org/10.1007/978-3-319-06025-5_31.

Garriga, E. and Melé, D. (2004) ‘Corporate social responsibility theories: Mapping the territory’, *Journal of business ethics*, 53(1), pp. 51–71.

Gerdes, A. and Øhrstrøm, P. (2015) ‘Issues in robot ethics seen through the lens of a moral Turing test’, *Journal of Information, Communication and Ethics in Society*, 13(2), pp. 98–109. Available at: <https://doi.org/10.1108/JICES-09-2014-0038>.

Good, I.J. (1965) ‘Speculations Concerning the First Ultraintelligent Machine’, in *Advances in Computers*. Elsevier. Available at: [https://doi.org/10.1016/S0065-2458\(08\)60418-0](https://doi.org/10.1016/S0065-2458(08)60418-0).

Gunkel, D.J. (2020) ‘Mind the gap: responsible robotics and the problem of responsibility’, *Ethics and Information Technology*, 22(4), pp. 307–320. Available at: <https://doi.org/10.1007/s10676-017-9428-2>.

Hakkarainen, K. and Sintonen, M. (2002) ‘The interrogative model of inquiry and computer-supported collaborative learning’, *Science & Education*, 11(1), pp. 25–43.

Hakli, R. and Mäkelä, P. (2019) ‘Moral Responsibility of Robots and Hybrid Agents’, *The Monist*, 102(2), pp. 259–275. Available at: <https://doi.org/10.1093/monist/onz009>.

Hakli, R.A. and Mäkelä, P.A. (2018) ‘Voiko robotti kantaa vastuuta?’, *Tiedepolitiikka*, 43(4), pp. 17–21.

Hallamaa, J., Haikarainen, R. and Kalliokoski, T. (2021) ‘Tekoälyteknologiat ja turvallisuus’, *Tieteessä tapahtuu*, 39(4), pp. 11–17.

Hintikka, J. (1990) ‘The Languages of Human Thought and the Languages of Artificial Intelligence’, *Language, Knowledge, and Intentionality: Perspectives on the Philosophy of Jaakko Hintikka*. Edited by L. Haapaanta, M. Kusch, and I. Niiniluoto, Acta Philosophica Fennica 49, pp. 307–330.

Hyvärinen, M., Nikander, P. and Ruusuvoori, J. (eds) (2017) *Tutkimushaastattelun käsikirja*. Tampere: Vastapaino.

Koskinen, I. (2019) ‘Antihumanismi, posthumanismi ja humanistiset tieteet’, *niin & näin*, pp. 51–61.

Krüger, O. (2021) ‘“The Singularity is near!” Visions of Artificial Intelligence in Posthumanism and Transhumanism’, *International Journal of Interactive Multimedia and Artificial Intelligence*, 7(1), pp. 16–23. Available at: <https://doi.org/10.9781/ijimai.2021.07.004>.

Lammenranta, M. (1987) 'Taideteos symbolina: Nelson Goodmanin taidefilosofia', Helsingin yliopiston yleisen kirjallisuustieteen, teatteritieteen ja estetiikan laitoksen monistesarja, 15.

Lucas, J.R. (1961) 'Minds, Machines and Gödel', *Philosophy*, 36(137), pp. 112–127.

Mäkelä, P. and Hakli, R. (2018) 'cover Bad Arguments for Responsibility Sharing', in M. Coeckelbergh et al. (eds) *Envisioning Robots in Society – Power, Politics, and Public Space*. IOS Press (Frontiers in Artificial Intelligence and Applications). Available at: <https://doi.org/10.3233/978-1-61499-931-7-275>.

Michie, D. (1961) 'Trial and error', in *Science Survey*. Harmondsworth: Penguin (2), pp. 129–145.

Michie, D. (1962) 'Puzzle-learning versus game-learning in studies of behaviour', *The Scientist Speculates*, pp. 90–100.

Nayar, P.K. (2013) *Posthumanism*. Oxford, United Kingdom: Polity Press. Available at: <http://ebookcentral.proquest.com/lib/helsinki-ebooks/detail.action?docID=1729549>.

Niiniluoto, I. (1990) *Maaailma, minä ja kulttuuri*. Otava.

Niiniluoto, I. (1999) 'Ihminen tekniikan pauloissa', in T. Airaksinen (ed.) *Minä. Ruoka. Henkisyys. Sukupuoli. Tunteet. Järki. Tekniikka*. Helsinki: Otava, pp. 40–67.

Pihlström, S. (2021) *Ihmisen maailma: Esseitä humanismista, totuudesta ja ajattelun tilasta*. Suomi: Niin & näin.

Ross, B. (2020) *The philosophy of transhumanism: A critical analysis*. Emerald Group Publishing. Available at: <https://ebookcentral-proquest-com.libproxy.helsinki.fi/lib/helsinki-ebooks/detail.action?docID=6183833>.

Russell, S.J. and Norvig, P. (2003) *Artificial intelligence - a modern approach*. 2nd edn. Upper Saddle River, New Jersey: Prentice Hall (Prentice Hall series in artificial intelligence).

Sparrow, R. (2007) 'Killer Robots', *Journal of Applied Philosophy*, 24(1), pp. 62–77. Available at: <https://doi.org/10.1111/j.1468-5930.2007.00346.x>.

Syropoulos, A. (2019) 'On TAE machines and their computational power', *Logica Universalis*, 13(2), pp. 165–170. Available at: <https://doi.org/10.1007/s11787-018-0196-5>.

Tuomela, R. (1982) *Joint social action*. Helsinki: Helsingin yliopiston filosofian laitos (Helsingin yliopiston Filosofian laitoksen julkaisuja).

Turing, A.M. (1936) 'On Computable Numbers, with an Application to the Entscheidungsproblem', *Proceedings of the London Mathematical Society*, s2-42(1), pp. 230–265. Available at: <https://doi.org/10.1112/plms/s2-42.1.230>.

Turing, A.M. (1939) 'Systems of logic based on ordinals', *Proceedings of the London Mathematical Society, Series 2*, 45, pp. 161–228.

Turing, A.M. (1945) 'Proposed electronic calculator, report for National Physical Laboratory', in B.J. Copeland (ed.) , 2005, *Alan Turing's Electronic Brain: The Struggle to Build the ACE, the World's fastest Computer*. Oxford University Press.

Turing, A.M. (1947) 'Lecture to the London Mathematical Society on 20 February 1947', in B.E. Carpenter and R.W. Doran (eds) *A. M. Turing's ACE report of 1946 and other papers*. Massachusetts: MIT Press (1986).

Turing, A.M. (1948) 'Intelligent machinery, report for National Physical Laboratory', *Machine Intelligence*. Edited by D. Michie and B. Meltzer, 7 (1969).

Turing, A.M. (1950) 'Computing Machinery and Intelligence', *Mind*, 50, pp. 433–460. Available at: <https://doi.org/10.1093/mind/LIX.236.433>.

Valkonen, M. (2019) 'Äly älä jätä', *TUNIL*, 2019(2), pp. 11–14.

Vinge, V. (1993) 'The coming technological singularity: How to survive in the post-human era', *Science fiction criticism: An anthology of essential writings*, pp. 352–363.

Von Wright, G.H. (1987) *Tiede ja ihmisjärki: suunnistusyritys*. Otava.

ETHICS IN AUTONOMOUS MILITARY TECHNOLOGY – A CALL FOR ETHICAL COMPETENCY

Mireille Farzan

Abstract In this paper I outline the ethical questions linked with the use of autonomous military technology. I address issues related to moral responsibility, transparency and automation bias. I focus on the emergence of these issues in instances where humans are controlling ('in the loop') and supervising ('on the loop') autonomous military devices. After discussing the ethical questions, I tie them into the military ethics concept of ethical competency and argue that it provides promising solutions for addressing some of the pressing ethical issues linked to the use of this technology.

Keywords: Autonomous military technology, autonomous weapons, moral responsibility, ethical competence, military ethics.

Introduction

The recent development of military technology is seeing an increase in autonomous technology. (Hellström 2013.) The development and potential use of autonomous military technology raises a number of ethical questions when military tasks and decisions that carry heavy ethical implications are delegated to machines that are inadequate candidates for moral attitudes such as moral responsibility. One approach to the military tasks in question is through *prima facie* obligations. James Childress suggests the ethics in war to be a balance between a fundamental moral principle, to refrain from injuring or killing other human beings, and another equally significant moral obligation, to protect innocent people from being unjustly injured and killed. The heavy ethical implications of warfare can be seen as a result of being in a situation which requires overriding a fundamental moral principle in order to secure another significant one. These implications refer to instances where it may seem necessary to harm, injure and kill to protect innocent people from unjustified acts of war. (Childress 1978, 436.) I use the term *heavy* to refer to the significance of the overridden obligation, the weighty reasons that provide justification (i.e. defence), and the way in which the overridden obligation must continue to guide action (e.g. by encouraging to discriminate between active combatants and civilians and to only cause necessary and proportionate harm). (Ibid, 430-444.)

Childress further argues that one has the responsibility to explain the *intentions* behind the exceptional actions that war requires (ibid, 437). Another factor that has been recognized as essential for agents acting in war is taking responsibility for what is done (Walzer 1977, 287; Smidt 2000, 167-169; Toiskallio 2009, 57; Verweij 2007, 59). The use of autonomous weapons, however, introduces actors onto the arena of war that complicate the questions of intentions and responsibility. This is because machines are regarded to lack some key attributes we commonly attach to intentionality and responsibility. Attributes that are suggested to be missing in today's artificial agents include free will, consciousness, the ability to make and justify moral judgements, self-reflection, being appropriate sites for punishment/praise and the comprehension of

human morality and emotions. (Wallach et al. 2009, 35; Sparrow 2007, 74; De Jong 2020, 728; Leveringhaus 2018, 533). An additional viewpoint argues that even if machines were to carry some of these traits, the ways in which machines are programmed to have and express the traits may nevertheless undermine their autonomy and render them unfit candidates for moral attitudes. (Haksli & Mäkelä 2019, 269-271.) This has resulted in a debate over the morality of using autonomous weapons especially on questions involving moral responsibility, transparency, and automation bias.

Though military technology has often been subject to a plethora of ethical questions, what separates the questions mentioned above from those of the past is the nature of machine autonomy. Machine autonomy can be understood as a machine's ability to perform tasks independently with no or substantially reduced human intervention for extended periods of time. (Noorman & Johnson 2014, 61.) It has also been linked to an algorithm's machine-learning ability that allow it to acquire new information or use old information in new ways and essentially develop new skills, skills that might not have been a part of the initial programming phase. Machine autonomy can also be approached through the concept of 'autonomous power, where autonomy points to the amount and complexity of activities and decisions a machine can perform on its own'. (Hellström 2013, 101.) The use of machine autonomy in military technology, especially in weapons technology unveils autonomous weapons systems with abilities to search, detect, evaluate, track, and engage a target, this too without human intervention. The weapon can also have the capacity of selecting between different methods of engagement and use of force, providing lethal or non-lethal responses. (Ibid 100.)

In this paper I illustrate the ethical questions mentioned above from a human-robot collaboration point of view. I focus specifically on the questions regarding the controlling and supervision of autonomous military devices, otherwise known as human 'in the loop' and 'on the loop' scenarios. The term 'in the loop' refers to scenarios where humans control all or most of the aspects involved in machine behavior. (Marra & McNeil 2013, 1141). 'On the loop' refers to supervising and monitoring machine behavior while it executed given tasks (ibid. 1176). The questions will be approached from two point of views, control collaboration, where humans are fully or partially controlling the device, and supervising collaboration, where the supervising party has veto power over the devices automatic and autonomous functions. Autonomous devices that operate without human interaction will be left outside the scope of this paper due to limited space.

I will conclude by introducing the ethical competence framework, a concept in military ethics, and argue that it might provide promising solutions for the presented ethical questions. In the course of this paper, I will describe machines and algorithms as 'behaving', 'deciding', 'selecting' and performing other activities previously linked with human mental states. This is done for to ease to the flow of text for no stances is taken on the existence or nature of machine mental states. In addition, A distinction will be made between action and behavior. This distinction ties action onto notions of intentionality, agency, goals, aims and purposes. Behavior on the other hand will be tied to acts that lack broad intention and that are in a sense automatic. From this

point of view, intentional actions will be ascribed to humans and activities completed by machines will be referred to as machine behavior. This distinction is also made for highlighting the fact that wanted machine behavior is in most cases a result and an expression of human intention. This is an important factor for moral responsibility.

Ethics in autonomous military technology

There are several ethical questions linked to controlling and supervising autonomous military devices. In this paper I will focus on moral responsibility, transparency and automation bias. I will also address the meaning of ethically relevant human oversight. I will begin with the issue of moral responsibility.

Moral responsibility

Moral responsibility is important in the human-robot collaboration. Human-robot collaboration takes place when a human or group of humans are designing and working closely with a robot. This includes individuals, designers, users, operators, organizations, companies and institutions that come together to create and use the machine. (see Nyholm 2018, 1203.) It is assigned to those who are praiseworthy or blameworthy for the actions and decisions executed during the collaboration. Moral responsibility comes with a control condition as well as an epistemic condition (Fischer & Ravizza 1998, 13). These entail that in order for an agent to be morally responsible for an action, they are to have sufficient control over the action and sufficient knowledge on the meaning and consequences of said action. Moral responsibility is blurred in the human-robot collaboration of autonomous technology. One reason for this is the ‘many hands’ problem. (Taylor 2021). The humans involved in the collaboration consist of designers, programmers, users, operators and others who have an influential role over the machine and its behavior. In addition, the institutions, organizations, companies and societies these individuals work for might have values that are expressed in the collaboration.

In cases of unanticipated machine behavior, moral responsibility is blurred when it is unclear, what individual or collection of individuals is responsible for what aspect of the machine’s behavior. This is in large part because current social norms do not regard machines as moral agents and thus do not accept assigning moral responsibility to them. (see Hakli & Mäkelä 2019, 272. For discussion see Noorman & Johnson 2014, 55.) Moral responsibility is blurred even more when the autonomous device has the ability of learning from its experiences and altering its coding. Here, the machine can produce behavior that was not initiated by the humans involved (Matthias 2004, 176). The inability to assign responsibility to robots and devices moves responsibility to the shoulders of the human individuals or organizations who are a part of the collaboration. Here, humans and organizations become morally responsible for the behavior of the machine even when it involves activities and decisions made completely or mostly by the machine (Nyholm 2017, 1213) This is the case even when the machine behavior in question is uncontrollable and unpredictable to the humans involved. (Matthias 2004.)

Moral responsibility is especially important in the military use of autonomous technology. This is due to military methods. When goals such as national security are approached through civilian tactics, the methods often involve different forms of diplomacy and co-operation. Military methods, however, involve the use of military force and in many cases, the use of lethal and destructive force. It is the morally demanding military task of balancing the obligation to defend people from unjust military threats, and the obligation to refrain from causing unjust harm, that makes moral responsibility for military methods and actions paramount. (Childress 1978, 436; Walzer 1977, 287.)

It is evident that since different individuals are involved in the developing and operating the same autonomous device, different forms of responsibilities fall upon different agents. (Champagne & Tonkens 2015, Hakli & Mäkelä 2019, Nyholm 2018.) Here, I will address the moral responsibility of those in control and supervision collaboration with autonomous machines. Moral responsibility for the behavior of a remote-controlled autonomous device resembles the responsibility one has for one's own actions. This is when the device is operating without malfunction and the remote control feature allows the user full control over the device. The issue of responsibility becomes interesting when control over the device is not remote-control but rather the selecting or approving of alternative actions or decisions provided by the machine.

This scenario can be approached through Tomas Sheridan's 'spectrum of autonomy', a tool for assessing the autonomy level of an artificial agent. Here, machines become more autonomous when the decision making authority begins to shift from the human to the machine. At the lower end of autonomy, humans make all of the decisions while increased machine autonomy transfers the decision making authority to the machine (Marra & McNeil 2013, 1156), e.g. a navigational system that 'decides on' or selects the fastest route to a destination without additional input from the driver. When the decision making authority is the machine, autonomy is increased when the human operators 'veto authority' decreases. The veto authority refers to the human's position to approve, disapprove or stop the machines autonomous functions (ibid). According to this spectrum, a fully autonomous self-driving car would be one that selects the routes, does the driving and navigating, and does not allow the human passenger to override or alter its course. The cases I address here are those of semi-autonomous devices that fall somewhere in the middle of Sheridan's spectrum. These machines make the decisions or selections, but human operators remain in the position to override, approve or discontinue machine operations.

I suggest that the moral responsibility of the controlling or selecting individuals is blurred in cases where they are not provided with enough or sufficient information over what they are selecting and the immediate outcomes of their choices. In a supervision collaboration, where humans observe the robot's behavior and have the power to discontinue unwanted behavior, the moral responsibility of the supervising party seems to be tied to recognizing unwanted behavior and discontinuing it. Here, if the device engages in unwanted behavior, the supervising party is not morally responsible for the behavior, rather they are morally responsible for failing to identify the behavior or failing to discontinue it. (See Sheridan 2021, 755-756.) Responsibility gaps immerge

when everyone has done their part correctly, but the machine still expresses unwanted behavior.

The term ‘responsibility gap’ has been most notably explored by Andreas Matthias in his paper ‘The responsibility gap: Ascribing responsibility for the actions of learning automata’. Matthias argues that designing machines to behave autonomously may result in situations where the machines in question express behavior that is not a direct outcome of human commands. (Matthias 2004, 176-177). Learning machines might learn new behavior that becomes unexpected to both the designers and the users (ibid. 178). While the objective of designing a learning autonomous machine is to have it react to new information as it comes and to perform autonomously without the need for constant human control, Matthias explores the possibility of unplanned machine behavior and asks, who is to be held responsible, when a machine expresses unwanted behavior (ibid. 176-177). This question rises in contexts where both the designers and users of the machine could not predict or control the machine (ibid. 181-182).

Matthias’ responsibility gap has then been applied to military technology, more specifically autonomous weapons. (Gunkel 2020, Johnson 2003, Nyholm 2018, Leveringhaus 2018.) While the ethical issues surrounding autonomous weapons are largely debated upon (Arkin 2008, Asaro 2012, Krishnan 2016, Johnson 2013.), the issue of the responsibility gap in these weapons asks to identify those morally responsible for unwanted machine behavior. While all weapons, and technological artefacts for that matter, have the possibility of malfunctioning, the debate over the responsibility gap in autonomous weapons focuses on instances where the weapon *deliberately* selects a course of action that is against the intentions of both designers and users. More specifically, the risk of autonomous weapons deliberately breaking the ethical and legal rules of war e.g., harming those who are not meant to be intentionally harmed at war. In his influential work ‘Killer robots’, Robert Sparrow argues that the problem with autonomous weapons is that if they were to break the rules of war, those who have designed or who command the weapon cannot be held morally responsible, thus explicating the nature of the responsibility gap in the military context. (Sparrow 2007, 69-71.)

Transparency

Transparency can be seen as access to a device’s algorithmic decision-making process. (De Laat 2018, 528.) This access helps determine why the algorithm selects one course of action over another. This access further aids in explaining, predicting, and controlling the behavior of the device. (Santoro et al. 2008, 302-305.) Ethical questions on the transparency of an autonomous device deal with the implications of the inability to access important information from the device. Transparency is important in the use of autonomous military technology because it helps identify the means with which military goals are obtained. Obscured transparency leads to a limited understanding on how and why different functions will be carried out. This is the case when the device is such that its developers and users cannot ‘have a look at the information that is stored inside the network, and, even more importantly; – see what information is not represented inside it.’ (Matthias 2004, 178-179.) This issue is made more

problematic when the device's main function, as a weapon, is to issue destructive or lethal force. In operating autonomous weapons, it seems morally important to be able to understand and predict the ways in which the user's intentions will be expressed through the weapon's behavior.

The importance of *how* military objectives are obtained is highlighted in the just war tradition, specifically the *jus in bello* principles. Principles that aim to guide the conduct of warfare go against the 'any means necessary' approach, where the ends justify the means and the *how* and *why* of military action is deemed irrelevant. (See Lazar 2010.) From this perspective, transparency into the decision-making process of autonomous algorithms helps the humans involved ensure important ethical values. Transparency can be obscured due to the innate nature of the device. (See De Laat 2018, 525-526.) Autonomous devices with machine learning capabilities are able to learn from their experience. This ability, the ability to alter its own code, can turn autonomous devices into "black boxes" where the designer and user are unable to identify the things the program has learned. This is especially true in cases where the program itself is unable to understandably communicate this information to the designers or users. (Matthias 2004, 178-179.) Allowing machines to learn from their experience enables them to improve their reactivity, flexibility, and ability to analyze data in ways unreachable to humans. The ethical questions linked to (obscured) transparency point to the explicability and controllability of the behavior of the device.

In the use of autonomous devices, transparency allows the controlling party to understand exactly what it is they are controlling. If the device offers alternatives for the controlling party to select from, transparency involves some level of insight on how and why specific alternatives are presented and some specific knowledge on the outcomes of each selection. Lack of sufficient knowledge opens the door to scenarios where the controlling party does not fully understand what they are selecting. This might be the case if the autonomous algorithm provides alternatives without providing enough understandable information to support them. A similar stand can be taken when supervising autonomous devices. Transparency adds the ability to understand what it is one is supervising. Lack of relevant knowledge regarding what is being supervised hinders making risk and ethical evaluations over situations at hand.

Automation bias

Automation bias is a key topic in the ethical use of automatic and autonomous devices. Automation bias can be observed as errors in commission and omission. Errors in commission occur when the operator follows the flawed directives of an automated system in instances where earlier training and other more valid and reliable indicators point to the unreliability of the directives. Omission refers to the operator's failures to take correct action in situations where the automatic algorithm neglects to point out the need for action. (Skitka et al. 1999, 993.)

Commission and omission are caused by reduced focus due to relying on the automated process and the perceived authority of the computer. This phenomenon has been linked to the tendency to agree to do the wrong thing when ordered to do so by

an authority. (Ibid.) These errors are also linked to overestimating the infallibility or accuracy of the computer. (Ibid, 1001.) This has the potential of manifesting itself as a fear of contradicting an infallible authority and increased likelihood for commission. Both errors in omission and commission have been seen to occur more often in the presence of an automatic decision-making aid than in cases without such aid. (Ibid, 999.)

When controlling an autonomous military device, and in cases where an autonomous algorithm provides suggestions from which the operators makes selections, commission might lead the operator to select courses of actions that are in direct contradiction with earlier training or personal moral values. In cases of oversight, automation bias leads to omissions. Here, omission errors prevent the operator from reacting to things they are supposed to react to. (Parasuraman & Manzey 2010, 391.) If this is the case in the use of military technology, supervising parties might omit discontinuing unwanted machine behavior even when they have valid reasons to do so.

The ethical relevance of human oversight

The sections above illustrate some ethical issues that arise in the use of autonomous military devices. Concerns are often raised about autonomous weapons using lethal force without human involvement. This paper aims to show, however, that having humans control and supervise autonomous technology will not remove ethical problems altogether. This is to say that control and supervision over this technology misses its point if it is not done in the *right* way. Thus, it is worth highlighting, that even the tightest form of human control over the use of lethal force does not ensure ethical conduct. This can be illustrated through a modification of Ekelhof's soldier in a room. (Ekelhof 2017, 328.)

Picture a soldier in a room who controls a button that fires a weapon. The soldier is instructed to press the button whenever a light goes on in the room. The light is operated by an autonomous algorithm that evaluates threats and turns on the light when it calculates a need to use force. Though the algorithm identifies and evaluates the threats, the soldier alone has the power to control the use of force. The soldier has no other ways of perceiving the outside world other than from the light. Now picture that the light is being monitored by another soldier in a separate room through camera surveillance. The second soldier has a monitor, operated by the same autonomous algorithm, that identifies potential targets and notifies the imminence of threats. When the threat is at hand, the otherwise blank monitor displays the words "imminent threat", and the light in the first room is turned on. The second soldier makes sure that the light is turned on when the monitor displays the words. She does so by being ready to turn on the light herself in cases where the words are displayed on the monitor and the light doesn't go on. The words on the monitor and the light are the second soldier's only way of perceiving the outside world.

The illustrated thought experiment aims to highlight one key aspect; the discourse around human control and oversight over autonomous weapons needs accuracy if it is to lead to ethical conduct. In the thought experiment, the first soldier has complete

control over the use of force. The second soldier is supervising a key aspect of the performance of the autonomous algorithm and has the power to step in at will. However, neither has adequate understanding on the use of force that is being controlled and the operations of the algorithm that are being supervised. The soldier in control of the use of force has inadequate knowledge of why he is firing the weapon and at what. Their presence cannot ensure ethical conduct because they are unable to express their ethical insight. The nature of control and oversight they have, leaves them no room to make ethical evaluations on the situations they are in or room to assess the meaning, nature and consequences of their actions.

The thought experiment goes to show that using autonomous weapons in an ethical way requires placing humans in the *right* positions and having them perform the *right* tasks. The right position and tasks are those that allow users to use their moral judgement and deliberation. Here, they can see the moral implications of their actions and those of the human machine collaboration and play to their strengths as moral agents. (see Verweij 2007, 58-59 and Hakli & Mäkelä 2019, 260.) In the example presented above, the right position and tasks for the soldiers is that which allows them to see the information that results in a decision to use force, and to monitor how this information is processed. One way of identifying the positions and tasks for ethical control and supervision of autonomous military devices, is with the help of the '*ethical competence*' framework.

Ethical competence

In military ethics, ethical competence is used to describe a level of competence a soldier must have in order to assess the ethical dimensions of her work. (Also see Verweij 2007 and Wortel & Bosch 2011.) *Ethical competence* is a sub concept derived from the wider military concept of *action competence*. Action competence describes an individual's potential to take action and is an element that is present in the execution of that action. (Toiskallio 2009, 48.) Action competence is an umbrella concept with sub areas such as physical competence, social competence, psychological competence, and ethical competence. (Ibid, 49.) This paper focuses on the nature and implementation of ethical competence. Ethical competence can be seen as the glue that keeps the other aspects of action competence together. The physical, social and psychological aspects are expressed through ethical competence in the form of deliberation and decision making. (Ibid.) The ethical competence framework is highly useful due to its holistic nature and its focus on cultivating the individual rather than the collective. (Ibid, 51.) As such, it provides tools to the individual that are applicable in highly diverse situations and contexts. These tools and skills are important in facing the novel and potentially complex issues of automated warfare.

The ethical competence framework can provide guidelines on the nature of control and supervision that makes an ethical difference. This takes us from simply using autonomous military devices, to using them in the *right way*. This is essential when autonomous machines cannot tackle issues of morality autonomously. (Wallach et al. 2009, 34-36.) Here, when it is up to the controlling and supervising party to do so, there is a call for ethically competent operators in autonomous military technology.

The ethical competence framework highlights the importance of accountability and responsibility. (Toiskallio 2009, 57.) In what follows I will regard accountability as responsibility. This is because they, as well as other forms of moral responsibility (e.g. attributability and answerability) seem to share the same key conditions; they arise when an agent acts freely, and with sufficient knowledge about the significance of her actions. (see Smith 2012, 589 and Shoemaker 2011 for discussion.)

Ethically competent individuals have the ability and willingness to be accountable for their work. (Verweij 2007, 59.) This can be elevated to the ability and willingness to be morally responsible for their actions and omissions. Applied to those who control and supervise autonomous military devices, moral responsibility requires them not only to have control over their devices, but also to have enough epistemic understanding on what their devices are doing and how. This also involves having an understanding on important ethical values and being able observe whether the values are being implemented in the use of autonomous devices. (Ibid.) In the thought experiment of the two soldiers in the room, the soldiers' inability to observe the broad picture, their situation, the threats and the outcomes, makes them unable to utilize their potential awareness of important ethical values, put them into action or, make sure they are being implemented. Due to their positioning, the soldiers in the room cannot express their ethical competence nor can they be accountable for their work with the device. Here, even if those with immediate control lack a broader picture, the morally responsible use of autonomous devices requires that humans at other levels are able to ensure that the broad picture is in accordance with important values. Highlighting concepts such as accountability also helps plug potential responsibility gaps. I argue that for the most part, this gap can be plugged with a willingness to be accountable. Since moral responsibility requires both sufficient control and knowledge over what one is responsible for, a willingness to be accountable might translate into a willingness to prefer predictable and controllable autonomous devices.

Ethically competent operators have transparency into the inner workings of the devices they use and are able to observe its functions. Transparency allows them to understand and predict the machine behavior they are responsible for. (Santoro et al. 2008, 302.) In cases where transparency and observation are impossible or impractical for those with immediate control and supervision, other humans involved in the operation express ethical competency by demanding and having transparency into the device and observing that the device operates according to important ethical values. Adding transparency is key in adding accountability, and accountability is key in sustaining ethical competency. (Toiskallio 2009, 49.) On the ethical issue of automation bias, ethical competency and thus a heightened sense of accountability has the potential of reducing errors in commissions and omission amongst those who operate autonomous devices. This is because accepting moral responsibility for one's work helps operators of highly automated functions avoid errors in commission caused by automation bias. (Parasuraman & Manzey 2010, 396.)

Here, highlighting ethical competence in those who operate autonomous weapons provides a guideline to addressing the ethical issues raised above. When ethically competent operators are willing to be morally responsible for their work with autonomous

military devices, and other ethically competent individuals are willing to be morally responsible for using autonomous devices to further the military goals they have set, the problem of moral responsibility and responsibility gaps becomes easier to approach. A willingness to be responsible does not suffice. It must be backed up with the intentional actions of seeking the control, understanding and transparency that is required for moral responsibility. Assumed and ensured responsibility also safeguards against automation biases and the ethical problems they raise.

Conclusion

In this paper I outline ethical questions related to the use of autonomous military devices and offer a potential solution. The topic is narrowed down to uses of autonomous devices that requires human control and supervision or ‘human in and on the loop’ scenarios. The ethical questions illustrated are those of moral responsibility, transparency and bias. The potential solution provided is the use of the ethical competence framework. The ethical competence framework and its emphasis on accountability and morally responsibility provides a guideline on why and how operators of autonomous devices should be accountable and morally responsible for their work. This also illuminates the level of transparency that is needed for said responsibility. In addition, it helps prevent the ethically problematic aspects of bias in autonomous devices. Highlighting the importance of ethical competence might motivate individuals to seek better control over autonomous devices and have better information over their functions. I also argue that this and general willingness to be accountable will help plug any responsibility gaps. The importance of the ethical competence expressed through human operators and the current challenges in creating artificial moral agents would suggest an additional point. It suggests the benefits of adding human involvement in the use of ‘fully’ autonomous military devices, even in cases where the device is technically capable of completing its tasks without human involvement. The ethical competence framework highlights the fact that technical ability in completing military tasks is necessary but in no means sufficient. Ethical insight is needed if military goals are to be obtained in ethical ways.

References

Arkin, R.C. (2008) ‘Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture part I: Motivation and philosophy’, in *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 121–128. Available at: <https://doi.org/10.1145/1349822.1349839>.

Asaro, P. (2012) ‘On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making’, *International Review of the Red Cross*, 94(886), pp. 687–709. Available at: <https://doi.org/10.1017/S1816383112000768>.

Champagne, M. and Tonkens, R. (2015) 'Bridging the Responsibility Gap in Automated Warfare', *Philosophy & Technology*, 28(1), pp. 125–137. Available at: <https://doi.org/10.1007/s13347-013-0138-3>.

Childress, J.F. (1978) 'Just-War Theories: The Bases, Interrelations, Priorities, and Functions of Their Criteria', *Theological Studies*, 39(3), pp. 427–445. Available at: <https://doi.org/10.1177/004056397803900302>.

Ekelhof, M.A.C. (2017) 'Complications of a Common Language: Why it is so Hard to Talk about Autonomous Weapons', *Journal of Conflict and Security Law*, 22(2), pp. 311–331. Available at: <https://doi.org/10.1093/jcsl/krw029>.

Fischer, J.M. and Ravizza, M. (1998) *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press (Cambridge Studies in Philosophy and Law). Available at: <https://doi.org/10.1017/CBO9780511814594>.

Gunkel, D.J. (2020) 'Mind the gap: responsible robotics and the problem of responsibility', *Ethics and Information Technology*, 22(4), pp. 307–320. Available at: <https://doi.org/10.1007/s10676-017-9428-2>.

Hakli, R. and Makela, P. (2019) 'Moral Responsibility of Robots and Hybrid Agents', *Monist*, 102(2), pp. 259–275. Available at: <https://doi.org/10.1093/monist/onz009>.

Hellström, T. (2013) 'On the moral responsibility of military robots', *Ethics and Information Technology*, 15(2), pp. 99–107. Available at: <https://doi.org/10.1007/s10676-012-9301-2>.

Johnson, A.M. and Axinn, S. (2013) 'The Morality of Autonomous Robots', *Journal of Military Ethics*, 12(2), pp. 129–141. Available at: <https://doi.org/10.1080/15027570.2013.818399>.

Johnson, D. (2014) 'Technology with No Human Responsibility?', *Journal of Business Ethics*, 127(4), pp. 707–715. Available at: <https://doi.org/10.1007/s10551-014-2180-1>.

de Jong, R. (2020) 'The Retribution-Gap and Responsibility-Loci Related to Robots and Automated Technologies: A Reply to Nyholm', *Science and Engineering Ethics*, 26(2), pp. 727–735. Available at: <https://doi.org/10.1007/s11948-019-00120-4>.

Krishnan, A. (2016) *Killer Robots: Legality and Ethicality of Autonomous Weapons*. London: Routledge. Available at: <https://doi.org/10.4324/9781315591070>.

de Laat, P.B. (2018) 'Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?', *Philosophy & Technology*, 31(4), pp. 525–541. Available at: <https://doi.org/10.1007/s13347-017-0293-z>.

Lazar, S. (2010) 'The Responsibility Dilemma for Killing in War : A Review Essay', *Philosophy & Public Affairs*, 38(2), pp. 180–213.

Leveringhaus, A. (2018) 'What's So Bad About Killer Robots?', *Journal of Applied Philosophy*, 35(2), pp. 341–358. Available at: <https://doi.org/10.1111/japp.12200>.

Marra, W.C. and McNeil, S.K. (2013) 'Understanding the Loop: Regulating the Next Generation of War Machines', *Harvard Journal of Law & Public Policy*, 36, pp. 1139–1185.

Matthias, A. (2004) 'The responsibility gap: Ascribing responsibility for the actions of learning automata', *Ethics and Information Technology*, 6(3), pp. 175–183. Available at: <https://doi.org/10.1007/s10676-004-3422-1>.

Noorman, M. and Johnson, D.G. (2014) 'Negotiating autonomy and responsibility in military robots', *Ethics and Information Technology*, 16(1), pp. 51–62. Available at: <https://doi.org/10.1007/s10676-013-9335-0>.

Nyholm, S. (2018) 'Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci', *Science and Engineering Ethics*, 24(4), pp. 1201–1219. Available at: <https://doi.org/10.1007/s11948-017-9943-x>.

Parasuraman, R. and Manzey, D.H. (2010) 'Complacency and bias in human use of automation: an attentional integration', *Human Factors*, 52(3), pp. 381–410. Available at: <https://doi.org/10.1177/0018720810376055>.

Santoro, M., Marino, D. and Tamburrini, G. (2008) 'Learning robots interacting with humans: from epistemic risk to responsibility', *AI & SOCIETY*, 22(3), pp. 301–314. Available at: <https://doi.org/10.1007/s00146-007-0155-9>.

Sheridan, T.B. (2021) 'Human Supervisory Control of Automation', in *Handbook Of Human Factors And Ergonomics*. John Wiley & Sons, Ltd, pp. 736–760. Available at: <https://doi.org/10.1002/9781119636113.ch28>.

Shoemaker, D. (2011) 'Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility', *Ethics*, 121(3), pp. 602–632. Available at: <https://doi.org/10.1086/659003>.

Skitka, L.J., Mosier, K.L. and Burdick, M. (1999) 'Does automation bias decision-making?', *International Journal of Human-Computer Studies*, 51(5), pp. 991–1006. Available at: <https://doi.org/10.1006/ijhc.1999.0252>.

Smith, A. (2012) 'Attributability, Answerability, and Accountability: In Defense of a Unified Account', *Ethics*, 122, pp. 575–589. Available at: <https://doi.org/10.1086/664752>.

Sparrow, R. (2007) 'Killer Robots', *Journal of Applied Philosophy*, 24(1), pp. 62–77. Available at: <https://doi.org/10.1111/j.1468-5930.2007.00346.x>.

Taylor, I. (2021) 'Who Is Responsible for Killer Robots? Autonomous Weapons, Group Agency, and the Military-Industrial Complex', *Journal of Applied Philosophy*, 38(2), pp. 320–334. Available at: <https://doi.org/10.1111/japp.12469>.

'Toimintakyky sotilaspedagogiikan käsitteenä' (2009) in Toiskallio, J., Mäkinen, J., and Johtamisen ja sotilaspedagogiikan laitos, *Sotilaspedagogiikka: sotiluuden ja toimintakyvyn teoriaa ja käytäntöä*. Helsinki: Maanpuolustuskorkeakoulu, pp. 48–73. Available at: <http://www.doria.fi/handle/10024/74178> (Accessed: 21 September 2022).

Verweij, D. (2007) 'Military Ethics: A Contradiction in Terms', in J. Toiskallio (ed.) *Ethical Education in the Military*. Helsinki: National Defence University, pp. 43–62.

Verweij, D., Hofhuis, K. and Soeters, J. (2007) 'Moral Judgement within the Armed Forces', *Journal of Military Ethics*, 6(1), pp. 19–40. Available at: <https://doi.org/10.1080/15027570701228545>.

Wallach, W., Allen, C. and Smit, I. (2008) 'Machine morality: bottom-up and top-down approaches for modelling human moral faculties', *AI & SOCIETY*, 22(4), pp. 565–582. Available at: <https://doi.org/10.1007/s00146-007-0099-0>.

Walzer, M. (1977) *Just and unjust wars: a moral argument with historical illustrations*. New York: Basic Books, a member of the Perseus Books Group.

Wortel, E. and Bosch, J. (2011) 'Strengthening Moral Competence: A "Train the Trainer" Course on Military Ethics', *Journal of Military Ethics*, 10(1), pp. 17–35. Available at: <https://doi.org/10.1080/15027570.2011.562372>.

USING SIMULATION TO EVALUATE SARS-COV-2 TRANSMISSION IN FINNISH DEFENCE FORCES' LEAVE ARRANGEMENTS

Ville Rissanen, Ruslan Lagashkin, Einar Eidstø, Tenho Korhonen, Emil Toivonen, Jaakko Pentikäinen, & Jouko Vankka

Keywords: coronavirus, simulation, conscription, COVID-19, SARS-CoV-2

Introduction

As a safety procedure during the SARS-CoV-2 pandemic, the Finnish Defence Forces (FDF) significantly changed conscripts' leave arrangements and training implementations. New leave arrangements, combined with other precautions, have been successful at preventing the disease spreading widely within the organization so far. To evaluate the FDF intervention, we have implemented an agent-based SEIR model². In particular, we simulate two leave arrangements to see if one is better at preventing infections among conscripts than the other.

Method

Related Work

Some prior military-related studies in modelling infectious diseases (such as different types of influenza) have also used SEIR-models or at least some variants of it (Chowell, Nishiura and Bettencourt, 2007; Sohn and Boulier, 2012). Besides, there has been research into transmission risks in facilities (Cuevas, 2020). We did not find any research of SARS-CoV-2 modelling in a military conscription context. Agent-based SEIR-models have been used extensively in modelling SARS-CoV-2 intervention measures (Kai *et al.*, 2020; Silva *et al.*, 2020; Thompson *et al.*, 2020). However, we did not find disease prevention measures similar to those implemented by the FDF from any other large conscription-based militaries.

Leave arrangements

The simulation attempts to comply with the same principles the FDF is following currently (Finnish Defence Forces' Defence Command, 2020). The simulated population is divided into four subgroups: conscript groups *A*, *B*, *C*, and the rest of the population (i.e. *civilians*). Conscript groups are equal in size (i.e. 1/3rd of all conscripts). While in service, conscript groups should not interact with each other nor *civilians*. While on leave, a conscript group may interact with *civilians* (and other conscript groups on leave).

² Each person is assigned one of the following states: Susceptible(S), Exposed(E), Infected(I) or Recovered(R), with every person being represented as an agent (A).

Such that: $N = A_S + A_E + A_I + A_R$, where N is the total population.

Simulation

The simulation logic is such that three conscript groups alternate their leaves according to their leave arrangements so that most of the time there are two conscript groups in service and one on leave. There is a short time window during group exchanges when two groups are on leave simultaneously.³ Conscripts are assumed to move freely during their leave, just as the general population.

While in service, conscripts who show infection symptoms, along with nearby *exposed* individuals are quarantined. At the same time, some conscripts may experience and spread the disease without appearance of symptoms. Such *asymptomatic* conscripts would not be automatically quarantined. On the other hand, if a conscript on leave develops infection symptoms, they are ordered in self-quarantine. The model does not take into account quarantines caused by other diseases than SARS-CoV-2.

For the simulation, we used available data from the European Centre for Disease Prevention and Control and the Finnish Institute for Health and Welfare at the time of writing. The simulation parameters are summarised in Table I.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Reproduction number in the general population (R_0)	1.8 [10]
Close contact range	≤ 1.5 m [9]
Incubation period (uniform)	1-14 days [9] (avg. 7)
Symptomatic/Asymptomatic period (uniform)	7-14 days [9] (avg. 12)
Quarantine length	14 days
Asymptomatic cases	45% [11]
Amount of conscripts	3000
Simulation duration	165 days
Chance of infection per close contact	17.3% [9]

We assumed that infected individuals are spreading disease during the incubation period, while not being aware of their condition (Aguilar *et al.*, 2020; Huang *et al.*, 2020). In 45% of cases (European Center for Disease Prevention and Control (ECDC), 2020) (worst-case scenario) agents do not develop symptoms i.e. undergo the disease asymptotically. These carriers are equally infectious. The only way for asymptomatic carriers to get quarantined is for them to get quarantined with a symptomatic carrier. Then all neighbouring agents would be marked as *exposed* and quarantined en masse.

³ For example, if group *A* leaves on Thursday, group *B* arrives on Friday. This is done in order to facilitate cleaning work for disinfecting the areas used by the previous group.

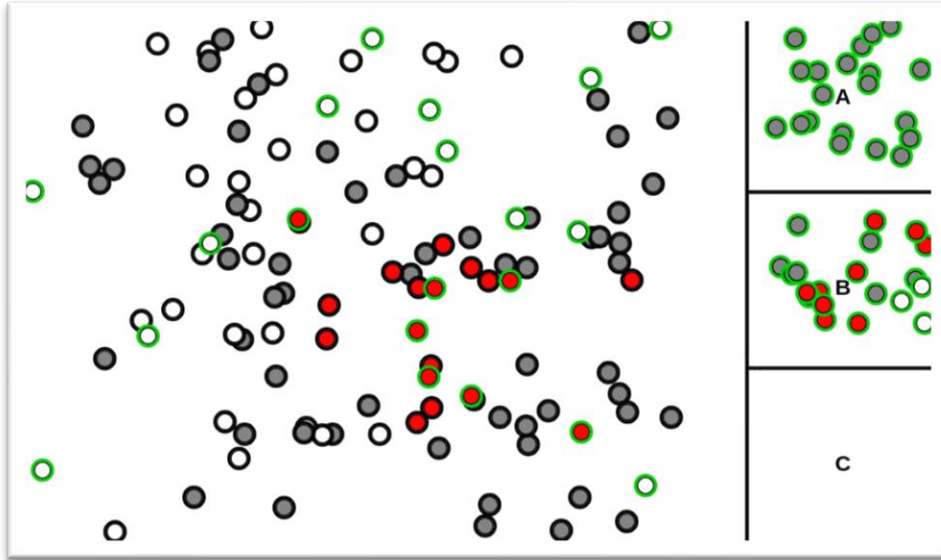


Fig. 1. A simplified graphical version of the simulation setup with reduced amount of agents. Agents move in random patterns in their assigned areas and thus having close contacts with other agents facilitating the transmission of the infection. White-susceptible, Red-infected, Gray-recovered. Green edge of an agent denotes conscripts from the rest of the population. (Quarantine not visible.)

In the simulation, all persons will make a full recovery and become immune to reinfection. This appears to be a reasonable assumption since:

1. Conscripts are younger than 28 years old (19.1 years old on average (Santtila *et al.*, 2018)) and have no severe pre-existing conditions (Finnish Defence Forces, 2020).
2. SARS-CoV-2 biological immunity has been estimated to last 6-12 months (Thompson *et al.*, 2020), while the conscription military service lasts one year at maximum.

We simulated two different leave arrangement scenarios: the present leave arrangement (27 days of service, followed by 15 days of leave) and a shorter one (13+8 days) used briefly before the present arrangement.

The choice of service/leave proportion is constrained by the number of conscript groups. From three groups *A*, *B*, *C* one group is on leave and two in service. Both scenarios were tested with and without a 14-day quarantine procedure.

Results

The simulation results are presented in Table II and III. The positive effect of shorter service- and leave period (13+8 vs. 27+15 days) is quite pronounced ($p \approx 0.08$) for decreasing the total number of infected conscripts, given that quarantine measures are implemented. Without quarantines, the difference effectively disappears (p -value of ~ 0.53) with a significantly higher amount of conscripts becoming infected than with quarantine.

TABLE II. SIMULATION RESULTS I

Leave Arrangement Used (Days in Service) + (Days on leave)	Results (20 simulation runs per arrangement and quarantine variation)				
	14 Day Quarantine Used	Amount of Conscripts Infected	Difference	Worst Case Scenario: Able ^a Conscripts at Minimum	Difference
13+8	Yes	0.36%	3.26%	53.2%	40.9%
13+8	No	3.62%		94.1%	
27+15	Yes	0.57%	4.82%	69.4%	27.5%
27+15	No	5.39%		96.9%	

^a Conscripts not symptomatic or in quarantine,

The use of quarantines is statistically beneficial over both leave arrangements with p-values less than 0.05 using the non-parametric two-tailed Mann-Whitney U Test. The principal explanation is the following: quarantines cut off diseases in service, thus leaving more conscripts healthy.

Figure 2 and 3 (see Appendix) present the difference of incubation periods using two rotations (13+8 and 27+15) with and without asymptomatic carriers. The incubation periods for the simulations for this figure had zero variance. With shorter incubation periods, such as the common influenza, the two rotations have no significant difference. With longer incubation periods the difference becomes significant as the asymptomatic carriers spread the disease for a longer time continuously with a longer in-service period. In Figure 3 there is no significant difference between the leave arrangements when asymptomatic cases do not occur.

TABLE III. SIMULATION RESULTS II

Leave Arrangement Used (Days in Service) + (Days on leave)	Results (Quarantine Used, 50 simulation runs per leave arrangement)				
	Amount of Infected Conscripts on average in total	Amount of Infected Conscripts on average in service	Amount of Infected Conscripts on average on leave	Ratio of Infected Conscripts* leave/service of all infected Conscripts	p-value for the difference between two ratios of Infected conscripts leave/service
13+8	12.93	7.98	4.95	0.62	0.07
27+15	23.33	18.62	4.71	0.253	
Difference between leave arrangements	-10.4	-10.64	0.24	0.367	

*Conscripts not symptomatic or in quarantine,

The ratio of conscripts falling ill on leave and in service is higher with 13+8 arrangement with difference of 0.367 and p-value of 0.07.

The FDF employs about 12,000 employees and trains 22,000 conscripts annually (*About us - Finnish Defence Forces*, 2020). This simulation only analyzes the movement

of conscripts. The employees or reservists in refresher exercises are not modeled, which is a significant limitation of this simulation. It should be also noted that this study does not take into account aspects related to training arrangements in conscripts' service. In this study the reproduction number was constant, however in the future we intend to expand it in the future by using a time-varying reproduction number.

References

About us - Finnish Defence Forces (2020) *Puolustusvoimat*. Available at: <https://puolustusvoimat.fi/en/about-us> (Accessed: 27 August 2020).

Aguilar, J.B. *et al.* (2020) 'Investigating the Impact of Asymptomatic Carriers on COVID-19 Transmission', *medRxiv* [Preprint]. Available at: <https://doi.org/10.1101/2020.03.18.20037994>.

Chowell, G., Nishiura, H. and Bettencourt, L.M. (2007) 'Comparative estimation of the reproduction number for pandemic influenza from daily case notification data', *Journal of the Royal Society Interface*, 4(12), pp. 155–166.

Cuevas, E. (2020) 'An agent-based model to evaluate the COVID-19 transmission risks in facilities', *Computers in biology and medicine*, 121, p. 103827.

European Center for Disease Prevention and Control (ECDC) (2020a) *Rapid Risk Assessment: Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK – eleventh update*. Available at: <https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-eueea-and-uk-eleventh>.

European Center for Disease Prevention and Control (ECDC) (2020b) *Rapid Risk Assessment: Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK – tenth update*. Available at: <https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-pandemic-tenth-update>.

Finnish Defence Forces (2020) *Varusmiespalveluksessa ei ole lähtökohtaisesti ole vakavan koronavirustaudin riskiryhmään kuuluvia*. Available at: <https://puolustusvoimat.fi/-/varusmiespalveluksessa-ei-lahtokohtaisesti-ole-vakavan-koronavirustaudin-riskiryhmaan-kuuluvia> (Accessed: 31 August 2020).

Finnish Defence Forces' Defence Command (2020) *Muutoksia varusmiesten koulutus- ja lomajärjestelyihin*. Available at: <https://puolustusvoimat.fi/-/muutoksia-varusmiesten-koulutus-ja-lomajärjestelyihin> (Accessed: 21 August 2020).

Finnish Institute for Health and Welfare (THL) (2020) *Ennusteita bidastamistoimien vaikutuksesta*. Available at: <https://thl.fi/fi/web/infektiotaudit-ja-rokotukset/ajankohtaista/ajankohtaista-koronaviruksesta-covid-19/tilannekatsaus-koronaviruksesta/koronavirusepidemian-mallinnusta> (Accessed: 25 September 2020).

Huang, L.-S. *et al.* (2020) 'Taking Account of Asymptomatic Infections in Modeling the Transmission Potential of the COVID-19 Outbreak on the Diamond Princess Cruise Ship', *medRxiv*, p. 2020.04.22.20074286. Available at: <https://doi.org/10.1101/2020.04.22.20074286>.

Kai, D. *et al.* (2020) 'Universal masking is urgent in the COVID-19 pandemic: SEIR and agent based models, empirical validation, policy recommendations', *arXiv:2004.13553* [Preprint]. Available at: <https://arxiv.org/abs/2004.13553>.

Oran, D.P. and Topol, E.J. (2021) 'Prevalence of asymptomatic SARS-CoV-2 infection', *Annals of Internal Medicine*, 174(2), pp. 286–287.

Santtila, M. *et al.* (2018) 'Physical fitness in young men between 1975 and 2015 with a focus on the years 2005-2015', *Medicine and science in sports and exercise*, 50(2).

Silva, P.C. *et al.* (2020) 'COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions', *Chaos, Solitons & Fractals*, 139, p. 110088.

Sohn, K. and Boulier, B.L. (2012) 'Estimating Parameters of the 1918-19 Influenza Epidemic on U.S. Military Bases', *The Journal of Applied Business and Economics*, 13, pp. 30–42.

Thompson, J. *et al.* (2020) 'Modelling the estimated likelihood of eliminating the SARS-CoV-2 pandemic in Australia and New Zealand under public health policy settings: an agent-based-SEIR approach.', *Available at SSRN 3588074* [Preprint].

Appendix

Conscripts infected (all carriers symptomatic)

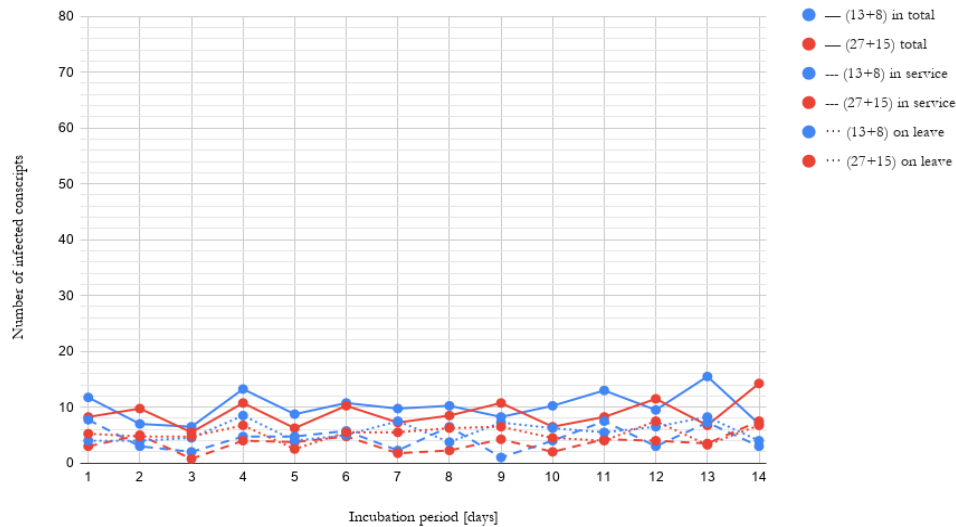


Figure 2. Red lines belong to the longer 27+15 rotation arrangement, while blue – to the shorter 13+8. Solid lines represent the total number of infected conscripts and are combined from those infected on leave (dotted line) and in service (dashed line). The results are averaged over 20 simulation runs for every period. Longer incubation period of the disease (over 9 days) leads to more infected conscripts in the case of longer service periods ($p \approx 0.08$)

Conscripts infected (asymptomatic carriers allowed)

with 45% asymptomatic infections

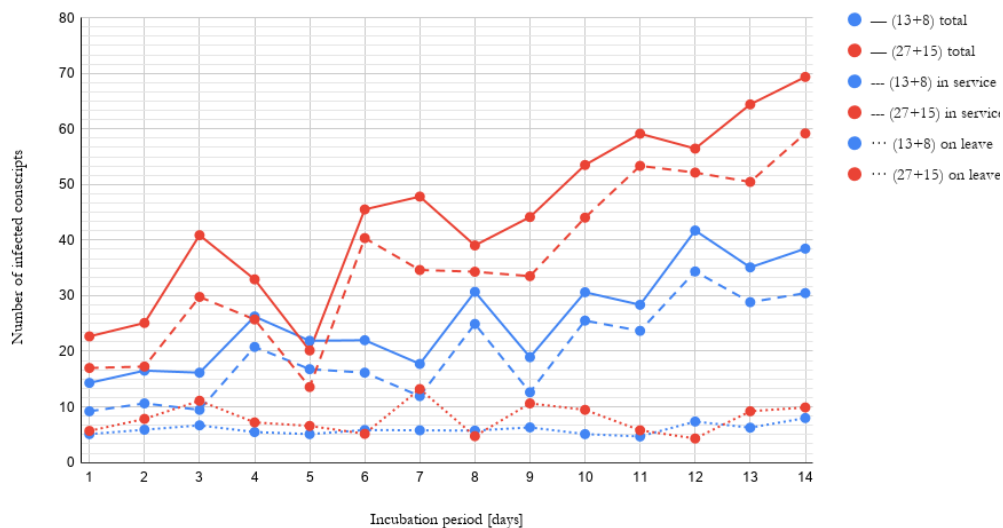


Figure 3 Red lines belong to the longer 27+15 rotation arrangement, while blue – to the shorter 13+8. Solid lines represent the total number of infected conscripts and are combined from those infected on leave (dotted line) and in service (dashed line). The results are averaged over 20 simulation runs for every period. There is no significant difference between rotations

SUMEEA LOGIIKKA SOTAVÄESSÄ

Riitta Penttinen

Ihmisen aivot, ajattelu ja logiikka mallina

1 980- ja 1990-luvuilla ryhdyttiin Suomessakin tutkimaan ja kehittämään ihmisen ajattelua ja aivojen toimintaa matkivia teknisiä ratkaisuja. Suomalaiset ja tanskalaiset innostuivat tästä ja olivatkin edellä muuta Eurooppaa ja länsimaita. Näkökulmia tähän oli kaksi: toisaalta tutkittiin ajattelun, päätöksentekemisen tai päättelmissen prosessia ja toisaalta aivojen solutason toimintaa. Näihin molempiin, ajatteluun ja solutason toimintaan, kehitettiin ohjelmallisia ratkaisuja, jotka kykenevät matkimaan ihmisen aivojen toimintaa. Ajattelun ja päätöksenteon sovellukset ovat esimerkiksi asiantuntijajärjestelmiä, monikriteerianalyysiä ja sumeaa logiikkaa, solutason tapahtumia mallintavat erilaiset neuroverkot (ks. esim. Ek 2018; Ertel 2017; Ferdaus 2020; Haikonen 2017; Kelleher 2020; Kosko 1993; Mattila 2002; Penttinen 1998; Penttinen, Ala-Siuru, Plomp & Vormisto 2000; Siukonen & Neittaanmäki 2019.). Ne ovat osa tekoälyn kehittämistä ja tutkimusta, mutta niiden lisäksi tekoäly sisältää paljon muutakin.

Vaikka aikaa on kulunut jo tovi ja sitä myötä tutkimusta ja kehitystyötä on tehty runsaasti, paljon on vielä tehtävää. Tietoteknisissä sovelluksissa nämä aivoissa samanlaisesti tapahtuvat prosessit ovat edelleen erillisiä, toisistaan irrallaan tapahtuvia prosesseja. Tämän lisäksi on vielä paljon asioita ja ilmiöitä, joita emme ihmisen päättelystä ymmärrä, joten tutkimus jatkuu.

Sumea logiikka, asiantuntijajärjestelmät ja neuroverkot kuuluvat kaikki moderniin säätötekniikkaan, jossa niitä on hyödynnetty erilaisissa teollisuuden(kin) prosessien ohjauksessa. Alkuvuosina niiden käyttöä tutkittiin paperi- ja sellutehtaissa, sotilasalalla esim. tukikohdan vesiensuojelussa sekä monimutkaisten prosessien mallinnuksessa tai niihin liittyvässä päätöksenteossa. Myöhemmin mukaan ovat tulleet esimerkiksi autojen ABS-jarrut, riisinkettimet ja kameroiden optinen säätö. (ks. esim. Penttinen 1994; Penttinen 1998; Carter ym. 2021.)

Sääntöpohjaisia päättelylogiikoita ovat erilaiset asiantuntijajärjestelmät, monikriteerianalyysit sekä sumea logiikka sovelluksineen. Niissä kaikissa yhteisenä tekijänä ja lähtökohtana on jonkinlainen säännöstö, jota prosessoimalla tullaan lopulliseen johtopäätökseen. Tässä artikkelissa keskitytään nimenomaan näihin sääntöpohjaisiin järjestelmiin. Esimerkiksi neuroverkoista löytyy oivallisia julkaisuja ja oppaita ja niistä on tehty Suomessa paljon urauurtavaakin tutkimusta (kts. esim. Ek 2018; Ertel 2017; Ferdaus 2020; Haikonen 2017; Kelleher 2020; käytännössä kaikki suomalaiset yliopistot).

Tässä artikkelissa tarkastellaan sääntöpohjaisten järjestelmien eroja käytännönläheisesti esimerkkien kautta. Esimerkit eivät ole täydellisiä ja kaikki mahdolliset vivahteet kattavia, vaan niiden tavoitteena on havainnollistaa yksinkertaisesti ja selkeästi valittujen järjestelmien eroja. On hyvä huomata, että säännöstöjä voidaan laatia eri tavoin ja

niiden pohjalta voidaan tehdä päätelmiä eri tavoin saman yläkäsitteen alla. Tässä artikkelissa ei mennä näihin yksityiskohtiin, niitä sivutaan vain, mikäli se tuottaa yleiskuvan saamiseen lisäarvoa.

Asiantuntijajärjestelmät

Asiantuntijajärjestelmä on yksinkertaisin muoto sääntöpohjaisesta järjestelmästä. Siinä jokaiselle muuttujalle ja sen vaihtoehtoiselle arvolle on muodostettu oma sääntönsä - jokaiselle tilanteelle on siis oma sääntönsä ja yksi sääntö sopii vain yhteen tilanteeseen. Näin sääntökannan koko kasvaa hyvin helposti huomattavan suureksi. (Hyvönen, Karanta, Syrjänen & Carlson 1993; Penttinen 1994.)

Kiinnostus asiantuntijajärjestelmiin Suomessa oli varsin suurta 1980-1990-luvuilla, sittemmin mielenkiinto on suuntautunut enemmän muiden kehittyneiden menetelmien ja neuroverkkojen suuntaan. Asiantuntijajärjestelmien hyödyntämisen alkuaikoina julkaistiin paljon myös suomenkielisiä tekstejä niiden perusteiden lisäksi menetelmien soveltamisesta erilaisiin kohteisiin (ks. esim. Hyvösen Eeron julkaisut vuodelta 1985 asiantuntijajärjestelmistä, niiden kehitystyökaluista ja hyödyntämisestä teollisuudessa). Kirjastojen järjestelmistä hakusanalla "asiantuntijajärjestelmä" löytyy runsaasti tuon aikakauden julkaisuja, tutkimuksia ja tutkielmia ja selailemalla tuloksia saa helposti hyvän kuvan siitä soveltamisen laajuudesta ja luovuudesta, joka tähän aikaan liittyi.

Erinomainen esimerkki asiantuntijajärjestelmästä on sairauksien diagnosointiin kaaavailtu järjestelmä. Lääkärin diagnostiikan logiikka on erittäin kiehtova teema, jossa tekoälyjärjestelmät myös toimivat hyvin. Tällöin potilaan kertomien oireiden perusteella lääkäri tai järjestelmä päätelee sairauden laadun ja sitä kautta tarvittavan hoidon. Oletuksena on, että potilas osaa kertoa kaikki oireet jättämättä mitään olennaista pois - tai lääkäri tahi järjestelmä osaa kysyä puuttuvia tietoja. Toinen oletus on se, että jokainen sairaus oireilee täysin samalla tavalla. Tiettyyn nimettyyn tautiin liittyy aina kuume ja kurkkukipu sekä nenän vuotaminen - ja jos näistä jokin puuttuu, kyse ei ole tästä taudista, tai jos näiden lisäksi on vielä jotain muuta, tauti ei välttämättä ole tämä. Sinänsä tekoälyä on käytetty monipuolisesti terveydenhuollossa (Vähäkainu & Neittaanmäki 2018).

Asiantuntijajärjestelmän sääntöjä havainnollistetaan tässä seuraavalla esimerkillä, jossa pohditaan sitä, miten tulisi varustautua lähdettäessä ulos. Lähtökohtana on siis tilanne, jossa lähdettäessä ulos katsotaan säätilaa ja päätetään, otetaanko mukaan *sateenvarjo*, *aurinkolasit* ja/tai *kaulaliina*.

Seuraavassa esimerkissä tarkasteltavia kriteereitä ovat sade (ei sada, tihkuttaa, sataa), auringonpaiste (paistaa, paistaa pilvien lomasta, ei paista) ja lämpötila (lämmin, viileä, kylmä). Tilanteesta riippuen otetaan mukaan *sateenvarjo*, *aurinkolasit* ja/tai *kaulaliina*. Sääntöjä on 27 kappaletta.

Taulukko 1. Esimerkki asiantuntijajärjestelmän sääntökannasta.

Jos ei sada ja aurinko paistaa ja on lämmin, niin ota mukaan aurinkolasit.
*Jos sataa vähän (tibkuttaa) ja aurinko paistaa ja on lämmin, niin ota mukaan aurinko-
lasit.*
Jos sataa ja aurinko paistaa ja on lämmin, niin ota mukaan sateenvarjo ja aurinkolasit.
*Jos ei sada ja aurinko paistaa hieman (pilvien lomasta) ja on lämmin, niin ota mukaan
ota mukaan aurinkolasit.*
...
...
...

Jos ei sada ja aurinko ei paista (on pilviä) ja on kylmä, niin ota mukaan kaulaliina.
*Jos sataa vähän (tibkuttaa) ja aurinko ei paista (on pilviä) ja on kylmä, niin ota mukaan
sateenvarjo ja kaulaliina.*
*Jos sataa ja aurinko ei paista (on pilviä) ja on kylmä, niin ota mukaan sateenvarjo ja
kaulaliina.*

Tässä tarkasteltavia asioita oli vain kolme ja jokaiselle niistä oli olemassa kolme vaihtoehtoa. Jos ajatellaan esimerkiksi sairauksien diagnostiikkaa, muuttujia on huomattavasti enemmän ja ne voivat saada huomattavasti erilaisempia arvoja. Lisäksi joidenkin sairauksien kohdalla potilaan tilanteesta ei kerro suinkaan yksittäiset arvot, vaan niiden kombinaatiot - ja lisäksi kyse voi olla tilanteesta, jossa mikään hälytysraja ei rikkoudu, vaan jokin arvo on pikkuisen yläkanttiin ja jokin toinen hieman alhainen. Käytännössä tämä tarkoittaa sääntökannan koon paisumista sekä sen huomioimista, miten reagoidaan puuttuvaan kriteeriin - todetaanko, ettei diagnoosia tai päättelyä voida tehdä, vai todetaanko, että tulos perustuu vajaisiin tietoihin ja että olisi suositeltavaa hankkia seuraavat tiedot.

Monikriteerianalyysi

Sääntöpohjaisten järjestelmien kehittyneempiä versioita ovat monikriteerianalyysi ja sumea logiikka, joissa sääntökantaa tarkastellaan hieman erilaisista näkökulmista. Monikriteerianalyysi on varsin yksinkertainen ja arjessa tuttu päätöksentekotapa, joka ei kuitenkaan sovellu aivan kaikkeen tekemiseen. Siinä tunnistetaan päätöksentekoon vaikuttavat kriteerit, jotka on yleensä laitettu tärkeysjärjestykseen. Lisäksi kriteereille on mietitty vaihtoehtoiset arvot. Menetelmän käytössä loppu on puhdasta matemaatiikkaa: kun lasketaan yhteen kriteerien arvot kerrottuna kyseisen kriteerin painoarvolla, saadaan vaihtoehdon sama yhteispistemäärä. Tilanteesta riippuen voittajavaihtoehto on se, mikä sai pienimmän tai suurimman pistemäärän. (Greco yms. 2016; Penttinen 1998; Sarin & Weistroffer 2016.)

Taulukko 2. Esimerkki monikriteerianalyysin kriteereistä painokertoimineen.

Kriteeri	Kriteerin painokerroin			
Sade	3	Ei sada (0)	Tihkuttaa (2)	Sataa (3)
Aurinko	1	Paistaa (1)	Pilvien lomasta (2)	Ei paista (3)
Lämpötila	2	Lämmin (1)	Viileä (2)	Kylmä (3)

Taulukkoa käytetään siten, että kunkin kriteerin vallitseva arvo kerrotaan kriteerin painokertoimella ja nämä lasketaan yhteen. Mitä enemmän pisteitä, sitä enemmän otettava välineitä mukaan eli parempi keli = vähemmän pisteitä.

Taulukko 3. Esimerkki taulukon 2 soveltamisesta taulukon 1 esimerkkiin.

Alla esimerkinomaisesti sääntöjen tilanteista tehtyjä laskelmia:

Jos ei sada ja aurinko paistaa ja on lämmin, niin ota mukaan aurinkolasit.

$$\Rightarrow 3 * 0 + 1 * 1 + 2 * 1 = 3$$

Jos sataa vähän (tihkuttaa) ja aurinko paistaa ja on lämmin, niin ota mukaan aurinkolasit.

$$\Rightarrow 3 * 2 + 1 * 1 + 2 * 1 = 9$$

Jos sataa ja aurinko paistaa ja on lämmin, niin ota mukaan sateenvarjo ja aurinkolasit.

$$\Rightarrow 3 * 3 + 1 * 1 + 2 * 1 = 12$$

Monikriteerianalyysiin tässä kuvattu esimerkki ei ole kovin hyvä, soveltuvampi on esimerkiksi auton valinta, jossa valintakriteereinä voisivat olla auton hinta, tehokkuus ja näyttävyys. Päätöksentekijästä riippuen näiden keskinäinen painotus voi vaihdella runsaastikin ja kriteerien arvot voivat saada pisteitä hyvin eri tavoin. Tulosta voidaan haakea joko tavoittelemalla mahdollisimman vähän tai paljon pisteitä.

Taulukko 4. Esimerkki monikriteerianalyysin soveltamisesta.

Kriteeri	Kriteerin painokerroin	Kriteerin arvot	Vaihtoehtoiset ajoneuvot		
			Auto X	Auto Y	Auto Z
Hinta	3	Kallis (1), keskihintainen (2), halpa (3)	Kallis 1	Kallis 1	Halpa 3
Tehokkuus	2	Tehokas (3), keskiverto(2), olematon (1)	Keskiverto 2	Tehokas 3	Keskiverto 2
Näyttävyyys	1	Komea (3), nuorisomalli (2), perheauto (1)	Perheauto 1	Nuorisomalli 2	Komea 3

Taulukon pisteytyksestä nähdään, että tavoitteena on siis halpa, tehokas ja komea ajoneuvo. Mikäli etsittäisiin turvallista perheautoa, kriteerit voisivat olla hieman erilaisia ja samoin niiden pisteytys. Tämän taulukon tuloksen perusteella auto X saa 8 pistettä ($3*1+2*2+1*1$), auto Y 11 pistettä ($3*1+2*3+1*2$) ja auto Z 10 pistettä ($3*1+2*2+1*3$).

Taulukon täyttämistä ja vaihtoehtojen pisteyttämistä voidaan ohjata erilaisin säännöin. Voidaan esimerkiksi ottaa kantaa siihen voiko sama arvo olla kahdella eri vaihtoehdolla vai pitääkö jokaisen kriteerin osalta laittaa vaihtoehdot paremmuus järjestykseen.

Menetelmänä monikriteerianalyysi on yksinkertainen ja varsin selkeä. Se on arjessa tuttu meille kaikissa valintatilanteissa. Tiedostamatta teemme valinnat sen mukaisesti: Mitä otan tänään lounaaksi? Otanko halpaa? Ratkaiseeko kotimaisuus tai raaka-aineet? Entä terveellisyys? Tai ruokailun nopeus?

Sumea logiikka

Sumea logiikka lähestyy mallinnettavaa ilmiötä hieman toisella tavalla. Siinä jokaiselle muuttujalle on arvovaihtoehdot ja vaihtoehtoa vastaava lopputulos. Päätelyn lopputulema muodostuu näiden kutakin muuttujaa koskevien sääntöjen yhteisvaikutuksesta. Yhteisvaikutus ei kuitenkaan ole automaattisesti suoraan sääntöjen yhdistelmä tai niiden vaatimien toimenpiteiden summa tai keskiarvo. Se voi olla esimerkiksi joukko toimintaohjeita tai vaihtoehtoja, joista sitten käyttäjä valitsee lopulliset. (Kosko 1993; Lowen & Roubens 1993; Penttinen 1994; Penttinen 1998.)

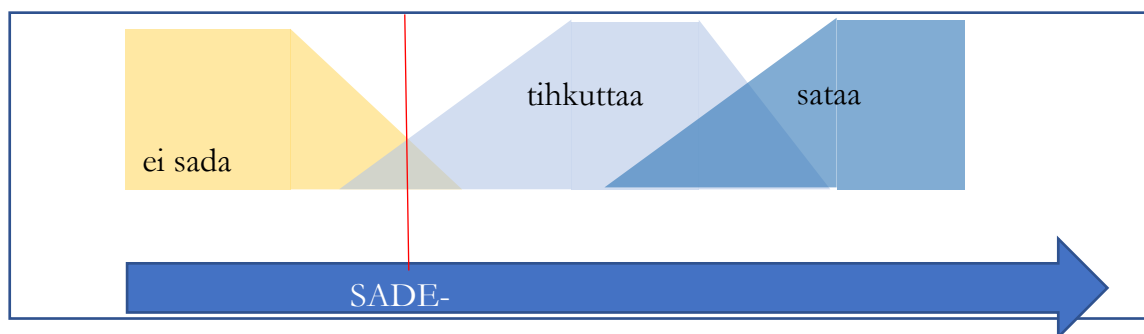
Edellä esitellyn esimerkin osalta säätilan arvioinnissa ja tarvittavien varusteiden pohdinnassa sääntökanta on varsin pieni (taulukko 5).

Taulukko 5. Esimerkki sumean logiikan sääntökannasta.

Jos ei sada, älä ota sateenvarjoja mukaan.
Jos sataa vähän (tihkuttaa), älä ota sateenvarjoja mukaan.
Jos sataa, ota sateenvarjo mukaan.
Jos aurinko paistaa, niin...
Jos aurinko paistaa pilven lomasta, niin...
Jos aurinko ei paista, niin...
Jos on lämmin, niin...
Jos on viileää, niin...
Jos on kylmä, niin...

Näiden muutaman säännön yhdistelmät tuottavat halutun lopputuloksen.

Sumeassa logiikassa on huomattava, että samalle muuttujalle voi olla yhtä aikaa voimassa useita eri sääntöjä ja myös negatiivisia sääntöjä voidaan hyödyntää. Negatiivisella säännöllä voidaan rajata jokin vaihtoehto pois - aivan kuten arkiajattelussakin. Toisin sanoen esimerkiksi sateen määrälliset rajat (sataa, tihkuttaa, ei sada) voivat liittyä. Aivan kuten arkiajattelussa sateen alkaessa ilma tuoksuu sateelle, pisaroita tipahtelee harvakseltaan, mutta ei vielä edes oikein tihkuta - tällöinhän ei oikein ole täysin ”ei sada”-tilanne, mutta ei vielä kunnolla ”tihkuta”. Toisin sanoen kumpikaan ei ole täysin (100 %) voimassa, mutta on vähän sinne päin. Tätä kuvataan esimerkiksi kuvion 1 tavalla, jossa muuttujan numeeriset arvot muutetaan yksinkertaisesti tiettyyn joukkoon kuulumisen arvoksi. (Kosko 1993; Lowen & Roubens 1993; Penttinen 1998.)



Kuvio 1. Esimerkki sumean logiikan muuttujan sataa arvoista "ei sada", "tihkuttaa" ja "sataa". Arvoalueet tai joukot menevät päällekkäin, joten raja-alueella muuttuja kuuluu useampaan joukkoon yhtä aikaa. Punainen viiva on esimerkki tilanteesta, jossa sademäärä mitattavana arvona on vähäinen, jolloin säätila ei ole aivan täysin "ei sada", mutta ei vielä oikein kunnolla "tihkutaan".

Sumean logiikan hyödyntämisessä on erilaisia tapoja. Tutkijoiden kesken on käyty keskustelua esimerkiksi siitä, kuinka moneen joukkoon muuttuja voi samanaikaisesti kuulua tai millaisia yhteenlaskettuja arvoja muuttujat voivat saada. Tilanteesta ja tutkijasta riippuen voidaan rakentaa järjestelmä siten, että muuttuja voi kuulua yhtä aikaa vain tiettyyn määrään joukkoja tai jokaiseen sovelluksessa olevaan joukkoon. Samoin osa tutkijoista rajoittaa joukkoihin kuulumista siten, että muuttujan joukkoihin kuulumisen saa aina arvon 1 riippumatta siitä, onko kyse yhteen tai useampaan joukkoon kuulumisesta. Luonnossa, todellisuudessa ilmiöt kuuluvat useisiin joukkoihin yhtä aikaa

ja yhteenlasketut arvot voivat hyvinkin ylittää 100% tai arvon 1, joten mikäli tavoitellaan luontoa matkivaa tilannetta, tämä on varsin ymmärrettävä lähtökohta. Esimerkkinä olkoon eräs lokakuinen päivä syksyltä 2020, jolloin kävelylenkin alussa taivaalta ei satanut yhtään mitään ja lenkin loppuessa satoi yhtä aikaa vettä (selkeitä pisaroita), räntää (sohjoisia lumiklönttejä), lumihiutaleita (hiutaleita, joista sakarat erottuivat) ja rakeita (kopsahtelivat silmälaseihin ja olivat pallomaisia). Alkutilanne oli siis selkeästi ”ei sada” ja 100% tätä. Lopputilanne oli täysin satamista (arvo 1), mutta sateen eri olomuotojen summa oli jotain muuta - jos tarkastellaan vain sadetta, olisin antanut sille arvon 0,5 (tai 50%), niin paljon vesipisaroita oli (siis verrattaessa tavalliseen sateeseen, ei siis osuus senhetkisestä sateesta), vastaavasti räntää tuli varsin runsaasti eli 0,5, lumihiutaleita esiintyi vähemmän eli ehkä vain 0,2 ja rakeita enemmän 0,4. Toisin sanoen sade oli erittäin kastelevaa, mutta ei pelkkää vettä. Ja jos olisi tullut vain lumihiutaleita, kyse olisi ollut kuvankauniista ensilumen satamisesta. Näin lenkin loppupuolella keli oli kyllä sadetta, mutta varsin tiheää sellaista, jossa vesi tuli eri olomuodoissa.

Tavoitteena sumeassa logiikassa on huomioida epävarmat ja epäselvät tilanteet erilaisen arvojen raja-alueilla. Siinä siis hyväksytään, että on/off-tilanteiden lisäksi on olemassa myös muitakin vaihtoehtoja - vähän, sinnepäin, hitusen. Lämpöpatteri voi olla lämpenemässä, jolloin se ei vielä ole lämmin, mutta se ei enää ole täysin kylmäkään. Ruoka voi olla menossa pilalle, se ei siis ole ihan vielä täysin syömäkeltvotonta, mutta ei enää ihan tuoretta - vähän siis kuin kuivettunut vanha leipä, joka ei ole vielä homeessa. Kahvi voi olla jäähtynyttä, mutta ei vielä täysin kylmää eikä enää vastakeitetyn kuumaa - ja siksi se vaatii erilaisen tavan käsittelylleen: sen voi juoda nopeasti.

Perinteinen arkinen keittiön uuni toimii siten, että se lämpenee ensin hieman yli asetetun lämpötilan ja lopettaa toviksi lämmityksen. Lämmitys napsahtaa takaisin päälle, kun uunin sisälämpötila on laskenut tarpeeksi. Taas uunia lämmitetään tovi, hieman yli halutun ja lämmitys napsahtaa pois. Näin uunissa todellisuudessa lämpötila seilaa valmistajan suunnitteleman alueen sisällä. Sumealla logiikalla uunin lämpötilan säätö tapahtuisi hieman hienovaraisemmin: kun lämpötila alkaa lähestyä tavoitetta, lämmitystehoa lasketaan, ja kun lämpötila laskee, lämmitystehoa nostetaan varovasti. Näin pyritään löytämään sellainen alue tai tila, jossa lämmitysteho olisi mahdollisimman vakio ja samalla uunin lämpötila pysyisi haluttuna. Tästä koituu suuremmissa teollisissa prosesseissa säästöjä energiankulutuksen osalta, samoin toiminnan pysyessä tasaisena laitteiden ja ympäristön kuormitus pysyy yleensä hallitumpana ja tasaisempana - toiminnan tasoittuessa tai vakioituessa sitä voidaan sitten ryhtyä optimoimaan haluttuun suuntaan. Sumeaa logiikkaa voidaan hyödyntää erittäin monipuolisesti sellaisenaan tai yhdistettynä muihin menetelmiin. (Carter yms. 2021; Lowen & Roubens 1993)

Sotaväen sovellukset

Asiantuntijajärjestelmien, monikriteerianalyysin ja sumean logiikan osalta sotavoimat ovat tarttuneet kehityksen mahdollisuuksiin jo hyvin varhain. Sumean logiikan ensimmäisiä sotilassovelluksia ovat olleet tukikohtien vesihuollon järjestelmät, joista käyttö on levinnyt nopeasti muualle. Sittemmin sumeaa logiikkaa on tutkittu ja hyödynnetty esimerkiksi maamiinojen paikannuksessa. (Aura 2018; Gader, Keller & Nelson 2001;

Frigui & Gader 2006; Muranen yms. 2003; Penttinen 1994; Saarinen yms. 1998.) Sitä on hyödynnetty sekä tiedustelussa analysoinnissa että joukkojen optimaalisen sijoittelun taistelussa parhaan suorituskyvyn takaamiseksi (Kewley & Embrechts 1998; Ragsdale yms. 1997). Siitä on kehitetty sovellus hankinnan päätöksenteon tueksi myös sotilaskäyttöön suunnattujen koulutukseen tarkoitettujen lentokoneiden arviointiin (Sánchez-Lozano & Rodríguez 2020). Sitä ja kehittyntä säätötekniikkaa on sovellettu joukkojen ja aseiden automatisoituun johtamisjärjestelmään ja aseiden ohjaamiseen (esim. Kazaryan 2010; Wang yms. 2011). Vastaavaa kehitystä ja tutkimusta on luonnollisesti tehty muiden teknologioiden kohdalla.

Yhteistä kaikille näille sovelluksille on tilanne, jossa päätös - oli sitten kyse prosessin ohjaukseen liittyvästä päätöksestä tai vaikka hankinnasta - on tehtävä vaihtelevien ja keskenään ristiriitaistenkin sekä epävarmojen lähtötietojen perusteella (Kazaryan 2010; Sánchez-Lozano & Rodríguez 2020). Lisäksi päätösvaihtoehdoilla tunnistetaan olevan monenlaisia vaikutuksia tilanteeseen ja sen tulevaan kulkuun. Ei siis välttämättä ole olemassa yhtä oikeaa ratkaisuja, vaan on joukko ratkaisuja, jotka aiheuttavat uusia tilanteita muuttaen vallitsevia olosuhteita. Toinen yhteinen ulottuvuus on päätökseen vaikuttavien tekijöiden sanallistaminen. Toisin sanoen päätöksentekijät pystyvät kuvaamaan ja kertomaan ne asiat, jotka vaikuttavat päätöksentekoon ja miten ne vaikuttavat siihen. Näin ollen päätöksenteon logiikka pystytään kuvaamaan ja mallintamaan se tavalla tai toisella.

Mahdollisia (sotilaallisia) sovelluskohteita ovat esimerkiksi tilanteet, joissa epämääräisistä havainnoista pitää tunnistaa olennainen tai jossa tilanteen kehityksessä tapahtuu vain pieni mutta olennainen muutos. Tällaisissa tapauksissa sekä sumea logiikka että neuroverkot voivat yhdessä tai erikseen tarjota mielenkiintoisia mahdollisuuksia toiminnan nopeuttamiseen tai (laadun) vakauttamiseen. Samalla on kuitenkin hyvä pitää mielessä, että näiden järjestelmien luotettavuus ja laatu ovat vahvasti kytköksissä niihin sääntöihin tai opetusdataan sekä mahdollisiin ennakko-oletuksiin, joiden pohjalta ne on rakennettu. Tilanteen muuttuessa sopivasti esimerkiksi ajauduttaessa ennakko-oletusten tai rajoitusten ulkopuolelle järjestelmä ei enää välttämättä toimi loogisesti - tämä koskee kaikkia tekoälysovelluksia. Taitava ja huolellinen järjestelmän rakentaja osaa varautua näihin tilanteisiin ennakoita ja rakentaa järjestelmään mahdollisesta virheestä ilmoittavan mekanismin - oli sitten kyse sääntöpohjaisista järjestelmistä tai neuroverkoista.

Kehittyneen säätötekniikan näkymiä

Kehittyneiden säätötekniikan menetelmien ja keksintöjen soveltamista uusille alueille tapahtuu koko ajan, siitä yhtenä esimerkkinä on tekoälyn hyödyntäminen esimerkiksi mainonnan optimoinnissa (Merilehto 2018). Sovelluksien ideoita kokeillaan varsin luovasti kohteesta toiseen. Tekoälyn kehittäminen on edelleen ajankohtaista ja kehitystä on syytä seurata, jotta tiedämme, missä mennään. Myös niiden vaikutuksia taistelukenttään yleisesti on syytä pohtia. (Cummings 2017; Fawkes & Menzel 2018; Hainonen 2017; Rabkin & Yoo 2017; Rutkowski yms. 2020; Svenmark yms. 2018).

Sääntöpohjaiset järjestelmät soveltuvat hyvin monenlaisiin tilanteisiin, mutta luonnollisestikaan eivät kaikkeen. Järjestelmien ja sovellusten kehittäminen käynnistyi varsin vauhdikkaasti ja alkuvuosina saatiinkin otettua isoja edistysaskelia. Nyt kehitys on tietyllä tavalla hidastunut, sillä kehitystyö kohdistuu menetelmien soveltamiseen haastavampiin tilanteisiin. Yksittäisten menetelmien ja teknologioiden lisäksi tutkitaan niiden yhteiskäyttöä erilaisina kombinaatioina. Multikriteerianalyysiin paneutuvat esimerkiksi Greco kumppaneineen (2016), Sarin ja Weistroffer (2016) sekä Tudorica kumppaneineen (2021). Menetelmien yhteiskäyttöä on tutkittu kohteen paikan ja tilan lyhytaikaisessa ennustamisessa yhdistämällä sumeaa logiikkaa neuroverkkoihin Romanenkon tutkimusryhmässä (2021). Sotilassovelluksissa nanoteknologiaan ovat keskittyneet esimerkiksi Kumar ja Dixit (2019). Myös neuroverkot ja miehittämättömät alukset ovat tutkimuksen kohteena (Irfan yms. 2020).

Tekoälyn kehittämisen alkuvuosien huuma on vaihtunut realistisempaan otteeseen, jossa myös mahdolliset ongelmat niin eettisyyden kuin tekniikoidenkin osalta on huomioitu aiempaa paremmin.

Lähteet

Aura, E. (2018) *Fuzzy logic and unmanned surface vehicles : Implementing collision avoidance in Python*. Master's Thesis. Available at: <https://www.semanticscholar.org/paper/Fuzzy-logic-and-unmanned-surface-vehicles-%3A-in-Aura/3768b5385dd142feba64c6f5bf7161d4eeda0ce0> (Accessed: 20 September 2022).

Carter, J. *et al.* (eds) (2021) *Fuzzy Logic: Recent Applications and Developments*. 1st ed. Cham: Springer International Publishing. Available at: <https://doi.org/10.1007/978-3-030-66474-9>.

Cummings, M. (2017) *Artificial Intelligence and the Future of Warfare*. Chatham House for the Royal Institute of International Affairs.

Ek, R. (2018) *Automatic Image Annotation Using Transfer Learning on Convolutional Neural Networks*. Master's Thesis. Åbo Akademi. Available at: <https://www.doria.fi/handle/10024/158799> (Accessed: 20 September 2022).

Ertel, W. (2017) *Introduction to Artificial Intelligence*. 2nd ed. Berlin, Heidelberg: Springer (Undergraduate Topics in Computer Science). Available at: <https://doi.org/10.1007/978-3-319-58487-4>.

Fawkes, A. and Menzel, M. (no date) 'The Future Role of Artificial Intelligence: Military Opportunities and Challenges', *Journal of the JAPCC*, 2018(27), pp. 71–77.

Ferdous, J. (2020) *Diagnosis of breast cancer using neural networks with different logistic regressions*. Master's Thesis. Åbo Akademi. Available at: <https://www.doria.fi/handle/10024/177454> (Accessed: 20 September 2022).

Figueira, J., Greco, S. and Ehrgott, M. (2005) *Multiple Criteria Decision Analysis, State of the Art Surveys*. 2nd ed. New York: Springer (International Series in Operations Research & Management Science, 233).

Frigui, H. and Gader, P. (2006) 'Detection and Discrimination of Land Mines based on Edge Histogram Descriptors and Fuzzy K-Nearest Neighbors', in *2006 IEEE International Conference on Fuzzy Systems. 2006 IEEE International Conference on Fuzzy Systems*, pp. 1494–1499. Available at: <https://doi.org/10.1109/FUZZY.2006.1681906>.

Gader, P.D., Keller, J.M. and Nelson, B.N. (2001) 'Recognition technology for the detection of buried land mines', *IEEE Transactions on Fuzzy Systems*, 9(1), pp. 31–43. Available at: <https://doi.org/10.1109/91.917112>.

Hyvönen, E. *et al.* (1993) *Tekoälyn ensyklopedia*. Helsinki: Gaudeamus : Suomen tekoälyseura (STeS).

Irfan, A. *et al.* (2020) 'Towards Verification of Neural Networks for Small Unmanned Aircraft Collision Avoidance', in *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*. 2020 IEEE/AIAA 39th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA: IEEE, pp. 1–10. Available at: <https://doi.org/10.1109/DASC50938.2020.9256616>.

Kazaryan, B.I. (2010) 'On Implementing Intellectualized Control Technology in Systems of Automated Command and Control of Troops and Weapons', *Military Thought*, 19(4), pp. 105–115.

Kelleher, J. (no date) 'Syväoppiminen: Kuinka tekoäly toimii', *Terra Cognita*. Available at: <https://www.terracognita.fi/tuote/syvaoppiminen/> (Accessed: 20 September 2022).

Kewley, R.H. and Embrechts, M.J. (1998) 'Fuzzy-genetic decision optimization for positioning of military combat units', in *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)*. SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218), pp. 3658–3663 vol.4. Available at: <https://doi.org/10.1109/ICSMC.1998.726634>.

Kosko, B. (1993) *Sumea logiikka. Fuzzy thinking*. 1. ed. New York: Hyperion.

Kumar, N. and Dixit, A. (2019) *Nanotechnology for defence applications*. Cham: Springer.

Lowen, R. and Roubens, M. (eds) (1993) *Fuzzy logic: state of the art*. Dordrecht: Springer (Theory and decision library Series D).

Mattila, J.K. (2002) *Sumean logiikan oppikirja: johdatusta sumean matematiikkaan*. 3. uus. p. Helsinki: Art House.

Merilehto, A. (2018) *Tekoäly: Matkaopas Johtajalle*.

Muranen, J. *et al.* (2003) ‘A Practical Study of Some Virtual Sensor Methods as Part of Data Fusion’, in V. Palade, R.J. Howlett, and L. Jain (eds) *Knowledge-Based Intelligent Information and Engineering Systems. KES 2003*, Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 1046–1051. Available at: https://doi.org/10.1007/978-3-540-45224-9_139.

Penttinen, R. (1994) *Meesaununin sumea säätö*. Diplomityö. Oulun Yliopisto, Sääätötekniikan laitos.

Penttinen, R. (1998) *A decision support system for water protection in agriculture*. Doctoral thesis. Oulun Yliopisto, Systemiteknikan laboratorio.

Penttinen, R. *et al.* (2000) ‘Rule-Based classification for a case-based system: 9th Finnish Artificial Intelligence Conference’, in *STeP 2000 - millennium of artificial intelligence. The 9th Finnish Artificial Intelligence Conference*, Finnish Artificial Intelligence Society FAIS (Publications of Finnish Society of Information Processing Science).

Rabkin, J.A. and Yoo, J. (2017) *Striking power: how cyber, robots, and space weapons change the rules for war*. First American edition. New York, New York: Encounter Books.

Ragsdale, D.J. *et al.* (1997) ‘A fuzzy logic approach for intelligence analysis of actual and simulated military reconnaissance missions’, in *Computational Cybernetics and Simulation. 1997 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2590–2595 vol.3. Available at: <https://doi.org/10.1109/ICSMC.1997.635325>.

Romanenko, I. *et al.* (2021) ‘Development of estimation and forecasting method in intelligent decision support systems’, *Eastern-European Journal of Enterprise Technologies*, 110(4), pp. 38–47. Available at: <https://doi.org/10.15587/1729-4061.2021.229160>.

Rutkowski, L. *et al.* (eds) (2020) *Artificial Intelligence and Soft Computing - 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12-14, 2020, Proceedings, Part I, Artificial Intelligence and Soft Computing - 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12-14, 2020, Proceedings, Part I*. Cham: Springer. Available at: <https://doi.org/10.1007/978-3-030-61401-0>.

Saarinen, J., Syrjärinne, J. and Lahtinen, T. (1998) *Sumeaan logiikkaan perustuva seuranta-algoritmi ilmatilannekuvan muodostamiseen*. Helsinki: Maanpuolustuksen tieteellinen neuvottelukunta.

Sánchez-Lozano, J.M. and Rodríguez, O.N. (2020) ‘Application of Fuzzy Reference Ideal Method (FRIM) to the military advanced training aircraft selection’, *Applied Soft Computing*, 88. Available at: <https://doi.org/10.1016/j.asoc.2020.106061>.

Sarin, R. and Weistroffer, H.R. (2016) ‘Introduction to Multi-criteria Decision Analysis and Support Systems Minitrack’, in *49th Hawaii International Conference on System Sciences (HICSS)*, pp. 1515–1516. Available at: <https://doi.org/10.1109/HICSS.2016.191>.

Siukonen, T. and Neittaanmäki, P. (2019) *Mitä tulisi tietää tekoälystä*. Jyväskylä: Docendo.

Svenmark, P. et al. (2018) *Possibilities and Challenges for Artificial Intelligence in Military Applications*. Stockholm: Swedish Defence Research Agency.

Tietoisuus, tekoäly ja robotit (2017). Art House.

Tudorica, A., Banacu, C.S. and Colesca, S.E. (2021) ‘A Literature Review Regarding The Application Of Multi-Criteria Analysis In Transport Infrastructure Projects’, *Management Research and Practice*, 13(2), p. 36+.

Vähäkainu, P. and Neittaanmäki, P. (2018) ‘Tekoäly terveydenhuollossa’, *Informaatioteknologian tiedekunnan julkaisuja / Jyväskylän yliopisto* [Preprint], (2018, 45). Available at: <https://jyx.jyu.fi/handle/123456789/57682> (Accessed: 20 September 2022).

Wang, Z. et al. (2011) ‘Intelligent scheduler design for networked control systems of guided weapon’, in 2011 6th IEEE Conference on Industrial Electronics and Applications. 2011 6th IEEE Conference on Industrial Electronics and Applications, pp. 2454–2459. Available at: <https://doi.org/10.1109/ICIEA.2011.5976005>.

TEKOÄLYSOVELLUSTEN JA ALGORITMIEN KYBERTURVAL- LISUUSRISKIT

Janne Mutanen

Jobdanto

Algoritmit ja tekoälyratkaisut muuttavat maailmaa. Nyt olemme nähneet vasta näiden ratkaisujen kehityskaaren alun. Vaikka jokapäiväisessä elämässämme huomaammekin algoritmien ja tekoälyn vaikutuksen, ovat nämä ratkaisut usein taustavoimia, jotka vaikuttavat organisaatioiden päätöksenteossa, käyttämisemme palveluissa ja laitteissa sekä teknologioissa monilla yhteiskunnan toiminnan osa-alueilla. Algoritmeista ja tekoälystä puhutaan niin tutkimuskirjallisuudessa kuin arkisessa keskustelussakin. Usein näitä sanoja käytetään kuvaamaan asioita, joiden kuvaamiseen asiantuntijat ja tutkijat eivät niitä välttämättä käyttäisi. Julkisessa keskustelussa algoritmi-sanalla viitataan usein esimerkiksi kaupallisten tai muiden toimijoiden tuottamien palveluiden taustalla vaikuttaviin lukuisiin algoritmeihin ja muihin ratkaisuihin sekä toimintamalleihin. Tekoäly-sanalla puolestaan saatetaan viitata johonkin tietojärjestelmään, joka ei välttämättä tekoälyn yleistä määritelmää täyttäisikään. Tämän artikkelin näkökulma on riskienhallinnallinen ja ennen kaikkea kyberturvallisuuteen liittyvä. Tämän vuoksi artikkelissa ei pyritäkään tekoälyn tai algoritmien puhdasoppiseen määrittelyyn tai ilmiöiden syvälliseen tarkasteluun. Pikemminkin tavoitteena on tarkastella näitä ilmiöitä kokonaisvaltaisena kehityssuuntana ja kiinnittää huomio niihin liittyviin riskeihin.

Algoritmit ja tekoäly

Algoritmin määritelmä

Lähtökohdan algoritmeihin liittyvien riskien tarkastelulle saamme sanakirjamäärittelystä (Merriam-Webster 2021), jonka mukaisesti voimme todeta algoritmin olevan vaiheittainen menettely jonkin ongelman ratkaisemiseksi tai johonkin lopputulokseen pääsemiseksi. Tässä määrittelyssä ”menettely” tarkoittaa yksityiskohtaisia ja selkeitä ohjeita siitä, miten jonkin asian on määrä tapahtua tai miten se on määrä suorittaa. Tällainen yksinkertaistettu määrittely vaikuttaa olevan kaukana tutkimuksesta, jossa arvioidaan esimerkiksi sosiaalisen median palveluiden käyttäjiä luokittelevia (Zarouali & Dobber 2020) tai pankin luottoriskejä arvioivia (Arora & Kaur 2020) algoritmeja. Toisaalta julkisella keskustelulla sosiaalisen median palveluiden tai hakukone- tai datayhtiöiden ”algoritmeista” ei äkkiseltään tunnu olevan merkittävästi kosketuspintaa algoritmin sanakirjamäärittelmään. Tosiasiassa kysymys on kuitenkin siitä, että esimerkiksi jokaisen organisaation toimintaprosessin tai digitaalisen palvelun taustalla vaikuttaa algoritmi tai tyypillisesti suuri joukko erilaisia algoritmeja, joista kullakin on oma tavoitteensa ja jotka yhdessä pyrkivät kohti jotakin suurempaa ja yhteistä tavoitetta.

Tekoälyn määritelmä

Myös tekoälyn määrittelyssä turvaudumme ensin sanakirjaan. Merriam-Websterin (2021) määritelmän mukaan tekoäly tarkoittaa koneen kyvykkyyttä jäljitellä ihmisen älykäästä toimintaa sekä tietojenkäsittelytieteen osa-aluetta, joka tarkastelee tätä ilmiötä. Kuten algoritmitkin, on myös tekoälysovellusten hyödyntäminen yleistymässä. Lääketieteen alalla tekoälyratkaisujen kehitys on vilkasta (Briganti & Le Moine 2020). Esimerkiksi kirurgiassa hyödynnetään koneoppimista, keinotekoisia neuroverkkoja, luonnollisen kielen käsittelyä sekä konenäköratkaisuja (Hashimoto et al. 2018). Psykiatrian tutkimuksessa edistetään koneälyratkaisuja, jotka perustuvat luonnollisen kielen automaattiseen käsittelyyn tai tekoälyn hyödyntämiseen erilaisten biomarkkereiden tarkastelussa häiriöiden luokitteluksi (Brunn et al. 2020).

Algoritmit ja tekoäly tässä artikkelissa

Tämän artikkelin, kyberturvallisuusriskien hallinnan ja yhteiskunnallisen keskustelun kannalta on mielekästä määritellä algoritmit ja tekoäly toisin, kuin vaikkapa näitä ilmiöitä koskevassa tutkimuskirjallisuudessa. Koska tämä artikkeli ei tarkastele algoritmeja, tekoälyä tai tekoälysovelluksia sinänsä, voidaan edellä mainitut asiayhteyttä varten määritellä esimerkiksi alla olevasti.

Algoritmit (arkikielessä) ovat ihmisen luomia järjestelmäkokonaisuuksia, jotka saavat syötteenä dataa ja pyrkivät syöttödataan pohjautuen pääsemään niille ennalta asetettuun tavoitteeseen. Algoritmit antavat tulosteena syöttödataan pohjautuvan ja tavoitteeseensa liittyvän tuloksen. (”Taikalaatikko”). Tekoäly sekä erilaiset tekoälysovellukset lisäksi toimivat autonomisesti (toimivat ilman ohjausta, oppivat, tekevät asioita, joita ei ole ennalta ohjelmoitu).

Kun edistyneitä algoritmeja ja tekoälyratkaisuja hyödynnetään, on erääksi keskeiseksi menestystekijäksi tunnistettu käyttäjän ymmärrys älykkään ratkaisun tietojenkäsittelyprosessista (Branley-Bell et al. 2020). Edellä todetaan algoritmien ja tekoälysovellusten olevan käyttäjän kannalta ”taikalaatikoita”, jotka tuottavat ratkaisuja ja vastauksia ongelmiin. Sovelluksen käyttäjän luottamuksen (sovellusta kohtaan) on todettu edesauttavan käyttökokemusta ja ratkaisun käytettävyyttä yleisesti (Shin 2020). Algoritmien ja ihmisten vuorovaikutuksen tutkimusta haistaa selkeästi se, että algoritmit ovat tyypillisesti suljettuja ja omistusoikeuden alaisia. Suljetun algoritmin (toiminnan ja tiedon käsittelyn periaatteet sekä tekninen toteutus eivät ole tiedossa) toimintaa sekä perusteita algoritmin tekemille ratkaisuille vuorovaikutustilanteissa on hyvin hankala selittää (Hargittai et al. 2020).

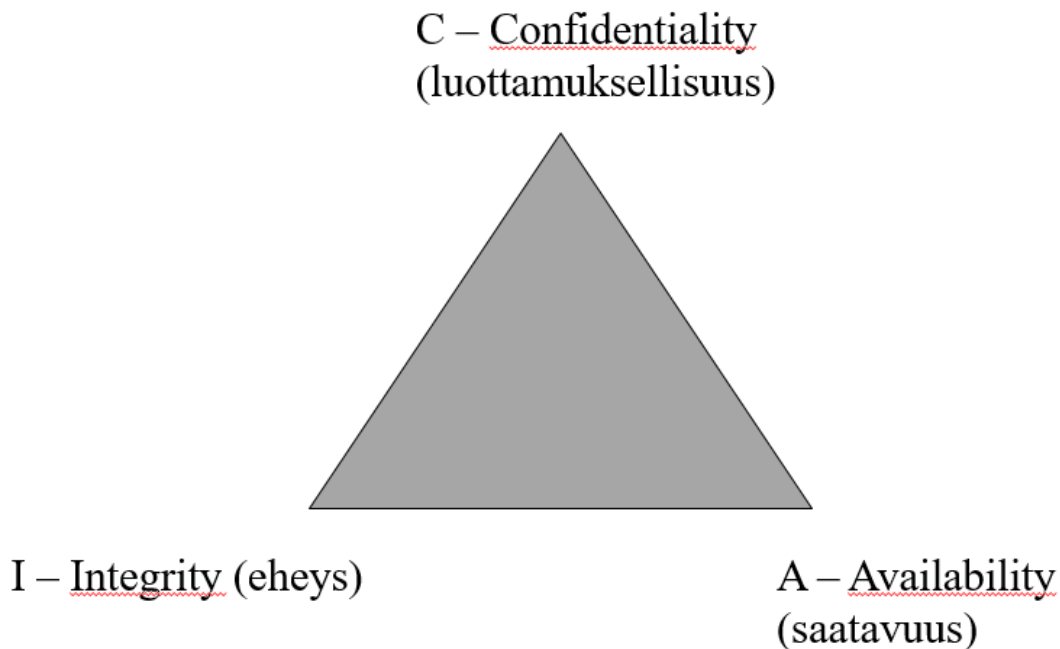
Tietoturvallisuuden CIA-malli

Tietoturvallisuuden peruskäsite määritellään usein CIA-mallin (CIA-kolmio, CIA triad) kautta (esimerkiksi Chai 2021). Termi on vakiintunut käyttöön tietoturvallisuutta koskevassa kirjallisuudessa, mutta sen historia on jokseenkin epäselvä. Mallin mukaan tietoturvallisuus rakentuu kolmesta komponentista – luottamuksellisuudesta (confidentiality), eheydestä (integrity) sekä saatavuudesta (availability). Todellisuudessa

esimerkiksi organisaation tai yksilön tietoturvapäättösten taustalla on riskien arviointi sekä painotukset edellä mainittujen tietoturvallisuuden komponenttien suhteen ja kesken. CIA-mallin määritelmän mukainen tietoturvallisuus on aina tilanne- ja asiayhteyssidonnainen asia. Mallin mukaisesti ei kuitenkaan ole mahdollista arvioida tietyn tilanteen tai asiayhteyden absoluuttista tietoturvallisuutta, vaan tietoturvallisuus perustuu em. komponenttien välillä tapahtuvaan optimointiin riskienhallinnallisin perustein.

Tietoturvallisuudella tarkoitetaan sitä, että tieto on eheää, luottamuksellista sekä saatavilla.

- Eheys: Tieto ei muutu ilman asianmukaisin valtuuksin tehtyjä toimenpiteitä.
- Luottamuksellisuus: Tietoa voivat käsitellä vain kyseisen tiedon käsittelyyn oikeutetut tahot.
- Saatavuus: Tieto on (oikeutettujen tahojen) saatavilla silloin kuin sitä tarvitaan.



Kuva 1. Tietoturvallisuuden CIA-malli

Kyberturvallisuudella tarkoitetaan digitaalisen ja verkottuneen toimintaympäristön turvallisuutta.

Mikäli esimerkiksi sairaala hallinnoi potilastietoja hakkaamalla ne kivitauluun ja tallentamalla kivitaulut kassakaappiin vahvojen lukkojen taakse, on niiden eheys ja luottamuksellisuus hyvällä tasolla, mutta tietojen saatavuus olisi potilaan hoitoa ajatellen heikkoa. Jos taas tiedot tallennetaan avoimelle ja julkiselle verkkosivulle, jonka palvelimia ylläpidetään korkean saatavuuden käytäntöjen mukaisesti, on tietojen saatavuus riippuvainen potilaita hoitavien ammattilaisten päätelaitteista ja niiden verkkoyhteyksistä. Eheys on tässä asetelmassa palvelinympäristön valvonnasta riippuvaista, mutta luottamuksellisuus olisi surkealla tasolla. Käytännössä tietoturvallisuutta kehitetään ja ylläpidetään pyrkimällä riskien analysoinnin perusteella tarkoituksenmukaiseen painotukseen luottamuksellisuuden, eheyden ja saatavuuden välillä.

CIA-malli pätee tietoturvallisuuden kaikissa asiayhteyksissä. Joissakin asiayhteyksissä ”taho” ei tarkoita ihmistä, vaan esimerkiksi sovellusta tai laitetta. ”Tiedolla” voidaan tarkoittaa vaikkapa anturin lähettämää mittausdataa. Esimerkiksi IoT-laitteiden (Internet of Things, esineiden internet) tietoturvallisuus on usein heikolla tasolla, koska monien arkisten fyysisen maailman laitteiden (kodinkoneet, ajoneuvot, kiinteistöt) tapauksessa ei ole osattu riittävästi huomioida, että niiden käyttöön liittyisi ”tietoa”, jonka ”luottamuksellisuudesta” tai ”eheydestä” olisi pitänyt huolehtia.

Algoritmien ja tekoölyn riskienhallintanaökökulma sekä keskeiset riskit

ISO/ISC 27005 -standardin (2018) lähtökohta on tieto-omaisuuseriin (asset) liittyvä riskienhallinta. Näitä omaisuuseriä voivat olla mitkä tahansa asiat, joilla on tietoarvoa organisaation kannalta tai jotka ovat merkityksellisiä organisaatiolle tärkeän tiedon kannalta. Näitä omaisuuseriä voivat olla esimerkiksi henkilöt, laitteet, dokumentit, ohjelmistot, infrastruktuuri tai data. Jokaisen omaisuuserän omistaja on tunnistettava, jotta voidaan määrittää omaisuuserän hallinnointiin liittyvät toimijat sekä näiden vastuut sekä tilivelvollisuus. Algoritmi tai tekoölysovellus voidaan nähdä standardin määritelmän mukaisesti omaisuuseräksi, jolla on oltava vastuullinen omistaja ja jonka riskejä tulee hallita. Tietojärjestelmän tai tiedon (kuten algoritmin tai tekoölysovelluksenkin) omistaja vastaa siitä, että kyseisen omaisuuserän luottamuksellisuuden, eheyden ja saatavuuden turvaamiseksi asetetaan asianmukaiset tietoturvakontrollit (NIST 2002).

Tekoölyä ja algoritmeja koskevat informaatioteknologiset lainalaisuudet sekä niihin liittyvät riskit

Kuten informaatioteknologiset järjestelmät yleisesti, ovat algoritmit ja tekoölyratkaisutkin alttiita häiriöille. Informaatioteknologisia järjestelmiä vastaan on mahdollista hyökätä maantieteellisestä sijainnista riippumatta. Kuten tänä päivänä tiedämme, on mahdollista aiheuttaa vakaviakin häiriöitä järjestelmiin puhtaasti etähyökkäyksellä internetin kautta (esimerkiksi E-ISAC, 2016). Järjestelmän hyökkäyspinta-alaan vaikuttavat esimerkiksi järjestelmän tulo- ja lähtöpisteet, kanavat sekä ei-luotettu data, jota järjestelmä tulostaa tai jota siihen syötetään. Joidenkin järjestelmän osien vikaantuminen tai virhetilanteet voivat aiheuttaa potentiaalisesti suurempia vahinkoja kuin joidenkin toisten. Jotkin järjestelmän komponentit voivat olla kriittisiä järjestelmäkokoisuuden toiminnan kannalta (Pratyusa & Wing 2010). Järjestelmän useat osat voivat olla osa sen hyökkäyspinta-alaan. Ohjelmakoodi, rajapinnat, palvelut, protokollat ja käyttäjärajapinnan tapahtumat vaikuttavat hyökkäyspinta-alaan (Theisen et al 2018). Kun järjestelmän kompleksisuus kasvaa tai kun erillisiä järjestelmiä yhdistetään suuremmiksi kokonaisuuksiksi, kasvaa myös hyökkäyspinta-ala. Myös tietojärjestelmien verkottuneisuus lisää hyökkäyspinta-alaan ja mahdollistaa edellä mainitun etähyökkäämisen.

Ohjelmiston laadussa on yleisesti haasteita tietoturvallisuuden kannalta. Hallitsematon avoimen lähdekoodin käyttö, päivitysmenettelyiden puute, turvattomat oletusasetukset sekä ohjelmistokehityksen menetelmien ohittaminen aiheuttavat turvallisuusriskejä valmiiseen ohjelmistoon (Murerwa 2018). Toisaalta tekoölyratkaisuja kehitetään usein

ilman asianmukaista harjaantumista ohjelmistotyön laukukäytäntöihin tai yleisiin ohjelmistotuotannon menetelmiin. Tekoälysuunnittelu voidaankin nähdä kokonaan uutena ammatillisena osa-alueena tai oppiaineena, jonka menetelmät ovat kehityskaarensa alussa (Lenarduzzi et al, 2021).

Algoritmien ja tekoälyratkaisujen häiriöherkkyyteen vaikuttaa järjestelmäkokonaisuuksien monimutkaisuus. Teknologia yleisesti (kuten algoritmit ja tekoälykin) kehittyvät toisaalta evolutiivisesti, toisaalta yhdistelemällä teknologioita (Arthur, 2009). Tämä tekee kehittyneistä järjestelmistä helposti erittäin monimutkaisia. Koska tietokoneet ovat luonteeltaan monikäyttöisiä (ohjelmoitavissa) ja verkottuneita (tietoliikenneverkkojen kautta), voi ongelmatilanne järjestelmäkokonaisuuden yhdessä osassa levitä helposti muihin osiin. Siten tietojärjestelmien ongelmatilanne on tyypillisesti kompleksinen ja hämmentävä. Tilanteen selvitykseen liittyy niin teknisiä kuin inhimillisiäkin ulottuvuuksia (Davis et al 1992). Tilanteesta selviämiseen tarvitaan vaihtoehtoisia (tietojärjestelmistä riippumattomia) toimintatapoja, sillä tyypillisessä tietojärjestelmäpoikkeamassa järjestelmä on käyttökelvoton (Russell et al 2014).

Tekoälyjärjestelmän tai algoritmin erityispiirre on se, että järjestelmä voi itsessään toimia teknisesti, mutta sovelluksen tuottamat tulokset voivat olla virheellisiä syöttödatasta johtuen. Tämä voi jäädä ratkaisua käyttävältä henkilöltä huomaamatta, mikäli järjestelmä toimii näennäisen oikein. Lisäksi järjestelmä voi antaa vähäisen syöttödatan muuntumisen (tahallisesti tai tahattomasti) seurauksena ihmisen tulkinnan kannalta täysin virheellisiä vastauksia (Goodfellow et al 2014, Papernot et al 2016). Kun algoritmia tai tekoälysovellusta käytetään tilanteessa, jossa tulosten arviointi asiantuntijoiden toimesta on hankalaa tai mahdotonta, on kysymys luottamuksesta teknologiaa kohtaan. Sokea luottamus aiheuttaa riskin (Hurlburt 2017). Toisaalta keinoälyratkaisuissa on tyypillisesti uudenlaisia ominaisuuksia verrattuna perinteisiin teknologioihin, joten luottamus ratkaisua kohtaan muodostuu hyvin eri tavalla, kuin perinteistä ja helpommin ymmärrettävää teknologiaa kohtaan (Siau & Wang 2018).

Datan laatuun ja eheyteen liittyvät riskit

Kuten algoritmi tai tekoälysovellus, myös data, jolla oppivaa algoritmia tai keinoälyratkaisua opetetaan, on omaisuususerä, jonka omistamiseen ja hallintoihin liittyy riskejä, joita täytyy hallita. Riskienhallinnan kannalta on haastavaa tunnistaa vastuutahot sellaisissa tilanteissa, joissa yhdistellään omaisuususeriä tai jotka muokkautuvat (oppivat) erilaisten omaisuususerien yhdistelyn myötä. Datan hallinnassa on tyypillistä se, että erilaisia ja eri lähteistä peräisin olevia dataeriä yhdistellään uusiksi datakokonaisuuksiksi, joilla voi olla täysin uusia käyttötarkoituksia. Tällaisen uuden datakokonaisuuden omistajuutta (ja vastuutahoa) voi olla vaikea määrittellä. Yhdistelyn myötä syntyneen uuden dataomaisuuserän vastuiden ja omistajuuden määrittely voi johtaa tilanteeseen, jossa alkuperäisen datan omistaja ei voi enää vaikuttaa siihen, miten hänen omistaansa dataa hallinnoidaan osana uutta dataomaisuuserää.

Koska tekoälyratkaisut ja algoritmit ovat usein suljettuja ja omistusoikeuden alaisia, niiden toimintaan liittyvät algoritmiset harhat voivat jäädä huomaamatta. Algoritmisella harhalla tarkoitetaan sitä, että algoritmiin (erityisesti oppivaan algoritmiin) raken-

netaan sisään ympäröivän yhteiskunnan tai algoritmin laatijoiden oletuksia ja sosiaalisia kaavoja. Algoritmin tekemät päätökset tietyissä valintatilanteissa heijastelevat algoritmin laatijoiden käsityksiä ihanteellisista päätöksistä. Algoritmisen harhan taustalla ovat ne asenteet ja sosiaaliset rakenteet, jotka piilevät datassa, jolla oppivaa algoritmia opetetaan. (Johnson 2020). Datan opettama algoritmi ei välttämättä tee asiayhteydessä optimaalisia ratkaisuja, vaan se toimii tavalla, joka mukailee datamassasta ilmi käyviä toimintatapoja. Esimerkiksi rikoksia ennustavissa algoritmeissa on havaittavissa selkeästi sekä ennako-oletusten että taustadatan vaikutus (Mayson 2019, Silva & Kenney 2019). Toisaalta on huomioitava, että itse algoritmi tai tekoälysovellus ei sinänsä välttämättä aiheuta vääristymää, vaan kysymys on juuri datasta, jolla algoritmia tai tekoälysovellusta opetetaan ja joka tuo mukanaan menneisyyden vääristymiä. (Bagaric et al. 2019, Danks & London 2017). Riippumatta algoritmin (tai koneoppivan ratkaisun) mahdollisesta algoritmisesta harhasta, on kuitenkin tyypillistä, että algoritmin toimintaperiaate käyttäjälle tuntematon ("taikalaatikko"). Diagnoosia tekevä lääkäri tai partioajelua suunnitteleva poliisi joutuu oman asiantuntemuksensa ja taustatietojensa lisäksi luottamaan algoritmiin tai koneälyratkaisuun, joka kertoo ajatuksia, mutta ei tyhjentävästi perustele niitä.

Yksityisyyteen ja lainsäädäntöön liittyvät riskit

EU:n yleinen tietosuojasetus (GDPR, General Data Protection Regulation) asettaa rekisterinpitäjälle vastuita ja velvollisuuksia henkilötiedon käsittelyn osalta. Tietosuojasetuksen 5 artiklan mukaisia tietosuojaperiaatteita (alla) on noudatettava henkilötietojen käsittelyssä. (GDPR, 5 artikla).

- Lainmukaisuus, kohtuullisuus ja läpinäkyvyys
- Käyttötarkoitussidonnaisuus
- Tietojen minimointi
- Täsmällisyys
- Säilytyksen rajoittaminen
- Eheys ja luottamuksellisuus

Tietosuojasetus takaa rekisteröidylle omiin tietoihinsa oikeuksia (alla), jotka on huomioitava myös algoritmien ja tekoälyratkaisujen suunnittelussa sekä datan hyödyntämisessä henkilötietojen osalta. (GDPR, 13-22 ja 34 artikla).

- Saada läpinäkyvää tietoa (tietojen keräys suoraan rekisteröidyltä tai muualta)
- Saada pääsy tietoihin
- Vaatia tietojen korjaamista
- Tulla unohdetuksi
- Käsittelyn rajoittamiseen
- Luottaa, että rekisterinpitäjä ilmoittaa oikaisuista, poistoista ja rajoituksista luovutettujen tietojen vastaanottajille
- Siirtää tiedot järjestelmästä toiseen
- Vastustaa tietojen käsittelyä

- Vastustaa automatisoituja yksittäispäätöksiä ml. profilointi (oikeus olla joutumatta sellaisen päätöksen kohteeksi, joka perustuu pelkästään automaattiseen käsittelyyn, kuten profilointiin)
- Tulla informoiduksi tietoturvaloukkauksista

Näiden oikeuksien toteutuminen voi käytännössä olla haastavaa, jos henkilötietoa on yhdistelty tai jos sitä on hyödynnetty tekoälyratkaisun opettamiseen. Pääsy tietoihin tai tietojen korjaaminen voi olla mahdotonta järjestää. Mikäli tieto on luovutettu eteenpäin (mahdollisesti kolmansille osapuolille), sen poistaminen tai tietojen käsittelyn rajoittaminen voi olla mahdotonta. Lisäksi joidenkin algoritmien tai tekoälyratkaisujen nimenomainen tarkoitus voi olla automatisoitujen yksittäispäätösten tekeminen.

Kyberturvallisuuteen liittyvän lainsäädännön laatiminen on haastavaa. Kyberturvallisuuden säätely edellyttää laaja-alaista (hallinnonalojen rajat ylittävää) yhteistyötä. Toisaalta kyberturvallisuuden kenttä on kompleksinen ja alati muuttuva. Tämä voi olla haaste, sillä usein lainsäädäntötyössä ei ole riittävää asiantuntemusta kokonaisuuden hahmottamiseen. On riski, että lainsäädäntöä ohjaavat tiedettyä enemmän poliittiseen päätöksentekoprosessiin liittyvät asiantuntijat – esimerkiksi julkisen hallinnon asiantuntijat, lobbaajat ja liike-elämä. (Grant et al 2010). Kyberturvallisuuteen ja informaatioteknologiaan liittyvää lainsäädäntöä on runsaastikin, mutta teknologian nopea kehitys aiheuttaa merkittäviä haasteita lainsäädännön ajantasaisuudelle (Galliess & Baumgarten 2020). Suomessa laissa terveydenhuollon laitteista ja tarvikkeista (720/2021) laitteen käyttötarkoituksella tarkoitetaan ”käyttöä, johon lääkinällinen laite valmistajan merkinnöissä, käyttöohjeessa tai myyminenestämistä koskevassa aineistossa annettavien tietojen mukaan on tarkoitettu”. Lainsäädännössä käyttötarkoituksenmukaisuus asettaa merkittäviä vastuita laitteiden valmistajille. Tämä voi olla haastavaa etenkin silloin, kun laite on lähtökohtaisesti rakennettu oppimaan ja kehittymään.

Lähteet

Analysis of the Cyber Attack on the Ukrainian Power Grid (2016). Electricity Information Sharing and Analysis Center, E-ISAC.

Arora, N. and Kaur, P.D. (2020) ‘A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment’, *Applied Soft Computing*, 86, p. 105936. Available at: <https://doi.org/10.1016/j.asoc.2019.105936>.

Arthur, W.B. and Pietiläinen, K. (2010) *Teknologian luonne: mitä se on ja millainen on sen evoluutio*. Helsinki: Terra Cognita.

Bagaric, M., Hunder, D. and Stobbs, N. (2020) ‘Erasing the Bias Against Using Artificial Intelligence to Predict Future Criminality: Algorithms are Color Blind and Never Tire’, *U. Cin. L. Rev.*, 88(4), pp. 1036–1081.

Branley-Bell, D., Whitworth, R. and Coventry, L. (2020) ‘User Trust and Understanding of Explainable AI: Exploring Algorithm Visualisations and User Biases’, in *Human-Computer Interaction. Human Values and Quality of Life. HCII 2020*, Cham: Springer International Publishing (Lecture Notes in Computer Science, vol 12183), pp. 382–399.

Briganti, G. and Le Moine, O. (2020) ‘Artificial Intelligence in Medicine: Today and Tomorrow’, *Frontiers in Medicine*, 7(27). Available at: <https://doi.org/10.3389/fmed.2020.00027>.

Brunn, M. et al. (2020) ‘The Future is Knocking: How Artificial Intelligence Will Fundamentally Change Psychiatry’, *Academic Psychiatry: The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, 44(4), pp. 461–466. Available at: <https://doi.org/10.1007/s40596-020-01243-8>.

Calliess, C. and Baumgarten, A. (2020) ‘Cybersecurity in the EU The Example of the Financial Sector: A Legal Perspective’, *German Law Journal*, 21(6), pp. 1149–1179. Available at: <https://doi.org/10.1017/glj.2020.67>.

Chai, W. (no date) *Confidentiality, Integrity and Availability (CIA Triad)*. Available at: <https://www.techtarget.com/whatis/definition/Confidentiality-integrity-and-availability-CIA> (Accessed: 11 September 2021).

Danks, D. and London, A. (2017) ‘Algorithmic Bias in Autonomous Systems’, *IJCAI*, 17, pp. 4691–4697. Available at: <https://doi.org/10.24963/ijcai.2017/654>.

Davis, G.B. et al. (1992) ‘Diagnosis of an information system failure: A framework and interpretive process’, *Information & Management*, 23(5), pp. 293–318. Available at: [https://doi.org/10.1016/0378-7206\(92\)90059-O](https://doi.org/10.1016/0378-7206(92)90059-O).

Euroopan parlamentin ja neuvoston asetetus (EU) 2016/679, annettu 27 päivänä huhtikuuta 2016, luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta (yleinen tietosuojasäätös) (ETA:n kannalta merkityksellinen teksti) (2016) OJ L. Available at: <http://data.europa.eu/eli/reg/2016/679/oj/fin> (Accessed: 21 September 2022).

Goodfellow, I.J., Shlens, J. and Szegedy, C. (2015) ‘Explaining and Harnessing Adversarial Examples’. arXiv. Available at: <https://doi.org/10.48550/arXiv.1412.6572>.

Grant, J. (2010) ‘Will There Be Cybersecurity Legislation?’, *Journal of National Security Law & Policy*, 4(1), pp. 103–118.

Hargittai, E. et al. (2020) ‘Black box measures? How to study people’s algorithm skills’, *Information, Communication & Society*, 23(5), pp. 764–775. Available at: <https://doi.org/10.1080/1369118X.2020.1713846>.

Hashimoto, D.A. *et al.* (2018) ‘Artificial Intelligence in Surgery: Promises and Perils’, *Annals of Surgery*, 268(1), pp. 70–76. Available at: <https://doi.org/10.1097/SLA.0000000000002693>.

Hurlburt, G. (2017) ‘How Much to Trust Artificial Intelligence?’, *IT Professional*, 19(4), pp. 7–11. Available at: <https://doi.org/10.1109/MITP.2017.3051326>.

ISO Information Technology-Security Techniques-Information Security Risk ISO/IEC 27005:2018 (2018). Available at: <https://www.iso.org/cms/ren-der/live/en/sites/isoorg/contents/data/standard/07/52/75281.html> (Accessed: 21 September 2022).

Johnson, G.M. (2021) ‘Algorithmic bias: on the implicit biases of social technology’, *Synthese*, 198(10), pp. 9941–9961. Available at: <https://doi.org/10.1007/s11229-020-02696-y>.

Kirkpatrick, K. (2017) ‘It’s not the algorithm, it’s the data’, *Communications of the ACM*, 60(2), pp. 21–23. Available at: <https://doi.org/10.1145/3022181>.

Lenarduzzi, V. *et al.* (2021) ‘Software Quality for AI: Where We Are Now?’, in D. Winkler *et al.* (eds) *Software Quality: Future Perspectives on Software Engineering Quality. International Conference on Software Quality*, Cham: Springer International Publishing (Lecture Notes in Business Information Processing), pp. 43–53. Available at: https://doi.org/10.1007/978-3-030-65854-0_4.

Manadhata, P.K. and Wing, J.M. (2011) ‘An Attack Surface Metric’, *IEEE Transactions on Software Engineering*, 37(3), pp. 371–386. Available at: <https://doi.org/10.1109/TSE.2010.60>.

Mayson, S.G. (2018) ‘Bias In, Bias Out’, *Yale Law Journal*, 128, pp. 2218–2301.

Murerwa, E.M. (2018) ‘Effect of poor software development practices on computer software security’, *International Journal of Education and Research*, 6(10), pp. 15–20.

Papernot, N. *et al.* (2017) ‘Practical Black-Box Attacks against Machine Learning’, in *Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security, Abu Dhabi, UAE*, pp. 506–519. Available at: <https://doi.org/10.48550/arXiv.1602.02697>.

Russell, K., Maennle, D. and Higgins, M. (2014) ‘Planning for Laboratory Functioning After Information System Crash’, *American Journal of Clinical Pathology*, 142(suppl_1), pp. A199–A199. Available at: <https://doi.org/10.1093/ajcp/142.suppl1.199>.

Shin, D. (2020) ‘How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance’, *Computers in Human Behavior*, 109, p. 106344. Available at: <https://doi.org/10.1016/j.chb.2020.106344>.

Siau, K. and Wang, W. (2018) ‘Building Trust in Artificial Intelligence, Machine Learning, and Robotics’, *Cutter Business Technology Journal*, 31(2), pp. 47–53.

Silva, S. and Kenney, M. (2019) ‘Algorithms, platforms, and ethnic bias’, *Communications of the ACM*, 62(11), pp. 37–39. Available at: <https://doi.org/10.1145/3318157>.

Stoneburner, G., Goguen, A. and Feringa, A. (2002) *Risk Management Guide for Information Technology Systems*. NIST Special Publication (SP) 800-30. National Institute of Standards and Technology. Available at: <https://doi.org/10.6028/NIST.SP.800-30>.

Theisen, C. *et al.* (2018) ‘Attack surface definitions: A systematic literature review’, *Information and Software Technology*, 104, pp. 94–103. Available at: <https://doi.org/10.1016/j.infsof.2018.07.008>.

Zarouali, B. *et al.* (2020) ‘Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media’, *Communication Research* [Preprint]. Available at: <https://doi.org/10.1177/0093650220961965>.

TEKOÄLY OPPIMISEN TUKENA – KEHITYKSEN VAIHEITA JA NÄKÖKULMIA

Antti Rissanen

Jobdanto

Tekoälyyn perustuvia järjestelmiä löytyy lähes kaikilta elämänalueilta. Pidämme joitakin toteutuksia niin luontevina ja integroituneina arjen toimintoihin, ettemme pysähdy pohtimaan, miten ennen tulimme toimeen ilman näitä innovaatioita. Tässä teemaluvussa tarkastellaan tekoälyn suhdetta oppimiseen ja koulutukseen. Kirjallisuudessa ja erilaisissa blogikirjoituksissa on tätä suhdetta käsitelty kunkin kirjoittajan omista lähtökohdista. Useat kirjoittajat lähestyvät aihealuetta mahdollisuutena ja saattavat valikoiden viitata erilaisiin muistioihin, joita kirjoittamisen ammattilaiset ovat vaikkapa EU:n hallinnolle tuottaneet. Suoranaista laaja-alaista tutkimusta, joka sitoisi teeman osat yhteen, on suppeasti. Koulutus nähdään yhtenä keskeisistä tavoista saada menestystä niin yksilönä kuin kansakuntana. Perinteinen paikkaan, opettavaan henkilöstöön, kirjalliseen aineistoon tai laboratorioihin perustuva opetus on kallista, mutta myös tuloksiltaan ennakoitavissa. Oppikirjakirjojen kustannustointiminta on ollut viime vuosiin kirjakustantamojen tukijalka. Monet ideat uusista digitalisaatioon perustuvista oppimateriaaleista ja niitä hyödyntävistä opetustapahtumista ovat olleet EU-tasoisia monikansallisia tutkimushankkeita. Rahoituksen luonteesta johtuen raportit saattavat osaltaan korostaa tekoälyä hypenä, jonka rahoitukseen myös opetussektorin tulisi osallistua. Näin siitäkin huolimatta, että itse tutkimuksesta tai siitä johdetuista tuotteista ei olisi suurta hyötyä opiskelijalle tai opettajalle.

Tekoäly oppimisessa

Teknisenä, toteutettavissa olevana käsitteenä tekoäly (Artificial Intelligence) on mainittu ensimmäisen kerran vuonna 1956 järjestetyn konferenssin nimessä. Konferenssikutsun alussa mainitaan myös opetuksellinen näkökulma (Moor 2006). NATO:n symposiumissa 1984 O'Shea ym. arvioivat tietokoneavusteinen opetuksen olevan hyvä valinta asiantuntijajärjestelmäsovelluksiin. Artikkelissaan he kuvaavat, miten rakentaa suunnittelutyökaluja, joiden avulla opettajat voivat muotoilla oman tietokoneavusteisen järjestelmän ilman, että käyttäjillä olisi "erityisiä ohjelmointitaitoja korkealaatuisten tietokonepohjaisten koulutusmateriaalien luomiseksi". Mallissa on viisi komponenttia, joita voidaan muokata ja vaihdella todellisen sovelluksen ja annettavan ohjeen tavan mukaan. Osat ovat: (1) Opetusjärjestelmän valvoja, (2) Opiskelijahistoria, (3) Opiskelijamalli, (4) Opetusstrategia ja (5) Opetusgeneraattori. (O'Shea ym. 1984)

Vuonna 2020 alkaneen koronapandemian yhteydessä havahduttiin, että digitalisaation mahdollisuuksista vain pieni osa oli poimittu luokkaopetukseen. Ehkä muutoksen tarvetta oli aiemmin tarkasteltu siten, että laitetekniikka ja tietoliikenteen mahdollisuudet oli haluttu integroida tuttuun opettajakeskeiseen toimintaan koulurakennuksessa. Etäopetuksen alettua opetusportaaleja muutettiin jo osin unohtuneiden kirjeopistojen kaltaisiksi digiopistoiksi. Kun palaamme syksystä 2022 uudestaan kouluihin, niin

tekoäly ei näy (edelleenkaan) robottina luokassa, tai oppilaan kotona, vaan se on tehokas työkalu, joka on osa monia opetukseen ja oppimisen tukeen liittyviä sovelluksia. Siksi yksittäiset algoritmit tai niitä toteuttavien komponenttien suorituskyky ei ehkä tuota suurta ahaa elämystä, mutta yhdessä ne luovat Industry 4.0:ksi kutsutun toimintaympäristön (esim. Rojko, 2017). Useiden erilaisten ja rinnakkaisten teknologioiden ansiosta opiskelu ei ole sidottu paikkaan eikä aikaan, mutta toisaalta käyttäjällä olisi oltava sekä soveltuva päätelaite, että riittävän tasokas nettiyhteys. Taito enenevään itseohjautuvuuteen ja mahdollisesti motivaation vaihtelusta johtuvat vaikeudet ovat yksilön oppimisen haasteita.

Opetuksen yhteydessä tarjotaan opiskelijalle mahdollisuus hankkia uutta tietoa, ymmärrystä tiedosta, käyttäytymistä tukevia malleja, taitoja, arvoja ja asenteita. Opetus ja oppiminen eivät ole sama asia, sillä ne kuvaavat vuorovaikutusta eri osallistujien näkökulmasta. Opetussuunnitelma on kuvaus siitä mitä osallistujalle pyritään tarjoamaan ja erityisesti, miten se toteutetaan. Oppija voidaan altistaa opetukselle, mutta muutos tiedoissa tai taidoissa on kunkin oppijan oma yksilöllinen tila. Oppimisen tuottamat muutokset arvoihin ja asenteisiin kestävät pitkään. Toisaalta tutkimukset ovat osoittaneet joidenkin taitojen ja myös muistinvaraisten tietojen heikkenevän nopeasti, jos niitä ei aktiivisesti ylläpidetä. Ihmisen oppiminen alkaa jo varhaisella iällä vuorovaikutuksella ympäristön kanssa. Vuorovaikutteista etäopetusta on pidetty uutena innovaationa. Toisaalta Australian harvaan asutuilla alueilla etäopiskelu oli ainut mahdollisuus, minkä avulla alakoululaiset pystyivät kotona asuen osallistumaan ohjattuun opetukseen. Alkeellisista VHF-yhteyksistä ja lähes yksilötason ohjauksesta on jo siirrytty pois. Etäisyys ei nykytekniikalla ole ongelma, sen sijaan yksilötasoinen ohjaava opetus on entistä kalliimpaa. Varhainen panostus opetus-portaaleihin ja etäopetuksen digitalisaatioon on Australiassa saanut syötteitä myös yhteisön yleisemmästä mielenkiinnosta oppimiseen. Moodle on omalla sarallaan australialaisen tietokoneavusteisen opetuksen helmi. Lähes ilmaisena ja modulaarisena järjestelmänä se antaa hyvät edellytykset opetusportaalle, johon voidaan tuoda aineistojen analyysiä algoritmien avulla.

Algoritminen ajattelu

Kirjallisuushavaintojen perusteella tekoälyn opetuskäyttö on pohjimmiltaan keskittynyt tietoon perustuvaan lähestymistapaan. Tämä ei tarkoita simulaattorien ja virtuaalitodellisuuden perustuvien ratkaisujen väheksymistä. Niiden käyttö on massaopetukseen keskittyneiden opetuksen tutkijoiden arjen ulkopuolella ja oman alansa kehitys urassa. Nämä järjestelmät korostavat yksilön aitoa soveltavaa osaamista. Siksi ne ovat harjoituskeskeisiä ja oppimista todentavia toteutuksia. Simulaattoriin tai virtuaalitodellisuuden laboratorioon mennään lähes aina hyvin valmistautuneina, soveltamaan tietoja ja taitoja, joita on jo muutoin omaksuttu.

Vielä vuosituhanen vaihteessa suppean tiedeyhteisön ulkopuolella tekoälyn mahdollisuuksiin suhtauduttiin skeptisesti. Mm Osmo Wiio näki kehityksen keskittyneen aasialaisten kielten tietokoneella kirjoittamiseen liittyvien ongelmien ratkaisuun. Toisaalta jo vuonna 2005 DARPA:n kilpailussa autonominen ajoneuvo oli haastekilpailun pääteema (DARPA 2014) (Montemerlo 2006). Modernin simuloinnin tai tietojenkäsittelytieteen ulkopuolella työskentelevät ihmiset mieltävät tekoälyn ilmenevän

parhaiten tai ainakin osin fyysisinä robotteina, jotka tekevät perinteiset fyysisistä suorituskykyä tai vähäistä koulutusta vaativat ammatit tai niiden keskeisimmät työnkuvat turhiksi.

Entä voisimmeko siirtyä yleiseurooppalaiseen oppimateriaaliin? Nykyisin EU:n regulaatio edellyttää kulutustuotteille kansallisia, siis meillä suomen- ja ruotsinkielisiä dokumentteja, joissa ohjataan kuluttajaa oikeaan käyttötapaan. Esimerkkeinä voi mainita vaikkapa seuraavat vihkoset: käyttöohje, varoitukset väärästä käytöstä, takuu ja sen ehdot, tai turvallisen hävittämisen dokumentti. Nykyään, kun terminologia on vakiintunut, konekääntäjät hoitavat tehokkaasti kansallisten ohjeiden teon. Esimerkiksi Microsoftin suomenkieliset dokumentit ovat kattavasti laadukkaita konekäännöksiä. Varovaisuusperiaatteen mukaan käyttäjää varoitetaan mahdollisista konekäännöksen tuottamista riskeistä. Hyvä ja selkokielinen alkuperäinen teksti tai artikulaatio osana puhetta on tarpeen, jotta konekäännöksessä alkuperäinen viesti tulisi tunnistetuksi. Kun Minecraft pelistä innostunut tarhaikäinen lapsi haluaa edetä ja saada lisää ominaisuuksia peliinsä, on avainsanan oikea ääntäminen opittava yrittämällä. Puheen tunnistaminen on käytössä monissa sovellutuksissa. Maailman valtakieliä osaavat Alexa ja Siri ovat hyviä esimerkkejä puheen tunnistuksesta yhdistettynä tiedon louhintaan. Jos pyyntö tunnistetaan oikein, niin ne usein vastaavat luontevasti kysyjälle. Kansallinen kulttuuri osana koulumaailmaa ja myös kielikuvien merkitys ovat liian isoja haasteita isolle integraatiolle oppivelvollisuuskoulutuksessa, mutta yliopistot voisivat jopa hyötyä tietyistä kehitysaskelista opetusmateriaalin tuottamisessa.

Tekoälyn toteuttamisesta

On esitetty useita opetukseen liittyviä mielikuvia siitä, mihin tekoälyä voitaisiin käyttää? Blogikirjoittajien mukaan seuraavat teemat voidaan nostaa soveltuviksi. Tekoälyohjaus toimii osana opettamista. Kirjallisuudessa pohditaan Alexan ja Sirin kaltaisia algoritmeja ja tietovarantoja, jotka on edelleen rajattu ja viritetty opetettavan teeman ympärille. Käyttöliittymästä riippumatta ne tuottavat vastauksia opiskelijan kysymyksiin tai antavat palautetta mallisuorituksen perusteella. Tekoäly toimii henkilökohtaisena oppimisen avustajana. Teemasta kirjoittajat esittävät, että tällainen tieto saattaisi auttaa ohjaajaa sopeuttamaan opetustaan oppilasprofiilin mukaan. Ehkä spesifi algoritmi pystyisi tunnistamaan jokaisen opiskelijan oppimisvaikeudet. Ehkä toinen sovellus auttaisi häntä eteenpäin, kun oppimisvaikeudet kompensoitaisiin käyttämällä hänen vahvuuksiensa mukaisia oppimisstrategioita. Tekoäly mahdollistaa yksilöidyt harjoitustehtävät tai satunnaisuuden tuonnin rutiinitehtäviin. Tämä tarkoittaa, että hyvin yksinkertainen algoritmi voi luoda lähtöarvoiltaan uusia kysymyksiä ja myös arvioida kunkin vastauksen oikeellisuuden. Opetustavoitteiden kannalta näin voidaan varmentaa, että kukin oppija käy tehtävän aidosti läpi.

CASE tehtävä-automaatti

LUT-yliopisto kehitti menetelmää, jolla voidaan tuottaa sekä sanallisia että laskennallisia harjoitustehtäviä ilman vakiovastauksen jäykkyyttä. Kokeilimme tätä menetelmää noin 100 hengen ryhmässä hyvin oppimistuloksin trigonometrinen tehtävien kertamiseen. Ahkerat opiskelijat tekivät tehtäviä runsaasti yli minimiharjoittelun ja myös

kykenivät antamaan huomioita kokeilujärjestelyjen toimivuudesta. Muutaman opintopisteen kurssi ei ollut vaativa tai yli työllistävä, mutta näyttäisi siltä, että osa opiskelijoista kokee tällaisen koneen valvoman yksilökohtaisen harjoittelun kiusalliseksi lisätyöksi. (Saastamoinen & Rissanen 2018)

MOOC ja tekoäly

Mainittakoon tässä yhteydessä myös massiiviset avoimet verkkokurssit Massive Open Online Courses (MOOC). Tämän tuotteen ydinpiirteet on usein ymmärretty väärin. Alkuperäinen konsepti ei sisällä mitään suoraan tekoälyn ominaisuuksiin luettavaa piirrettä. Perus- MOOC toteutusten opetussisällöt ovat olleet ensisijassa videoita aidoista tai studiossa tehdyistä luennoista. Suomen ulkopuolella, kansainvälisesti katsoen kurssit eivät suinkaan ole maksuttomia (free of charge), vaan kalliita kurseja. Kurssin suoritus ja sertifikaatti eivät seuraa kurssimaksusta automaattisesti, vaan olakseen hyväksyttävissä suorituksena johonkin koulutusohjelmaan, on opiskelijan osaamista arvioitava ennen todistuksen antoa. Portaalin striimauksen eli opetuksen seuraamisen lisäksi, opiskelija tuottaa arvioitavissa olevan tuotoksen. Tekoäly on kustannustehokkain tapa saada sekä arvio että kirjoittaa ja lähettää sähköpostitse todistus kurssilaiselle. Niinpä Forbesin sivuilla Marr (2018) on todennut opettajien siirtyneen opettamaan tietokoneille oikeita vastauksia (siis mallivastausten lisäksi heidän tulee laatia vertailuaineistoja oikeista tai riittävän oikeista ratkaisuksista). Monet tunnetut yliopistot tarjoavat MOOC toteutuksia ensimmäisen vuoden rutiiniopinnoissa ja lisämaksullisessa tutkintoon johtamattomassa täydennyskoulutuksessa (Blackmon 2016). Osa näistä projekteista on myös puhtaasti kaupallisista lähtökohdista toimivia yrityksiä kuten Coursera ja Udemy.

Tekoäly oppilaan tukena

Koulutusta voidaan arvioida eri näkökulmista. Hallinto katsoo ensisijassa opetussuunnitelmaa ja sen toteutumista suhteessa resursseihin. Toki akkreditoitu organisaatio arvioi myös käytössä olevin mittarein oppilaspalautetta ja opetustulosten saavuttamista. Opiskelijan ja opettajan näkökulmasta on muitakin muuttujia. Seuraavaksi tarkastellaan oppilaan näkökulmaa. Koulutuksen termein on pohdittava kolmea osa-alueetta: oppimista, kognitiivista kehitystä ja varsinaista opetustapahtumaa. Kunkin muuttujan suhteen tekoälyteknologian käyttö tuottaa joitain etuja. Samalla on pohdittava, voiko opettajan roolin kaventuminen tai automaation osuuden kasvattaminen tuottaa uudenlaisia rajoitteita tai tuottaa jollakin osa-alueilla uuden tyyppisiä vääristymiä.

Portaali mahdollistaa opiskelun itsenäisesti omaan tahtiin. Huolella toteutettu verkkokoulutus voi tarjota runsaasti tietoa kaikille kiinnostuneille ja oppikirjojen tapaan voidaan luetun ymmärtämistä arvioida yksinkertaisin kertaustehtävin. Yksittäisen opiskelijan edistymistä voidaan seurata useimmissa opetusportaaleissa. Analysoitaessa laajemmin yksittäisen opiskelijan tai ryhmän menestymistä tarvitsemme algoritmeja sekä ymmärrystä siitä, miten oppilas optimoi vastauksiaan. Oppilas saattaa toivoa essee-tyyppisiin vastauksiin sekä läpinäkyvyyttä että ennakkotietoa pisteytyksestä, eikä tämän informaation tuottaminen ole aina yksinkertaista. Korkeampia osaamistavoitteita mittaavia tehtäviä on vaikea tarkastaa massana. Usein opettaja arvioi kokonaisuuden

ja palaa yksitellen lopullisten arvioiden tekemiseen. Vastaavasti vertaileva arviointi avainsanoineen olisi analytiikan tehtävä. Suomenkielisissä aineistoissa voisimme Voikon avulla lemmittämistä (muuntaa perusmuotoon) suomenkielisen tekstin ja löytää osaamisen eri tasoja kuvaavat lauseet (Pirinen 2015). Vaihtoehtoisesti voimme laatia kyllä/ ei vastauksia ja ehkä opettaa oppilaita kirjoittamaan tekstiä tekoälylle. Valitettavasti viimeksi mainitut vaihtoehdot ovat erinomaisia vähentämään korkeampien oppimistavoitteiden toteutumista.

Ihmisen kognitiiviset taidot ja niiden kuvaaminen on haasteellista. Nykynuoren peukalot tuottavat tekstiä, joka on vähän sinnepäin. Älykäs algoritmi korjaa sanan ja lauseen tasolla viestin. Ipad tuotteen myötä varsinkin kiinalaiset lapset kokivat uusia elämyksiä (selkeät ikonit ja kirjoittamisen helppous?). Onko tässä kysymys uudesta orientaatiosta aivoissa, intuitiivisesta ja jopa koukuttavasta käyttöliittymästä vai Wiion skeptisesti pohtiman tekoälyn vaikutuksesta kiinankielen merkistön toteutumiseen osana laitteen käyttöä. Tekoäly voi toimia kyvyttömyyden väistön apuna. Sokealle tai vaikeasti lukihäiriöiselle henkilölle tekoäly sovellus lukee ääneen mitä hyvänsä tekstimuotoista aineistoa. Näitä ominaisuuksia on useissa arjen sovellutuksissa (mm Adobe Reader tai uusimmat MsWord versiot). Toisaalta useat applikaatiot voivat muuntaa puheen tekstiksi ja siten ohittaa kirjoittamisen rajoitteet. Yhdistämällä nämä kaksi asiaa voidaan muodostaa uudenlainen puheyhteys kahden erikielisen ummikon välille. Näyttäisi siltä, että jopa hyvin henkilökohtaiset tuntemukset onnistuvat tällaisella sovelluksella.

Tekoäly antaa tukea olemassa oleville kyvyille. Urheileva ihminen saa palautetta fyysisen suorituskyvyn harjoituksistaan. Vertaamalla palautumista eri harjoituskerroista voidaan arvioida optimaalinen harjoituskuorma kullekin harjoitukselle. Jos henkilö on luonut itsestään riittävän hyvän virtuaalisen mallin (vrt. Digital Twins), saattaa mallia hyödyntävä laskuri ehdottaa harjoituksen aikaistettua päättämistä ja siirtymistä koordinoituihin verryttelyyn. (Tretter, 2021)

Tekoälyn avustava ohjaus saattaa nopeuttaa kognitiivisen alueen edistymistä. Toisaalta, motivoiko itseopiskelu ja mahdollisuus saavuttaa itsenäisesti asetettuja välitavoitteita enemmän kuin väsymättömän ohjaajan apu? Itseopiskelijoille tarkoitettu DuoLingo sovellutus vaatisi (ehkä) oman lukunsa. Olemme tehneet havaintoja useamman aikuisopiskelijan omatavoitteisen kielitaidon kehityshaasteista heidän käyttäessä joko maksullista tai ilmaista DuoLingo applikaatiota. Tämä sovellutus aktiivisesti muistuttaa harjoittelusta ja pyrkii sopeutumaan opiskelijan kykyyn kehittyä opeteltavassa kielessä. Päivittäinen harjoittelu ja tehtävien tekeminen on ainut tapa päästä seuraavalle opetuksen tasolle. (Loewen ym. 2019)

Tekoäly opettajan näkökulmasta

Siirryttäessä MOOC pohjaiseen peruskurssien toteutukseen, oppiaineen professorin mahdollisuus toistaa itseään isossa luentosalissa vähenee. Toki myös mahdollisuus optimaaliseen tapaan opetusvelvollisuudesta kuittaamisesta poistuu. Samalla tekoäly korvaa urakalla kokeita tarkastavat assistentit. Vastaavasti ilman erillistä uudistamista kurssin painopiste siirtyy hyvien vakiolauseiden mukaiseen vastaamiseen ja oppilas-

kunnan tenttiarkiston paluuseen. Sääntöpohjaisten järjestelmien avulla myös oppilaan ohjaus isoilla, hyvin vakioituilla kurssilla voidaan automatisoida. Ratkaisu löytyy chatboteista. Ensimmäiset tällaiset kaikkien ulottuvilla olevat sovellutukset olivat jo Office paketissa 20 vuotta sitten (Binder nD). Office Assistant oli älykäs Microsoft Officen käyttöliittymä, joka auttoi käyttäjiä interaktiivisen animoidun hahmon avulla.

Useimmat pienempien toimijoiden chatbotit ovat uudelleen koodattuja usein kysyttyjen asioiden (ns. UKK) tietokantoja. Niinpä kysyjän kysymyksen muotoilu oikeanlaiseksi on oma haasteensa. Ensiksi algoritmi ei voi edetä, koska sitä ei ole koulutettu kysyjän asioihin tai seuraavaksi se arvioi tietokannassa olevia vastauksia ilman todellista avainsanaa. Kiusallista on, että kaiken yrittämisen jälkeen, saatu vastaus voi olla jo vanhentunutta tietoa. Tarvittaessa, hyvin toteutettu chatbot yhdistää kysyjän oikealle ihmiselle. Käytännössä kuitenkin oletettavasti tärkeä, mutta ratkaisematon kysymys päättyy jonkun organisaation yleiseen sähköpostiin, josta se aikanaan poimitaan.

Useat toimijat yrittävät luoda chatbottien kaltaisia päätöksentekoa tukevia tuotteita, joiden tarkoituksena on korvata tai avustaa ihmisasiantuntijaa. Tällaisissa järjestelmissä on kaksi elementtiä: tietokanta ja päättelymoottori. Tietokantaan kertyy luokituksen avulla tietokoodia, jonka avulla tehdään päätösehdotelmat. (Heller ym. 2005)

Älykkäät ohjausjärjestelmät

Ns. Älykkäät ohjausjärjestelmät, Intelligent tutoring systems (ITS), pyrkivät parempaan vuorovaikutukseen. Ensimmäiset opetuksen ohjausjärjestelmäkokeilut tehtiin 80-luvun alussa (Nwana 1990). Tutkijat yrittivät luoda uudenlaista vuorovaikutusta opettajien ja opiskelijoiden välille. Luokahuoneissa suurin haaste opettajille on tasapainottaa heterogeenisyyttä. Ohjausjärjestelmien alueella on erotettavissa eri malleja parempaan ohjaukseen:

- "Domain model" mallia kutsutaan myös kognitiiviseksi tai "asiantuntijatietomalliksi". Tämän mallin tehtävänä on ymmärtää kaikki käsitteet ja taidot, joita tietyllä alalla käytetään. Lisäksi se tarjoaa oikeat vaiheet ongelmien ratkaisemiseksi.
- "Opiskelijamalli" käyttää opiskelijan näkökulmaa. Se käyttää tilastollisia malleja opiskelijan tietotason ymmärtämiseen. Se voi seurata edistymistä ja taitoja milloin tahansa. Malli voi myös seurata joitakin yksityiskohtia, kuten tehtyjen tehtävien lukumäärää, tarvittavaa aikaa ja oikeiden vastausten prosenttiosuutta.
- "Tutorimalli" käyttää opettajan näkökulmaa. Se yrittää havaita opiskelijan tiedon puutteita ja olla vuorovaikutuksessa opiskelijan kanssa palautteen kautta. Palautteiden ajoitus, mutta myös jotkut muut tekijät, kuten tuen laatu ja spesifisyys, vaihtelevat järjestelmäkohtaisesti. Jotkut alustat antavat palautetta välittömästi ja toiset koko tehtäväkokonaisuuden tai kurssin suorittamisen jälkeen. Palaute voi kertoa opiskelijan tekemistä virheistä tai antaa vihjeitä ja vastauksia joihinkin opiskelijoiden pyyntöihin.

Ongelma on mallien luotettavuus. Murphyn (2019) havainnoimat tutkimukset osoittivat, että ITS-ohjeet tuottivat samanlaisia tuloksia kuin yksityistunnit tai pienryhmät. Yleensä nämä tulokset viittaavat enemmän opintojen laatuun, alueisiin ja arvosanatoihin. Millaisia rajoitukset sitten olisivat? ITS-mallit rajoittuvat fyysiikkaan, matemaatiikkaan, lukutaitoon ja tietojenkäsittelytieteeseen. Nämä ovat tieteenaloja, jotka sisältävät algoritmeja, menetelmiä ja yleisesti strukturoituja rutiineja.

Automatisoidun esseen pisteytyksen (Automated Essay Scoring, AES) taustalla oli motiivi luoda sovellus, joka voisi arvostella kokeita, joissa selkeitä lauseita (Attali & Burstein 2006). Oletettavasti MOOC toteutukset kuten Coursera, Udacity, EdX jne. käyttävät tällaisia järjestelmiä kursseille ilmoittautuneiden opiskelijoiden arvioimiseen. Tekniikka on nimeltään Luonnollisen kielen prosessointi (Natural Language Processing NLP). Konkreettisesti AES käyttää valvottua oppimista poimittujen ominaisuuksien oppimiseen. Haittapuolena on, että tällaisissa järjestelmissä on mahdollista huijata kirjoittamalla järjettömiä asioita. Yhdysvalloissa julkaistun raportin mukaan yli puolet lukioista käyttää varhaisvaroitusjärjestelmiä opetuksen suunnittelun apuna. Ns. on-track-indikaattorit antavat varoituksen tietyn kynnyksen lähestyessä. Tätä seuraa asiantuntijoiden tuki ongelman ratkaisemiseksi. (Balfanz ym, 2017) Vaikka tämä järjestelmä toimii hyvin, väärän luokituksen seuraukset voivat olla merkittäviä. Yleensä tällaiset järjestelmät voivat tukea olemassa olevia sääntöpohjaisia järjestelmiä ja hallita suurta tarkkuutta vaativia tehtäviä. Ne kuitenkin tarvitsevat valtavan määrän opiskelijoita ja heidän tietojaan mallin kouluttamiseen.

Tekoälyn käytön rajoitukset

Tekoälyjärjestelmien tarkkuus riippuu muutamasta asiasta. Tietojen laatu yhdessä käytettyjen algoritmien kanssa voi toimia myös odottamattomalla tavalla. Mutta kuka on algoritmien takana? Inhimillinen tekijä voidaan vielä havaita koodauksessa, mutta entä opetetuissa neuraaliverkoissa.

Black Box-ongelma tunnetaan yleisesti mm. neuraaliverkoin toteutetuissa tukipalveluissa. Tätä voidaan kuvata järjestelmän kyvyttömyydellä todistaa tekemiään päätöksiä. Esimerkiksi, jos opiskelija on järjestelmän ennusteen mukaan taipuvainen keskeyttämään, järjestelmä ei voi selittää päätöstä. Nykyään tavoitellaan "selittävää tekoäly", jossa yrittää purkaa tarjotun ratkaisuun päätyminen mysteerin ja antaa läpinäkyvyyttä saadulle arviolle (Turek, 2018).

Johtopäätökset

Ymmärrämme kaikki, että koulutuksen teknologialla on edelleen rajoituksia. Olemme esitelleet työkalu nimeltä tekoäly (AI). Sen avulla tehtyjen palveluiden on täytettävä monia vaatimuksia, jotta voisimme antaa niille autonomisemman aseman. Tästä syystä monet koulutusalaan kiinnostuneet arvioivat, ettei vanhanaikaista opettajaa voida koskaan korvata robotilla. Joten meidän on mietittävä, haluammeko rakentaa opetuksen tueksi vahvan tekoälyn, joka tekee opettajan tehtävästä monipuolisemman, mutta poistaa heiltä sekä valvojan että vastuunkantajan roolin tarpeettomana taakkana. Lähitulevaisuudessa voisimme ensimmäisenä pohtia tekoälyn soveltamista tukityö-

kaluna. Sellaisten applikaatioiden tehtävänä on vapauttaa opettajat tukitöistä hallinnon suuntaan. Toki osa oppimistapahtumista on luonteeltaan sellaisia, että ne on mahdollista täysin automatisoida. Valinnoissa tulee ottaa huomioon myös oppijan erilaiset taipumukset ja mieltymykset erilaisiin oppimisympäristöihin. Kuitenkin tekoälyä on opetuksessa pidettävä joustavuuden lisääjänä sekä lisäoptiona.

Lähteet

Attali, Y. and Burstein, J. (2006) 'Automated Essay Scoring With e-rater® V.2', *The Journal of Technology, Learning and Assessment*, 4(3). Available at: <https://ejournals.bc.edu/index.php/jtla/article/view/1650> (Accessed: 21 September 2022).

Balfanz, R., Herzog, L. and Mac Iver, D.J. (2007) 'Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions', *Educational Psychologist*, 42(4), pp. 223–235. Available at: <https://doi.org/10.1080/00461520701621079>.

'Binder - Microsoft Office for Windows (97-2003)' (no date).

Blackmon, S.J. (2016) 'Through the MOOCing Glass: Professors' Perspectives on the Future of MOOCs in Higher Education', *New Directions for Institutional Research*, 2015(167), pp. 87–101. Available at: <https://doi.org/10.1002/ir.20156>.

Heller, B. *et al.* (2005) 'Freudbot: An Investigation of Chatbot Technology in Distance Education', in. *EdMedia + Innovate Learning*, Association for the Advancement of Computing in Education (AACE), pp. 3913–3918. Available at: <https://www.learntechlib.org/primary/p/20691/> (Accessed: 21 September 2022).

Loewen, S. *et al.* (2019) 'Mobile-assisted language learning: A Duolingo case study', *ReCALL*, 31(3), pp. 293–311. Available at: <https://doi.org/10.1017/S0958344019000065>.

Marr, B. (2018) *How Is AI Used In Education -- Real World Examples Of Today And A Peek Into The Future*, *Forbes*. Available at: <https://www.forbes.com/sites/bernard-marr/2018/07/25/how-is-ai-used-in-education-real-world-examples-of-today-and-a-peek-into-the-future/> (Accessed: 21 September 2022).

Montemerlo, M. *et al.* (2006) 'Winning the DARPA grand challenge with an AI robot', in *Proceedings of the 21st national conference on Artificial intelligence - Volume 1*. Boston, Massachusetts: AAAI Press (AAAI'06), pp. 982–987.

Moor, J. (2006) 'The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years', *AI Magazine*, 27(4), pp. 87–87. Available at: <https://doi.org/10.1609/aimag.v27i4.1911>.

Murphy, R.F. (2019) *Artificial Intelligence Applications to Support K–12 Teachers and Teaching: A Review of Promising Applications, Opportunities, and Challenges*.

10. RAND Corporation. Available at: <https://www.jstor.org/stable/resrep19907> (Accessed: 21 September 2022).

Nwana, H.S. (1990) ‘Intelligent tutoring systems: an overview’, *Artificial Intelligence Review*, 4(4), pp. 251–277. Available at: <https://doi.org/10.1007/BF00168958>.

O’Shea, T. *et al.* (1984) ‘Tools for Creating Intelligent Computer Tutors’, in *Proc. of the international NATO symposium on Artificial and human intelligence*, pp. 181–199.

Pirinen, T.A. (2015) ‘Omorfi — Free and open source morphological lexical database for Finnish’, in *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*. Vilnius, Lithuania: Linköping University Electronic Press, Sweden, pp. 313–315. Available at: <https://aclanthology.org/W15-1844> (Accessed: 21 September 2022).

Rojko, A. (2017) ‘Industry 4.0 Concept: Background and Overview’, *International Journal of Interactive Mobile Technologies (ijIM)*, 11(5), pp. 77–90. Available at: <https://doi.org/10.3991/ijim.v11i5.7072>.

Saastamoinen, K. and Rissanen, A. (2018) ‘Behind the benefits of using an automated exercise creation system’, *Procedia Computer Science*, 126, pp. 1023–1028. Available at: <https://doi.org/10.1016/j.procs.2018.08.038>.

The DARPA Grand Challenge: Ten Years Later (2014). Available at: <https://www.darpa.mil/news-events/2014-03-13> (Accessed: 21 September 2022).

Tretter, M. (2021) ‘Perspectives on digital twins and the (im)possibilities of control’, *Journal of Medical Ethics*, 47(6), pp. 410–411. Available at: <https://doi.org/10.1136/me-dethics-2021-107460>.

Turek, M. (2018) *Explainable Artificial Intelligence (XAI)*. Defense Advanced Research Projects Agency. Available at: <http://web.archive.org/web/20190728055815/https://www.darpa.mil/program/explainable-artificial-intelligence> (Accessed: 21 September 2022).

TEKOÄLYAVUSTETTU TILANNEJOHTAMINEN

Lauri Kilpeläinen

Jobdanto

Klassinen taskulaskin kykenee tekemään yhtä monta laskutoimitusta minuutissa, kuin sitä käyttävä ihminen. Kuinka monta päätöstä ihmistä käyttävä kone kykenisi tekemään sekunnissa? Tekoälyteknologia ja oppivat järjestelmät ovat jo integroitumassa osaksi yhteiskuntaa kaupallisten toimijoiden ratkaisuuina ja yhteiskunnan useiden tekoälyä hyödyntävien hankkeiden kautta. Tekoälyn toiminta koostuu tiedon keräämisestä, analysoinnista ja joko analyysin esittämisestä tai päätöksenteosta ennalta asetettujen parametrien mukaan, joka johtaa halutun toiminnan käynnistämiseen.

Tekoälytekniikan soveltuvuutta tilannejohtajan toiminnan tukemiseksi tutkittiin vuosina 2017-2020. Tutkimuksen tavoitteeksi muodostui johtokeskuksessa toimivien käyttäjien kognitiivisen kuormituksen keventäminen tekoälyavusteisella teknologialla. Tutkimuksessa mallinnettiin tilannejohtajan toimintaa ja toimintaympäristöä, sekä sitä verrattiin olemassaolevaan teknologian mahdollisuuksia tunnistettujen haasteiden ratkaisuiksi. Johtopäätöksenä tekoälyavustetulla johtamisella voidaan nykyteknologialla tukea tilannejohtajan toimintaa kuormitusta keventämällä ja tulevaisuudessa päätöksentekoaapuna. Mikäli oppiva järjestelmä toteutetaan toimimaan ihmisen rinnalla, voidaan järjestelmän nähdä toimivan vertaisena ja mahdollisesti sijaisena tilannejohtajan roolissa (Kilpeläinen, 2020).

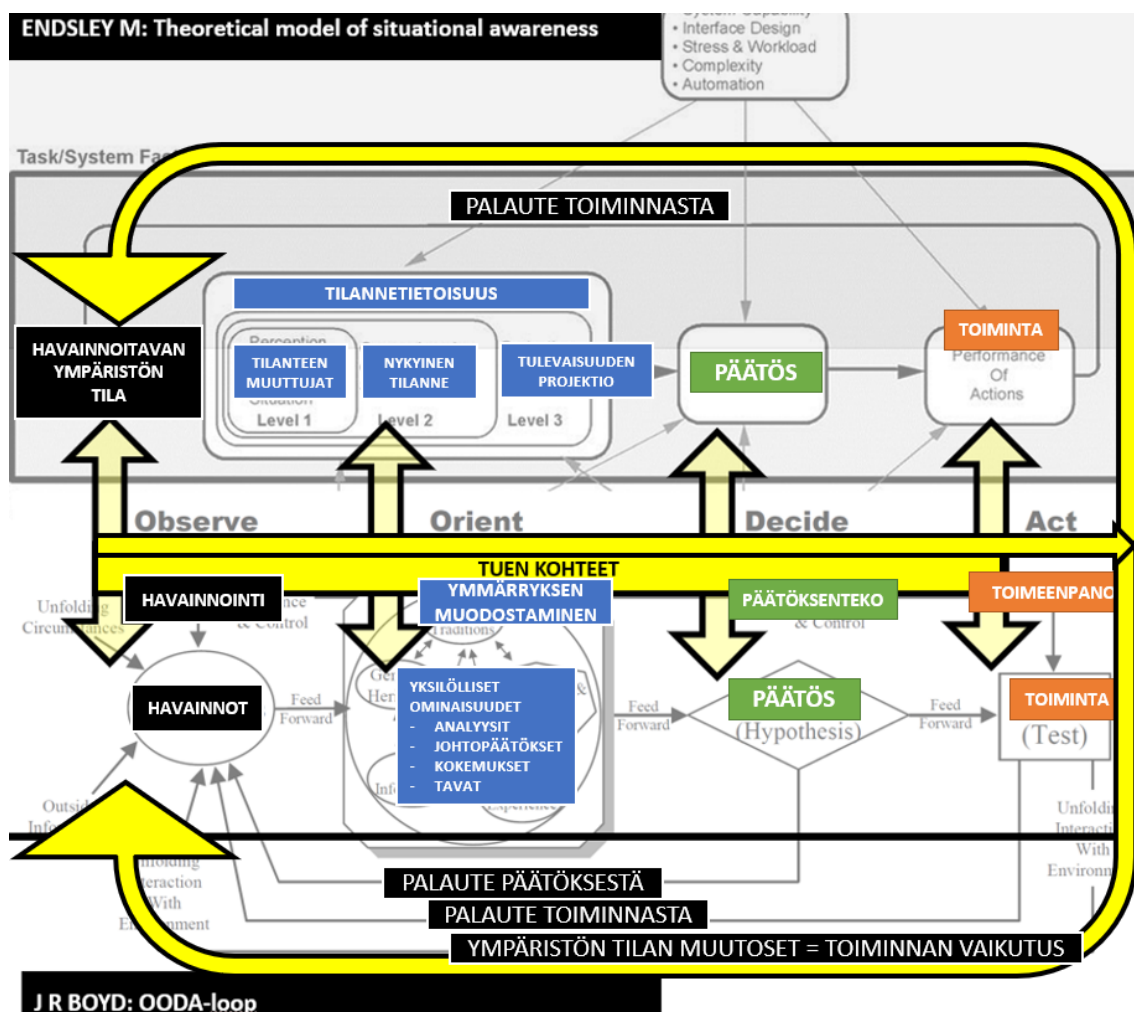
Tutkimuksessa havaittiin, että kuormituksen keventämisen tekoälytekniikalla voi aloittaa siirtämällä usein toistuvat tietojenkäsittelytehtävät ohjelmistorobotiikan hoidettavaksi, joka luo rajapintaa ihmisen ja koneen yhteistoiminnan edistämiseksi. Pelkkä ohjelmistorobotiikan hyödyntäminen ei tuota älykästä järjestelmää, koska järjestelmän täytyy luoda ymmärrys käsittelemästään tiedosta ja osata hyödyntää sitä asetettujen tavoitteiden saavuttamiseksi. Alkuvaiheessa tavoitteiden määrittely on ihmiskäyttäjän vastuulla, mutta oppivan järjestelmän kehittyessä tavoitteiden analysointia ja esittelyä voidaan siirtää järjestelmän tehtäväksi. Ihmiskäyttäjän rooli voi muuttua optimoitujen esityksien hyväksyjäksi, jolloin oppiva järjestelmä kykenee keräämään datan alkutilanteesta, suunnitelmasta, ihmisen syötteistä ja lopputuloksesta, jota voidaan hyödyntää järjestelmän kouluttamisessa kohti optimoitua toimintaa (Kilpeläinen, 2020).

Oppivan järjestelmän hyödyt tulevat esiin, kun järjestelmä otetaan toimimaan ihmiskäyttäjän rinnalle ja vuorovaikutusmenetelmät toteutetaan käyttäjälähtöisesti. Ihmisen rinnalla toimivalle järjestelmälle voidaan opettaa tilannejohtajaa avustavia tehtäviä, kuten analyysien tekoa ja tilannetuotteiden esittämistä. Merkittävämpää hyötyä voidaan tuottaa kouluttamalla järjestelmä toimimaan vertaisena tilannejohtajana, jolloin ihmisen ja koneen tehtävänjakoa täytyy tarkastella uudesta näkökulmasta. Oppivan järjestelmän etuina ovat mahdollisuus järjestelmän tai sen osan kopiointiin skaalattavuuden osalta ja simulaatioina toteutettavat koulutus ja kehityskierrokset. Oppivan järjestelmän kanssa voidaan harjoitella skenaarioita yleisimpien tilannekehyksien ulkopuolelta

ja sitä voidaan käyttää myös kopioituna vastustajana itseään vastaan harjoittelussa tai vastustajasta voidaan tehdä oma ylläpidettävä malli.

Tutkimuksen toteutus

Tilannejohtajien kuormitusta mitattiin tehtäväanalyysillä, The National Aeronautic and Space Administration Task Load Index (NASA TLX) ja Situational Awareness Rating Technique (SART) kyselyillä. Kyselyt selvittivät käyttäjän tehtävänäikaista kuormitusta ja kokemusta tilannetietoisuudesta (Stanton et al, 2005). Kerättyä tietoa käytettiin keskuksien toimintojen mallinnukseen ja tukikohteiden selvittämiseen. Tuen kohteiksi valikoituivat tilannejohtajan tehtäviin kuuluvat automatisoitavat toiminnot, toimijoiden rutiiniluonteiset tehtävät, vuorovaikutus ja toimeenpano (Kilpeläinen, 2020).



Kuva 1 Tilannetietoisuuden mallin (Mica Endsley, 2003) ja päätöksentekomallin (John Boyd, 1976) yhdistelmä käyttäjän toiminnan ja tuen kohteiden mallinnusta varten.

Tilannejohtajaa voitaisiin tukea havainnoinnin, tilanneymmärryksen, päätöksenteon ja toimeenpanon osa-alueilla. Osa-alueet voidaan liittää John Boydin OODA-loop johtamismalliin, joka on yleisesti käytetty sotilaallisessa toiminnassa sekä Mica Endsleyn

tilannetietoisuuden SITAW malliin. Tilannejohtajan toiminnan mallinnuksessa hyödynnettiin OODA- ja SITAW-mallien yhdistelmää tuen kohteiden määrittämiseksi, koska malleista on löydettävissä yhteneväisiä rakenteita kuvan mukaisesti.

Tilannejohtajan havainnointikyvyn kuormitus toiminnan ja vuorovaikutuksen kautta

Tilannejohtajan toiminta jakaantuu henkilökohtaiseen toimintaan ja vuorovaikutukseen ympäristön kanssa. Käyttäjä kerää tietoa jatkuvasti useista lähteistä, joita voivat olla eri järjestelmien syötteet, viestit, puhelut, kuvat ja näiden yhdistelmät. Tietoa voidaan ottaa vastaan yleensä vain yhden sensorisen toiminnon kautta tai useista lähteistä vaihdellen, jolloin tarkkaavaisuus ja muisti kuormittuvat (Oksama, 2011). Tiedon keräämisen lisäksi käyttäjä analysoi tietoa jatkuvasti, verraten sitä käskettyihin ja itse asetettuihin mielikuviiin, siitä minkälaista tietoa tulisi ottaa vastaan (Endsley, 2003). Käyttäjä jakaa keräämäänsä ja analysoimaansa tietoa, joko sellaisenaan tai sovitettuna tiedon tarvitsijan tavoitteisiin.

Seminaariesityksessä esiteltiin Baddeleyn työmuistin mallin (Alan Baddeley, 1974) modifikaatio, jota oli muokattu soveltumaan tutkimuksen mallinnustapaan. Työmuistin mallinnus esittää ihmisen muistin rajoitteita kuormituksen näkökulmasta. Kuormitus vaikuttaa käsiteltävien asioiden muistiin siirtoon ja sieltä palauttamiseen negatiivisesti. Rajoitteet ovat kohteita, joita voidaan tukea automaation keinoilla tilannejohtajan tietojenkäsittelyyn liittyen. Tilannejohtajan tukemisella automaation kautta voidaan siirtää kognitiivista kapasiteettia käytettäväksi yksilön tilanneymmärryksen muodostamiseen, vuorovaikutukseen ja päätöksentekoon. Tällä hetkellä kognitiivisten resurssien siirto päätöksentekoon olisi luonnollisin askel kohti tilannejohtamisen optimointia, koska perustoiminnan automaatio liittäisi järjestelmän osaksi toimintaa. Perustoimintojen automaatiosta, kuten tiedon kirjaamisesta, analysoinnista ja jakamisesta olisi luontaista kehittää järjestelmää kohti päätöksentekokykyä.

Tilannekeskusten ja johdettavien välinen toiminta kuormittaa käyttäjää vuorovaikutuksen määrän ja käytettävissä olevien johtamisyhteyksien osalta. Kuormituksen määrä vaihtelee ajankohtaan tai toiminnanvaiheeseen sidottuihin hetkiin. Mallinnuksen avulla tunnistettiin rinnakkaisen keskuksen ja johdettavien kuormittavan samoja vuorovaikutuksen väyliä, kuin alaisten, joten yksi käyttäjä kykenee toimimaan yhdessä vuorovaikutustilanteessa kerrallaan. Ruuhkautuneet vuorovaikutuksen väylät muodostuivat alaisten, sidosryhmien ja vertaisten päätöksentekijöiden vaatimasta kommunikoinnista, jolla toimeenpantiin käskyjä ja ylläpidettiin tilanneymmärrystä. Toistuvat vuorovaikutustilanteet varaavat käyttäjän kognitiivisia resursseja havainnointikyvyn ja työmuistin osalta, joka on pois tilanteenseurannan ja päätöksenteon prosesseista. Tilannejohtajan kuormitusta voidaan keventää sovittamalla vastaanotettavaa tietoa vapaana oleviin vuorovaikutuksen väyliin ja kohdentamalla yhteydenotot sopivimmalle käyttäjälle esitietojen ja tilanteenmukaisen toiminnan seurannalla. Kuormituksen keventäminen olisi mahdollista toteuttaa ohjelmistorobotiikan ja oppivan järjestelmän yhdistelmällä. Informaatiovirran käsittelyä voidaan myös keventää jakamalla haluttu tieto vastaanottajalle sopivimmassa muodossa. Käyttäjien toimintatapojen automatisella seurannalla voidaan opettaa järjestelmää ymmärtämään käyttäjäänsä ja sovittamaan informaatio tilanteeseen sopivaksi. Esimerkiksi eri muodoissa välitettävä infor-

maatio, kuten teksti, puhe ja kuvat voidaan kääntää vastaanottajalle sopivimpaan muotoon tekstistä puheeksi toiminnolla, joka löytyy jo lähes kaikista älypuhelimista.

Tilanneymmärryksen muodostus

Tilanneymmärryksen luontiin liittyy pyrkimys muodostaa projektiota tulevasta tilannekehityksestä (Endsley, 2003). Projisointiin käytetään eri lähteistä kerättyä ja kerääjän itse mielessään luokittelemaa informaatiota. Tulevan tilannekehityksen projisointi auttaa käyttäjää varautumaan vaihtoehtoisin kehityssuuntiin ja tekemään nopeampia päätöksiä, sekä ratkaisemaan tilanteeseen liittyviä ongelmia. Tilannejohtaja pyrkii ylläpitämään tilanneymmärrystään vertailemalla kerättyä tietoa tilanteesta rakennettuun tilanneymmärrykseen ja mielikuviin. Kerätty tieto sisältää suullisia ilmoituksia, keskustelua, tekstiä, piirroksia ja esityksiä, jotka liittyvät yksittäisiin tapahtumiin tai kokonaistilanteeseen yhteenvetojen muodossa. Ymmärryksen muodostaminen vaatii aikaa tiedon prosessoinnille, johon yksilölliset ominaisuudet ja opitut toimintatavat vaikuttavat kuormituksen lisäksi. Tilannejohtajat pyrkivät muodostamaan ja ylläpitämään tilanneymmärryksen kokonaisuutta henkilökohtaisesti sekä yhdessä. Havaintojen lähteet ja tilanneymmärryksen muodostamisen tavoitteet vaihtelevat keskuksessa toimivan tehtävän mukaisesti, mutta keskuksessa muodostettava kollektiivinen tilanneymmärrys liittyy yhteiseen tavoitteeseen tai tavoiteltavaan loppuasetelmaan.

Yksilökohtaisessa tilanneymmärryksen muodostamisessa käyttäjää voidaan tukea samoilla menetelmillä, kuin kuormituksen keventämisessä. Tukevalla järjestelmällä on kuitenkin oltava kyky muodostaa käsitys tavoiteltavista toiminnoista ja loppuasetelmasta, jotta se voi tukea tilanneymmärrystä tukevan tiedon esittämisessä ohjaamatta käyttäjää harhaan. Kyky voidaan tuottaa oppivalla järjestelmällä, jossa järjestelmälle opetetaan tilanneymmärryksen muodostamiseen liittyvien ilmiöiden analysointia ja tiedon tuottoa. Opetusmateriaalin tuotto vaatii ensin ihmiskäyttäjän toiminnan mallintamista koneelliseen koulutukseen sopivaksi.

Kollektiivisen tilanneymmärryksen muodostuksen haasteena ovat käyttäjien yksilölliset tavat toimia, kommunikoida ja ymmärtää tapahtumia samankaltaisesti. Eli ihmiset ymmärtävät asiat eri tavoilla, joka voi tuottaa jaettavasta tilanneymmärryksestä poikkeavaa ymmärrystä. Poikkeavan tilanneymmärryksen vaarana on tiedon toistuminen ja jakaminen virheellisenä, joka voi ohjata päätöksentekoa ja tavoiteasettelua harhaan. Rinnakkaisia käyttäjiä voidaan tukea seuraamalla käyttäjien toimintaa järjestelmän toimesta, vertailemalla tilannetta toimintaan ja sovittamalla analysoitua tietoa yksilöllisiin ominaisuuksiin sopivaksi, joka voi ehkäistä kognitiivisia vinoumia. Yksilöllisillä ominaisuuksilla tarkoitetaan ihmisen tiedonkäsittelytapoja kognitiotieteen näkökulmasta. Esimerkiksi aiempien kokemusten kautta tärkeäksi koettua tietoa havainnoidaan enemmän, kuin tuntematonta tietoa. Tärkeäksi koettu tieto voi kuitenkin olla todelliseen tilanteeseen nähden turhaa tai ohjata toimintaa epäedulliseen suuntaan

Päätöksenteko

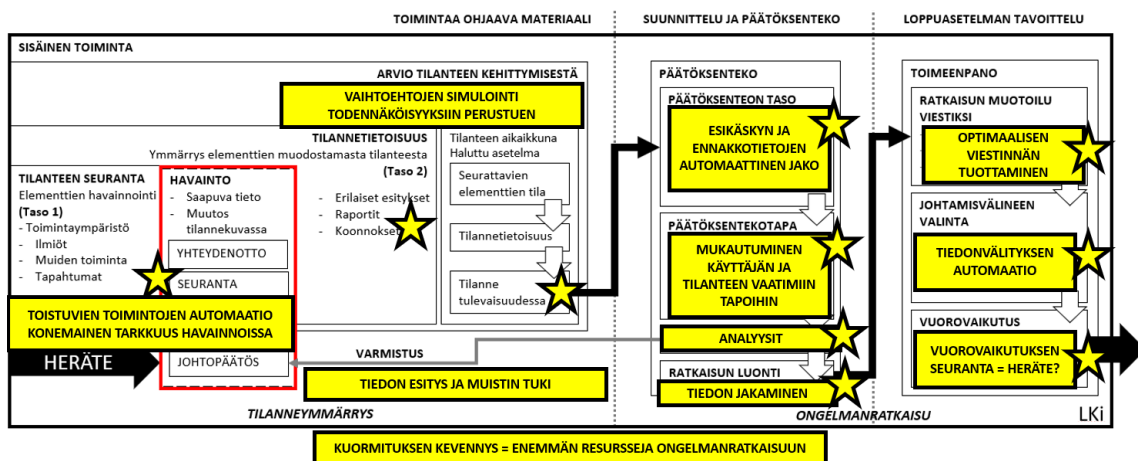
Päätöksenteko on ongelmanratkaisua, jossa toimintaa rajoittavina tekijöinä ovat aika ja käytettävissä olevan tiedon perusteella luotu ymmärrys todellisuudesta. Käyttäjää

tekee päätöksiä toiminnan jatkamisesta tai muuttamisesta esitiedon ja kerätyn tiedon analysoinnin sekä tilanteen kehityskulkujen vertailun perusteella. Käyttäjä voi toimeenpanna myös muiden tekemiä päätöksiä, jolloin tilanneymmärryksen vertaaminen tavoitteen mielikuvaan voi erota. Mielikuvien erot syntyvät yksilöiden välisistä vuorovaikutustavoista, henkilökohtaisista ominaisuuksista tilanteen ymmärtämiseen liittyen ja mahdollisista ennako-oletuksista johtuen. Erot tilanteen ymmärtämisessä voivat johtaa käsketyt suunnitelman toimeenpanoon väärällä tavalla tai väärään aikaan.

Tunnettujen päätöksentekomallien osalta rationaalisen ja intuitiivisen päätöksenteon mallit voivat soveltua parhaiten järjestelmän tuottamaan tukeen, koska päätöksentekotavat mukailevat menetelmiä, joita sovelletaan oppivien järjestelmien mallien kouluttamisessa. Intuitiivisen päätöksentekotavan tukeminen voi vaatia järjestelmältä tunteisiin liittyviä ominaisuuksia, joiden tekninen kypsyyt vaihtelee. Luovan päätöksentekotavan mallinnus korostaa tunteiden ja intuition merkitystä päätöksenteossa, joten tukevan järjestelmän tulisi ymmärtää ihmisen tunteiden toimintaa ja mahdollisesti kykyä muodostamaan niitä itse. Luovan päätöksentekotavan tukemisessa voitaisiin soveltaa laskennallisen luovuuden tieteenalaa. Tunneälykäs järjestelmä voi olla lähitulevaisuuden teknologiaa, koska se tukisi ihmisen ja koneen yhteistoimintaa vuorovaikutuksen ja työskentelymenetelmien osalta. Tunnistukseen perustuvassa päätöksenteossa uhkaksi ja mahdollisuudeksi tunnistettiin järjestelmän mahdollisuus tukea käyttäjän toimintaa tilannekehityksen kannalta edulliseen suuntaan tai vaihtoehtoisesti ohjata käyttäjää harhaan, mikäli järjestelmän parametrit ja tavoiteasettelu eroavat todellisuudesta. Kehitettävän järjestelmän haasteeksi voi muodostua käyttäjän tuntemus päätöksentekotapojen osalta ja mukautuminen toiminnan aikana muuttuviin päätöksentekotapoihin.

Päätöksentekoprosesseista tunnistettiin osa-alueita, joita voidaan mitata jo tämänhetkisellä teknologialla. Järjestelmällä voidaan myös mitata ympäristön tapahtumia ja ilmiöitä, jonka perusteella se voi arvioida päätöksentekoprosessia riittävien lähtötietojen kanssa. Päätöksentekotapojen tuntemuksella järjestelmä voi mukauttaa toimintaansa käyttäjälähtöisesti oppimisen kautta, kuten kuormituksen keventämisen ja tilanneymmärryksen tukemisen menetelmissä.

Tilannejohtamisen kokonaisuuden mallinnus OODA ja SITAW mallien yhdistelmänä



Kuva 2 Tilannejohtajan toiminta tilannejohtamisen näkökulmasta sovitettuna Boyd OODA-loop ja Endsley SITAW malliin (Kilpeläinen, Tekoäly ja logiikka näyttöesitys, 2021)

Kuvassa on esitetty OODA ja SITAW mallien yhdistelmän kautta mallinnettu tilannejohtajan toiminnan kokonaisuus. Toiminta käynnistyy herätteestä, joka voi olla joko tilanteesta tehty havainto, esikäsketty toiminta tai suora käsky käynnistää jokin toiminnan vaihe. Havainnoinnin osalta tilannejohtajaa voidaan tukea saapuvan tiedon esikäsittelyllä käyttäjälähtöisestä näkökulmasta tai esimerkiksi automaattisesti kriittiseksi tunnistetun tiedon korostamisena. Korostamistoiminnon tuottaminen vaatii tilanne-seurantakykyä järjestelmältä tai käyttäjän esiasettamia syötteitä. Havainnoinnin tuke-
misen vaikutus tiedonkäsittelyn automaation keinoin vapauttaa resursseja ongelmanratkaisuun, joka nopeuttaa toimeenpanoa ja havainnoinnin tiheyttä.

Herätteen käynnistämisen toiminnan jälkeen tilannejohtaja alkaa muodostamaan arviota tilanteen kehittymisestä vallitsevaan tilanneymmärrykseen perustuen. Tilanteen kehityksen arvio johtaa tulevaisuuden projektiioon seurattavien elementtien ja tilannetietoisuuden vertailevana analyysinä perustuen aiemmin opetettuihin toimintamalleihin tai yksinkertaisiin laskentatyökaluihin ja käskyrunkoihin. Tilannejohtaja joutuu tuottamaan arviotaan muuttuvassa tilanteessa, joka sisältää keskeytyksiä ja seurattavia uusia syötteitä. Tilannejohtajaa voidaan tukea tulevaisuuden projisoinnissa tosiasioihin perustuvien vaihtoehtojen simuloinnilla ja todennäköisyysanalyysillä, jotka voivat pohjautua vastustajan aiempaan tai arvioituun toimintaan. Simulointien ja analyysien tulisi tukea tilannejohtajan toimintaa helposti omaksuttavien esitystapojen ja muistia tukevien menetelmien avulla. Muistin tukemisella voidaan ehkäistä muistissa tapahtuvien vääristymien ja unohtusten vaikutusta verrattuna todelliseen mitattuun tai ilmoitettuun tilanteeseen nähden. Tilannejohtajan tilannearvion tuki voi vaatia käyttäjän yksilöllistä mallintamista, jotta tukitoimet keventävät kognitiivista kuormaa eivätkä lisää sitä.

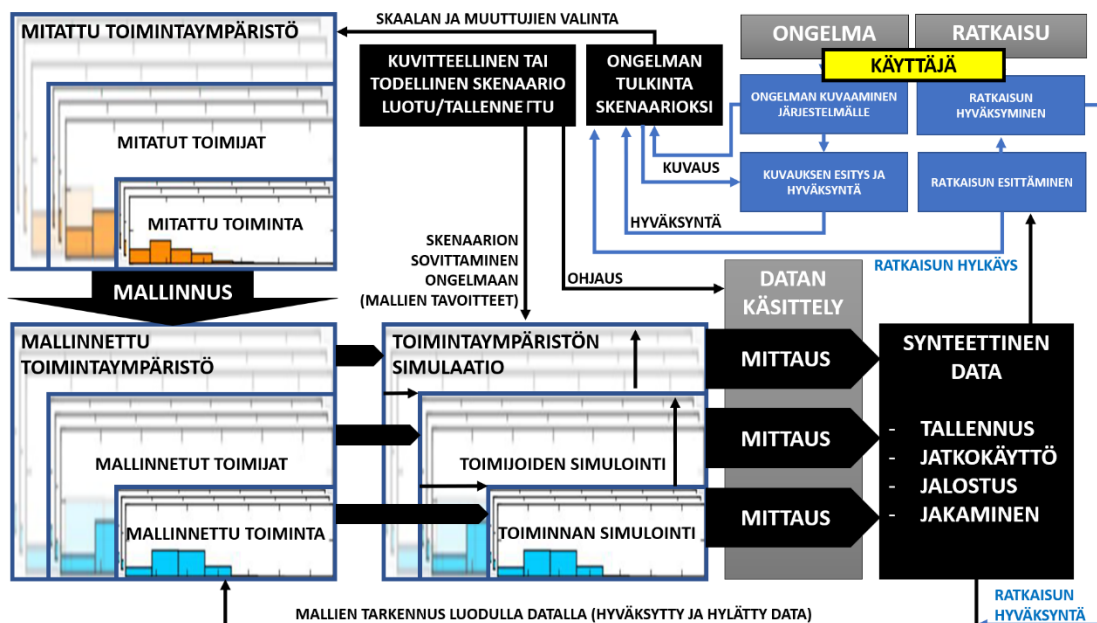
Tilannejohtajan tuottama tulevaisuuden projektiio voi tuottaa tarpeen uudelle päätöksentekokierrokselle, joka varaa tilannejohtajan havainnointikyvyn ja työmuistin pois tilanteenseurannasta. Johtopäätös tulevaisuuden projektion merkityksestä vaikuttaa tilannejohtajan toimintamahdollisuuksiin, jotka ovat tiedon jakaminen ylemmälle pää-

töksenteon tasolle tai päätöksenteko itse toimivaltuuksien rajoissa. Ylemmälle tasolle tietoa jaettaessa tulisi jakaa myös oma tilanneymmärrys, johon johtopäätös perustuu. Oppivalla järjestelmällä voidaan tukea ymmärryksen siirtoa sovittamalla ilmoituksen sisältö ja esitystapa ylemmän tason tilanneymmärryksen kokonaisuuteen. Toimivaltuuksien rajoissa tehtävässä päätöksessä tilannejohtajan on otettava huomioon suuri määrä muuttujia ja tehtävä analyysi päätöksen vaikutuksista tilannekehitykseen, josta johtaja on vastuussa. Tilannejohtajan kuormitusta voidaan keventää jakamalla tarpeelliseksi määritelty pohjatieto tarvitsijoille jo ennen päätöksentekoa ja esittämällä päätöksen vaikutukset tilannekehitykseen päätöksentekijälle.

Tilannejohtajan päätöksenteon jälkeinen toimeenpanovaihe sisältää päätöksen viestimisen toimeenpaneville tahoille, vuorovaikutusmentelmän valinnan ja toimeenpanon seurannan. Toimeenpanon viestintää ja vuorovaikutusmenetelmän valintaa voidaan tukea muotoilemalla ja sovittamalla viesti vastaanottajalle optimaaliseksi tilannesidonnaisesti. Viestinnän automaatiolla voidaan keventää tiedon jaon vaatimien toimintojen kuormitusta, kuten osoitteiden, nimien ja numeroiden muistamista. Toimeenpanon seuranta voidaan tukea automaatiolla, joka seuraa oletettuja muutoksia tilanteessa päätökseen sidottuna ja koostamalla toimeenpanoon liittyviä ilmoituksia.

Keinotekoisien mallien luominen tilannejohtajan tueksi

Toimintaympäristöstä voidaan mitata ja mallintaa ympäristössä esiintyviä fyysisiä ilmiöitä kuten tien kantavuutta, rakennuksen osien vahvuuksia tai esimerkiksi sään vaikutusta joukon liikkeeseen. Yksittäisiä mittauskohteita voivat olla räjähdysten vaikutus ympäristöön tai esimerkiksi yksilön fyysinen suorituskyky tehtäväkohtaisessa varustuksessa. Mallinnus tuo mahdollisuuden simuloida ympäristön ja mittauskohteiden vaikutuksia toisiinsa. Tilannejohtajista mitattavia ja mallinnettavia kohteita voivat olla toimintatavat, fyysiset ja kognitiiviset ominaisuudet ja rajoitteet, kuten työmuisti, tarkkaavaisuus ja tiedon analysointinopeus. Todellisista kohteista kerätyllä datalla voidaan testata, kehittää ja ylläpitää keinotekoisesti luotuja malleja, jotta ne vastaavat analysoitavaa tilannetta mahdollisimman realistisesti. Keinotekoisien mallien etuna on rajoittamattomat testimahdollisuudet uusien toimintatapojen kehittämiseksi, mutta haaste muodostuu keinotekoisien ja todellisen mallin todenmukaisuuden ylläpitämisen välille. Malleja voidaan tallentaa ja jatkokehittää ratkaistavien ongelmien mukaisesti sekä malleja voidaan kouluttaa yhdessä ja esimerkiksi toisiaan vastaan uuden koulutusdatan tuottamiseksi.



Kuva 3 Esimerkki järjestelmän opettamisesta (Kilpeläinen, Tekoäly ja logiikka näyttöesitys, 2021)

Kuvan esimerkissä järjestelmän kouluttamiseksi käyttäjä määrittelee ratkaistavan ongelman ja kuvaa sen järjestelmälle. Kuvaukseen käytettävä käyttöliittymä voi koostua käyttäjää seuraavista sensoreista ja päätelaitteista, joilla järjestelmä liittyy osaksi työympäristöä. Esimerkiksi kameroista ja mikrofoneista, jotka muuntavat kuva, tekstin ja puheen järjestelmän ymmärtämään muotoon. Ongelman kuvaamisen jälkeen järjestelmä esittää oman kuvauksensa, eli ymmärryksensä ongelmasta käyttäjälle, jonka jälkeen käyttäjä vertaa omaa ja järjestelmänsä ymmärrystä keskenään. Käyttäjä hyväksyy järjestelmän kuvauksen, jos ongelman kuvaus vastaa haluttua asetelmaa, jonka jälkeen järjestelmä aloittaa ongelman ratkaisun määrittämällä tarvittavat muuttujat ja skenaarion parametrit. Käyttäjä voi hylätä, muokata ja tarkentaa järjestelmän esittämää kuvausta, jolloin käyttäjän syötteitä hyödynnetään uusien mallien koulutuksessa.

Koulutustilanteen luontia varten voidaan valita joko mitattua todellisen tapahtuman historiallista dataa, josta luodaan haluttu mallinnus tai käytetään jo valmista ennalta tuotettua mallia, joka vastaa simuloitavaan ongelmaan. Skenaario luodaan valitulle mallille ja aloitetaan simulaatio, josta mitataan valitut muuttujat, sekä lopputulos ennalta asetetun tavoitteen mukaisesti. Järjestelmä voidaan opettaa tuottamaan sarjan skenaarioita ja analysoidaan mittauksia optimaalisen ratkaisun löytämiseksi. Simulaation jälkeen järjestelmä esittelee ratkaisun tai usean vaihtoehtoon sarjan käyttäjälle, joka hyväksyy tai hylkää ratkaisun. Järjestelmä käyttää tietoa ratkaisun hyväksynnästä tai hylkäyksestä koulutusdatan, luotujen mallien ja mittauksen uudelleenkäsitteilyyn ja kehittämiseen seuraavia skenaarioita varten.

Johtopäätökset

Tekoälytekniikka tunnistettiin monitieteiseksi kokonaisuudeksi, jota kehitetään eri tieteenalojen näkökulmista, mutta jota voidaan myös hyödyntää tieteenalojen kehityk-

sessä (Ailisto, 2018). Esimerkiksi tietojenkäsittely- ja neurotieteet hyödyntävät tekoälyratkaisuja sekä tuottavat tutkimusta, jota voidaan hyödyntää tekoälyjärjestelmien kehittämiseksi, eli tieteessä voidaan havaita takaisinkytkentä, kuin keinotekoisien neuroverkkojen kouluttamisessa. Vaikutusta saadaan kuitenkin hyödyntämällä ja kehittämällä koneoppimisen menetelmiä tieteenalasta riippumatta. Tekoälytekniikan etuna on datan tuottaminen ympäristöstä tai simulaatioista mitattaamalla uusien mallien koulutukseen. Järjestelmän kouluttamista voidaan toteuttaa aika- ja paikkarajoitteista riippumattomasti, sekä siihen voidaan yhdistää lisätyn todellisuuden ja mobiiliapplikaatioiden rajapintoja.

Kuormituksen keventäminen voisi olla ensimmäinen askel teknisen tuen kohteeksi, koska toistuvien tiedonkäsittelyä vaativien tehtävien siirtämisestä on jo toteutettu yhteiskunnassa esimerkiksi ohjelmistorobotiikan keinoilla (Kilpeläinen, 2020). Keskuksissa toimivien tilannejohtajien osalta kuormitusta ja toimintaympäristön toiminnallisuuksia mallinnettaessa ihmiskäyttäjän havaittiin toimivan jatkuvan informaatiotulvan alla. Suuri käsiteltävän informaation määrä heikentää havainnointikykyä ja kuormittaa lähimuistia, joka vie kognitiivisia resursseja pois analysointi- ja päätöksentekotilanteilta, jotka ovat ihmisen vahvuuksia koneeseen verrattuna. Käyttäjän kuormituksen keventämisen oletetaan luovan mahdollisuuksia laadukkaampaan päätöksentekoon ja tilanneymmärryksen ylläpitoon.

Ihmisen ja järjestelmän välisen toiminnan harmonisointi vaatii riittävän käyttäjätuntemuksen järjestelmän näkökulmasta, jotta ihmiskäyttäjälle voidaan tarjota oikea-aikaista informaatiota soveltuvassa muodossa. Esiitetty informaatio täytyisi synkronoida vallitsevaan tilannekehitykseen, joka asettaa vaatimuksia järjestelmän kykyyn arvioida tilannetta reaaliaikaisesti ja oppia siitä, eli järjestelmälle on luotava riittävä ymmärryskyky. Ymmärryskyvyn luominen voi vaatia laskennallisen luovuuden tai keinotekoisien tunteiden kehittämistä järjestelmään, koska ihmiskäyttäjän kanssa toimiminen vaatii ihmiselle luonnollisia vuorovaikutuksen keinoja myös tunteiden osalta.

Tekoälytekniikan kypsyyden on riittävä järjestelmien kehitystyön aloittamiseksi. Järjestelmälle olisi luotava riittävä oppimis- ja ymmärryskyky tunneälyn ja luovuuden osalta. Tunneälyn ja luovuuden kehityksen kautta voidaan luoda uusia ihmisen ja koneen välisiä vuorovaikutusmenetelmiä, sekä antaa mahdollisuus varsinaiseen älykkääseen päätöksentekoon, kun järjestelmä kykenee ymmärtämään tavoitteenasettajaa, tilannetta ja tuottamaan niistä oman tilanneymmärryksensä. Tuotettu järjestelmä tulisi nähdä vertaisena toimijana kehitystyön näkökulmasta, jotta ihmisen ja koneen potentiaali voitaisiin hyödyntää täysimääräisesti.

Tekoälyavustettua päätöksentekoa voisi testata prototyypillä, jonka voisi tuottaa pelillistämällä toimintaa tilannejohtajien toimintojen mallinnoiksi kautta ja mallintamalla siihen yleisimpiä tunnistettuja päätöksentekotapoja. Kehitystyöllä voisi tuottaa tekoälyjärjestelmien kanssa yhteensopivia päätöksentekotapojen uusia mallinnoiksi. Prototyypin kehityksessä tulisi pitää mukana järjestelmän koulutusnäkökulma, jolla ihmiskäyttäjän syötteitä voidaan hyödyntää järjestelmän opettamisessa.

Järjestelmän kehittämisessä on huomioitava järjestelmää opettaneiden käyttäjien toiminta, koska alkutilanteessa järjestelmässä ei ole eettisiä tai epäeettisiä toiminnallisuuksia, koska niiden voidaan olettaa kehittyvän oppimisen myötä opettajien vaikutuksesta. Vastuu lopputuloksesta on alkutilanteessa järjestelmän opettajilla, jonka jälkeen vastuu siirtyy toiminnallisuuksia ja ominaisuuksia hyväksyvälle taholle. Itseään kehittävän järjestelmän vastuun voidaan olettaa siirtyvän järjestelmälle, mikäli järjestelmä kykenee itsetietoisuuteen ja ymmärtämään toimintaansa vähintään käyttäjän tasoisesti. Käyttäjien aiemmin hankitun osaamisen hyödyntäminen käyttäjälle tuttujen menetelmien ja työkalujen osalta auttaa siirtämään ihmisen osaamista järjestelmälle. Järjestelmällä voidaan myös optimoida ja parantaa ihmisen toimintatapoja opettamalla käyttäjille toistensa tehokkaiksi todettuja tapoja tai luomalla vanhoista menetelmistä tehokkaampia oppivaa järjestelmää hyödyntäen.

Lähteet

Ailisto, H. (2017) Tekoälyn monet ulottuvuudet Havainnointi ja tilannetietoisuus. VTT.

Ailisto, H. (2018) Tekoälyn käsittekartta – VN-TEAS projekti. VTT.

Appelqvist, P. (2017) Tekoälyn monet ulottuvuudet- Symposiumin avainhuomioita. Puolustusministeriö, MATINE.

Artificial Intelligence Index Report 2019 (2019). Stanford: Stanford University.

Defense Advanced Research Projects Agency (2020) AI Next Campaign. Available at: <https://www.darpa.mil/work-with-us/ai-next-campaign>.

Endsley, M.R., Bolté, B. and Jones, D.G. (2003) Designing for situation awareness: An approach to user-centered design. CRC press.

Heljanko, K. (2017) Tekoälyn monet ulottuvuudet- Big Data platforms. Aalto Yliopisto.

Kaski, S. (2017) Tekoälyn monet ulottuvuudet- Koneoppimisen eri muodot. Aalto Yliopisto.

Kilpeläinen, L. (2020) Tekoälyavustettu taistelunjohto (SM1496). Maanpuolustuskorkeakoulu, TLIV.

Kilpeläinen, L. (2021) Tekoälyavustettu toiminta, näyttöesitys. Maanpuolustuskorkeakoulu.

Oksama, L. (ed.) (2011) Inhimillisten tekijöiden hallinta lentoteknisessä työssä. Tampere: Opetushallitus, Juvenes Print Oy. Rule, J.N. (2013) A Symbiotic Relationship: The OODA Loop, Intuition, and Strategic Thought. Carlisle: US Army War College. Russell, S.J. (2010) Artificial intelligence: a modern approach. 3rd ed. New Jersey: Pearson.

Stanton, N.A. et al. (2005) *A Practical Guide for Engineering and Design*. New York: Human Factors Integration Defence Technology Centre, Ashgate Publishing.

Tarkoma, S. (2017) *Tekoälyn monet ulottuvuudet- Data-analyysi tieteenalana*. Helsingin Yliopisto.

Toivonen, H. (2017) *Tekoälyn monet ulottuvuudet- Laskennallinen luovuus (ja ongelmanratkaisu)*. Helsinki: Helsingin yliopisto.

Tuominen, H. et al. (2019) *Tekoälyn perusteita ja sovelluksia*. Edited by P. Neittaanmäki and H. Tuominen. Jyväskylä: Jyväskylän Yliopisto.

TEKOÄLY JA PSYKOLOGINEN VAIKUTTAMINEN

Saara Jantunen

Keskustelua informaatiovaikuttamisesta vaivaa uskomus siitä, että vaikuttamiselta voisi suojautua hyvällä koulutuksella, medialukutaidolla ja olevalla rationaalinen ja järkevä ihminen, jollaiseksi toki kuka tahansa meistä haluaa identifioitua. Kuitenkin viimeistään 2020-luvulla Suomessakin on koronapandemian myötä saatu huomata, mitä laajamittainen tunteiden manipulointi on, ja kuinka disinformaatio ja salaliittoteoriat ovat vieneet mennessään myös itsellemme tuttuja ihmisiä, joita muuten olemme pitäneet varsin arvostelukykyisinä ihmisinä. Koronapandemia on kuitenkin vain yksi hyvä esimerkki siitä, miten ihmisten tunteet, moderni teknologia ja erinäisten toimijoiden tarve haitalliselle psykologiselle vaikuttamiselle muodostavat yhdistelmän, joka tuottaa jatkuvasti uusia, vaikeasti ennakoitavia vaikutuksia yhteiskunnassamme. Psykologinen vaikuttaminen perustuu kohdeyleisön tuntemukselle, ja siksi vaikuttamisen kohteisiin liittyvä data on nyt vaikuttamisen edellytys. Mitä yksilöidympää dataa yhteiskunnasta, sen kansalaisista ja ryhmittymistä on saatavilla, sen tehokkaammaksi vaikuttaminen voidaan hienosäätää.

Ei ole sattumaa, että demokratian jo 15 vuotta kestänyt globaali heikentymistrendi on osunut samaan aikaan, kun informaatioteknologian nopea kehittyminen. Vuonna 2020 yhtensä 75 % maailman väestöstä asui maassa, jossa demokratia heikentyi (Rupucci and Slipowitz, 2022). Autoritaariset valtiot, kuten Venäjä ja Kiina, käyvät jatkuvaa informaatiotaistelua demokraattisia normeja vastaan. Lehdistönvapaus, oikeusvaltioperiaate, ihmisoikeudet, yksityisyys ja kansalaisyhteiskunta ovat olleet hyökkäyksen kohteena, ja niiden tilalle on tarjolla autoritaaristen valtojen edustamia demokratian vastanormeja. Venäjä on operationalisoinut erityisesti perinteisten arvojen narratiivin, jota on hyödynnetty offensiiveissa länsimaita vastaan. Parhaana esimerkkinä toiminevat venäläiset narratiivit länsimaiden moraalisisesta mädänneisyydestä, jota sukupuolten ja eri ihmisryhmien tasa-arvo, vapaa liikkuvuus ja yksilönoikeudet edustavat. Kyseenalaistaakseen demokratian normit, on autoritaaristen valtojen tarjottava tilalle vaihtoehtoa. Tätä vaihtoehtoa markkinoidaan vahvasti offensiivisen informaatiovaikuttamisen keinoin. Vaikuttamisen alustana toimivat paitsi ulko- ja turvallisuuspolitiikka ja diplomatia, myös koko arkinen informaatioympäristömme medioineen, digitaalisine palveluineen ja viestintäsovelluksineen.

Tunteet ovat avain vaikuttamiseen

”Suomea ja suomalaisia suojelee korkea koulutustaso ja hyvä medialukutaito.”

”Sosiaalisessa mediassa pitää olla tarkkana, eikä uskoa kaikkea.”

”Olen rationaalinen ja analyttinen persoona, joten en lähde mukaan sosiaalisen median vouhotuksiin.”

Informaatiovaikuttamisesta keskustellaan hyvin toisella tavalla, kuin monista muista uhkista. Toisin kuin kyberuhilta tai verkkorikollisuudelta suojautumisen, informaatio-

tiovaikuttamiselta suojautumiseen liittyy vahvasti uskomus siitä, että koulutettu länsimaalainen, puhumattakaan koulutetusta pohjoismaalaisesta, olisi ominaisuuksiltaan automaattisesti enemmän immuuni informaatiovaikuttamiselle. Korostamme mielellämme rationaalisuutta ja analyttisyyttä henkilökohtaisina ominaisuuksina, jotka ovat koulutuksen avulla saavutettavissa.

Kognitiotieteen ja psykologian näkökulmasta me ihmiset emme kuitenkaan jakaudu järki- ja tunneihmisiin, vaan olemme kaikki pääasiassa tunneihmisiä. Tunteet eivät ole ihmiselle koulutus- tai valintakysymys, vaan ihmiseen esihistoriasta asti koodattu keino selviytyä. Ihmiselle, joka eli esihistoriallisina aikoina varsin erilaisissa toimintaympäristössä kuin nykyisin, oli oman selviytymisen kannalta tärkeää luottaa vaaroista varoittaviin vaistoihinsa, sillä se piti hänet todennäköisemmin hengissä, kuin analyttinen syväajattelu. Tästä syystä ihminen tekee edelleen suurimman osan päätöksistään automaattisesti ja intuitiivisesti, eli tunneperusteisesti. Toisin kuin analyttinen, intuitiota ja tunteita kyseenalaistava ajattelu, se mahdollistaa rutinoitumisen ja nopeat päätökset, jotka yleensä ovat riittävän oikeassa (Kahneman, 2011). Kuten Bargh and Chartrand (Bargh and Chartrand, 1999) toteavat:

“[M]ost of a person's everyday life is determined not by their conscious intentions and deliberate choices but by mental processes that are put into motion by features of the environment that operate outside of conscious awareness and guidance.”

Emootio on monivaiheinen ja monimutkainen tapahtuma, joka luo valmiuden toiminnalle. Tunteen perustana on kognitiivinen arvio tilanteesta, tapahtumasta tai olosuhteista. Tämä arvio laukaisee sarjan reaktioita, jotka yhdessä kognitiivisen arvion kanssa luovat emootion. Emootioilla on siis yleensä selvä syy ja aiheuttaja (Nolen-Hoeksema, Fredrickson, Loftus, *et al.*, 2009, pp. 396–397).

Kun ympäristö täyttyy informaatiosta, josta suuri osa on tavoitteellisen harhaanjohtavaa tai valheellista ja suunniteltu vetoamaan kohdeyleisöjen heikkouksiin, muuttuu ihmellinen tunneohjautuvuus haavoittuvuudeksi. Ihmisen aivot eivät yksinkertaisesti kykene jatkuvaan analyttisyyteen ja syväajatteluun, sillä ne ovat prosesseina kuluttavia ja raskaita. Näin ollen myöskään vaikuttamiselta suojautumista ei tulisi perustaa sen varaan, että ihmiset eläisivät modernissa informaatioympäristössä jatkuvassa valppauden ja analyttisyyden tilassa ja kykenisivät näin ohittamaan ihmiselle luontaisen tavan reagoida ärsykkeisiin intuitiivisesti. Koulutuksen ja monilukutaidon merkitystä ei pidä vähätellä, mutta niillä ei voida täysin ohittaa sitä, mikä on ihmiselle luontaista ja vaistonvaraista. Mikäli suomalaiset olisivat tavallista vastustuskykyisempiä kaikenlaiselle manipulaatiolle, ei esimerkiksi mainostoimistoilla olisi täällä töitä.

Kuluneina vuosina informaatioympäristössä tapahtunut vihamielinen vaikuttaminen on kohdistunut yhä enenevässä määrin tunnereaktioiden luomiseen asioissa, jotka eivät suoraan vaikuta liittyvän suurvaltapolitiittiseen kamppailuun. Esimerkiksi Venäjän sosiaalisessa mediassa toteuttamat vaikuttamiskampanjat ovat hyödyntäneet valheellisia ja provosoivia väitteitä vaalivilpistä, maahanmuutosta ja vähemmistöryhmien ja sukupuolten tasa-arvosta väestösuhteiden polarisoimiseksi ja yhteiskuntarauhan horjuttamiseksi länsimaissa (*United States of America v. internet Research Agency LLC*, 2018).

Lisäksi kampanjoissa on pyritty heikentämään liittoumien ja kumppanimaiden keskinäistä luottamusta (Nimmo *et al.*, 2020). Jälkimmäiset ovat perinteisempiä informaatioidankäynnin teemoja, jotka on niin ikään kohdistettu kohdeyleisön tunnekokemuksiin. Näissä kampanjoissa toistetaan jo kylmän sodan ajalta tuttuja näkemyksiä siitä, että Yhdysvaltoihin ei voi luottaa, Nato uhkaa Venäjää, ja että länsimaiden toiminta uhkaa eskaloida sodan, josta kaikki siviilit tulisivat kärsimään.

Jotta vaikuttaminen olisi tehokasta, on se kyettävä suuntaamaan mahdollisimman suoraan ja massamaisesti itse kohdeyleisöihin. Tässä teknologinen kehitys erottaa menneen maailman propagandan, tämän päivän informaatiovaikuttamisen ja tulevaisuuden tekoälypohjaisen psykologisen vaikuttamisen. Toistaiseksi tekoälyä hyödynnetään pääasiassa defensiivisesti, mutta tämä tulee muuttumaan ajan myötä.

Informaatiotilan muokkaaminen: botit, joukkoistukset ja epäautenttinen sisältö

Sosiaalisessa ja perinteisessä mediassa tehdyt joukkoistukset ovat voimakas vaikuttamisen keino ja osa internet-kulttuuria, jossa jäljitellään sodan logiikkaa: se, joka saa taakseen suuremman ja äänekkäämmän joukon, on taistelussa vahvemmilla. Tätä logiikkaa pyrkivät hyödyntämään myös poliittiset vaikuttajat ja tavan ”some-influensserit”, joilta esimerkiksi erilaiset kansalaisjärjestöt ostavat näkyvyyttä.

Joukkoistuksia voi kuitenkin käyttää hyvän sijaan myös haitantekoon ja jopa tuhoamiseen. Internetin keskustelupalstoilla julkaistaan tälläkin hetkellä kuvia esimerkiksi virkamiehistä ja pyydetään ihmisiä tunnistamaan heidät ja ilmoittamaan heihin liittyviä tietoja, jotta heihin voi kohdistaa häirintää. Tällaiselle manuaaliselle joukkoistamiselle on jo nyt tekoälyyn perustuvia vaihtoehtoja, jotka murtavat nopeasti kenen tahansa yksityisyyden. Verkossa on esimerkiksi palveluja, joihin voi syöttää valokuvan ihmisestä, joka halutaan tunnistaa.

Joukkoistamista voidaan hyödyntää myös laajempiin kohdeyleisöihin vaikuttamisessa. Tähän mennessä julkisuudessa esillä olleet sosiaalisen median vaikutusoperaatiot ovat pääosin perustuneet bottiverkostojen rakentamiseen ja niiden hyödyntämiseen vaikutusoperaatioissa, sekä käsin kirjoitettujen sosiaalisen median viestien levittämiseen. Vaalivaikuttamiseen osallistunut ja sen myötä suurennuslasin alle joutunut Pietarin Internet Research Agency lienee yksi tunnetuimmista tällaisista tapauksista. Nämä sisällöt tunnistetaan sosiaalisen median verkostoissa melko hyvin, sillä ne ovat laadultaan sellaisia, että niissä näkyy tekijöidensä käsiala tyypillisiä kielioppivirheitä ja aksenteja myöten. Tämä heikentää materiaalin uskottavuutta ja vaikuttavuutta. Autenttiselta vaikuttavan sisällön luominen käsin on vaativaa. Kenties tästä syystä mm. Venäjä on panostanut joukkoistamiseen, jossa se pyrkii erilaisin peitetyin keinoin mobilisoimaan valitsemiaan kohdeyleisöjä tuottamaan materiaalia itse. Juuri tällaisessa vaikuttamisessa erityisesti negatiivisten tunnereaktioiden herättäminen kohdeyleisöissä on tärkeä osa mobilisointia. Kun raivostunut somekansa nousee barrikadeille ja tuottaa sopivaa sisältöä ja viestintää itse, on se myös uskottavampaa, sillä tällainen sisältö on autenttista. Kuten Meta toteaa raportissaan: sitä mukaa kun pahantahtoisten toimijoiden on yhä vaikeampi operoida laajojen valetiliverkostojen avulla, ne yrittävät huijata ihmisiä levittämään ja vahvistamaan sisältöään (Meta, 2022). Sosiaalisen median yhtiöt, kuten

Meta, ovat kääntyneet tekoälyn puoleen yrityksissään kitkeä alustoilta vahingollista sisältöä. Toistaiseksi tulokset ovat kuitenkin olleet kokonaisuutena vaatimattomia. Meta kuitenkin on ilmoittanut käyttävänsä tekoälyä miljoonien valetileiksi epäiltyjen tilien blokkamiseen ja sulkemiseen joka päivä, sillä valetilejä pidetään vaikutusoperaatioiden kannalta täysin keskeisinä (Facebook, 2021).

Vaikuttamisoperaatioiden edellyttämä uskottavuus, kuten kulttuuristen tyylien ja kielipillisen tarkkuuden saavuttaminen, on niin ikään vielä haasteellista. Kuitenkin tekoälyn opettaminen autenttiselta vaikuttavan sisällön tuotantoon on helpompaa, kuin saada liukuhihnalta some-postauksia tuottava nuori työntekijä kirjoittamaan uskottavasti vieraalla kielellä. Tekoäly luo täysin uudenlaisia mahdollisuuksia informaatiotilan täyttämiseen. Yksilöön ja massoihin vaikuttamisen lisäksi tekoälyllä voidaan vaikuttaa informaatiotoimintaympäristöön tavoilla, jotka vaikeuttavat tilannekuvan muodostamista ja viestintää. Informaatioympäristöä voidaan esimerkiksi täyttää (floodata) tekoälyn generoimilla sisällöillä. Tämä voi tarkoittaa mitä tahansa sisältöä: kuvia, videota, tekstiä, ääntä, jne. Kiinalainen Xinhua hyödyntää jo nyt tekoäly-uutisankkureita uutislähetysissään. Sivustolla nimeltä NotRealNews.net puolestaan julkaistaan tekoälyn tuottamia, perättömiä uutisia. On siis jo nyt mahdollista luoda tekoälypohjaisia ekosysteemejä, joissa tuotetaan, levitetään ja julkaistaan uutisia, jotka joko ovat paikkansapitäviä tai sitten eivät. Vaikka toistaiseksi tällainen uutistuotanto sisältää myös hauskaakin kömpelöä kieltä, näin ei tule olemaan kauaa. Lisäksi palvelut ovat laajalti saatavilla: OpenAI poisti vuoden 2021 lopulla jonotuksen GPT-3 ohjelmointirajapinnasta. Tämä tarkoittaa sitä, että kyseisen tekstintuottomallin saa käytettäväksi ja kehitettäväksi pyytämällä. Sisällön generoimista voidaan käyttää niin offensiivisesti kuin defensiivisesti psykologisessa vaikuttamisessa.

Väärentäjien ja väärennösten tunnistajien kilpajuoksu

Kun keskustellaan tekoälykehityksen roolista osana informaatiovaikuttamista, ensimmäisenä esiin nousevat yleensä deep fake -videot. Riskit ja uhat, joita deep fake -videoilla voidaan aiheuttaa, tunnistettiin nopeasti. Erityisesti valtionpäämiehistä tehdyt audio-visuaaliset väärennökset on nähty tulenarkoina. Valtionpäämiehillä on valtaa, jota voidaan pyrkiä hyväksikäyttämään väärennöksillä.

Arkisessa käytössä olevat sosiaalisen median kasvoja muokkaavat filterit ovat esimerkki siitä, kuinka tällainen teknologia on jo arkipäiväistynyt. Filtereiden avulla kuka tahansa sosiaalisen median käyttäjä voi dramaattisesti muokata ulkonäköään tai sijoittaa itsensä täysin toiseen ympäristöön virtuaalisesti. Samalla hän mahdollisesti tallentaa kuvansa ja tietonsa kolmannen osapuolen palveluun, joka käyttää kuvia kehittääkseen tekoälyä edelleen ja taltioi niitä tietokantoihin, joissa kuka tahansa voi käydä maksum vastaan tarkastamassa kuvan ihmisen henkilöllisyyden.

Informaatio- ja psykologisen vaikuttamisen näkökulmasta tekoälykehitys ei palvele vain valtiollisia tarpeita, vaan toimii suorana jatkumona muulle informaatioteknologian kehitykselle, joka mahdollistaa nykyisenkaltaisen elämäntyylin digitaalisine palveluineen. Kun tekoälyä soveltavat palvelut arkipäiväistyvät sosiaalisen median filtereiksi ja erilaisiksi tiedonhakupalveluiksi, kokee tekoälyyn perustuva maalittaminen ja

vaikuttaminen samanlaisen "demokratisoitumisen", kuin muukin sosiaalisessa mediassa tapahtuva joukkoistettu häirintä. Nykymalli, jossa yksilön henkilöllisyys on vahvasti sidoksissa viestivälineisiin ja käyttäjätileihin sormenjälkien, kasvokuvien, puhelinnumeroiden, sähköpostiosoitteiden ja muiden tunnisteiden kautta, rikkoo palveluiden käyttäjän yksityisyyttä tavoilla, jotka mahdollistavat hyvinkin vahingollisen vihamielisen vaikuttamisen ja tietojen hyväksikäytön.

Poikkeusoloissa ja sotatilassa kohteen verkostoon, kuten ystäviin ja perheeseen, voidaan kohdistaa vaikuttamista. Näin esimerkiksi Venäjä on toiminut Ukrainassa, jossa ukrainalaisten rintamasotilaiden perheenjäsenille on lähetetty tekstiviestejä, joissa sotilaan on väitetty kuolleen. Maalikuussa 2022 Ukraina ilmoitti, että ukrainalaiset käyttävät Clearview-kasvojentunnistusohjelmaa tunnistamaan kaatuneet venäläissotilaat, jotta heidän omaisilleen voidaan toimittaa tieto kaatumisesta.

Suurin osa meistä on jakanut kuvia ja tietoja erilaisten digitaalisten alustojen käyttöön, jolloin tietomme on osa dataa, josta erilaiset palvelut tuottavat koosteita tiedon ostajille. Yksityisyyden ja tietosuojan romahtaminen tekoälypalveluiden myötä voi tulevaisuudessa yhdistyä tekoälypohjaiseen vaikuttamiseen, jossa rintamasotilaan omaiselle ei vain lähetetä tekstiviestiä tai kuvaa, vaan heidän puhelinnumeroihinsa ohjataan puheluita, joissa tekoäly kykenee keskustelemaan omaisten kanssa siten, että omaiset uskovat keskustelewansa oikean ihmisen kanssa – tai jopa heille tutun ihmisen kanssa, jonka ääni on kloonattu deep fake -teknologialla.

Henkilötietoja ja metadataa voidaan käyttää paitsi identiteettivarkauksiin, myös esimerkiksi viranomaisten, toimittajien, yritysten avainhenkilöiden tai kenen tahansa häirintään ja vainoamiseen. Tai sitten yksinkertaisesti harhauttamiseen: kuinka moni toimittaja menisi vipuun keskusteltuaan tietämättään kloonatun äänen kanssa, ja kertoisi sitten yleisölle saaneensa tietoja ”korkealta taholta”?

Tekoälybotit tulevat keräämään ja käyttämään ihmisten tuottamaa metadataa heitä vastaan sekä vaikuttaakseen heidän käyttäytymiseensä. Kenties keskustelit huomamattasi tekoälybotin kanssa keskustelupalstalla ja täydennät näin tietokantaa, joka sisältää jo vuosien mittaan kertyneen keskusteluhistoriasi, selaustietosi, tykkäyksesi ja verkostosi.

Toisaalta tekoälyä voidaan hyödyntää myös vaikuttamisen paljastamisessa. Helmikuussa 2022 The New York Times haastatteli kahden eri tutkimusryhmän tutkijoita, jotka olivat selvittäneet QAnon-salaliittoteorian ja sitä seuranneen liikkeen perustajan henkilöllisyyttä. Tutkijoiden mukaan henkilöllisyydet kyettiin selvittämään 98 ja 99 prosentin varmuudella stylometrian, eli kirjoittajan tyyliä arvioivan tilastoanalyysin avulla. Toinen ryhmistä pilkkoi salaliittoteorian perustajan, eli Q:n, viestit kolmen merkin pituisiin jonoihin ja tilastoi, kuinka usein kukin jono esiintyi Q:n teksteissä. Toinen tutkimusryhmä käytti sen sijaan tekoälyä Q:n tekstien maneerien analysoimiseen. Vastaavaa koneoppimisalgoritmia on käytetty aiemminkin. Esimerkiksi salanimellä julkaistu rikosromaani *Käen kutsu* on tunnistettu J.K. Rowlingin kirjoittamaksi (Kailio, 2022).

Keskeistä tekoälyyn perustuvassa vaikuttamisessa on siis se, millä tahdilla vaikuttamisen ja sen tunnistamisen mahdollistavat teknologiat kehittyvät ja millä nopeudella ne siirtyvät laajemmin käytettäviksi. Niin kauan, kun esimerkiksi väärennetyn sisällön tunnistamiseen tarvitaan koulutettu tutkijaryhmä, on väärennettyä sisältöä tuottava tekoäly vahvemmillä. Olipa kyse disinformaation tunnistamisesta tai attribuutiosta, kaiken keskiössä on data, jota väistämättä tuotamme osallistuessamme ja kommunikoidessamme digitalisoituneessa informaatioulostuvuudessa.

Laaja-alainen mikrotargetointi: data vs. yksityisyys

Vaikka psykologinen vaikuttaminen on kuulunut sodankäynnin ja poliittisen vaikuttamisen keinovalikoimaan läpi historian, vasta internet ja henkilökohtaisten viestintävälineiden käytön lisääntyminen on tehnyt siitä niin tehokkaan, että se voidaan rinnastaa asejärjestelmiin. Tekoäly on luonnollinen osa tätä kehityskaarta. Robotiikka, autonomisuus ja tekoäly ovat asejärjestelmien nykyinen kehitysaste, joten tekoälyn hyödyntäminen myös psykologisessa vaikuttamisessa on osa tätä kehitystä.

SCL Group ja Cambridge Analytica ovat tästä kehityksestä hyviä esimerkkejä. Ne tunnetaan pääasiassa Yhdysvaltain vuoden 2016 presidentinvaaleihin ja Brexit-äänestykseen liittyvistä kampanjoinnista, joissa yritys hyödynsi sosiaalisen median käyttäjien dataa vaalikampanjan mikrotargetoinnissa. Näkemykset kampanjan onnistumisesta ovat moninaisia. Yleisesti ajatellaan, että profiloituyrityksistä huolimatta kampanja ei lopulta määrittänyt vaalien lopputulosta, mutta todennäköisesti vaikutti siihen (Privacy International, 2017). SCL Groupilla oli ennen vuotta 2016 pitkä historia nimenomaisesti vaaleihin liittyvässä kampanjoinnissa, ja se on kehittänyt metodejaan erityisesti kehittyvissä maissa. Kohdeyleisöstä tehdyt analyysit perustuivat alun perin melko pinnallisille tiedoille, kuten asuinpaikalle. Vasta sosiaalisen median tuottamaa dataa hyödyntämällä kyettiin kohdeyleisön persoonallisuuden profilointiin massamaisesti: Cambridge Analytica kykeni keräämään 50-60 miljoonan ihmisen Facebook-datan muutamassa kuukaudessa ja suorittamaan mikrotargetoinnin tekoälyavusteisesti (Warren, 2021).

Todistamme siis täysin uudenlaista yksityisyyden alasajoa. Yksityisyyden käsite asettuu räikeään ristiriitaan tämän päivän ”datakapitalismin” kanssa, joka on muotoutunut yhä tarkemman datankeruun ympärille. Yksilön identiteetti on yhä enemmän kytköksissä jatkuvan digitaalisen vuorovaikutuksen tarpeeseen, jossa hän luo itsestään narratiiveja digitaalisille alustoille. Tämän päivän ihminen kokee tarvetta määritellä ja tarinallistaa elämänsä osa-alueita, joita ei ennen ole koettu tarpeelliseksi määritellä ja tarinallistaa, ja tuotteistaa elämänsä osa-alueita, joita ei ennen ole ollut tarvetta tuotteistaa (Törnberg and Uitermark, 2022). Monesti tämä palvelee yksilön psykologisia perustarpeita, kuten hyväksynnän, merkityksellisyyden ja sosiaalisen liittymisen tarpeita, jolloin ajatus tällaisesta identiteettiprojektista luopumisesta tuntuu vieraalta. Tämä takaa siis jatkossakin jatkuvan yksityisasioiden datavirran verkkoon.

Kun yksilön tiedot on kertaalleen verkkoon syötetty, ei niitä saa sieltä pois, vaikka oma elämäntilanne muuttuisi, tai vaikka normaaliolot muuttuisivat poikkeusoloiksi tai jopa sotatilaksi. Tällöin kaikkea saatavilla olevaa dataa voidaan aktiivisesti hyödyntää

massamaisessa, yksilöön kohdistuvassa vaikuttamisessa. Siinä missä yksilön maalittaminen ja häneen vaikuttaminen on perinteisesti ollut aikaa vievää käsityötä, lisääntyvien datamassojen myötä tekoäly mahdollistaa nopean, yksilötasolle kohdennetun ja laajoihin massoihin ulottuvan vaikuttamisen.

Juuri tämä on tekoälyn vahvuus. Tekoäly tietää ihmistä paremmin, millainen viestintä kohdeyleisöön tai kohdehenkilöön tehoaa ja millaisista viesteistä tulee varmemmin viraaleja sosiaalisessa mediassa. Tulevaisuudessa tekoälyllä generoitu sisältö paitsi harhauttaa, myös mobilisoi ja lamauttaa kohdeyleisöjä tunteita manipuloimalla, sillä se tuntee kohteensa identiteetin tarkkuudella. Tämä kaikki vaatii kuitenkin datataloutta, mikä on ydinbisnestä yhtiöille, jotka keräävät asiakkaidensa ja käyttäjiensä tietoja ja myyvät sitten palvelujaan mm. valtiollisille toimijoille.

Kognitiivinen ylivoima

Sotilasorganisaatioita tekoäly luonnollisesti kiinnostaa paitsi tiedusteluanalyysin ja psykologisen vaikuttamisen vuoksi, myös omien suorituskykyjen takia. Tekoälyllä voidaan kehittää omaa kognitiivista suorituskykyä, millä voi olla dramaattinen vaikutus vastustajan taistelutahtoon ja moraaliiin.

Tulevaisuuden konflikteissa taistelukentän uskotaan laajentuvan perinteisistä fyysisestä ja informaatioulottuvuudesta tietoisuuteen. Tällöin dominanssi fyysisissä suorituskyvyissä ei riitä taistelumenestykseen. Aivo-tietokoneiliittymät mahdollistavat tulevaisuudessa paitsi tekoälyn hyödyntämisen johtamisessa, myös esimerkiksi dronen tai robotin operoinnin ihmisaivojen avulla (Kania, 2019). Robotiikkaa, eksoskeletooneja ja lääketieteellistä ”parantelua” on jo pitkään tutkittu ja kehitetty keinona parantaa yksittäisten sotilaiden suorituskykyä. Kuitenkin ajatus hyökkäyksestä sellaista vihollista vastaan, jonka suorituskyvyt tukeutuvat voimakkaasti tekoälyyn, voi olla pelotteena täysin uudella tasolla. Jos esimerkiksi tietoturvayritykset hyödyntävät tekoälyä yritysten verkkoliikenteen poikkeavuuksien tunnistamiseen, voi kynnys yritykseen kohdistuvalle kyberhyökkäykselle nousta. Vastaavasti tavanomaisista inhimillisistä rajoitteista vapaa johtamisjärjestelmä voi heikentää vihollisen hyökkäyshalukkuutta. Kognitiivisesti ylivoimainen vastustaja kykenee nopeampiin ja parempiin päätöksiin, ja tähän pyritään tulevaisuudessa fuusioimalla ihmisaivot ja hermoverkot tekoälyyn.

Kognitiivisella dominanssilla on niin ikään sekä offensiivinen että defensiivinen potentiaali. Offensiivisia kykyjä kehittävät valtiot ovat yhtä lailla huolissaan kehittämänsä teknologian uhkista ja riskeistä omalle turvallisuudelleen. Kuten muussakin vaikuttamisessa, tätä kehitystä leimaa teknologian ja sen vastateknologioiden kilpajuoksu.

Kuten sotilasprofessori, everstiluutnantti Aki-Mauri Huhtinen aikanaan totesi: ”Kaksikymmentä vuotta sitten panikoimme siitä, että sähköpostimme alkoi ilmestyä roskapostia. Sähköpostipalveluita tuottavat yhtiöt kehittivät kuitenkin nopeasti filtrit, jotka lähes poistivat roskapostiongelman. Tämä sama ilmiö toistuu uudestaan ja uudestaan uusien teknologioiden kehittyessä.” Tämä on syytä pitää mielessä aina, kun puhumme uusien teknologioiden muodostamista uhkista. Kaiken keskiössä on kuitenkin lopulta ihmis mieli, joka on yhtä aikaa toimintakykymme vahvin ja heikoin lenkki.

Bargh, J.A. and Chartrand, T.L. (1999) 'The unbearable automaticity of being', *American Psychologist*, 54(7), pp. 462–479. Available at: <https://doi.org/10.1037/0003-066X.54.7.462>.

Facebook (2021) *Threat Report: The State of Influence Operations 2017-2020*. Available at: <https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf>.

Kahneman, D. (2011) *Thinking, fast and slow*. First paperback edition. New York: Farrar, Straus and Giroux (Psychology/economics).

Kailio, A. (2022) *Tekoäly paljasti internetin pahamaineisimman salaliittoteorian isän 98 % varmuudella, Tivi*. Available at: <https://www.tivi.fi/uutiset/tekoaly-paljasti-internetin-pahamaineisimman-salaliittoteorian-isan-98-varmuudella/fe709db0-8a0e-4e5b-94f2-6ed80810ba85> (Accessed: 21 September 2022).

Kania, E.B. (2019) 'Minds at War: China's Pursuit of Military Advantage through Cognitive Science and Biotechnology', *PRISM*, 8(3), pp. 82–101.

Meta (2022) *January 2022 Coordinated Inauthentic Behavior Report*. META. Available at: <https://about.fb.com/news/2022/02/january-2022-coordinated-inauthentic-behavior-report/>.

Nimmo, B. *et al.* (2020) *Secondary Infection*. Graphika. Available at: <https://secondary-infektion.org/downloads/secondary-infektion-report.pdf>.

Nolen-Hoeksema, S. *et al.* (eds) (2009) *Atkinson & Hilgard's introduction to psychology*. 15. ed. Hampshire: Cengage Learning EMEA.

Privacy International (2017) *Cambridge Analytica Explained: Data and Elections*. Available at: <http://privacyinternational.org/long-read/975/cambridge-analytica-explained-data-and-elections> (Accessed: 21 September 2022).

Repucci, S. and Slipowitz, A. (2022) *The Global Expansion of Authoritarian Rule*. Freedom House. Available at: <https://freedomhouse.org/report/freedom-world/2022/global-expansion-authoritarian-rule> (Accessed: 21 September 2022).

Törnberg, P. and Uitermark, J. (2022) 'Tweeting ourselves to death: the cultural logic of digital capitalism', *Media, Culture & Society*, 44(3), pp. 574–590. Available at: <https://doi.org/10.1177/01634437211053766>.

United States of America v. internet Research Agency LLC (2018). Available at: <https://www.justice.gov/archives/sco/file/1035477/download>.

Warain, S. (2021) 'Role of AI in Psychological Warfare', *Geek Culture*, 29 June. Available at: <https://medium.com/geekculture/role-of-ai-in-psychological-warfare-73ec0c98515a> (Accessed: 21 September 2022).

Kirjoittajat

Mireille Farzan, FM, väitöskirjatutkija, Turun yliopisto

Ippo Halonen, FT, Helsingin yliopisto

Saara Jantunen, ST, erityisasiantuntija Valtioneuvoston kanslia, viestintäosasto

Eero Kallio, Komentaja, Merisotakoulu

Lauri Kilpeläinen, SM, Maasotakoulu, Sotatieteiden opetusryhmä

Riitta Penttinen, TkT, tutkimuspäällikkö, Ilmavoimat

Arto Mutanen, FT, dos. Merisotakoulu & Maanpuolustuskorkeakoulu

Janne Mutanen, DI.

Antti Rissanen FT, tutkija, Sotatekniikan laitos

Ville Rissanen, Ruslan Lagashkin*, Einar Eidstø**, Tenho Korhonen*, Emil Toivonen*** & Jaakko Pentikäinen, Tutkimusavustajina / tutkimusryhmässä vuonna 2020 Sotatekniikan laitoksella. Muut sidokset: *Aalto yliopisto, **Helsingin yliopisto, *** Åbo Akademi

Kalle Saastamoinen, FT, tutkija, Sotatekniikan laitos

Jouko Vankka, TkT, sotatekniikan professori, Sotatekniikan laitos

Maanpuolustuskorkeakoulu

Sotatekniikan laitos
PL 7, 00861 HELSINKI

Puh. +358 299 800

www.mpkk.fi

ISBN 978-951-25-3332-9 (nid.)

ISBN 978-951-25-3333-6 (pdf)

ISSN 2737-0615 (verkkajulkaisu)

SOTATAIDON YTIMESSÄ



Puolustusvoimat
The Finnish Defence Forces