

# **Tekoälyn käytön eettisiä kysymyksiä sotilasjohtamisessa, erityisesti käytettäessä tappavia autonomisia asejärjestelmiä** (vertaisarvioitu)

**Matti Häyry**

## **Abstract**

The ethics of warfare and military leadership must pay attention to the rapidly increasing use of artificial intelligence and machines. Who is responsible for the decisions made by a machine? Do machines make decisions? May they make them? These issues are of particular interest in the context of Lethal Autonomous Weapon Systems (LAWS). Are they autonomous or just automated? Do they violate the international humanitarian law which requires that humans must always be responsible for the use of lethal force and for the assessment that civilian casualties are proportionate to the military goals?

The article analyses relevant documents, opinions, government positions, and commentaries using the methods of applied ethics. The main conceptual finding is that the definition of autonomy depends on what the one presenting it seeks to support. Those who want to use lethal autonomous weapons systems call them by another name, say, automated instead of autonomous. They impose standards on autonomy that machines do not meet, such as moral agency. Those who wish to ban the use of lethal autonomous weapon systems define them broadly and do not require them to do much more than to be a self-standing part of the causal chain.

The article's argument is that the question of responsibility is most naturally perceived by abandoning the most controversial philosophical considerations and simply stating that an individual or a group of people is always responsible for the creation of the equipment they produce and use. This does not mean that those who press the button, or their immediate superiors, are to blame. They are doing their jobs in a system. The ones responsible can probably be found in higher military leadership, in political decision-makers who dictate their goals, and, at least in democracies, in the citizens who have chosen their political decision-makers.

## Johdanto

Sodankäynnin ja sotilasjohtamisen etiikassa on otettava huomioon tekoälyn ja koneiden lisääntyvä käyttö. Kuka on vastuussa koneen tekemistä päätöksistä? Tekeekö kone päätöksiä? Saako se tehdä niitä? Kysymykset ovat erityisen kiinnostavia tarkasteltaessa tappavia autonomisia asejärjestelmiä. Ovatko ne autonomisia vai vain automatisoituja? Rikkovatko ne sitä kansainvälisen ihmisoikeuslain periaatetta, että ihmisen on aina oltava vastuussa tappavan voiman käytöstä ja siviiliuhrien oikean määrän arvioinnista suhteessa tavoiteltuihin sotilaallisiin päämääriin?

Artikkelissa analysoidaan soveltavan etiikan metodein asiaa koskevia dokumentteja, lausuntoja, viranomaiskantoja ja kommentaareja. Käsitteellinen päälöydös on, että autonomian määrittelyn sanelee keskustelussa se politiikka, jota itse kukin haluaa tukea. Ne, jotka haluavat käyttää tappavia autonomisia asejärjestelmiä, kutsuvat niitä jollain muulla nimellä, esimerkiksi automatisoiduiksi. He asettavat autonomisuudelle vaatimuksia, joita koneet eivät täytä, kuten moraalisen toimijuuden. Ne, jotka haluavat kieltää tappavien autonomisten asejärjestelmien käytön, määrittelevät ne paljon väljemmin eivätkä edellytä niiltä juuri muuta kuin omatoimista kausaalista osallisuutta seurausten aiheuttamisessa.

Artikkelin väite on, että vastuukysymys hahmottuu luontevimmin, kun luovutaan kiistanalaisimmista filosofisista pohdinnoista ja todetaan vain, että joku ihminen tai jokin ihmisjoukko on aina vastuussa tuottamiensa ja käyttämiensä laitteiden aikaansaannoksista. Tämä ei tarkoita, että syylliset nimettäisiin napin painajista tai heidän sotilasjohtajistaan. Nämä toimivat useimmiten melko pitkälti järjestelmän ehdoilla. Oikeat vastuulliset löytyvät korkeammalta sotilasjohdosta, poliittisista päättäjistä, jotka sanelevat heidän päämääränsä, ja ainakin toimivissa demokratioissa kansalaisista, jotka ovat valinneet poliittiset päättäjänsä.

## Tehtävä

Sotilasjohtaminen nojautuu jo tietyssä määrin koneisiin ja autonomisiin järjestelmiin. Jotkut näistä perustuvat tekoälyyn, joka näyttää tekevän päätöksiä ihmisten puolesta. Tulevaisuudessa tällaisten koneiden ja järjestelmien käyttö voi huomattavasti lisääntyä. Ne eivät kaikki liity sotilaalliseen päätöksentekoon tai sotilasjohtamiseen. Tappavat autonomiset asejärjestelmät (*Lethal Autonomous Weapon Systems*, LAWS) ovat kuitenkin lähes pelkästään sotilaskäyttöön tarkoitettu sovellus (vaikka niitä voitaisiin periaatteessa käyttää myös lainvalvonnassa). Valitsemani tehtävä on ottaa selvää, minkälaisia eettisiä kysymyksiä

meidän pitäisi ottaa huomioon pohtiessamme LAWSien ja vastaavien tekoälypohjaisten järjestelmien ottamista käyttöön.

## **Yleinen taustaetiikka ja sen laajennukset**

Isaac Asimovin ”robotiikan lait” tarjoavat hyvän lähtökohdan koneiden ja tekolyn etiikkaan, vaikka ne ovatkin fiktiivistä alkuperää, novellista *I, Robot* (1950). Asimovin lait ovat:

1. Robotti ei saa teoillaan vahingoittaa ihmisiä eikä tekemättä jättämisillään antaa ihmisten vahingoittua.
2. Robotin on noudatettava ihmisten antamia käskyjä, paitsi silloin, kun niiden noudattaminen rikkoisi ensimmäistä lakia.
3. Robotin on suojeltava omaa olemassaoloaan, kunhan tämä ei ole ristiriidassa ensimmäisen ja toisen lain kanssa.

Yksityiskohtaisempia esityksiä aiheesta ovat viime aikoina julkistaneet muiden muassa British Engineering and Physical Sciences Research Council ja Arts and Humanities Research Council robotiikan periaatteissaan (EPSRC 2011) sekä Satya Nadella omassa sääntöluettelossaan (Nadella 2016). Näissä on otettu huomioon tappamisen kieltö, inhimillisen vastuun vaatimus, turvallisuuden korostus, robottien pitäminen riittävän etäisinä ja konemaisina, läpinäkyvyyden ja ymmärrettävyyden edellytys sekä inhimillisen autonomian, ihmisarvon, yksityisyyden ja yhtäläisyyden kunnioitus. Tällaiset yleiset säännöt eivät tietenkään koske vain tekoälyn tai tekoälypohjaisten laitteiden sotilaskäyttöä. Siihenkin ne silti soveltuvat.

## **LAWS ja yleinen etiikka**

LAWS näkyvät törmäävän ensimmäiseen esteeseen mikä tahansa tavanomaisista lähestymistavoista valitaankin. Ne surmaavat ihmisiä, kun niitä käytetään oikeaan tarkoitukseensa (= ”Lethal”). Yhdistyneet Kansakunnat (Gayle 2019), Euroopan Parlamentti ja monet tekoälyn kehittäjät ja tutkijat ovatkin pyrkineet niiden kieltämiseen, samoin roomalaiskatolinen kirkko (josta lisää alla).

Kaikkeen sodankäyntiin liittyy kuitenkin ihmisten vahingoittamista ja surmaamista, ja siksi perinteiset oikeutetun sodan teoriat voivat pelastaa LAWSit, joskin tarkoin reunaehdoin. Kaksoisvaikutuksen opin mukaan hyvään päämäärään (esim. valtion kansalaisten suojaamiseen sotilaalliselta aggressiolta) saa

tiettyjen ehtojen vallitessa pyrkiä siitä huolimatta, että samalla aiheutuu pahaakin (McIntyre 2019; Lazar 2020). Ehtoja ovat, että paha vaikutusta (siviilien loukkaantumista ja hengenmenetystä) ei aktiivisesti tarkoiteta (ellei heitä jostain syystä menehdy iskuissa, heitä ei lähdetä erikseen tavoittelemaan) ja että hyvä ja paha vaikutus ovat oikeasuhtaisia.

Kun näitä kahta *jus ad bellum* – ehtoa sovelletaan LAWSien *jus in bellum* -oikeutukseen, esiin nousee kaksi tarkennusta:

- LAWSien täytyy erottaa taisteluun osallistuvat ja osallistumattomat mahdollisimman luotettavasti.
- LAWSien taisteluun osallistumattomille aiheuttamien vahinkojen pitää olla oikeassa suhteessa saavutettavaan sotilaalliseen tavoitteeseen.

Näitä vaatii myös kansainvälinen humanitaarinen oikeus. Henkilömiinojen käyttökielto Ottawan sopimuksessa perustuu tälle logiikalle, vaikkakaan miinoja ei ilman tekoälyn osallisuutta luokitella LAWSeiksi.

## **Laajennetusta huolesta erityiskysymykseen**

Roomalaiskatolinen kirkko vastustaa LAWSien käyttöä, koska autonomisten järjestelmien pelätään epäinhimillistävän sodankäyntiä. Valottaessaan kirkon kantaa LAWSien epäeettisyydestä, Alice de la Rochefoucauld esitti joukon kysymyksiä:

Kykenevätkö koneet korvaamaan ihmisen elämää ja kuolemaa koskevassa päätöksenteossa ja onko tämä yhteensopivaa kansainvälisen humanitaarisen oikeuden kanssa? Voivatko koneet olla vastuussa kansainvälisen oikeuden loukkauksista? Voiko koneen toiminta, eettisesti ottaen, korvata ihmisen kyvyn moraaliseen harkintaan? (de la Rochefoucauld 2017.)

Yksiselitteinen vastaus kaikkiin näihin on de la Rochefoucauldin mukaan ”Ei”, ja siitä seuraa, että LAWSien ja muiden niiden kaltaisten koneiden hyödyntäminen tekisi sodankäynnin epäpersoonalliseksi ja epäinhimilliseksi. Kun jätetään syrjään metafysiset sävyt (perinteisen sodankäynnin romantisoitu ”inhimillisuus”) konkreettinen huolenaihe tässä lienee, että yhä useampi viaton siviili kuolee eikä ketään voida pitää siitä vastuussa.

Näin muotoiltuna ongelmaan näyttäisi ainakin teoriassa olevan ilmeinen ratkaisu. Jos kansainvälisen humanitaarisen oikeuden loukkauksia sattuu,

niistä ovat vastuussa ne ihmiset, jotka tuottavat ja käyttävät LAWSeja. He voivat vaikuttaa järjestelmän toimintaan pitkäänkin joko päätösten tekijöinä tai niiden valvojina ja mahdollisina kumoajina, mutta tällä ”in the loop”-, ”on the loop”- tai ”out of the loop”-positiolla ei ole normatiivista merkitystä. Koneet eivät pysty moraaliseen harkintaan. Ihmiset pystyvät. Viimeinen ihminen, joka tekee merkityksellisen, autonomisen päätöksen on vastuussa.

Vai voiko asia olla näin yksinkertainen? Pitäisikö de la Rochefoucauld’n kysymyksiä katsoa tarkemmin ja ottaa ne kirjaimellisina eikä retorisisina?

## **Inhimillinen autonomia ja auktoriteetti**

Moraalifilosofiassa tunnustetaan kaksi *inhimillisen* autonomian muotoa. Individualistisen käsityksen mukaan päätökset ovat autonomisia jos ja vain jos toimijat tekevät ne perustuen omaan harkintaansa, uskomuksiinsa ja arvoihinsa eivätkä heitä rajoita ulkoiset sosiaaliset tai kulttuuriset voimat. Relationaalisempien näkemysten mukaan päätökset ovat autonomisia jos ja vain jos toimijat tekevät ne perustuen omaan harkintaansa, uskomuksiinsa ja arvoihinsa ottaen asianmukaisesti huomioon oikeat sosiaaliset ja yhteisölliset normit ja arvot. (Häyry & Takala 2016.) On oma kysymyksensä, pitäisikö autonomisten päätösten myös olla hyvin informoituja. Ainakin pätevän suostumuksen kohdalla terveydenhuollossa ja tieteellisessä tutkimuksessa nämä vaatimukset pidetään erillään, ja yksilöiden päätöksiä pidetään pätevänä vain, jos he tekevät ne vapaasti (ilman ulkoista voimaa tai pakottamista) ja informoidusti ja autonomisesti. Autonomiaan ei siis ainakaan aina kuulu hyvän tietämyksen vaatimus.

Päätöksentekoon sotilasjohtamisessa kuuluu tyypillisesti kahdenlaisen auktoriteetin yhdistyminen (Flathman 1980). Ensimmäinen näistä on *asiantuntijuuden* auktoriteetti, joka perustuu tietoon. Joukkoyksiköissä ja sitä isommissa yhtymissä aselajipäälliköt tarjoavat sen informaation ja tiedon, joka operatiota koskevaan päätökseen tarvitaan. Komentaja esikuntineen yhdistää eri lähteistä tulevan tiedon ja tekee (autonomisesti) päätöksen, jonka velvoittavuus perustuu komentajan *aseman* auktoriteettiin. Tilanne on hieman erilainen pienemmissä yksiköissä ja muuttuu radikaalisti tultaessa niihin salamavalintoihin, joiden tekijä on kirjallisuudesta kuuluisa ”strateginen korpraali” ja joilla voi paikallisuudestaan huolimatta olla maailmanlaajuisia seurauksia (Krulak 1999; Lovell & Baker 2017). Asiantuntijuus on edelleen läsnä, mutta se on huomattavasti pienemmässä roolissa kuin korkeamman tason kabinettiratkaisuissa.

## Koneen autonomia – inklusiivinen tulkinta

Aseiden ja asejärjestelmien autonomialle on erilaisia tulkintoja, joiden suunnan sanelevat niiden tekijöiden normatiiviset sitoumukset ja käytännön tarpeet. Ne, jotka haluavat rajoituksia ja kieltoja, tulkitsevat käsitteen inklusiivisesti (teh-däkseen mahdollisimman monista asejärjestelmistä ”autonomia”), kun taas ne, jotka vastustavat jyrkkiä kieltoja, luottavat eksklusiivisuuteen (rajoittaakseen ”autonomian” mahdollisimman harvoihin järjestelmiin).

Rauhantutkija Mark Gubrud on inklusiivisuuden puolestapuhuja (Gubrud 2014). Deskriptiivisenä lähtökohtanaan hän pitää Yhdysvaltojen Puolustusministeriön määritelmää, jonka mukaan autonominen asejärjestelmä (*Autonomous Weapon System*, AWS) on:

Asejärjestelmä, joka, kun se on aktivoitu, voi tunnistaa [select] kohteita ja sitoutua [engage] niihin ilman inhimillisen käyttäjän lisäpuuttumista. (DoD 2012.)

Asejärjestelmän autonomia tarkoittaa tässä sitä, että kun inhimillinen päätös on tehty, kone ”ottaa ohjat” ja tekee uusia ”tunnistamis”- ja ”sitoutumis”-päätöksiä, joita käyttäjähimiset eivät enää voi hallita. Voisimme toki sanoa, että tällaisissa tapauksissa vastuu kuuluu ketjun viimeiselle ihmiselle. Gubrud ei kuitenkaan omaksu tätä kantaa. Hän näyttää ajattelevan, että inhimillinen vastuu haihtuu, kun kone tekee omia päätöksiään. Koska hän myös uskoo, että ”aseiden ja konfliktien on aina oltava ihmisten hallinnassa”, hänen normatiivinen johtopäätöksensä on selvä. Autonomiset asejärjestelmät pitäisi lähtökohtaisesti kieltää, ja jokaisen, joka haluaa käyttää niitä, on esitettävä valinnalleen erillinen oikeutus.

Gubrud jatkaa arvostelemalla puolustusministeriön esitystä erottaa toisistaan (hyväksyttävät) puoliautonomiset asejärjestelmät (*Semi-Autonomous Weapon Systems*, SAWS) ja (tuomittavat) kokonaan autonomiset asejärjestelmät (*Fully Autonomous Weapon Systems*, FAWS tai AWS). Joissakin puoliautonomisissa asejärjestelmissä ihmisten on tehtävä vain viimeinen ”sitoutumis”-päätös. Gubrud huomauttaa, että sellaiset järjestelmät ovat muunnettavissa kokonaan autonomisiksi liian helposti erottelun turvallisuuden kannalta. ”Fire and forget”-ratkaisuisissa minimaalinenkin inhimillinen osallisuus on poistettu, kun laukaisun jälkeen kone tunnistaa maalinsa ja sitoutuu siihen itsenäisesti. Puolustusministeriön kannan mukaan laukaisuun johtavat laskennat muodostavat ratkaisevan inhimillisen päätöksen, Gubrud on eri mieltä ja tästä eteenpäin kiista muuttuu semanttiseksi – sanojen merkityksiin liittyväksi. Ovatko puoliautonomisten asejärjestelmien ”tunnista”- ja ”sitoudu”-toiminnot autonomia koneen tekemiä päätöksiä vai eivät? Yksi mahdollisuus olisi, että ne ovat tiettyssä

mielessä autonomisia (kone toimii ilman ulkoista hallintaa) mutta eivät aitoja päätöksiä (siinä mielessä, että jonkun pitäisi olla niistä vastuussa).

## **Koneen autonomia – eksklusiivinen tulkinta**

Britannian puolustusministeriö käyttää tiukempaa seulaa määritellessään asejärjestelmien autonomisuuden:

Autonominen järjestelmä pystyy ymmärtämään korkeamman tason intentionaalisuutta ja suuntautuneisuutta. Tämän ymmärryksen ja ympäristön havainnoinnin avulla sellainen järjestelmä kykenee ryhtymään asianmukaiseen toimintaan saadakseen aikaan toivotun tuloksen. Se pystyy päättämään toimintavaihtoehdon, monien vaihtoehtojen joukosta, tukeutumatta ihmisten valvontaan ja hallintaan, vaikka nämä vielä voivatkin olla läsnä. Vaikka autonomisen miehittämättömän ilma-aluksen toiminta pääpiirteissään on ennustettavissa, sen yksittäiset toimet eivät ehkä ole. (MoD 2017, 13.)

Viimeinen, ennustettavuutta koskeva lause asettaa vastakkain autonomiset ja automaattiset järjestelmät. Automaattiset tuottavat vain tuloksia, jotka tiedämme, kunhan tunnemme säännöt, joita järjestelmä on säädetty noudattamaan. Autonomiset ylittävät nämä ja tuottavat ennustamattomia tuloksia, joista vastuu kuuluu koneälylle. Mikäli hyväksymme sen normatiivisen lähtökohdan, että koneiden ei pidä tehdä moraalisesti tai laillisesti vastuunalaisia päätöksiä, aidot autonomiset asejärjestelmät pitäisi kieltää. Mutta kieltö voidaan kiertää määrittelemällä niiltä näyttävät asejärjestelmät ”kaukoautomatisoiduiksi” (Remote Automatic). Niin ministeriö näyttää tekevän nimeämällä miehittämättömät ilma-alukset juuri tähän luokkaan kuuluviksi, toteamalla, että kyseessä on vain yksi tulkinta kansainvälisestä oikeudesta, mutta pitäen kiinni siitä, että tulkinta on oikeutettu.

Määritelmän ensimmäinen lause tukee ajatusta edellyttämällä, että aidosti autonomiset järjestelmät pystyvät ”ymmärtämään korkeamman tason intentionaalisuutta ja suuntautuneisuutta”. Ei tietenkään ole itsestään selvää, mitä ”ymmärtäminen” ja ”korkeamman tason intentionaalisuus ja suuntautuneisuus” tarkoittavat, mutta ne tuntuvat vaativan ominaisuuksia, joita vain autonomisilla inhimillisillä toimijoilla voi olla. Kynnys asettuu niin korkealle, ettei mikään sieluton ja mieletön kone voi ylittää sitä.

## Tunnista, sitoudu ja koneen auktoriteetti

Sekä LAWSien arvostelijat että kannattajat näyttävät olevan sitä mieltä, että niillä voisi olla kummankinlaista komentajan ratkaisuun liittyvää auktoriteettia. Sen jälkeen, kun koneet ovat jättäneet taakseen inhimillisen valvonnan, ne tunnistavat kohteensa itsenäisesti eli niillä on *asiantuntijuuden auktoriteettia*. Ja kun tunnistus on tehty, kone tuntuisi tekevän vielä lisäpäättöksen sitoutumisesta kohteeseen. Tämä toinen askel, kone käyttämässä *aseman auktoriteettia*, olisi eettisesti hälyttävä, mutta se voi olla antropomorfinen korjattavissa oleva virhetulkinta.

Asiantuntemuksen auktoriteetti ”tunnista”-vaiheessa pitäisi oikeastaan ilolla antaa hyvin toimivien koneiden hoidettavaksi. Vaihtoehdot voivat olla vaikkapa räjähtävä tykinammus, jonka on vastuullisesti lähettänyt matkaan ihminen, ja ihmisen ulottumattomissa oleva miehittämätön ilma-alus, jolla on vastaava tulivoima. Kumpi olisi parempi kohdata? Kun tykinammus saavuttaa kohteensa – minut, viattoman siviilin – se räjähtää ja surmaa. Kun oikein ohjelmoitu drooni tulee perille, se tarkastaa vielä kerran statukseni, lukee sen oikein ja menee pois. Kansainvälisen humanitaarisen oikeuden kirjain ja ehkä henkikin olisivat paremmassa turvassa konevaihtoehdossa, ilman inhimillistä tekijää.

Aseman auktoriteetti ”sitoudu”-vaiheessa huolestuttaakin eetikkoja enemmän. ”Saako kone korvata ihmisen elämää ja kuolemaa koskevassa päätöksenteossa?” kuten de la Rochefoucauld asian ilmaisi. Hän ei kuitenkaan näyttänyt tarkoittavan sellaista arviota kuin esimerkkinä drooni teki, vaan jotain enemmän. Mutta mitä se olisi? Ehkä inhimillisten tunteiden mukana oloa. Ihminen voisi sääliä viattomia siviiliuhreja ja keskeyttää tehtävän. Toisaalta ihminen voisi surun tai raivon vallassa päättää päinvastoin. Oli miten oli, eetikot tuntuvat kaipaavan koneillekin jonkinlaista inhimillistyyppistä tahdonaktia, jonka myötä moraalinen vastuu tulisi osaksi valintaa. Tämä tekee asiasta sotkuisen.

## Utilitaristinen sähköpaimenetiikka vaihtoehtona

Tarve saada mukaan analyysiin tahdonakti, joka tekee päätöksestä moraalisesti vastuullisen, on peräisin kansainvälisen humanitaarisen oikeuden taustaideologioista, jotka ovat aristoteelinen ja kantilainen etiikka, luonnonoikeusteoria ja ihmisoikeusajattelu (Häyry 2018, 203–204). Näille on olemassa vaihtoehto, nimittäin utilitaristinen tai pragmaattinen filosofia, joka keskittyy motiivien ja intentioiden sijasta toiminnan seurauksiin (Häyry 1994a, 164–166). Mikäli kansainvälisen humanitaarisen oikeuden tavoitteeksi katsotaan se, että LAWSit (muiden sodankäynnin teknologioiden ja käytäntöjen muassa) eivät surmaa



suhteetonta määrää taisteluun osallistumattomia, vastuu voidaan määritellä pelotteen näkökulmasta. Lailliset oikeudet ja velvollisuudet ovat (tämän käsityksen mukaan) tarkoituksenmukaisia fiktioita. Kysymys onkin, miten vastuunkantajat pitäisi nimetä, jotta pelote ehkäisisi parhaiten LAWSien käyttöä taisteluun osallistumattomia vastaan. Rangaistaanko ketjun viimeistä ihmislenkkiä? Mahdollista, mutta luultavasti epäoikeudenmukaista, koska asema organisaatiossa todennäköisesti pakottaa heidät painamaan nappia. Rangaistaanko napinpainajien lähimpiä johtajia? Mahdollista, mutta samantapaisin varauksin. Rangaistaanko korkeaa sotilasjohtoa? Mahdollista, mutta edelleen sama varaus. Rangaistaanko poliittisia johtajia? Mahdollista, mutta kuka heitä rankaisisi? Kansainvälisen oikeuden täytäntöönpanosta ei aina ole takuita.

Komentoketjun – sodan aloittamisesta tappavan koneen laukaisuun – ulkopuolelta löytyy vielä muita vaihtoehtoja. Insinöörit, jotka laativat algoritmin? Äänestäjät, joiden ansiosta poliittiset johtajat ovat asemissaan? Vihamieliset ulkomaalaiset, joiden takia sota on välttämätön? Globaalikapitalistit, joiden teot ja tekemättä jättäminen loivat olosuhteet sodan väistämättömyydelle? Kulluttajat, jotka valinnoillaan edistivät globaalikapitalismia? Mahdollisuuksia on, mutta jos pidetään tavoite mielessä, empiirisen tutkimuksen pitäisi pystyä tunnistamaan parhaat pelotteen kohteet ja muodot. Huomattakoon, että tämä (utilitaristinen) rangaistusteoria ei ole kiinnostunut syyllisyydestä, syvemmästä moraalisesta vastuusta tai edes viattomuudesta. Lailliset sanktiot pitäisi vain määritellä niin, että niiden täytäntöönpano ja valvonta onnistuu ja että ne muodostavat tehokkaimman mahdollisen pelotteen taisteluun osallistumattomien suhteetonta surmaamista vastaan. Ehkä tämä kannustaisi kehittämään asejärjestelmiä, joiden erottelukyky olisi paljon nykyistä parempi. Silloin ”sitoutuminen” voitaisiin nähdä vain viimeisenä ”tunnistuksena” – ja vaikka sen tekisi kone, siihen ei tarvitsisi liittää liikaa filosofista hölynpölyä.

## **Taisteluun osallistumattomat, taisteluun osallistuvat, oikeasuhtaisuus ja tappamisen välttämättömyys**

Jokunen outside-the-box -kysymys jää jäljelle. Miksi keskittyä vain taisteluun osallistumattomien suojelemiseen? He ovat haavoittuvia ja monissa tapauksessa suhteellisen viattomia, mutta niin ovat monet taisteluun osallistuvatkin. Sitä paitsi taisteluun osallistumattomat voivat hyvinkin osallistua sotaponnistukseen, usein oman tahtonsa mukaisesti. Miksi heidän siinä tapauksessa pitäisi välttyä vahingolta (paitsi että kansainvälinen humanitaarinen oikeus sanoo niin)? (Lazar 2020; McMahan 2011.)

Tämä herättää uudelleen kysymyksen oikeasuhtaisuudesta, kansainvälisen humanitaarisen oikeuden toisesta ehdosta. Milloin vahinko taisteluun osallistumattomille on oikeassa suhteessa sotilaallisiin tavoitteisiin? Milloin vahinko taisteluun osallistuville on oikeassa suhteessa niihin? Mitä ovat sotilaalliset tavoitteet? Miten niistä voidaan tehdä yhteismitallisia operaatioiden aiheuttamien vammojen ja hengenmenetysten kanssa?

Ajattelemalla asiaa pragmaattisesti ja käytännönläheisesti tästäkin päästään eteenpäin. Yksi mahdollisuus olisi haastaa ”tappavuus”. Miksi asevoimien hallitusten tuella kehittää LAWSeja, kun ne voisivat pyrkiä tuottamaan ei-tappavia INLAWSeja (*Incapacitating Non-Lethal Autonomous Weapon Systems*), pysyvästi lamaanuttavia PINLWSeja (*Permanently Incapacitating Non-Lethal Autonomous Weapon Systems*) tai tilapäisesti lamaanuttavia TINLAWSeja (*Temporarily Incapacitating Non-Lethal Autonomous Weapon Systems*)? Erityisesti viimeinen näistä kuulostaa houkuttelevalta, siitä huolimatta, että olemassa olevat säädökset biologisia ja kemiallisia aseita vastaan asettavat lähestymistavalle haasteita (Allhoff 2016). Jos kansainvälisen oikeuden päämääränä olisi minimoida sekä taisteluun osallistumattomien että niihin osallistuvien vammautumista ja hengenmenetyksiä, niin vaikka sodankäyntiä pidettäisiin tarpeellisena, tällaisia säädöksiä pitäisi ehkä miettiä uudelleen ja päivittää.

## Yhteenveto ja johtopäätökset

Etikot ovat olleet huolissaan siitä, että tekoälyavusteiset LAWSit ja muut AWSit tekevät päätöksiä, joita vain ihmisten pitäisi tehdä. Käyttämällä sellaista teknologiaa emme vain leiki Jumalaa (Häyry 1994b) vaan myös annamme koneiden tehdä niin. Tarkastelustani nouseva näkemys kuitenkin on, että tämä voi olla liioiteltu metafyyinen huolenaihe. Koneet tekevät tunnistustehtäviä ja niillä on jossain mielessä ihmisistä riippumatonta asiantuntijuuden auktoriteettia, mutta tämä ei ole välttämättä uhka. Koneiden ihmisistä riippumatta tekemät sitoumista koskevat ”päätökset” on todennäköisesti parasta hahmottaa lopullisina tunnistusvalintoina, ei aseman auktoriteetin käyttönä, eikä niihin liity mystisiä intentioiden tai tahtomusten osoituksia. Sellaiset kuuluvat vain inhimillisiin valintoihin, ja ne tekevät ihmiset moraalisesti vastuullisiksi tekemisistään ja tekemättä jättämisistään. Kone ei ole koskaan vastuussa aiheuttamastaan vahingosta – vastuu on ihmisen jossain kohdassa ketjua.

Filosofinen ongelma ratkeaa tällä, ja jäljelle jää asian laillinen puoli. Kansainvälinen humanitaarinen oikeus edellyttää, että LAWSit erottavat jollain tarkkuudella taisteluun osallistuvat ja osallistumattomat ja sitoutuvat kohteeseensa vain jos jälkimmäisille aiheutuva vahinko on oikeassa suhteessa sotilaallisten

tavoitteiden saavuttamiseen. Mikäli omaksutaan utilitaristisen rangaistusteorian peloteajattelu, keitä pitäisi rangaista ja mistä rikoksesta? Kansainvälisen oikeuden henki määrittelee kaksi rikostyyppiä, joista kumpikin liittyy sekä järjestelmien kehitykseen että niiden käyttöön. Ensinnäkin LAWSien on tunnistettava taisteluun osallistumattomat, ehkä erityisesti viattomat.<sup>1</sup> Toiseksi, LAWSien ei pitäisi sitoutua kohteisiinsa ilman perusteellista ja hyväksyttävää oikeasuhtaisuuden arviointia. Sellaisten LAWSien, jotka eivät täytä näitä ehtoja, kehittäminen ja käyttö ovat asiaan liittyviä rikoksia, joista pitäisi seurata rangaistus.

Kaikki maat eivät tunnista näitä rikoksia, koska ne käyttävät omia määritelmäänsä ja tulkitsevat kansainvälistä humanitaarista oikeutta tavalla, joka tukee niiden omaa automaattisten tai autonomisten asejärjestelmien käyttöä. Tälle ei juuri nyt ole tehtävissä paljoakaan. Kun edes Yhdistyneissä Kansakunnissa ei ole asiasta sopua, mikään kansainvälinen voima ei voi ohjata niitä oikealle tielle. Jos kuitenkin jokin sopimus tulevaisuudessa saadaan aikaan, kansainvälinen humanitaarinen oikeus voisi sanella, että poliittisten, politiikka- ja bisnespäätösten tekijät, jotka tilaavat, kehittävät, tuottavat ja käyttävät riittämättömiä erittelyjä ja arviointeja tekeviä LAWSeja, pitäisi asettaa siitä vastuuseen.

## Kiitokset

Artikkeli on kirjoitettu Maanpuolustuskorkeakoulun ja Pääesikunnan tilauksesta taustaraportiksi hankkeeseen Multinational Capability Development Campaign Future Leadership. Kiitän tilaajia. Kiitän myös taloudellisesta tuesta Suomen Akatemiaa (hanke Biotalous ja oikeudenmukaisuus, SA 307467) ja Maa- ja metsätalousministeriötä (hankkeet Oikeudenmukaisuuden rooli biotaloudellisessa päätöksenteossa sekä Oikeudenmukainen johtamismalli systemiseen ja kestävään siirtymiseen kohti biotaloutta).

## Viitteet

- 1 Viattomuuden arviointi, niin kaukaa haetulta kuin se kuulostaakin, voisi nyt tai lähitulevaisuudessa ehkä olla mahdollista tekemällä pika-analyysit mahdollisen kohteiden verkkokäyttäytymisestä. Meitä luokitellaan jo kaupallisen potentiaalimme mukaan, joten ei varmaankaan olisi täysin poissuljettua, että sama toimisi sotilaallisia ja humanitaarisia tarkoituksia varten.

## Lähteet

- Allhoff, Fritz (2016). The paradox of nonlethal weapons. *Stanford Law School Biosciences Blog* 10 March. <https://law.stanford.edu/2016/03/10/the-paradox-of-nonlethal-weapons/>, (29.2.2020).
- de la Rochefoucauld, Alice (2017). Introduction to the documents of the Holy See. Teoksessa Alice de la Rochefoucauld & Stefano Saldi (toim.), *The Humanization of Robots and the Robotization of the Human Person: Ethical Reflections on Lethal Autonomous Weapons Systems and Augmented Soldiers*. Caritas in Veritate Foundation Working Papers. <http://www.fciv.org/downloads/WP9-Book.pdf>, (29.2.2020).
- DoD – Department of Defense (2012). *Autonomy in Weapon Systems*. Directive Number 3000.09, November 21, 2012 Incorporating Change 1, May 8, 2017. <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf>, (29.2.2020).
- EPSRC – Engineering and Physical Sciences Research Council (2011). *Principles of Robotics*. <https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>, (29.2.2020).
- Flathman, Richard (1980). *The Practice of Political Authority: Authority and the Authoritative*. Chicago: University of Chicago Press.
- Gayle, Damien (2019). UK, US and Russia among those opposing killer robot ban. *The Guardian* 29 March, <https://www.theguardian.com/science/2019/mar/29/uk-us-russia-opposing-killer-robot-ban-un-ai>, (29.2.2020).
- Gubrud, Mark (2014). Autonomy without mystery: Where do you draw the line? *1.0 Human* 9 May. <http://gubrud.net/?p=272>, (29.2.2020).
- Häyry, Matti & Tuija Takala (2016). Coercion. Teoksessa Henk ten Have (toim.), *Encyclopedia of Global Bioethics*. Cham: Springer, 595-605.
- Häyry, Matti (1994a). *Liberal Utilitarianism and Applied Ethics*. London: Routledge.
- Häyry, Matti (1994b). Categorical objections to genetic engineering – A critique. Teoksessa Anthony Dyson & John Harris (toim.), *Ethics and Biotechnology*. London: Routledge, 202–215.
- Häyry, Matti (2018). Doctrines and dimensions of justice: Their historical backgrounds and ideological underpinnings. *Cambridge Quarterly of Healthcare Ethics*, 27(2), 188–216.
- Krulak CC. (1999). The strategic corporal: Leadership in the three-block war. *Marine Corps Gazette*, 83(1), 23.
- Lazar, Seth (2020). War. Teoksessa Edward Zalta (toim.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/spr2020/entries/war/>, (29.2.2020).
- Lovell David & Deane-Peter Baker (2017). *The Strategic Corporal Revisited: Challenges Facing Combatants in 21<sup>st</sup> Century Warfare*. Cape Town: UCT Press.
- McIntyre, Alison (2019). Doctrine of double effect. Teoksessa Edward Zalta (toim.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/spr2019/entries/double-effect/>, (29.2.2020).
- McMahan, Jeff (2011). Who is morally liable to be killed in war? *Analysis*, 71(3), 544–559.
- MoD – Ministry of Defence (2017). *Unmanned Aircraft Systems*. Joint Doctrine Publication 0-30.2 (JDP 0-30.2), dated August.
- Nadella, Satya (2016). The partnership of the future. *Slate*, 28 June. <https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-work-together-to-solve-societys-challenges.html>, (29.2.2020).