

AVTONOMNA OROŽJA IN EROZIJA MORALNE ODGOVORNOSTI: SYSTEMATIZACIJA NASILJA IN IZGINJANJE PRAVIČNOSTI

TOMAŽ GRUŠOVNIK

Univerza na Primorskem, Pedagoška fakulteta, Koper, Slovenija
tomaz.grusovnik@pef.upr.si

Sinopsis Poglavlje skuša pokazati, da avtonomna orožja predstavljajo resno nevarnost za erozijo moralne odgovornosti v sodobnem vojskovanju. Tako psihologija kot zgodovina vojskovanja kažeta, da je človeku lasten odpor do ubijanja sovražnih vojakov, tako imenovani 'nizek nivo zadetkov'. Ta ovira ubijanja sovražnikov je bila v veliki meri presežena s fizično razdaljo med vojaki in njihovimi tarčami, tj. s pomočjo artilerije, bombardiranja in brezpilotnikov. Avtonomna orožja pa ne le dodatno večajo to razdaljo, pač pa jo predstavljajo na povsem novo raven – distanca zdaj postane konceptualna, saj je odločitev za ubijanje prenesena na tovrstne sisteme. Ta vrzel odgovornosti ima za posledico tako nižanje praga za uporabo sile, zaradi česar so vojaška posredovanja verjetnejša, kot tudi manko pravičnosti, saj je težko imeti posamezne ljudi za odgovorne za vojne zločine, ki jih zagrešijo stroji. Razširjanje avtonomnih sistemov na druga področja – varovanje reda in miru, pravo, klanje in procesiranje mesa, pa tudi diagnostiko in šolstvo – lahko ima podobno za posledico sistemsko zatiranje, saj se lahko parcialni interesi uveljavljajo brez dejanske človeške odgovornosti.

Ključne besede:

avtonomno orožje,
umetna inteligenca,
moralna
odgovornost,
moralno vršilstvo,
vrzel odgovornosti

AUTONOMOUS WEAPONS AND THE EROSION OF MORAL RESPONSIBILITY: SYSTEMATISATION OF VIOLENCE AND DISAPPEARANCE OF JUSTICE

TOMAŽ GRUŠOVNIK

University of Primorska, Faculty of Education, Koper, Slovenia
tomaz.grusovnik@pef.upr.si

Abstract The chapter argues that autonomous weapons represent a real danger for erosion of accountability and moral responsibility in modern warfare. Indeed, the psychology as well as history of war shows us that humans exhibit aversion to killing enemy soldiers, a so-called “low hit ratio”. This obstacle to killing enemies was largely overcome with physical distance between soldiers and their targets, i.e., with the help of artillery, bombing and UAVs. Autonomous weapons, however, do not only extend this distance but move it to an entirely different level – the distance now becomes conceptual, as the decision to kill is transferred to those systems. This responsibility gap results in lowering the bar for using force, making military interventions likelier, as well as in lack of justice, since it is hard to hold individual humans accountable for potential war crimes committed by machines. The spread of autonomous systems to other areas – policing, law, slaughter and meat production, as well as diagnostics and schooling – can similarly result in systemic oppression because partial interests can be enforced without any real human responsibility.

Keywords:

autonomous
weapons,
artificial
intelligence,
moral
responsibility,
moral agency,
responsibility gap

1 Opredelitev avtonomnega orožja in ravni avtonomnosti

27. marca 2020 se je naposled zgodilo to, česar so se bali vojaški strategji, zagovorniki človekovih pravic, politiki in širša javnost: na bojišču v Libiji so bili v napadu uporabljeni brezpilotniki, ki so zmožni sami poiskati tarčo in jo napasti. Odbor za Libijo Varnostnega sveta Združenih narodov je tako nedavno v pismu poročal sledeče:

Logistični konvoji in umikajoče se HAF¹ so posledično izsledile in daljinsko napadle brezpilotna bojna letala ali smrtonosni avtonomni oboroženi sistemi, kot so STM *Kargu 2* (glej prilogo 30) in druge samodejne naprave baražnega ognja.² Smrtonosni avtonomni oboroženi sistemi so bili programirani za napad na tarče brez potrebe po podatkovni povezavi z operaterjem in orožjem: dejansko prava 'streljaj, pozabi in najdi' zmožnost. (Panel of Experts on Libya established pursuant to resolution 1973 (2011) 2021: 17)

Čeprav gre za šokantne novice, vredne kakšnega znanstveno-fantastičnega filmskega scenarija, pa žal ne gre za posebno presenečenje. Uporaba avtonomnih orožij je namreč čedalje bolj zaželeno, med razlogi, zaradi katerih vojske držav po svetu uporabljajo avtonomna orožja,³ pa je njihova računska moč (stroji so sposobni upoštevati in preračunati veliko več podatkov kot človek, zaradi česar lahko situacijo prej 'osmislijo' in so tudi natančnejši ter varčnejši), sposobnost delovanja v človeku sovražnem okolju (ne le vojaškem, pač pa tudi v biološko-fizičnem) ter sposobnost delovanja v 'brezpovezavnem načinu', torej tam, kjer ni podatkovne povezave med sistemom in operaterjem oziroma upravljalcem, zaradi česar klasični brezpilotniki odpovedo. Zraven vseh prednosti, ki jih tovrstno orožje ponuja, se v literaturi omenja, da bo uporaba le-teh postala pravzaprav neizbežna: zaradi izjemno kompleksnih situacij bo namreč človeško procesiranje informacij na bojišču in posledično odločanje postalo preprosto prepočasno in neučinkovito: napad jate brezpilotnikov, flote duhov (avtonomnih plovil) in tropa robotskih psov bodo lahko odvrnili le sorodno kompleksni obrambni sistemi. Poleg tega se pogosto – in morda

¹ Sile, povezane s Haftarjem, libijskim vojaškim poveljnikom, op. T. G.

² Angl. *loitering munition*.

³ Celotna besedna zveza, ki označuje ta koncept, je sicer 'smrtonosni avtonomni oboroženi sistem', angl. *Lethal Autonomous Weapons System* (LAWS). V nadaljevanju bom za ta koncept uporabljal izraz 'avtonomno orožje', ki se je prijel tudi v slovenščini.

paradokсно – velikokrat omenja humanitarna prednost uporabe takšnih orožij, k čemur se vrnemo v naslednjem odseku.

Toda kaj sploh je 'avtonomno orožje'? Pojem 'avtonomnega orožja' ima več različnih definicij, v glavnem pa lahko rečemo, da gre za sistem strojne in programske opreme, ki je na podlagi senzorjev zmožen prepoznati tarče in jih uničiti brez človekovega posredovanja. Članek, objavljen v zbirki povzetkov iz srečanja Mednarodne komiteja Rdečega križa na temo raznolikih vidikov avtonomnega orožja, sicer podaja sledečo definicijo, ki jo prevzemajo različni avtorji in institucije:

Avtonomni oboroženi sistem je tisti, ki se lahko priuči svojega delovanja oziroma ga prilagodi kot odgovor na spreminjajoče se okoliščine v okolju, v katerem je uporabljen. Resnično avtonomni sistem ima umetno inteligenco, ki je sposobna implementacije mednarodnega humanitarnega prava. (Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects 2014: 64).

V konkretnem smislu gre lahko za premične sisteme (brezpilotniki, kvadro- in multikopterji, kopenski roboti in vozila ter vodne in podvodne aplikacije) oziroma za stacionarne priprave (avtonomna mitraljezna gnezda, t. i. *sentry guns*, kakršna najdemo na izraelsko-palestinski meji). Avtonomna orožja so lahko smrtonosna, torej namenjena ubijanju ljudi, lahko pa so tudi namenjena uničevanju sovražnikovih izstrelkov in avtonomnih sistemov. Poleg njihove potencialne smrtonosnosti je v moralnem smislu še najpomembnejše razlikovanje stopnje avtonomije takšnega orožja, ki jo določimo glede na vlogo človeškega operaterja pri delovanju take naprave. To razlikovanje, ki ga kot najbolj smiselno v navezavi na pojem 'avtonomije' izpostavlja tudi Kaja Smolnikar (2019) v svoji magistrski nalogi, lahko povzamemo po publikaciji Human Rights Watch na sledeč način:

Človek-v-zanki: sistemi lahko izberejo tarče oziroma izbrane tarče uničijo zgolj na podlagi človekovega povelja;

človek-nad-zanko: sistemi lahko izberejo tarče in jih uničijo samodejno, a so nenehno pod nadzorom človeškega operaterja, ki ima možnost preklica dejanja;

človek-izven-zanke: sistemi najdejo in uničijo tarče, ne da bi to delovanje usmerjal ali spremljal človek. (Human Rights Watch 2012: 2)

Med 'popolnoma avtonomna orožja' lahko tako prištevamo v celoti le zadnjo kategorijo, drugo pa pogojno, in sicer v tistih primerih, kjer je človeški nadzor zgolj še načelen, saj je dogajanje na bojišču preveč kompleksno in hitro za dejansko človekovo smiselno odločanje. V slednjem primeru možnost preklica sicer zmeraj obstaja, a je v praksi skoraj že enako odpoklicu sistemov tretje kategorije. Druga kategorija je torej precej nejasna cona, mejni primer, kjer je stopnja avtonomije sistema lahko odvisna od zelo partikularnih in kontingentnih okoliščin, tudi konec koncev od podatkovne povezave med nadzornikom in sistemom, saj si je moč zamisliti situacijo, v kateri odsotnost slednje de facto izniči možnost preklica. Zaradi tega bomo slednjo prištevili med avtonomna orožja, saj je konec koncev znotraj te kategorije stroj zmožen ubiti človeka brez operaterjevega posredovanja. V nadaljevanju bomo torej z 'avtonomnim orožjem' mislili tiste sisteme, ki delujejo na podlagi načel 'človek-nad-zanko' in 'človek-izven-zanke'.

2 Etiški izzivi, povezani z avtonomnimi orožji

Že zgoraj smo nakazali, da nekateri avtorji med prednosti avtonomnih orožij uvrščajo humanitarne razloge. Po njihovem mnenju naj bi uporaba avtonomnih orožij tako pomagala spoštovati mednarodna humanitarna načela, saj lahko zmanjša število tako vojaških kot civilnih žrtev, nepotrebno trpljenje, pa tudi vojne zločine, ker ne spodbuja krvoločnosti, je natančnejša, kar praviloma rezultira v manj civilnih žrtvah, hkrati pa tudi zaščiti vojake (glej Zawieska 2017: 51). Ronald Arkin tako denimo zagovarja uporabo avtonomnega orožja iz moralnega vidika, saj bi se z njegovo pomočjo lahko izognili številnim dejavnikom vojnih zločinov:

Možne razlage vojnih zločinov vsebujejo: visoke zavezniške izgube, ki spodbujajo maščevanje; pogoste zamenjave v poveljniški hierarhiji, ki vodijo k šibkemu vodenju; razčlovečenje nasprotnika s pomočjo poniževalnih vzdevkov; slabo izurjene in neizkušene enote; nejasno določenega sovražnika; nejasne ukaze, kjer se lahko namera ukaza napačno interpretira kot nezakonita; mladost in neizkušenost enot; zunanji pritisk, denimo zahteva po številčnosti sovražnikovih žrtev; ugodje, ki ga prinaša moč ubijanja ali občutje frustracij. Obstaja očiten prostor za izboljšave in

avtonomni sistemi nam bi lahko pomagali nasloviti nekatere od teh problemov. (Arkin 2013: 2)

Marco Sassoli je enako optimističen glede izboljšanja humanitarne situacije na bojiščih z vpeljavo avtonomnih orožij. Prepričan je, da bodo roboti, ki jih ne bo strah za lastno preživetje in ki ne bodo pobijali 'preventivno' in ki tudi sicer ne bodo podvrženi čustvom, denimo jezi ali kakšni drugi strasti, delovali bistveno natančneje in bolj upravičeno:

Morda zaradi tega, ker sem v tolikih oboroženih konfliktih bil soočen s številnimi kršitvami, ki so jih povzročila človeška bitja, a nikdar s hudodelstvi, ki bi jih zagrešili roboti (čeprav je treba priznati, da v oboroženih konfliktih, ki sem jim bil priča, niso obstajali), moje prvo občutje ni skepticizem, pač pa upanje za boljše spoštovanje mednarodnega humanitarnega prava. Samo človeška bitja so lahko nehumana in samo človeška bitja lahko načrtno izberejo, da se ne bodo držala pravil, ki so bila vzpostavljena zato, da bi jim sledili. Zdi se mi bolj verjetno pričakovati (in zagotoviti), da bo mednarodnemu humanitarnemu pravu sledila oseba, ki bo sestavila in pripravila avtonomno orožje na mirnem delovnem mestu kot pa vojak na bojišču ali v sovražnem okolju. Roboti ne morejo sovražiti, ne občutijo strahu, ne morejo biti lačni ali žejni in utrujeni ter nimajo nagona po preživetju. 'Roboti ne posiljujejo'. Hkrati lahko zaznajo več informacij in jih procesirajo hitreje kot človeško bitje. Ko bodo orožja, ki sprožajo kinetično energijo, postala čedalje hitrejša in bolj kompleksna, se lahko zgodi, da bodo ljudje preprosto preplavljeni z informacijami in odločitvami, ki jih bodo morali sprejeti, da bi jih usmerili. Ljudje pogosto druge ubijajo zato, da bi se izognili lastni smrti. Robot pa lahko preloži uporabo sile do zadnjega, najprimernejšega trenutka, ko je bilo ugotovljeno, da sta tarča in napad legitimna. (Sassoli 2014: 310)

Seveda je jasno, da vsi niso tako navdušeni nad uporabo avtonomnih orožij in da obstaja zelo veliko pomislekov glede uporabe takšnih sistemov. Že omenjeni Ronald Arkin tako našteva probleme, povezane z odgovornostjo za vojne zločine (kar je tudi ključna točka tega poglavja, h kateri se še vračamo v nadaljevanju), tako tudi z zniževanjem praga za uporabo sile, pa z nepripravljenostjo podeliti robotom pravico veta in zavrnitve izvršitve ukaza, neavtoriziranim razširjanjem tehnologije,

posledicami za kohezijo v vojski, elektronsko varnostjo, iznakaženjem misij.⁴ Human Rights Watch se za razliko od Arkina poskuša strožje držati moralne in humanitarne problematike ter jo posledično ne pomeša z moralnimi izzivi, našteje pa sledeče zagate, ki bi jih avtonomno orožje lahko imelo v navezavi na skladnost z mednarodnim humanitarnim pravom in sledečimi načeli:

Razlikovanje: Avtonomno orožje bi lahko naletelo na težave pri razlikovanju med bojevniki in civilisti. V zadnjem času se je namreč korenito spremenila narava vojskovanja, saj so tradicionalne oborožene konflikte med državami nadomestili spopadi urbanih skupin, ki jih je zelo težko ločiti od civilistov;

Sorazmernost: Bojazen je, da avtonomni oboroženi sistemi ne bi znali preceniti, kdaj uporaba vojaške sile povzroči več škode za civiliste, kot pa prinese koristi na bojišču. Sorazmernost je močno odvisna od konteksta in številnih dejavnikov, ki jih je nemogoče kvantificirati na način, da bi bila situacija primerna za pretres z algoritmom;

Vojaška nujnost: Algoritem ne bi bil zmožen presoditi, kdaj je kak sovražni vojak postal hors de combat, torej nezmožen bojevanja, kar pomeni, da mu po vojnem pravu pripada poseben status. Lahko bi se torej zgodilo, da bi avtonomno orožje likvidiralo de facto onesposobljenega sovražnega vojaka, saj denimo ne bi prepoznalo, da je ranjen do te mere, da ni več sposoben za boj in bi mu pravzaprav moral pripasti status vojnega ujetnika;

Martensova klavzula: To obče načelo pravi sledeče: »v primeru, ko neka situacija ni urejena s pravili veljavnega pogodbenega prava, civilno prebivalstvo in pripadnike oboroženih sil ščitijo načela mednarodnega prava, ki izhajajo iz običajev, sprejetih med civiliziranimi narodi, zakoni človečnosti in zahtevami javne vesti« (Sancin, Švarc in Ambrož 2009: 27)⁵ Algoritmi nikdar ne bodo zmožni odločanja v skladu s 'civiliziranimi

⁴ Angl. *mission creep*; gre za razširjanje intervencij preko prvotno predvidenega obsega, glej Arkin 2014: 35).

⁵ Dokument vsebuje tudi tehnične razlage drugih tukaj uporabljenih načel.

običaji', prav tako nimajo nikakršne 'javne vesti', zato niso primerni vršilci⁶ na bojišču.

Poleg težav, ki bi jih avtonomna orožja lahko imela z upoštevanjem zgornjih načel, Human Rights Watch našteje še sledeče moralne zagate, ki bi lahko nastopile z uporabo avtonomnih orožij.

Pomanjkanje človeških občutij: Zagovorniki moralnosti uporabe avtonomnih orožij pogosto navajajo argument, da bi zaradi odsotnosti strahu, interesa po preživetju, jeze ter maščevanja robotski sistemi lažje sledili načelom mednarodnega humanitarnega prava kot pa človeški borci. Toda hkrati pozabljajo, da so čustva tudi eden glavnih zaviralcev uporabe sile: ljudje imamo namreč spontano averzijo do ubijanja in nasilja, kar je naravna varovalka pred prekomerno uporabo sile, ki pri avtonomnem orožju umanjka. K 'psihološkemu otopivanju' in 'emocionalnemu umikanju' kot predpogoju ubijanja se še vrnemo v nadaljevanju.

Zniževanje praga za uporabo sile: Ker lahko avtonomna orožja vojske uporabljajo učinkovito, tudi prikrito in brez tveganj za življenje vojakov, se lahko zgodi, da bodo silo uporabile prej, kot bi jo sicer, če bi bilo treba napovedati konvencionalno vojno.

Zmanjševanje odgovornosti: Če avtonomno orožje po nesreči ubije civilista, ni jasno, kdo je odgovoren – je to morda poveljnik, ki je izdal ukaz, ali pa programer, izdelovalec ali celo robot sam? Na to ni enostavnega odgovora, kajti poveljnikov ne morejo odgovarjati za dejanja, ki jih niso želeli in na katera niso imeli vpliva. Tako nastane 'vrzel odgovornosti', kar pa je velik problem, saj ima odgovornost na bojišču vsaj dve funkciji: odvrča od povzročanja bodoče škode za civiliste, žrtvam ali njihovim svojcem pa nudi možnost pravičnega povračila.⁷

⁶ Besedo 'vršilec' uporabljam za prevod angl. *agent*. Ta izbira, ki sem jo prvič zasledil pri kolegici Branislavi Vičar v sklopu njenih preiskovanj etike živali, se mi zdi bistveno bolj posrečena kot pojem, ki sem ga uporabljal do zdaj, namreč 'akter'. Za *agency* posledično uporabljam pojem 'vršilstvo'.

⁷ Za vse glej Human Rights Watch 2012: 30–42.

Natanko te zagate so hkrati tiste, ki so najtesneje povezane z vprašanjem krivde ob napačni oziroma moralno sporni uporabi avtonomnega orožja, tako da se jim v nadaljevanju natančneje posvečamo.

3 Erozijska odgovornost in izginjanje pravičnosti s sodobnega bojišča

Naposled prihajamo do središčnih tez pričujočega poglavja: pokazati bomo poskušali, da enega največjih izzivov uvedbe avtonomnega orožja predstavlja omenjena vrzel odgovornosti⁸ oziroma to, kar v naslovu poimenujemo erozija moralne odgovornosti. Posledica tega je na eni strani večja pripravljenost uporabiti orožje oziroma zniževanje praga uporabe sile, na drugi pa sistematizacija nasilja, s čimer imamo v mislih proces, podoben zgoraj omenjenemu 'iznakaženju misij', kjer bi lahko prišlo do tega, da bi bili avtonomni oboroženi sistemi nenehno vseprisotni z nalogo 'dušenja uporov'. Nadaljnja posledica takega procesa bi lahko bila izguba pravičnosti, in sicer retributivne pravičnosti, saj z izginjanjem moralnega vršilstva tudi izginja odgovornost, tako da že načelno umanjka točka prevzema odgovornosti za morebitne civilne žrtve.

Da bi razumeli, kako se avtonomna orožja skladajo z doktrino modernega vojskovanja, si moramo ogledati tako psihologijo kot z njo povezano kratko zgodovino bojevanja. S psihološkega stališča je gotovo najzanimivejše dejstvo, da ima velika večina ljudi averzijo do ubijanja. Nasilje je nekaj, čemur se najraje izognemo, to pa toliko bolj, če bi mi morali biti tisti, ki naj bi ga povzročali. Da bi torej zmogli postati nasilni – kar vojaki vsekakor morajo biti –, se je treba naučiti postaviti sočutje in empatijo do drugega človeškega bitja v oklepaj, kar je proces, ki ga v primeru mesarjev klavcev Melanie Joy imenuje »psihološko otopevanje« (Joy 2010: 144), Erich Fromm pa v primeru medčloveškega nasilja »emocionalno umikanje« (Fromm 2013: 164). Tudi če smo tega začasno sposobni, je velika verjetnost, da bo izvrševanje nasilja na nas pustilo globoke posledice. Te psihološke posledice so dobile celo ime: izvršiteljevi travmi, ki nastane zaradi nasilja, pravimo »perpetration-induced traumatic stress« (MacNair 2002).

⁸ Angl. *accountability gap*.

Zaradi te človekove averzije do nasilja in ubijanja postane razumljivo, da so vojaki na frontah večinoma načrtno streljali v zrak, da bi zgrešili nasprotnika, ki so ga želeli le prestrašiti, nikakor pa tudi ubiti. Kot v svoji monografiji o psihologiji ubijanja pripoveduje vojaški psiholog Dave Grossman, je ta pojav tako pogost, da je izčrpno dokumentiran. V resnici se je človeštvo v velikem delu svoje zgodovine bojevalo predvsem na podlagi take taktike ustrahovanja s 'poziranjem',⁹ ki je izključevalo ubijanje nasprotnika in s pomočjo katerega Grossman nadgradi tradicionalni 'boj ali beg' model, ki ima poslej štiri kategorije: boj, beg, poza, podreditev. Aleksander Veliki naj bi na svojem pohodu tako izgubil vsega 700 mož, pri čemer je večina njegovih sovražnikov bila ubitih po bitki, medtem ko je poročnik George Roupell moral v prvi svetovni vojni s sabljo tolči svoje može po hrbtu in jim v strelskem jarku groziti, da merijo nižje.¹⁰ Vse od plemenskih vojn do evropskih napoleonskih vojn je veljalo, da si moral sovražnika prestrašiti. Napredek tehnike vojskovanja pa se kaže tudi v poznavanju načinov, s katerimi je mogoče ljudi prepričati, naj sovražnike ubijajo (in da v svoji taktiki nazadnje celo računajo s tem, da bo sovražnik iz upora do ubijanja najprej zastraševal!):

Res, na zgodovino vojskovanja lahko gledamo kot na zgodovino čedalje učinkovitejših mehanizmov za omogočanje in pogojevanje ljudi, da premagajo svoj vrojen upor do ubijanja drugih sorodnih človeških bitij. V številnih okoliščinah so se visoko usposobljeni sodobni vojaki borili s slabo usposobljenimi gverilskimi silami in težnja slabo pripravljenih sil k temu, da se nagonsko zatečejo k zastraševalnim mehanizmom (recimo streljanje v zrak), je pomenila pomembno prednost za bolj izurjene sile. (Grossman 2009: 13)

Eden teh načinov, kako prepričati ljudi, da bodo pobijali soljudi, je umik vojaka iz bojišča in uvedba tehnologije bojevanja na daljavo. Med prvimi takimi orožji je bila artilerija dolgega dometa in pa bombardiranje, pozneje pa termovizija (primer Grossman 2009: 11, 169–170) in seveda brezpilotni letalniki. Grossman povzema: »Večino dejavnikov, ki omogočajo ubijanje na bojišču, je moč ugledati v difuziji odgovornosti« (Grossman 2009: 193).

⁹ Angl. *posturing*.

¹⁰ Za številne primere od antike do vietnamske vojne glej Grossman 2009: 6–17.

V tej točki postane očitno, da so avtonomna orožja kot nalašč ustvarjena za krpanje tiste 'šibkosti', ki jo ima človeški vojak, namreč averzije do ubijanja in travme, ki izvira iz nje, če do jemanja življenj pride. Smemo reči celo tole: *Če so artilerija, bombardiranje in brezpilotniki vnesli fizično distanco do nasprotnika, ki je omogočila obsežnejše in lažje ubijanje z manj slabe vesti, potem avtonomno orožje vnaša konceptualno distanco do tarče, ki vprašanje vesti odpravlja v celoti.*

Konceptualna distanca, ki nastane med poveljniki in vojaki, avtonomnim orožjem ter vojaškimi ali celo civilnimi žrtvami, erodira odgovornost, saj ni več znano, kdo naj bi za dejanja odgovarjal. So to proizvajalci, programerji ali poveljniki, nadzorniki oziroma politiki? Najtežje je odgovornost obesiti programerjem, kajti algoritmi, ki so vprogramirani v takšne sisteme, so običajno plod dolgega razvoja celih skupin strokovnjakov. Prvi algoritmi za samodejno zaznavanje so bili sprogramirani kot pomoč slabovidnim, medtem ko lahko danes predstavljajo pomemben element 'osmišljanja' informacij iz senzorjev avtonomnih sistemov. Proizvajalca lahko imamo že nekoliko lažje za odgovornega, saj je tisti, ki v prvi vrsti sploh materialno omogoči tovrsten instrument. Še najbolj intuitivno je, da imamo v primeru človeka-izven-zanke za odgovorne politike, ki so avtorizirali uporabo sile in pa poveljnike, čeprav tudi tukaj prihaja do zapletov: moč si je namreč zamisliti, da so avtorizirali uporabo avtonomnega orožja, ki pa je nato samo bodisi zatajilo bodisi izvršilo nekaj, česar ljudje niso predvideli. Za konkretne avtonomne 'odločitve' sistema (uničenje natanko *te* tarče, povzročitev natanko *te* škode) torej ne moremo kriviti ljudi, saj jih tudi ne moremo kriviti za dejanja, ki jih zagrešijo njihovi podrejeni in za katera poveljniki niso vedeli. Politike in poveljnike lahko imamo tako odgovorne le (še) za odobritev občin ciljev misij.¹¹

V primeru človeka-nad-zanko se nam najbrž intuitivno zdi, da lahko imamo vsekakor za odgovornega nadzornika, ki bi lahko preprečil spodrselj avtonomnega orožja oziroma ki bi moral prevzeti odgovornost za vse smrti, ki jih avtonomno orožje povzroči, saj je potihom omogočal delovanje sistema s tem, ko ga ni preklical. Toda tudi tukaj nastane težava, saj je, kot opozarjajo mnogi strategji in tehniki, sodobno bojišče tako zapleteno, da ga ljudje le stežka osmislimo, zaradi česar se ravno zatekamo k uporabi umetne inteligence in avtonomnih sistemov. Kadar je torej situacija na bojišču izjemno zapletena in avtonomno orožje deluje v skladu s

¹¹ Za podobno stališče prim.: Human Rights Watch 2012: 42.

svojimi programiranimi načeli, postaja funkcija človeškega nadzornika čedalje bolj nerelevantna, s tem pa se manjša tudi njegova odgovornost. Najbrž je celo tako, da bi nadzornik ravno bil odgovoren v primeru, kadar bi preventivno prekinil delovanje avtonomnega sistema, a bi se naknadno izkazalo, da je bila to preuranjena odločitev, ki je ogrozila ali celo povzročila izgubo prijateljskih sil. V tem primeru bi mu lahko očitali, da je preprečil uporabo dobrega, delujočega in preverjenega sistema, ki bi lahko obvaroval življenja. Tako si je lahko zamisliti, da bodo operaterji in nadzorniki raje sledili predlogom algoritmov oziroma prepustili dogajanje avtonomnim sistemom, kot pa tvegali, da bi se s svojo omejeno zmožnostjo presoje zoperstavili umetni inteligenci.

Če pa ne moremo imeti več nikogar za zares odgovornega za spodrsaljaje ali nepredvidena 'dejanja' avtonomnih sistemov, potem lahko postane uporaba sile pogostejša. Obstaja namreč utemeljena bojazen, da se bodo države z uporabo avtonomnih orožij lažje odločale za vojaška posredovanja, saj bo umanjala ključna zavora, ki posameznike na poziciji moči zaustavi pred napovedjo vojne: nevarnost izgube in velikih 'stroškov', tako materialnih kot moralnih: »Voditelji bi se morda manj obotavljali vojne napovedi, če bi se tveganje za njihove vojake zmanjšalo ali izničilo« (Human Rights Watch 2012: 39–40). Kar se tiče posameznikov na bojišču pa smo tudi videli, da je za vojake ključnega pomena premagovanje averzije za ubijanje, ki je toliko večja, kolikor manjša je razdalja med vojakom in nasprotnikom: ker avtonomno orožje – kot smo večkrat poudarili – vnaša konceptualno distanco med vojaka in nasprotnika, si je moč zamisliti, da bo imel vojak veliko manj pomislekov sprožiti avtonomni sistem kot pa puško.

Avtonomna orožja bodo torej skoraj zagotovo znižala prag uporabe sile, toda morda bo to zgolj prva faza uvedbe tovrstnih instrumentov na sodobno bojišče in bojišče prihodnosti. V drugi fazi bi uporaba avtonomnih orožij lahko postala čedalje bolj preventivna. Brezpilotniki, ki bi nenehno patroljirali v zraku, flote duhov, ki bi nenehno plule po oceanih in morjih, pa tudi policijski kopenski roboti bi lahko bili uporabljani čedalje bolj 'preventivno', na način, da bi 'nevtralizirali' točke upora režimu, *še preden* bi se oboroženi spopad sploh razplamtel. Grožnja s silo bi na tak način lahko postala sistemska, ljudje pa bi živeli pod 'avtonomnim ščitom', ki bi jih nenehno 'varoval' pred 'nevarnostmi'. Jasno je, da bi takšen skorajda že Orwellovski svet bil največje možno tveganje za demokracijo, saj bi resne alternative vladajočim režimom tako rekoč strojno izničil, *še preden* bi se pojavile. Če bi ob tem uporabili

še različna 'diagnostična' orodja, torej umetno inteligenco, ki s pomočjo naprednih algoritmov nenehno 'rangira' oziroma razvršča in kategorizira posameznike ter označuje tiste, ki so 'potencialno nevarni' sistemu, potem je jasno, da bi svet, v katerem bi pristali, bil še hujši od tistega, ki si ga je zamislil Orwell.

Dodajmo še, da se z erozijo moralne odgovornosti s sodobnega bojišča iz njega umika tudi pravičnost. Če namreč ni nihče neposredno kriv za spodrsrljaje in napake oziroma če pravni sistem in javna vest celo namensko delujeta na tak način, da pristajata na skrivanje odgovornosti za avtonomnimi sistemi, potem svojci žrtev ne morejo več najti zadoščenja, saj za smrti njihovih bližnjih naenkrat 'ni nihče kriv'. Svojci žrtev in drugi oškodovanci tako izvisijo, kar nakazuje na krizo retributivne oziroma povračilne pravičnosti, saj ni nikogar, ki bi se moral opravičiti, odgovarjati za nasilje oziroma plačati odškodnino za povzročeno škodo. Pravzaprav lahko rečemo, da se pravičnost z bojišča umika skupaj s človekom kot moralnim vršilcem.

4 Avtonomni sistemi in sistemsko nasilje

Znanstvenofantastični filmi zahodne produkcije so v dvajsetem stoletju največjo težavo umetne inteligence prepoznavali v njeni zoperstavitvi človeku in družbi: roboti niso postali le zavestna bitja, pač pa so s svojo strojno inteligenco presegli človeka in si ga podjarmili, morda na podoben način, kot si je človek podjarmil živali in naravo. Toda čas, z njim pa tehnološki in družbeni razvoj, je pokazal, da je nevarnost vdora umetne inteligence v človekov življenjski svet povsem drugačna: človekovo samoiniciativno podrejanje zahtevam informacijske družbe. Če pomislimo na primer pametnih telefonov, hitro uvidimo, da nam jih ni moral nihče vsiliti, nasprotno, še preveč radi se podrejamo uporabi aplikacij, ki nas – če si lahko dovolimo ta izraz – učijo razmišljanja, analognega funkcioniranju algoritmov. Umetna inteligenca nas torej ni osvojila, ampak prej zasvojila in nas prisilila spremeniti naše mišljenjske, čustvene in druge odzive ter vedênje. Ne zavzemajo nas roboti, ampak se robotiziramo sami. Karolina Zawieska tako pomenljivo pravi, da »zdaj je mogoče razpravljati o tem, ali se lahko stroji držijo človeških etičnih načel, ne zaradi tega, ker bi avtonomni sistemi postali podobni ljudem, pač pa zato, ker človeška bitja nase čedalje bolj gledajo kot na stroje« (Zawieska 2017: 53). V zaključku poglavja tako poskušamo pokazati, da uvajanje avtonomnih sistemov ni težavno zgolj za sodobno bojišče, kjer morda sicer res pride najočitneje do izraza

njegova moralna problematika, pač pa nasploh pomeni korenit poseg v družbo, in sicer predvsem v smeri sistematizacije nasilja in represije.

Zraven vojske in bojevanja bodo roboti prav gotovo v naslednjih letih prevzeli klavniška opravila. Tudi tukaj¹² imamo namreč opraviti s travmatskim stresom kot posledico opravljanja takšnega dela in robotizacija klavnega procesa bi tega ne le pospešila, ampak naredila tudi bolj 'humanega' in 'higienskega'. S tem bi se nasilje nad živalmi že skoraj v celoti robotiziralo in mehaniziralo, s čimer bi zelo očitno vkorakali v popolno živinorejsko distopijo, kjer bi vsi samo še prodajali in konzumirali mesne izdelke, medtem ko odgovornosti za smrti živali ne bi rabil več prevzemati nihče.

Težave, povezane z avtonomnimi sistemi in strojnim odločanjem, lahko zasledimo tudi na številnih drugih področjih, kjer njihova uporaba iz vidika izginjanja moralnega vršilstva ni nič manj sporna, pa čeprav je – vsaj na prvi pogled – manj pretresljiva. Specifično v navezavi na erozijo odgovornosti in posledično izginjanje pravičnosti je vpeljava avtonomnih sistemov problematična znotraj policijskega delovanja. Že dolgo je tako, da se policisti in drugi uradniki radi sklicujejo na to, da »oni samo opravljajo delo« ali »izvršujejo ukaze«, kar je natanko mehanizem distanciranja od lastne odgovornosti in poskus bega pred dejstvom, da so sami neizbežno moralni vršilci in prav oni kot posamezniki tisti, ki napišejo kazen, zavrnejo vstop v državo ali izvršijo kakšno drugo dejanje, ki ima neprijetne, včasih celo brezčutne in krivične posledice za posameznice in posameznike. Skupaj s Henryjem D. Thoreaujem lahko takšnemu uradniku, ki se sklicuje na svojo nemoč, medtem ko dejansko izvršuje neko dejanje, zmeraj odvrnemo: »Pusti službo« (Thoreau 2016: 24)! S tem izpostavimo njegovo hoteno nevednost glede dejstva, da je zmeraj on sam kot posameznik vršilec dejanja, da ima zmeraj izbiro, pa čeprav se tega ne zaveda ali noče zavedati.¹³ Če smemo povzeti, lahko zato rečemo, da avtonomni sistemi in uporaba rangirnih algoritmov naslavlja natanko to človekovo potrebno po distanciranju od odgovornosti: pričakovati je, da bo njihova vpeljava pripomogla k še bolj sistematiziranemu razpuščanju individualne odgovornosti in da se bodo posamezniki skorajda z olajšanjem odrekli moralnemu vršilstvu ter se pri moralnem odločanju raje sklicevali na protokole, algoritme in komisije, saj se, kot

¹² Za analizo travmatskih posledic za mesarje klavce glej Grušovnik 2016: 142–47.

¹³ Za analizo hotene nevednosti, tudi glede prevzemanja lastne odgovornosti, glej mojo razpravo v Grušovnik 2020: poglavje o Sartru razkriva, da je prepričanje, da 'ne moreš ničesar spremeniti', primer 'slabe vere'.

pravi Friderik Klampfer: »/.../ naporno in zagatno moralno presojo raje nadomesti z deklamiranjem določil poklicnih etičnih kodeksov /.../« (Klampfer 2010: 17).¹⁴ Pravniki, zdravniki in učitelji se bodo za vzrok specifičnih sodb, diagnoz in ocen lahko preprosto sklicevali na algoritme in se na ta način otresli svoje odgovornosti.

Že zgoraj smo opisali primer nadzornika avtonomnega sistema, delujočega po načelu človek-nad-zanko, ki se v nekem trenutku odloči preklicati operacijo, zaradi česar nastanejo prijateljske izgube. Takemu posamezniku bi lahko očitali, da je imel na voljo najboljšo oceno situacije na bojišču, ki so jo pripravili najnaprednejši algoritmi, a se je raje odločil, da bo izvršitev dejanja preklical, zaradi česar mu lahko skušamo naprtiti krivdo za prijateljske izgube. Ker obstaja takšna nevarnost, je verjetnost, da bo nadzornik raje slepo sledil predlogom umetne inteligence, kot pa se s svojo omejeno zmožnostjo presoje zoperstavil algoritmom, velika. To pomeni, da bodo ljudje raje, kot bi tvegali samostojno odločanje in s tem prevzemanje odgovornosti, slepo sledili algoritmom. To pa seveda ne velja zgolj za vojaške situacije, pač pa tudi za šolstvo, sodstvo in zdravstvo. Lahko si zamislimo zdravnika, ki mu algoritem navrže eno diagnozo, do katere pa je na podlagi lastnih izkušenj skeptičen in je ne sprejme. Nato pa se izkaže, da je algoritem vendarle imel prav. Je v takem primeru zdravnik kriv, ker je 'zanemaril' rezultate najboljšega diagnostičnega orodja, ki mu je bilo na voljo? Tudi če odgovora na to vprašanje nimamo, vidimo, da se zdravniku bolj 'splača' slepo slediti diagnostičnemu orodju, kot pa tvegati prevzem odgovornosti za lastno moralno vršilstvo in s tem tvegati krivdo ter celo morebitne odškodninske tožbe. Povsem enako seveda velja za sodnika, ki bi se uprl nakazani sodbi, ali učitelja, ki bi preglasil strojno predlagano oceno.

Tovrstni procesi – erozija moralne odgovornosti zaradi ponikanja individualnega in osebnega moralnega vršilstva v avtonomnih sistemih – bi lahko predstavljali nekaj takšnega kot dodatno sistematizacijo nasilja, kjer bi algoritmi, o katerih bi lahko odločala le peščica močnih posameznikov, diktirali smer družbenega razvoja. Lahko bi prišlo do systemskega zatiranja in uveljavljanja parcialnih interesov pod pretvezo 'digitalizacije', za kar pa – ker bi vse izvrševali 'optimizirani' avtonomni sistemi – nihče ne bi bil odgovoren.

¹⁴ Za podobno poanto glej Grušovnik 2021: 9.

Viri in literatura

- Arkin, R. (2013.) »Lethal Autonomous Systems and the Plight of the Non-Combatant«. *Artificial Intelligence and Simulation of Behaviour Quarterly*, 137.
- Arkin, R. (2014.) »Ethical Restraints of Lethal Autonomous Robotic Systems: Requirements, Research, And Implications«. V *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*«. Ženeva: International Committee of Red Cross, brošura s povzetki prispevkov, str. 33–37.
- »Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects«. (2014). Ženeva: International Committee of the Red Cross, brošura s povzetki prispevkov.
- Fromm, E. (2013). *Anatomija človekove uničevalnosti*. Ljubljana: Mladinska knjiga.
- Grossman, D. (2009). *On Killing: The Psychological Cost of Learning to Kill in War and Society*. New York: Little, Brown, and Company.
- Grušovnik, T. (2016). *Etika živali*. Koper: Založba Annales.
- Grušovnik, T. (2020). *Hotena nevednost*. Ljubljana: Slovenska matica.
- Grušovnik, T. (2021). »Uvodnik: vloga filozofije pri razumevanju etike danes«. V Grušovnik, T. in Pirc, G. (urd.), *Etika Danes, Filozofsko raziskovanje in razumevanje etike v sodobni družbi*. Koper: Založba Annales.
- Human Rights Watch. (2012). *Losing Humanity – The Case against Killer Robots*. Cambridge, Massachusetts: International Human Rights Clinic. URL = https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf.
- Joy, M. (2009). *Why We Love Dogs, Eat Pigs, and Wear Cows? – An Introduction to Carnism*. San Francisco: Conari Press.
- Klampfer, F. (2010). *Cena žinjlenja – razprave iz bioetike*. Ljubljana: Krtina.
- MacNair, R. (2002). *Perpetration-Induced Traumatic Stress: The Psychological Consequences of Killing*. Praeger Publishers: Westport.
- Sancin, V., D. Švarc in M. Ambrož. (2009). *Mednarodno pravo oboroženih spopadov – strokovno delo za potrebe Slovenske vojske*. Ministrstvo za obrambo Republike Slovenije, Knjižnično-informacijski in založniški cente.
- Sassoli, M. (2014). »Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified«. *International Law Studies*, 90, str. 308–340.
- Smolnikar, K. (2019). *Zakonitosti uporabe avtonomnega orožja v mednarodnem pravu* (magistrsko delo). Univerza v Ljubljani: Pravna fakulteta.
- Thoreau, H. D. (2016). *Državljanska nepokorščina*. Ljubljana: LUD Šerpa.
- UN. Panel of Experts Established pursuant to Security Council Resolution 1973 (2011). (2021). »Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council«. United Nations Security Council.
- Zawieska, K. (2017). »An Ethical Perspective on Autonomous Weapon Systems: Perspectives on Lethal Autonomous weapon Systems«. *UNODA (United Nations Office for Disarmament Affairs) occasional papers*, 30.