O medzere v zodpovednosti za autonómne smrtiace roboty[1]

Ivan Koniar
Katolícka univerzita v Ružomberku
ivan.koniar@ku.sk

Abstrakt:
Jednou z ústredných etických obáv, ktoré vyvoláva vyhliadka na použitie vojenských autonómnych smrtiacich robotov, je otázka morálnej zodpovednosti za ich konanie. Robert Sparrow tvrdí, že vojenské roboty vybavené schopnosťou učiť sa by boli natoľko nezávislé a samostatné, že by umožnili ľudským aktérom odmietnuť zodpovednosť za ich konanie, čím by vznikla situácia, ktorú Andreas Matthias označil ako „medzeru v zodpovednosti". Medzera v zodpovednosti je stav, keď nie je nikto zodpovedný za konanie učiacich sa autonómnych robotov. Tento stav je dôsledkom neschopnosti ľudí plne kontrolovať a predpovedať konanie takýchto technológií. V článku argumentujem, že uvedený záver je nesprávny, pretože autonómne technológie nezbavujú ľudí zodpovednosti za dôsledky ich použitia. Kontrola a predvídateľnosť nie sú nevyhnutnou podmienkou pripísania zodpovednosti. Vzhľadom na riziká, ktoré použitie takýchto zbraní predstavuje, sú tí, ktorí ich vytvárajú alebo používajú, morálne zodpovední za ich konanie. Hoci nasadenie autonómnych smrtiacich zbraní nemusí byť dobrý nápad, „medzera v zodpovednosti" ich ešte nerobí nemorálnymi.

Kľúčové slová: morálna zodpovednosť, smrtiace autonómne zbrane/roboty, kontrola, riziko

DOI: https://doi.org/10.46854/fc.2021.4r.795

Úvod

Otázka vývoja, výroby a použitia autonómnych zbraňových systémov a najmä smrtiacich autonómnych zbraní priťahuje čoraz väčšiu pozornosť. Predovšetkým nárast používania vojenských robotov a vývoj v oblasti umelej

1 Článok vznikol ako súčasť riešenia grantovej úlohy VEGA č. 1/0496/18.

inteligencie a robotiky podnecuje debaty o budúcnosti vojny a etických problémoch autonómnych zbraní. Vojensky najsilnejšia krajina sveta vidí tieto technológie ako kľúčové pre víťazstvo v budúcich konfliktoch a počet národov, ktoré vyvíjajú takéto zbrane, narastá.[2]

Zástancovia autonómnych zbraňových systémov na jednej strane tvrdia, že takéto stroje by štátom umožnili lepšie chrániť svojich občanov a znížili by riziko zabitia a zmrzačenia (vlastných) vojakov. Z dlhodobého hľadiska môžu byť roboty lacnejšie ako ľudskí vojaci, keďže nepotrebujú plat, dôchodok, bývanie, stravu alebo nemocnice. Prekonajú človeka, pokiaľ ide o rýchlosť, presnosť a schopnosť fungovať bez odpočinku. Niektorí autori sa dokonca domnievajú, že autonómne roboty sa budú správať humánnejšie a etickejšie ako ľudskí vojaci, keďže strach, únava, hnev, frustrácia alebo pocity pomsty neoslabia ich úsudok.[3]

Na druhej strane autonómne zbrane vyvolávajú obavy z budúcnosti, v ktorej „roboti zabijaci" ohrozujú ľudstvo, a viacero mimovládnych organizácií vyzvalo na to, aby boli smrtiace autonómne zbrane zakázané.[4] Kritici poukazujú na riziká zníženia prahových hodnôt zapojenia sa štátov do ozbrojených konfliktov alebo na to, že ich použitie by mohlo konflikty predĺžiť.[5] Viacerí spochybňujú schopnosť robotov rozlišovať medzi legitímnymi a nelegitímnymi cieľmi a problematizujú možnosť naprogramovať ich tak, aby spĺňali požiadavky medzinárodného humanitárneho a vojnového práva.[6] Iní poukazujú na to, že ich nasadenie môže byť neprijateľné, pretože roboty by jednoducho nemali mať možnosť rozhodovať o živote a smrti ľudských bytostí.[7]

Jedným z najdiskutovanejších problémov použitia autonómnych zbraňových systémov je otázka morálnej zodpovednosti za dôsledky konania takýchto technológií. Otázka pripísania zodpovednosti je dôležitým aspektom

2 Pozri napr.: Unmanned Systems Integrated Roadmap 2017–2042. Office of the Secretary of Defense. Dostupné na: https://www.defensedaily.com/wp-content/uploads/post_attachment/206477.pdf; [cit. 24. 10. 2021]. Boulanin, V. – Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems*. Stockholm, SIPRI 2017. Dostupné na: https://www.sipri.org/publications/2017/other-publications/mapping-development-autonomy-weapon-systems; [cit. 24. 10. 2021].

3 Arkin, R., The Case for Ethical Autonomy in Unmanned Systems. *Journal of Military Ethics*, 9, 2010, No. 4, s. 333; Wallach, W. – Allen, C., *Moral Machines: Teaching Robots Right from Wrong*. Oxford, Oxford University Press 2009.

4 V roku 2012 organizácia *Human Rights Watch* spolu s viacerými ďalšími mimovládnymi neziskovými organizáciami iniciovala kampaň *Stop Killer Robots*. O zákaze autonómnych zbraní sa opakovane diskutovalo na pôde OSN a autonómne zbrane boli agendou napr. na Svetovom ekonomickom fóre v Davose alebo Mníchovskej bezpečnostnej konferencii v roku 2016.

5 Singer, P., *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. London, Penguin 2009.

6 Sharkey, N., Saying No! To Lethal Autonomous Targeting. *Journal of Military Ethics*, 9, 2010, No. 4, s. 369–383.

7 Johnson, M. A. – Axinn, S., The Morality of Autonomous Robots. *Journal of Military Ethics*, 14, 2013, No. 2, s. 129–141.

uvažovania o etike vojny. Z pohľadu teórie spravodlivej vojny je zásadným predpokladom vedenia spravodlivej vojny to, že niekto bude zodpovedný za smrť vojakov a civilných osôb zabitých v jej priebehu.[8] Podľa niektorých autorov však môže použitie autonómnych zbraní viesť k situácii, na ktorú sa v literatúre odkazuje ako na „medzeru v zodpovednosti". Ide o (hypotetickú) situáciu, keď nikto nebude zodpovedný za konanie autonómneho systému. Základom tejto obavy je predpoklad, že ľudia nebudú schopní správne kontrolovať počítačové technológie vybavené umelou inteligenciou a softvérom strojového učenia, pretože správanie týchto strojov bude príliš zložité a nepredvídateľné. Termín „medzera v zodpovednosti" zaviedol Andreas Matthias ako súčasť všeobecnej kritiky učiacich sa autonómnych technológií. Podľa Matthiasa medzera v zodpovednosti nastáva vtedy, keď je autonómny systém naprogramovaný tak, aby prispôsoboval svoje správanie svojmu prostrediu, čím sa jeho prevádzka nedá úplne predvídať.[9]

V kontexte debaty o smrtiacich autonómnych robotoch tento argument rozpracoval Robert Sparrow. Tvrdí, že pri vojenských robotoch budúcnosti nastane medzera v zodpovednosti v dôsledku rozsahu autonómie, ku ktorej sa stále autonómnejšie technológie dopracujú. Domnieva sa, že čím viac je systém autonómny, tým má väčšiu schopnosť robiť iné voľby, než sú predpovedané jeho programátormi. V určitom okamihu už nebude možné, aby programátori kontrolovali a predpovedali správanie systému práve pre mieru autonómie tohto systému. Sparrow zároveň využíva medzeru v zodpovednosti ako základ argumentu proti používaniu autonómnych zbraňových systémov. Tvrdí, že ich používanie je nemorálne, pretože žiadny človek nemôže byť zodpovedný za to, čo urobia.[10]

Cieľom tohto článku je ukázať, že autonómne zbraňové systémy nevedú k medzere v zodpovednosti, a odmietnuť tvrdenie, že autonómne zbrane predstavujú neriešiteľný problém pre tradičné spôsoby pripísania morálnej zodpovednosti. Podľa zástancov medzery v zodpovednosti prvok nepredvídateľnosti a strata kontroly znemožňujú pripísanie zodpovednosti za konanie autonómnych technológií ľudským aktérom. Domnievam sa, že nesplnenie podmienok predvídateľnosti a kontroly ešte nemusí znamenať, že za konanie

8 Pripísanie zodpovednosti je nevyhnutné najmä pre uplatnenie princípov *ius in bello* teórie spravodlivej vojny. Napríklad princíp *diskriminácie*, ktorý vyžaduje, aby bojujúci rozlišovali medzi legitímnymi a nelegitímnymi cieľmi, predpokladá, že môžeme určiť, kto je zodpovedný za útoky, ktoré by tento princíp porušovali. Inak povedané, uplatňovanie princípov *ius in bello* vyžaduje, aby sme mohli identifikovať osoby zodpovedné za konanie, ktoré majú tieto princípy riadiť. To tiež znamená, že ak je povaha zbrane taká, že je nemožné vyvodiť zodpovednosť za obete, ktoré spôsobí, jej použitie vo vojne je nemorálne.

9 Matthias, A., The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics and Information Technology*, 6, 2004, No. 3, s. 175–183.

10 Sparrow, R., Killer Robots. *Journal of Applied Philosophy*, 24, 2007, No. 1, s. 62–77.

autonómnych zbraňových systémov nemôže byť zodpovedný žiadny človek. Oveľa dôležitejším faktorom pre pripísanie zodpovednosti v prípade autonómnych zbraní sú riziká vzniknuté nasadením takýchto zbraní. V závislosti od situácie budú za konanie autonómnych zbraní zodpovední ich dizajnéri alebo používatelia.

V článku budem pokračovať nasledovne. Najprv sa budem venovať výzvam, ktoré pre náležité pripísanie morálnej zodpovednosti predstavujú moderné výpočtové technológie a najmä technológie so schopnosťou učiť sa. Potom predstavím problém medzery v zodpovednosti a argumentáciu, ktorou medzeru v zodpovednosti obhajujú Matthias a Sparrow. V tretej časti uvediem, prečo sú predpoklady obidvoch autorov nedostatočné na preukázanie vzniku medzery v zodpovednosti, a následne predstavím, prečo sú ľudskí aktéri zodpovední za konanie smrtiacich autonómnych zbraní.

Zodpovednosť a moderné výpočtové technológie

V posledných dvoch dekádach sa na bojiskách vojnových konfliktov dramaticky rozšírila úloha robotických zbraní. Roboty menia spôsob, akým sa vo vojnách bojuje, a to vytváraním nových možností na vykonávanie prieskumov, podporných akcií či bojových operácií. Väčšina robotov používaných v súčasnosti na bojové operácie nie je plne autonómna. Niektoré činnosti síce vykonávajú samostatne, do veľkej miery však ide o diaľkovo ovládané alebo človekom priamo kontrolované zariadenia.[11] Hoci takéto zbrane sú z mnohých dôvodov kontroverzné, otázka morálnej zodpovednosti za činy spáchané takýmito zbraňami sa dá pomerne jednoznačne zodpovedať. Autonómne zbraňové systémy predstavujú oveľa vážnejšiu výzvu na určenie morálnej zodpovednosti, pretože takéto zbrane sú po svojej aktivácii schopné v širokej miere konať samostatne a nezávisle od ľudského zásahu. Autonómne smrtiace roboty sú schopné vyhľadať, vybrať a zaútočiť na ciele bez priamej kontroly alebo dohľadu operátora v reálnom čase.

Na najširšej úrovni analýzy nie je ťažké porozumieť obavám ohľadne morálnej zodpovednosti, ktoré vyvoláva vyhliadka na stále autonómnejšie, nielen vojenské, technológie. Zrejme pre veľkú časť ľudí (ako aj filozofických teórií) sa akceptovanie morálnej zodpovednosti bude spájať s nasledujúcimi podmienkami. (1) Možnosť slobodne sa rozhodnúť konať určitým spôsobom, nakoľko nemá zmysel pripisovať osobe zodpovednosť za činy, ktoré boli určené alebo vynútené vonkajšími silami. (2) Osoba by taktiež mala mať vedomosť o relevantných okolnostiach situácie, v ktorej koná, a byť schop-

11 Jednými z najznámejších príkladov sú drony ako Predator a Reaper.

ná predvídať a zvážiť možné dôsledky svojich činov. (3) V neposlednom rade by malo existovať kauzálne prepojenie medzi osobou a výsledkom konania, z ktorého by plynulo, že osoba má vplyv na výsledok udalostí. Zdá sa však, že moderné výpočtové technológie komplikujú aplikovateľnosť všetkých týchto podmienok.

(1) Slobodne sa rozhodnúť konať nejakým spôsobom je pravdepodobne najdôležitejšou podmienkou pripísania morálnej zodpovednosti, no tiež jednou z najspornejších. Máme sklon nepripisovať osobám morálnu vinu, ak nemali inú možnosť ako konať, a spravidla ospravedlňujeme konanie, ktoré boli ľudia donútení urobiť. V morálnej filozofii môže sloboda rozhodovať taktiež znamenať, že aktér má slobodnú vôľu.[12] Keď sa osoba rozhodne konať, vykonáva svoju vôľu, respektíve vykonáva svoju schopnosť voľby. A len vtedy, keď osoba vykonáva túto schopnosť ako výraz vlastného chcenia a bez inej rozhodujúcej sily, môžeme tento čin nazvať slobodným. Táto sloboda je vo filozofickom zmysle často označovaná ako autonómia.

V praxi sa však ukazuje, že prisúdiť autonómiu ľuďom nie je úplne priamočiare úsilie. Osobám pripisujeme autonómiu v stupňoch a dospelý sa všeobecne považuje za autonómnejšieho ako dieťa. Naša autonómia ako jednotlivcov sa líši, pretože sme rozdielne ovplyvňovaní a manipulovaní silami okolo nás, ako napríklad rodičmi, známymi či médiami. Vnútorné fyzické alebo psychologické vplyvy, ako sú závislosť alebo duševné problémy, sa taktiež vnímajú ako faktory obmedzujúce autonómiu osoby. No a výpočtové technológie pridávajú ďalší rozmer zložitosti pri určovaní, či niekto môže konať slobodne, pretože ovplyvňujú to, aké možnosti výberu ľudia majú.

Pravdepodobne najširšou aplikačnou oblasťou využitia výpočtovej techniky je automatizácia rozhodovacích procesov.[13] Automatizácia môže centralizovať a zvýšiť kontrolu nad viacerými procesmi, a tak obmedziť možnosti rozhodovania ľudských operátorov. V súčasnosti je mnoho rozhodovacích procesov vo verejnej správe či bankovníctve riadených algoritmami. Takéto technológie sú v dôsledku ich vyššej efektivity zámerne naprogramované tak, aby v podstate obmedzili právomoci ľudských aktérov, respektíve priamo nahradili ľudské rozhodovanie. Rovnako autonómne technológie majú potenciál ovplyvňovať slobodu rozhodovania. V prípade dopravnej nehody autonómneho – samo sa riadiaceho – vozidla bude riešenie vzniknutej situácie priamo závisieť od algoritmov zabudovaných vo vozidle, a nie od jeho ľudských užívateľov.

12 Fischer, J. M., Recent Work on Moral Responsibility. *Ethics*, 110, 1999, No. 1, s. 93–139.

13 Noorman, M., Computing and Moral Responsibility. In: Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, 2018. Dostupné na: https://plato.stanford.edu/entries/computing-responsibility/; [cit. 7. 10. 2021].

(2) So zmysluplným pripísaním zodpovednosti za rozhodnutie súvisí i to, že osoba by mala mať dostatočné vedomosti na zváženie dôsledkov svojich činov a dokázať predvídať, aké následky by jej konanie mohlo spôsobiť. Je nespravodlivé niesť zodpovednosť za konanie, ak by osoba nemohla vedieť, že konanie bude viesť k poškodeniu alebo ublíženiu. Používanie autonómnych technológií však môže obmedziť schopnosť používateľov porozumieť a zvážiť dôsledky svojich činov. Tieto technológie nie sú úplne bez chýb a za ich rozhraním sa ukrývajú čoraz zložitejšie procesy. V neurónových sieťach s tzv. hlbokým učením nie sú výsledky učenia uložené na jednom konkrétnom mieste, ale sú skôr „kódované silou viacnásobných spojení", teda podobným spôsobom, akým sa to pravdepodobne deje v ľudskom mozgu. Vzhľadom na skutočnosť, že umelé neurónové siete môžu mať milióny spojení a byť usporiadané do mnohých vrstiev, nie je ľahké spätne analyzovať procesy na pochopenie toho, čo neurónová sieť urobila a ako presne sa generujú výstupné výsledky.[14]

Technológie strojového učenia sa tak čelia problému predvídateľnosti ich konania. Vstupy a výstupy systému sú síce pozorovateľné, ale proces, ktorý vedie od vstupu po výstup, je neznámy alebo veľmi ťažko pochopiteľný. Používatelia často vidia iba časť z výpočtov, ktoré program vykonáva, a väčšinou majú iba čiastočné znalosti o predpokladoch, modeloch a teóriách, na ktorých sú založené informácie na obrazovke počítača alebo správanie robota. Nepriehľadnosť fungovania počítačových systémov tak môže prekážať pri posudzovaní platnosti a relevantnosti informácií a môže užívateľovi brániť správne predvídať následky konania.

(3) V konečnom dôsledku na to, aby osoba mohla byť morálne zodpovedná za konkrétnu udalosť, musí mať možnosť na túto udalosť uplatniť nejaký kauzálny vplyv. Je zrejme nesprávne viniť niekoho za nehodu, ak ju nemohol odvrátiť tým, že by konal odlišne, alebo ak nemal kontrolu nad udalosťami, ktoré viedli k incidentu. Príčinná súvislosť sa v prvom rade problematizuje tým, že skúmanie sledu udalostí, ktoré viedli k zlyhaniu komplexných výpočtových technológií, zvyčajne vedie mnohými smermi, pretože takéto incidenty sú zriedka výsledkom jedinej chyby. Naopak technologické nehody sú zvyčajne výsledkom hromadenia chýb, nedorozumení alebo nedbanlivého správania jednotlivcov zapojených do vývoja, používania a údržby počítačových systémov.[15]

Autonómne technológie však môžu narúšať tradičné príčinné súvislosti medzi činmi osoby a prípadnými následkami aj iným spôsobom. Podľa Davida Gunkela sú autonómne technológie zariadením, na ktoré sa nedá nahliadať

14 Castelvecchi, D., Can We Open the Black Box of AI? *Nature*, 538, 2016, No. 7623, s. 22.
15 Noorman, M., Computing and Moral Responsibility, c.d.

ako na nástroj. Priamo porušujú definíciu inštrumentu premiestnením konania z človeka na umelú entitu. Takéto mechanizmy potom nie sú iba nástrojmi, ktoré majú ľudia používať, ale skôr zastávajú miesto ľudského aktéra.[16] Príkladom sú autonómne vozidlá, ako *Google Car* a iné, ktoré nie sú navrhnuté alebo určené na nahradenie dopravného prostriedku, ale v podstate nahrádzajú úlohu vodiča. Avšak prvá fatálna nehoda prototypu autonómneho auta spoločnosti *Uber* ukázala problematickosť pýtania sa na to, kto nehodu spôsobil.[17]

Podobne počítačový program *AlphaGo*, navrhnutý na hranie japonskej doskovej hry go, opakovane preukázal schopnosť poraziť ľudského majstra v tejto hre. Ak sa však spýtame, kto porazil majstra Leeho Sedola, je otázne, či možno toto víťazstvo pripísať programátorom, ktorí *AlphaGo* dizajnovali. Bola porážka ľudského protihráča skutočne dôsledkom ich spoločného úsilia? Toto vysvetlenie je značne problematické aplikovať práve na taký program, ako je *AlphaGo*, ktorý bol zámerne vytvorený na to, aby robil veci, ktoré presahujú znalosti a kontrolu jeho ľudských dizajnérov.[18] V skutočnosti v správach o tejto udalosti nebola za víťaza označovaná spoločnosť Google ani inžinieri a programátori spoločnosti Deep-Mind, ale samotný program *AlphaGo*.

A zdá sa, že prvé prípady odmietania zodpovednosti za správanie autonómnych technológií možno nájsť už dnes. V marci 2016 firma Microsoft uviedla v rámci služby Twitter chatovacieho bota nazvaného *Tay*. Po šestnástich hodinách online fungovania ho stiahla, pretože *Tay* začal postovať bigotné a rasistické tweety. Podľa vyjadrenia Microsoftu to bolo spôsobené tým, že sa užívatelia snažili zneužiť schopnosti *Tay* tak, aby odpovedal nevhodným spôsobom. Ani programátori, ani spoločnosť však podľa tohto vyjadrenia nie sú zodpovední za túto *hate speech*. Neskôr firma Microsoft vydala druhé miernejšie stanovisko, v ktorom sa ospravedlnila za nezamýšľané urážlivé tweety a uviedla, že *Tay* bude späť vtedy, keď bude spoločnosť Microsoft schopná lepšie predvídať zlý úmysel. Avšak aj podľa týchto slov je

16 Gunkel, J. D., Mind the Gap: Responsible Robotics and the Problem of Responsibility. *Ethics and Information Technology*, 22, 2020, No. 4, s. 310.

17 20. Novembra 2019 *National Transportation Safety Board* vydala správu o vyšetrovaní, ktorá nehodu pripísala na vrub veľkému množstvu ľudských chýb a kultúre bezpečnosti spoločnosti *Uber*. Dostupné na: https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1903.pdf; [cit. 24. 10. 2021].

18 Thore Graepel, jeden z tvorcov *AlphaGo*, tvrdí, že aj keď bol *AlphaGo* naprogramovaný, nemali sme tušenie, s akými ťahmi príde. Jeho ťahy sú vznikajúci fenomén, my sme vytvorili iba súbory údajov a výcvikové algoritmy. Ťahy, ktoré *AlphaGo* potom urobil, nie sú naše. Dostupné na: https://www.wired.com/2016/03/googles-ai-wins-pivotal-game-two-match-go-grandmaster/; [cit. 24. 10. 2021].

spoločnosť Microsoft zodpovedná skôr za to, že nedokázala predvídať zlý úmysel, a nie za urážlivé odpovede, ktoré vytvoril jej softvér.[19]

Google Car, *AlphaGo* i *Tay* sú prípadmi umelých aktérov, ktorí sú vybavení algoritmami strojového učenia. Termín „umelý aktér" označuje technologický artefakt, ktorý vníma prostredie a v mene klienta koná a plní úlohy, ktoré obecne sú v danom prostredí zadané. Predovšetkým výskum v oblasti umelej inteligencie umožňuje vytvárať softvérové systémy, ktoré vykazujú istú úroveň inteligencie na plnenie úloh a učenie sa novým úlohám bez ľudského vedenia, dozerania alebo zásahu. Takíto umelí aktéri môžu mať čisto softvérovú podobu – ako napríklad internetové boty, vyhľadávače či počítačové vírusy – alebo môžu byť integrovaní do fyzických entít, ako sú roboty.

Strojové učenie možno chápať ako počítačový program, ktorý prostredníctvom svojej skúsenosti zlepšuje výkon pri nejakej úlohe.[20] Softvér dokáže prostredníctvom prílivu veľkého množstva údajov rozoznať vzory v týchto údajoch a učiť sa z nich. Niektorí umelí učiaci sa aktéri majú schopnosť samostatne – bez zásahu človeka upravovať pravidlá rozhodovania a prichádzať s novými vzormi a modelmi, ktoré sa dajú použiť na predpovedanie nových údajov. Napríklad algoritmy strojového učenia sa, ktoré sa používajú na rôzne klasifikačné úlohy, dokážu navrhnúť triedy, ktoré idú nad rámec pôvodných tréningových údajov, a definovať pravidlá rozhodovania pre spracovanie nových vstupov.[21] Učenie sa tak môže zohrávať významnú úlohu pri rozširovaní autonómie výpočtových artefaktov. Ak je artefakt schopný získať nové vzorce správania pomocou vhodného učenia sa, potom sa postupom času môže autonómia systému zvýšiť. Viacerí autori sa tak domnievajú, že učebné kapacity môžu poskytnúť robotom taký stupeň autonómie, v zmysle schopnosti samostatne rozhodovať, vďaka ktorému programátori stratia vplyv a kontrolu nad konaním takéhoto stroja. Stroju tak možno pripísať efektívne samostatné rozhodovanie o svojej činnosti a podľa niektorých i morálnu zodpovednosť za konanie.[22]

Medzera v zodpovednosti

Hlavná výzva pre otázku zodpovednosti autonómnych zbraňových systémov prichádza od tých, ktorí poukazujú na problematický aspekt strojového uče-

19 Lee, P., Learning from Tay's Introduction. *Official Microsoft Blog* (2016, March 25). Dostupné na: https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/; [cit. 7. 9. 2021].

20 Mitchell, M. T., *Machine Learning*. New York, McGraw Hill 1997, s. 2.

21 Domingos, P., A Few Useful Things to Know about Machine Learning. *Communications of the ACM*, 55, 2012, No. 10, s. 78–87.

22 Pozri napr.: Floridi, L. – Sanders, J., On the Morality of Artificial Agents. *Minds and Machines*, 14, 2004, No. 3, s. 349–379; Sullins, P. J., When Is a Robot a Moral Agent? *International Review of Information Ethics*, 12, 2006, No. 1, s. 23–30.

nia sa. Keďže sa niektorí umelí aktéri – roboty počas svojho fungovania dokážu učiť, tí, ktorí ich navrhli, a tí, ktorí ich nasadia, nemusia byť schopní pochopiť, ako tieto roboty dospeli k rozhodnutiam, a predpovedať, čo pri ich použití presne urobia. Existuje preto obava, že tým, ako sa roboty budú stávať čoraz viac autonómnymi, zodpovednosť za ich správanie nebude možné nikomu pripísať. Matthias charakterizuje túto situáciu ako medzeru v zodpovednosti. Uvádza, že na to, aby bola osoba považovaná za zodpovednú za konanie, musí mať kontrolu nad svojím správaním a dôsledkami konania. Len ak osoba pozná konkrétne skutočnosti, ktoré sa spájajú s jej konaním, a ak je schopná slobodne sa rozhodnúť konať, možno jej pripísať morálnu zodpovednosť za konanie.[23]

Podľa Matthiasa vývoj výpočtových, vysoko prispôsobivých a autonómne prevádzkovaných zariadení nevyhnutne vedie k strate kontroly obsluhy nad zariadením. Autonómne, učiace sa stroje vytvárajú novú situáciu, keď výrobca a operátor v zásade už nie je schopný predpovedať a kontrolovať budúce správanie stroja, a preto zaň nemôže byť morálne zodpovedný. Matthias uvádza, že: „(…) v súčasnosti sa vyvíjajú alebo už používajú stroje, ktoré dokážu rozhodnúť o postupe a konať bez ľudského zásahu. Pravidlá, podľa ktorých konajú, nie sú stanovené počas výrobného procesu, ale môžu byť zmenené počas prevádzky stroja samotným strojom. Toto nazývame strojové učenie sa.“[24]

Strojové učenie sa umožňuje, že rozhodnutia riadiaceho programu sa nezakladajú iba na predprogramovaných údajoch, ale aj na skutočnostiach, ktoré boli pridané do databázy stroja až po jeho spustení. Tieto údaje nie sú súčasťou pôvodného programu, ale predstavujú skutočnú skúsenosť získanú autonómnym strojom v priebehu jeho prevádzky. Výrobca tak môže odmietnuť zodpovednosť za konanie stroja, pretože stroj, ktorý bol schopný učiť sa, môže zmeniť parametre svojho programu v priebehu svojej prevádzky tak, aby sa lepšie prispôsobil svojmu prostrediu. Z tohto dôvodu už nie je možné, aby výrobca predpovedal alebo kontroloval správanie stroja v danej situácii.[25]

Podľa Matthiasa tak možno ukázať, že existuje rastúca trieda činností strojov, kde tradičné spôsoby zodpovednosti nie sú zlučiteľné s naším zmyslom pre spravodlivosť a morálnym rámcom spoločnosti. Nikto viac nemá natoľko dostatočnú kontrolu nad činnosťou stroja, aby bol schopný prevziať za neho zodpovednosť. Čelíme tak stále sa rozširujúcej medzere v zodpovednosti, ktorá predstavuje hrozbu ako pre konzistentnosť morálneho rámca spoločnosti, tak aj pre základy koncepcie zodpovednosti v práve.[26] Na pod-

23 Matthias, A., The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata, c.d., s. 175.
24 Tamže, s. 177.
25 Tamže, s. 176.
26 Tamže.

poru svojej pozície Matthias popisuje niekoľko vyvíjaných alebo už používaných systémov, ktoré majú príslušné charakteristiky.[27]

V článku „Killer Robots" Sparrow uvádza podobný argument – tentoraz v kontexte autonómnych zbraňových systémov. Sparrow chápe autonómiu ako vec škály a píše, že existujúce autonómne zbraňové systémy zahŕňajú riadené strely, torpéda, roboty pre prieskum či bojové drony. Tieto zbrane majú len veľmi limitovanú úroveň autonómie, hoci podľa Sparrowa existujú zbraňové systémy, ktoré na svoju činnosť nevyžadujú takmer žiadnu ľudskú kontrolu. Sparrow však upozorňuje aj na ďalšiu generáciu inteligentných robotov. Odvolávajúc sa na počítačových vedcov, uvádza, že inteligentné zbraňové systémy budú schopné robiť vlastné rozhodnutia a budú tak robiť inteligentným spôsobom. Aj keď budú naprogramované tak, aby prijímali rozhodnutia podľa určitých pravidiel, ich kroky nebudú v zásade predvídateľné. Tieto systémy budú mať významnú kapacitu na utváranie a revidovanie svojho programu a možnosť poučiť sa zo skúseností. V praxi to bude znamenať, že činnosť týchto strojov sa rýchlo stane nepredvídateľnou.[28]

Sparrow sa následne pýta, kto by mal niesť zodpovednosť za vojnový zločin v situácii, ktorá zásadne zahŕňa rozhodnutie takéhoto autonómneho zbraňového systému. Berúc do úvahy programátorov, veliaceho dôstojníka a samotný stroj ako kandidátov na prevzatie zodpovednosti, Sparrow tvrdí, že žiadnemu z nich nie je možné zodpovednosť pripísať. Jeho vysvetlenie, prečo nie sú zodpovední programátori, ilustruje akceptovanie medzery v zodpovednosti. Sparrow píše: „(...) možnosť, že autonómny systém urobí rozhodnutia iné ako tie, ktoré predvídajú jeho programátori, spočíva v tvrdení, že je autonómny. Ak má dostatočnú autonómiu, aby sa poučil zo svojich skúseností a okolia, potom môže robiť rozhodnutia, ktoré odzrkadľujú jeho skúsenosť rovnako alebo viac ako počiatočné programovanie. Čím viac je systém autonómny, tým má viac kapacity na výber iných možností, ako predpokladajú jeho programátori. V určitom okamihu už nebude možné považovať programátorov a dizajnérov za zodpovedných za výsledky, ktoré nemohli kontrolovať ani predvídať. Spojenie medzi programátorom a výsledkami systému, ktoré sú základom priraďovania zodpovednosti, je prerušené autonómiou systému."[29]

Sparrow dodáva, že považovať programátorov za zodpovedných za konanie ich výtvoru, by bolo rovnaké, ako považovať rodičov za zodpovedných za konanie svojich detí, ktoré už viac nie sú v ich starostlivosti. V podobnom duchu Sparrow odmieta pripísať zodpovednosť za konanie autonómnych

27 Tamže, s. 178–181.
28 Sparrow, R., Killer Robots, c.d., s. 65.
29 Tamže, s. 70.

strojov vojenským veliteľom. Podľa Sparrowa autonómia strojov znamená, že rozkazy neurčujú činnosť stroja, hoci jeho správanie zjavne ovplyvňujú. Používanie autonómnych zbraní preto predstavuje riziko, že vojenský personál bude braný na zodpovednosť za činnosť strojov, ktorých rozhodnutia nekontroloval. S vývojom čoraz autonómnejších technológií toto riziko narastá a v určitom okamihu už nebude férové, aby veliaci dôstojník bol zodpovedný za činnosť stroja. Ak si stroje skutočne samostatne vyberajú svoje ciele, veliaci dôstojník nemôže niesť zodpovednosť za také činy stroja, ktoré by sa mohli označiť za vojnové zločiny.[30] Sparrow však nepripisuje zodpovednosť ani samotnému zbraňovému systému. Tvrdí, že na to, aby sme mohli aktéra považovať za morálne zodpovedného za svoje činy, musí mať schopnosť trpieť, inak nie sme schopní potrestať ho. Stroje tak nemôžu byť brané na zodpovednosť, pretože nemôžu byť potrestané, respektíve nemôžu byť potrestané, pretože nemôžu trpieť.[31]

Zdá sa, že Matthias a Sparrow predpokladajú, že do používania budú zavedené také technológie, pri ktorých ľudia už nebudú schopní predpovedať a kontrolovať ich správanie. Podľa Matthiasa existuje medzera v zodpovednosti vtedy, keď je stroj navrhnutý tak, aby prispôsoboval správanie svojmu prostrediu, čím sa jeho prevádzka nedá úplne predvídať a kontrolovať. Sparrowovo chápanie medzery v zodpovednosti je špecifickejšie. Zodpovednosť podľa neho súvisí s autonómiou v tom zmysle, že ak aktér koná samostatne, nie je možné, aby za jeho konanie niesol zodpovednosť niekto iný, navyše niekto, kto nemohol toto konanie predvídať a kontrolovať. Avšak, keďže umelým aktérom nie je možné pripísať morálnu zodpovednosť za ich konanie, čelíme tak medzere v zodpovednosti. K takémuto pohľadu na autonómne zbraňové systémy sa hlásia i niektorí ďalší autori.[32] Argumentácia Matthiasa a Sparrowa sa stala i jedným z argumentov celosvetovej kampane *Stop Killer Robots*, ktorá usiluje o zákaz vývoja a používania autonómnych zbraní.[33]

30 Tamže, s. 71.

31 Tamže, s. 71–72.

32 Napr. Roff, H. M., Killing in War: Responsibility, Liability and Lethal Autonomous Robots. In: Allhoff, F. – Evans, N. G. – Henschke, E. (eds.), *Routledge Handbook for Ethics and War: Just War Theory in the 21st Century*. London, Routledge Press 2013, s. 352–364.

33 Správa *Losing Humanity,* ako aj niektoré ďalšie správy organizácie *Human Rights Watch* tvrdia, že roboti zabijaci by okrem iného nespĺňali právo na opravný prostriedok, pretože zodpovednosť v prípade porušenia práva by bola nejasná. Správa *Mind the Gap: the Lack of Accountability for Killer Robots* definuje plne autonómne smrtiace zbrane ako zbraňové systémy, ktoré vyberajú ciele a útočia bez zmysluplnej ľudskej kontroly. Pozri: http://www.stopkillerrobots.org.; [cit. 24. 10. 2021].

Odpoveď na medzeru v zodpovednosti

Existuje viacero prístupov k riešeniu problému medzery v zodpovednosti, ktorý vytvárajú učiace sa autonómne technológie.[34] Odpoveď, ktorú ponúka tento článok, je založená na spochybnení základného predpokladu argumentu medzery v zodpovednosti ohľadom spôsobu pripisovania zodpovednosti. S poukázaním na praxe pripisovania zodpovednosti, ktoré fungujú i napriek skutočnosti, že osoby neboli schopné predvídať a kontrolovať výsledok svojho konania. Na začiatok je potrebné poznamenať, že argument, ktorý Matthias a Sparrow predkladajú, nie je o tom, že nedokážeme určiť, kto je zodpovedný, ale o tom, že nikto nie je zodpovedný. Nejde teda o epistemický problém identifikovania osôb, ktoré sú zodpovedné za konanie – napríklad na spôsob problému „mnohých rúk". Pri medzere v zodpovednosti jednoducho žiadny človek nenesie morálnu zodpovednosť za konanie autonómnych technológií. Ich argument sa však predsa len odvoláva na epistemické (a praktické) obmedzenia, ktoré sa týkajú ľudských schopností predvídať a kontrolovať správanie sa učiacich sa robotov v prevádzkových situáciách.

Argumentácia Matthiasa a Sparrowa je postavená na troch základných premisách. Prvou je, že vzhľadom na schopnosť umelých aktérov učiť sa programátori a veliaci dôstojníci nemusia vedieť (t. j. predvídať), čo autonómny robot počas prevádzky urobí. Druhou je, že žiadna z uvedených osôb nie je schopná slobodne rozhodnúť a ovplyvniť konanie (t. j. kontrolovať) učiaceho sa robota, ktorý sa po jeho nasadení riadi samo-regulujúcim sa programom. Treťou je, že osoba môže byť braná na zodpovednosť za ujmu iba vtedy, ak má kontrolu nad vývojom udalostí v tom zmysle, že má vedomosti o skutočnostiach týkajúcich sa konania, ktoré vedie k ujme. Na základe týchto skutočností je osoba následne schopná slobodne konať a ovplyvniť vývoj udalostí. Záverom argumentu je, že ani programátori, ani výrobcovia alebo veliaci dôstojníci či operátori takýto typ kontroly nemajú. Vzniká tak nejaký druh morálneho vákua, ktoré nemožno prekonať našimi tradičnými pojmami zodpovednosti.

Podľa Matthiasa a Sparrowa sú *predvídateľnosť* – v zmysle poznania okolností konania – a *kontrola* – v zmysle slobody rozhodovať sa a ovplyvňovať

34 Schulzke, M., Autonomous Weapons and Distributed Responsibility. *Philosophy & Technology*, 26, 2013, No. 2, s. 203–219; Hellström, T., On the Moral Responsibility of Military Robots. *Ethics and Information Technology*, 15, 2013, No. 2, s. 99–107; Noorman, M. – Johnson, G. D., Negotiating Autonomy and Responsibility in Military Robots. *Ethics and Information Technology*, 16, 2014, No. 1, s. 51–62; Johnson, G. D., Technology with No Human Responsibility. *Journal of Business Ethics*, 4, 2015, No. 1, s. 707–715; Champagne, M. – Tonkens, R., Bridging the Responsibility Gap in Automated Warfare. *Philosophy & Technology*, 28, 2015, No. 1, s. 125–137; Santoni de Sio, F. – Hoven, J., van den, Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, 5, 2018, No. 15, s. 1–14.

konanie – ústrednými podmienkami na pripísanie zodpovednosti. Náš zmysel pre spravodlivosť nás však zrejme vždy nenúti považovať predvídateľnosť a kontrolu za nevyhnutné podmienky na pripísanie morálnej zodpovednosti. Pripísanie zodpovednosti nie je zásadne ohrozené v situáciách nedostatku poznania a kontroly. Pripísanie zodpovednosti totiž nevyhnutne nevyžaduje plné predvídanie situácie a akt priameho ovládania z pozície, ktorá je prepojená v priestore a čase. Zdá sa, že pripísanie zodpovednosti umožňuje oddelenie osoby a príslušných morálnych účinkov konania, na ktorom sa osoba zúčastňuje. Náš tradičný systém pripisovania zodpovednosti je omnoho robustnejší a Matthias a Sparrow nezohľadňujú celý kontext. Na obhájenie tejto pozície sa poďme bližšie pozrieť na obidve postulované podmienky.

Medzera v zodpovednosti znamená, že autonómny robot môže urobiť niečo, čo programátor priamo nenaprogramoval a používateľ nenariadil. Matthias a ani Sparrow však zároveň medzeru v zodpovednosti nespájajú s nejakým zlyhaním stroja, ide skôr o dôsledok adaptívnej povahy učiacich sa strojov. Prvok nepredvídateľnosti je na jednej strane presne to, čo sa požaduje od autonómnych technológií – s cieľom zaručiť ich pružnú reakciu. Programátori nedokážu programovať stroje pre každú možnú situáciu, a preto je rozumné ponechať stroj navigovať sa v zhode s okolitým prostredím samostatne. Na druhej strane, pokiaľ ide o použitie sily, nepredvídateľnosť sa stáva problematickou. Napríklad, ak nie je jasné, že robot využije smrtiacu silu iba na špecifikovaný cieľ, jeho nasadenie by porušilo vojnové konvencie a princípy *ius in bello* teórie spravodlivej vojny.

Je ale sporné, či táto nevedomosť implikuje medzeru v zodpovednosti. Napokon naprogramovaním zbrane tak, aby po svojom nasadení konala autonómne – samostatne a nezávisle, sa používateľ vedome a dobrovoľne vzdáva priamej kontroly nad zbraňou. A podobne ako v prípade nepredvídateľnosti je odovzdanie kontroly v zmysle schopnosti konať samostatne a nezávisle od ľudského operátora práve to, čo robí tieto technológie a obzvlášť zbrane zaujímavými. Avšak kontrola ako schopnosť slobodne sa rozhodnúť a ovplyvniť to, čo sa stane, je v takomto prípade značne obmedzená. Otázkou teda je, či môžu byť programátori a používatelia takýchto strojov zodpovední za to, čo sa stane počas nasadenia. A to napriek tomu, že odovzdali stroju kontrolu, vediac, že stroj môže adaptívnym a nepredvídateľným spôsobom interagovať so svojím prostredím.[35]

Sparrow obhajuje svoju verziu medzery v zodpovednosti využitím dvoch myšlienkových experimentov. Najprv si predstavme vysoko sofistikovaného smrtiaceho autonómneho robota, ktorý zabije skupinu vzdávajúcich sa vojakov. Takíto vojaci sú z pohľadu vojnového práva považovaní za nebojujúcich,

35 Leveringhaus, A., *Ethics and Autonomous Weapons*. Oxford, Palgrave Macmillan Ltd. 2016, s. 80.

a preto ich zabitie možno klasifikovať ako vojnový zločin. Sparrow zdôrazňuje, že tento čin nebol nejakým druhom chyby stroja, nedošlo k zlyhaniu zacielenia ani nedošlo k zámene rozkazov. Bolo to rozhodnutie autonómneho robota s plným poznaním situácie a dôsledkov. Robot sa podľa Sparrowa mohol samostatne dopracovať k tomu, že vzhľadom na okolnosti by v tomto prípade bolo príliš nákladné zajať vojakov a držať ich nažive. Naprogramované algoritmy zároveň umožňovali robotovi prepojiť premenné „životy ľudí" a „náklady", keďže takéto porovnanie bolo potrebné na celkové úspešné ukončenie cieľov jeho misie. Programátori však priamo nenaprogramovali to, aby zabil vzdávajúcich sa vojakov.[36]

Teraz si predstavme skupinu detských vojakov, ktorých ich veliteľ vyslal do dediny, aby tam zabili nepriateľských bojovníkov. Deti dokážu premôcť nepriateľa, avšak utrpia značné straty. Keď zistia, koľkí ich kamaráti zahynuli, frustrovaní touto situáciou zmasakrujú civilné obyvateľstvo v dedine. Sparrow poukazuje na to, že rovnako ako by sme deti zrejme nepovažovali za zodpovedné za to, čo urobili dedinčanom, je nemožné pripísať zodpovednosť autonómnemu stroju. Ani deti (z dôvodu veku), ani robot (z dôvodu neschopnosti trpieť) nie sú plnohodnotnými morálnymi aktérmi, hoci majú dostatočné schopnosti – úroveň autonómie na vykonanie týchto činov. Sparrow však zároveň nepripúšťa ani to, že za masaker dedinčanov je zodpovedný veliteľ, podobne ako nepripúšťa to, že za konanie učiaceho sa autonómneho robota je zodpovedný nejaký človek.[37]

Zdá sa tiež, že by s ním mal súhlasiť každý, kto by trval na podmienke predvídateľnosti a kontroly. Veliteľ totiž nevedel, že si deti vybijú zlosť na dedinčanoch, a po ich odchode do dediny ich už nemohol zastaviť. Rovnako ako veliteľ autonómneho robota nevedel, čo presne stroj urobí, a po jeho vyslaní nad ním stratil kontrolu. Avšak, je veliteľ detských vojakov skutočne zodpovedný len za smrť padlých v boji, ktorý sa uskutočnil na jeho priamy rozkaz, alebo aj za masaker, ktorý spáchali deti na ich vlastný popud? Myslím si, že veliteľ oddielu, rovnako ako ostatný, ktorí sa aktívne podieľali na manipulácii a vyzbrojovaní bojujúcich detí, nesú morálnu zodpovednosť za smrť každého človeka, ktorý bol pri incidente zabitý. Všetky tieto osoby sú zodpovedné, keďže do rúk detí vložili zbrane a vystavili ich extrémnym situáciám a na ich vek neadekvátnemu psychickému tlaku. Je nedbanlivosťou zveriť deťom nástroje na vykonanie úlohy, ktorá je sama o sebe prípustná, ak tieto nástroje na svoju činnosť vyžadujú istú technickú a morálnu spôsobilosť. Skutočnosť, že deti môžu mať na vykonávanie úlohy *takmer* dostatočnú technickú a morálnu spôsobilosť, nezbavuje dospelých (ktorí ich na túto úlohu delegovali)

36 Sparrow, R., Killer Robots, c.d., s. 66.
37 Tamže, s. 73–74.

zodpovednosti. Inými slovami povedané: blízka, ale nie úplná kompetencia detského vojaka nevytvára medzeru v zodpovednosti.[38] Zbrane by nemali byť zverené deťom, pretože im chýba najmä morálna spôsobilosť na ich správne používanie. Tí, ktorí tak urobia, zodpovedajú za akékoľvek zneužitie týchto zbraní.

Taktiež sa možno zmysluplne domnievať, že úroveň autonómie smrtiaceho vojenského robota nezbavuje tých, ktorí ho nasadili alebo naprogramovali, zodpovednosti za nespravodlivé zabitia, ktoré spáchal. V danom kontexte autonómia smrtiaceho robota znamená hlavne jeho spôsobilosť pracovať bez ľudského dozoru. Autonómny robot je stroj schopný vykonávať určité funkcie samostatne, bez potreby ľudského operátora a autonómia strojov v technickom zmysle znamená len to, že operátor sa stáva zbytočným. Úroveň autonómie stroja súvisí predovšetkým s úrovňou kontroly, ktorú stroj má nad vykonávaním rôznych procesov – v porovnaní s tým, koľko ľudského zásahu je potrebného.

Vyššia úroveň autonómie je pripisovaná tým systémom, ktoré sú ponechané na vykonávanie úloh samostatne a ktoré majú nad týmito procesmi väčšiu kontrolu. V kontexte výpočtových technológií je autonómia pozorovateľná a merateľná vlastnosť vzťahu medzi umelým aktérom a prostredím, a ako taká nemá žiadne morálne alebo normatívne konotácie. To, že autonómna technológia môže samostatne a dlhší čas pôsobiť v nejakom prostredí, nemá morálne dôsledky pre samotnú technológiu alebo človeka.[39] Autonómia je však relatívnym pojmom a v rámci oborov, ako i naprieč nimi, či už ide o robotiku, softvérové inžinierstvo, biológiu, kognitívne vedy alebo o filozofiu, majú odborníci rozdielny pohľad na to, kedy možno systém považovať za autonómny a čo autonómia presne znamená.

Obidva Sparrowom uvádzané príklady problematického pripísania zodpovednosti spadajú do širokej kategórie, ktorá sa týka napríklad rodičov detí, majiteľov domácich zvierat, vlastníkov výrobných zariadení či nehnuteľností. A zrejme aj všetkých tých prípadov, keď je ťažké určiť pôvod kauzálneho reťazca, ktorý vedie k poškodzujúcej udalosti. Rodičia a opatrovníci, ktorí nezabezpečia primerané vzdelanie, starostlivosť a dohľad, môžu byť za určitých okolností morálne zodpovední za škody spôsobené ich deťmi, hoci neexistuje jasný príčinný reťazec, ktorý by ich spájal s udalosťami. Podľa Marca Champagnea a Ryana Tonkensa sú rodičia aspoň čiastočne zodpovední

38 Simpson, W. T. – Müller, V., Just War and Robot's Killings. *The Philosophical Quarterly*, 66, 2015, No. 263, s. 306.

39 Noorman, M., Limits to the Autonomy of Agents. In: Briggle, A. – Waelbers, K. – Brey, A. E. P. (eds.), *Current Issues in Computing and Philosophy*. Amsterdam, IOS Press 2008. s. 68. Podrobnejšie k tejto téme: Noorman, M., *Mind the Gap. A Critique of Human/Technology Analogies in Artificial Agents Discourse.* Maastricht, Universitaire Pers Maastricht 2008, s. 103–139.

za prípravu svojich detí na okamih, keď opustia ich starostlivosť a stanú sa nezávislými. Súčasťou úlohy rodiča je povinnosť zadosťučiniť tejto zodpovednosti a vychovávať dieťa tak, aby sa naučilo správať sa morálne prijateľným spôsobom, a akceptovať aspoň čiastočnú zodpovednosť aj v prípadoch, keď tak intencionálne neučiní.[40]

Majitelia zvierat môžu byť na základe vzťahu vlastníctva taktiež morálne zodpovední v prípadoch, keď ich zverenec spôsobí materiálnu škodu, ujmu na zdraví či smrť, napriek tomu, že sa zvieratá dokážu učiť, interagovať s prostredím či vytvárať vlastné spôsoby komunikácie. Mateo Santoro, Dante Marino a Guglielmo Tamburrini poukazujú na to, že výrobcovia tovaru a zamestnávatelia sú zodpovední na základe ešte menej priamych príčinných súvislostí, ktoré sú zhrnuté v zásade rímskeho práva *ubi commoda ibi incommoda*.[41] V týchto prípadoch sa očakávaný zisk výrobcu alebo zákonného vlastníka považuje za dostatočný základ pre pripísanie zodpovednosti za škody spôsobené výrobkom. Podobná je i situácia zamestnávateľa, ktorý nedodržal normy náležitej starostlivosti alebo iné predpisy súvisiace s bezpečnosťou a zdravím zamestnancov.

Pri riešení problémov tohto typu sa pripísanie zodpovednosti nezačína od existencie jasného príčinného reťazca alebo od schopnosti kontrolovať a predvídať dôsledky konania. Rozhodujúce sú tie aspekty, ktoré sa týkajú identifikácie možných škôd, ich sociálnej udržateľnosti a spôsobu rozdelenia kompenzácií za tieto škody.[42] Vzhľadom na epistemické a praktické obmedzenia, ktoré súvisia s predvídateľnosťou a kontrolou konania učiacich sa robotov, možno problémy s pripísaním zodpovednosti za ich konanie zaradiť práve do tejto kategórie problémov. A ako sa zdá, existuje viacero fungujúcich koncepčných rámcov, ktoré umožňujú riešiť problémy zodpovednosti bez odvolania sa na požiadavku predvídania a kontroly.

Alex Leveringhaus sa na príklad domnieva, že osoba môže byť zodpovedná za nepredvídateľný výsledok, ak sa správala bezohľadne alebo nedbanlivo. Ak osoba na seba vzala nadmerné riziká alebo nezohľadnila ďalšie riziká a okolnosti konania, je zodpovedná za to, čo sa stalo, bez ohľadu na nedostatok úmyslu, poznania alebo kontroly. V kontexte príkladu detských vojakov Leveringhaus poznamenáva, že veliteľ je zodpovedný v prvom rade preto, lebo porušil práva týchto detí tým, že ich použil ako vojakov. Po druhé veliteľ nielenže uškodil deťom, ale navyše vystavil neprimeraným rizikám dedinčanov.

40 Champagne, M. – Tonkens, R., Bridging the Responsibility Gap in Automated Warfare, c.d., s. 134.

41 Santoro, M. – Marino, D. – Tamburrini, G., Learning Robots Interacting with Humans: from Epistemic Risk to Responsibility. *AI & SOCIETY*, 22, 2008, No. 3, s. 310.

42 Marino, D. – Tamburrini, G., Learning Robots and Human Responsibility. *International Review of Information Ethics*, 6, 2006, No. 12, s. 60.

Veliteľ postúpil kontrolu skupine vysoko nestabilných detí – a bez ohľadu na to neprimerane očakával, že deti sa budú v extrémnej situácii správať normálne.[43]

Ako ukazuje Sparrowov, príklad vraždiaceho robota, používanie robotov s takým typom algoritmov a mierou ďalších technických schopností môže byť nesmierne riskantným krokom. Keďže autonómne stroje môžu byť nepredvídateľné, ich použitie predstavuje riziko. Ak umožnime nepredvídateľnému ozbrojenému robotovi porovnať premenné „ľudský život" a „náklady", môže to byť rovnako neprimerane riskantné a bezohľadné, ako používať detských vojakov. Tak ako je veliteľ morálne zodpovedný za riziká, ktoré spôsobil zneužitím detských vojakov, sú dizajnéri a používatelia autonómnych zbraní morálne zodpovední za vytváranie rizík pre jednotlivcov na bojiskách vojnových konfliktov.[44]

Vo všeobecnosti sú technológie konštruované podľa stanovených noriem fungovania a testované na spoľahlivosť a predvídateľnosť. Od technológií sa zvyčajne očakáva vysoký stupeň spoľahlivosti a vyžaduje sa nízka miera rizika. Pri nenulovej tolerancii zlyhania je tak však stále možné namietať, že medzera v zodpovednosti predsa len existuje a to v prípade, keď technológie čelia podmienkam, ktoré nezodpovedajú stanovenému stupňu spoľahlivosti. Je ale otázne, či fungovanie mimo stanovenú úroveň spoľahlivosti je vo všetkých prípadoch skutočne spoľahlivým nástrojom pre odmietnutie zodpovednosti.[45] V reakcii na uvedenú námietku Thomas Simpson a Vincent Müller poukazujú na to, že niekedy nie je možné určiť, či technológia funguje v rámci očakávanej miery spoľahlivosti na základe jedného prípadu. Zoberme si legálne schválený liek, ktorý pri jednej z tisíc osôb vyvoláva bolestivé vedľajšie účinky. To, či bolestivá reakcia konkrétneho pacienta na požitý liek je v rámci alebo mimo rámec očakávanej spoľahlivosti, závisí od súhrnných údajov o tom, koľko ďalších osôb po požití tohto lieku takúto reakciu utrpelo. Podľa obidvoch autorov to zároveň znamená, že farmaceutická firma, ktorá predáva liek, o ktorom vie, že má bolestivé vedľajšie účinky, je v každom jednom prípade zodpovedná za svoj produkt.[46] Jeff McMahan v podobnom duchu tvrdí, že osoba, ktorá použitím technológie vystaví inú osobu riziku, hoci i bezvýznamnému, je zodpovedná za dôsledky tohto konania. McMahan uvádza, že ak opatrný vodič dobre udržiavaného auta počas bizarnej nehody stratí

43 Leveringhaus, A., *Ethics and Autonomous Weapons*, c.d., s. 81–82.

44 Tamže, s. 80.

45 Samozrejme, ak sa pod náporom prívalovej vody, ktorá niekoľkonásobne presahuje rozumne očakávanú úroveň za posledných dvesto rokov merania, zrúti dobre postavený most, za prípadné úmrtia spôsobené kolapsom mosta nebude nikto zodpovedný. Tak ako nie je nikto zodpovedný za smrť po zásahu blesku.

46 Simpson, W. T. – Müller, V., Just War and Robot's Killings, c.d., s. 309.

kontrolu nad šoférovaním a v dôsledku toho usmrtí chodca, je zodpovedný za jeho smrť. Vodič je zodpovedný, pretože sa dobrovoľne zapojil do činnosti vytvárajúcej riziko a hrozbu pre iných.[47]

Vyššie uvedení autori sa však zároveň domnievajú, že ako farmaceutická firma, tak i vodič auta nie sú morálne odsúdeniahodní za vzniknuté dôsledky. Firma ani vodič si nezaslúžia byť morálne obvinení a právne trestuhodní, avšak stále sú zodpovední za to, čo sa stalo. Nie je nezvyčajné, že osoba je morálne zodpovedná za konanie v zmysle povinnosti vysvetliť svoje konanie a prípadne prijať opatrenia na nápravu negatívnych účinkov, a to bez toho, že sa týmto konaním dopustila niečoho morálne odsúdeniahodného.[48] Nastolená námietka ohľadom existencie medzery v zodpovednosti, ktorú technológie vytvárajú v prípade, že fungujú mimo rozumne očakávanú mieru spoľahlivosti alebo primeranú mieru rizika, je tak skôr všeobecne akceptovanou medzerou v trestuhodnosti.[49] Táto medzera taktiež vzniká len v prípade, ak akceptujeme istú mieru tolerancie zlyhania, napríklad ak akceptujeme, že smrtiaci autonómny robot môže na každých sto zabitých bojujúcich zabiť jedného nevinného civilistu. Je však možné, že od autonómnych robotických zbraní (budúcnosti) budeme vyžadovať nulovú toleranciu zlyhania.

Ozbrojený konflikt ako oblasť použitia autonómnych smrtiacich robotov možno považovať za vysoko rizikovú sféru, a to najmä kvôli katastrofálnym následkom, ktoré môže mať použitie smrtiacej sily. Možno teda tvrdiť, že vzhľadom na to, že ide o vysoko rizikovú oblasť s vysokou mierou nebezpečenstva, už i samotné prijatie rizika v podobe použitia autonómnych smrtiacich robotov znamená, že používatelia a/alebo výrobcovia sú morálne zodpovední za následky ich konania. Odpoveď na otázku, komu možno pripísať morálnu zodpovednosť, sa síce bude líšiť od prípadu k prípadu, ale zvyčajne bude zahŕňať operátorov, vojenských veliteľov a politikov alebo výrobcov, programátorov a dizajnérov takejto zbrane. Výrobcovia by boli zodpovední za navrhovanie a stavbu robotov v rámci stanovenej úrovne spoľahlivosti a miery rizika, a ak by tak neurobili, zodpovedali by za nespravodlivé zabitia v dôsledku zlyhania funkčnosti. Užívatelia – politici, vojenskí velitelia a operátori by boli zodpovední za používanie robotov iba v rámci stanovenej miery spoľahlivosti a rizika. Ak ich vedome nasadia nad rámec nastavených

47 McMahan, J., The Basis of Moral Liability to Defensive Killing. *Philosophical Issues*, 15, 2005, No. 1, s. 394.

48 Podrobnejšie o vzťahu medzi zodpovednosťou a trestom pozri: Lucas, J. R., *Responsibility*. Oxford, Clarendon Press 1993, s. 86–123.

49 O medzere v trestuhodnosti pozri tiež: Simpson, W. T. – Müller, V., Just War and Robot's Killings, c.d., s. 307; Leveringhaus, A., *Ethics and Autonomous Weapons*, c.d., s. 84.

parametrov, sú oni zodpovední za následné zabitia v dôsledku nesprávneho použitia.[50]

Záver

Rozhodnutím delegovať rôzne kompetencie na umelých aktérov sa ľudia nezbavujú morálnej zodpovednosti za ich konanie. Pripísanie zodpovednosti v prípade autonómnych učiacich sa technológií môže byť komplikované, čo však neznamená, že za ne nie je zodpovedný žiadny ľudský aktér. Pokiaľ ide o pripísanie zodpovednosti za následky používania artefaktov ľuďom, existujú viaceré stratégie na riešenie prípadov neistoty alebo nedostatku kontroly, ktoré môžu jednotlivci pri používaní technológií zažívať. Vznikli rôzne právne koncepty, ktoré pripisujú zodpovednosť jednotlivcom alebo právnickým osobám a vyzývajú ich, aby napravili neželané udalosti, ktoré nemohli kontrolovať alebo predvídať. Rovnako existujú rôzne formálne a neformálne normy v rámci komunít, spoločenstiev či organizácií, ako sú rôzne profesijné etické kódexy a žité praxe. Takéto normy odzrkadľujú všeobecne akceptované predstavy o zodpovednosti a naznačujú, že aj keď osoba nemohla priamo kontrolovať alebo porozumieť dôsledkom svojich činov, jej rozhodnutia a konanie do istej miery ovplyvnili výsledok udalostí, a preto by mala prevziať zodpovednosť a minimálne podať isté vysvetlenie alebo prijať nejaké opatrenia.

Smrtiace autonómne zbraňové systémy, ako každá iná zbraň, predstavujú potenciálnu hrozbu a nebezpečenstvo pre iné osoby a vzhľadom na istú nepredvídateľnosť ich konania vytvárajú i istú mieru rizika. Tí, ktorí takéto zbrane vyvíjajú alebo používajú, sú za ne morálne zodpovední. Ich morálna zodpovednosť sa odvíja od rizík, ktoré použitie takýchto zbraní v kontexte vojnových konfliktov predstavuje. Hoci v praxi môže byť niekedy ťažké pripísať zodpovednosť, nie je to v žiadnom prípade dostatočným dôvodom na to, aby sme rezignovali na úlohu pripísania zodpovednosti, najmä ak čelíme vskutku nebezpečným technológiám.

50 Podobný záver tiež: Krishnan, A., *Killer Robots. Legality and Ethicality of Autonomous Weapons*. Farnham, Ashgate Publishing Ltd. 2009, s. 105; Schulzke, M., Autonomous Weapons and Distributed Responsibility, c.d., s. 203–219; Noorman, M. – Johnson, G. D., Negotiating Autonomy and Responsibility in Military Robots, c.d., s. 51–62; Simpson, W. T. – Müller, V., Just War and Robot's Killings, c.d., s. 302–322; Walsh, J. I., Political Accountability and Autonomous Weapons. *Research & Politics*, 2, 2015, No. 4, s. 1–6.

SUMMARY

On the Responsibility Gap for Autonomous Killer Robots

One of the primary ethical concerns raised by the prospect of using military autonomous killer robots is the question of the moral responsibility for their use. Robert Sparrow argues that military robots fitted up with the ability to learn would be so independent and self-contained that they would allow human actors to deny responsibility for their use, creating a situation that Andreas Matthias called a "responsibility gap." A responsibility gap occurs in situations where no one is responsible for the actions of autonomous learning robots. This situation results from the inability of people to fully control and predict the actions of these technologies. In the article, I argue that this conclusion is not a correct one because autonomous technologies do not mitigate people of responsibility for the consequences of using them. Control and predictability are not inevitable prerequisites for attributing responsibility. Given the risks that the use of such weapons presents, those who create or use these weapons are morally responsible for the weapons' actions. Even though the deployment of autonomous lethal weapons might not be a good idea, the "responsibility gap" does not by itself make them immoral.

Keywords: moral responsibility, autonomous killer weapons/robots, control, risk

ZUSAMMENFASSUNG

Über die Lücke in der Verantwortung für tödliche autonome Roboter

Eine der zentralen ethischen Befürchtungen, hervorgerufen durch die reale Aussicht auf die Anwendung tödlicher autonomer Waffensysteme, ist die Frage nach der moralischen Verantwortung für deren Wirken. Robert Sparrow behauptet, dass lernfähige Militärroboter in einem Maße unabhängig und selbstständig seien, dass sie es den beteiligten Human-Akteuren ermöglichten, die Verantwortung für das Tun dieser Roboter abzulehnen. Dadurch entstünde eine Situation, die Andreas Matthias als „Verantwortungslücke" bezeichnet. Eine Verantwortungslücke ist ein Zustand, in dem niemand für das Wirken lernfähiger autonomer Roboter verantwortlich ist. Dieser Zustand ist die Folge der Unfähigkeit des Menschen, die Handlungen dieser Technologien vollständig zu kontrollieren und vorherzusehen. Die Argumentation im Artikel zielt auf eine Ablehnung dieser These, da autonome Technologien den Menschen nicht von der Verantwortung für die Folgen ihrer Anwendung entbinden. Kontrolle und Vorhersehbarkeit sind keine unabdingbaren Voraussetzungen für die Zuschreibung von Verantwortung. In Anbetracht der Risiken, die sich aus der Anwendung derartiger Waffensysteme ergeben, sind diejenigen, die diese Systeme entwickeln oder einsetzen, moralisch für das Wirken dieser Waffen verantwortlich. Obwohl der Einsatz tödlicher autonomer Waffensysteme nicht unbedingt eine gute Idee sein muss, macht die „Verantwortungslücke" sie nicht unbedingt unmoralisch.

Schlüsselwörter: moralische Verantwortung, tödliche autonome Waffen/Roboter, Kontrolle, Risiko