FILOZOFIA
Volume 79, Number 6, 2024

DOI: https://doi.org/10.31577/filozofia.2024.79.6.4

O definícii autonómnych zbraňových systémov

IVAN KONIAR, Katolícka univerzita v Ružomberku, Filozofická fakulta, Katedra filozofie, Ružomberok, Slovenská republika

KONIAR, I.: On the Definition of Autonomous Weapon Systems
FILOZOFIA, 79, 2024, No 6, pp. 621 – 636

This article focuses on the ongoing debate on the definition of autonomous weapon systems. This debate is introduced using two approaches to the definition of autonomy: (1) autonomy as the human – machine relation, and (2) autonomy as the complexity of a machine's decision-making capabilities. The aim of the text is to critically review the current state of the debate and analyze the key issues related to each approach to defining autonomous weapons. I argue that in the current discourse, neither of these approaches offer definitions that could become the basis for regulating autonomous weapon systems.

Keywords: autonomous weapon system – human-machine relation – sophistication of the machine – autonomous – automated – automatic – artificial intelligence

1. Úvod

V súčasnosti prebieha široká diskusia ohľadom zadefinovania autonómnych zbraňových systémov. Počiatky tejto diskusie možno datovať do prvého desaťročia 21. storočia, a to ako medzi zástancami autonómnych zbraní (Arkin 2009), tak aj medzi tými, ktorí pred nimi varujú (Sparrow 2007). Diskusia sa zintenzívnila po roku 2012, keď ministerstvo obrany Spojených štátov (U.S. DoD) zverejnilo *Smernicu 3000.09 o autonómii v zbraňových systémoch*, ktorá ponúkla definíciu takýchto zbraní. Spolu so správou *Loosing Humanity: The Case against Killer Robots* organizácie Human Rights Watch (HRW) tieto dokumenty rozprúdili rozsiahlu diskusiu o definícii autonómnych zbraní a etických a právnych problémoch spojených s týmito zbraňami. Odvtedy sa do diskusie zapojilo množstvo vojenských a vládnych expertov, vedcov a odborníkov z mnohých oblastí. Téma autonómnych zbraňových systémov sa stala dôležitou agendou Medzinárodného výboru Červeného kríža (ICRC), Inštitútu OSN pre výskum odzbrojenia (UNIDIR) a v rámci Dohovoru o zákaze

alebo obmedzení použitia určitých konvenčných zbraní bola zriadená skupina vládnych expertov (CCW/GGE) na identifikovanie a zadefinovanie týchto technológií. Napriek intenzívnej diskusii sa zatiaľ nepodarilo nájsť spoločný prístup a dohodu k riešeniu problému definovania autonómnych zbraní (Taddeo – Blanchard 2022, 5).

Cieľom textu je kriticky preskúmať súčasný stav diskusie o definícii autonómnych zbraňových systémov a analyzovať kľúčové problémy, ktorým čelia snahy o ich definovanie. Táto diskusia bude predstavená prostredníctvom dvoch prístupov k definícii autonómnych zbraní: (1) autonómia ako vzťah medzi človekom a strojom a (2) autonómia ako zložitosť rozhodovacích schopností stroja. Obidva prístupy sa zameriavajú na dve odlišné spektrá autonómie, ktoré sú základom väčšiny súčasných snáh o definovanie autonómnych zbraňových systémov. Domnievam sa, že ani jeden z prístupov v súčasnom diskurze neponúka definície, ktoré by sa mohli stať základom pre reguláciu autonómnych zbraňových systémov.

V ďalšom texte najprv načrtnem, ako sa v súčasnosti uvažuje o autonómii zbraňových systémov. Následne podrobne predstavím dva vybrané prístupy k definícii autonómnych zbraňových systémov a v poslednej časti poukážem na hlavné nejasnosti a problémy obidvoch prístupov. V závere zhrniem argumenty v prospech zastávanej tézy.

2. Prístupy k definovaniu autonómnych zbraňových systémov

V počítačovej vede, umelej inteligencii a robotike sa termín „autonómia" vo všeobecnosti používa na označenie schopnosti programu, umelého aktéra alebo robota (stroja) pracovať nezávisle od ľudského vedenia (Totschnig 2020, 2473). Vytvorenie aktérov, ktorí sú v tomto zmysle autonómni, je v zásade hlavným cieľom týchto odborov. Donedávna sa tento cieľ dal dosiahnuť len obmedzením a kontrolou podmienok, v ktorých aktéri pracujú. Dnes sme však svedkami vytvárania umelých aktérov, ktorí sú navrhnutí tak, aby fungovali v reálnom svete a nekontrolovaných prostrediach. Zrejme najvýraznejšími príkladmi sú samo-riadiace autá a autonómne zbraňové systémy.

Autonómiu stroja možno chápať ako schopnosť robota po aktivácii pracovať bez akejkoľvek vonkajšej kontroly v niektorých alebo vo všetkých oblastiach svojej činnosti počas dlhšieho časového obdobia (ICRC 2014, 62). George A. Bekey (2005, 1) termínom „autonómny" označuje systém schopný fungovať v prostredí reálneho sveta bez akejkoľvek formy externej kontroly počas dlhšieho časového obdobia. Wendell Wallach a Colin Allen (2013, 126) sa domnievajú, že autonómna činnosť robota je jednoducho „aktivita bez dozoru".

V súčasnosti prevláda názor, že autonómia zbraní sa týka kritických funkcií zbraňových systémov, t. j. výberu cieľa a použitia sily. Autonómny zbraňový systém je zastrešujúci pojem, ktorý zahŕňa zbraňové systémy s autonómiou v kritických funkciách. To znamená, že po aktivácii ľudským operátorom procesy výberu cieľa a útoku preberá samotný zbraňový systém pomocou svojich senzorov, programu a zbrane (ICRC 2016, 71). Na základnej úrovni je to autonómia kritických funkcií, ktorá odlišuje autonómne zbraňové systémy od všetkých ostatných zbraní, vrátane tých, v ktorých sú tieto funkcie diaľkovo ovládané ľudským operátorom. Treba ešte poznamenať, že *smrtiace* autonómne zbraňové systémy predstavujú špecifickú podmnožinu autonómnych zbraní, ktoré sú určené na použitie proti ľudským cieľom.

Existuje množstvo definícií autonómnych zbraňových systémov, ktoré navrhli štáty, medzinárodné organizácie, skupiny zastupujúce občiansku spoločnosť a akademici z rôznych odborov. Medzi zainteresovanými stranami však stále chýba dohoda na definícii autonómnych zbraní. Diskusia o tejto otázke však nie je roztrieštená na nespočetné konkurenčné koncepcie. Môžeme identifikovať dva dominantné prístupy k definícii autonómnych zbraňových systémov.[1] Prvý prístup pozostáva z definícií, ktoré sú formulované na základe vzťahu medzi človekom a strojom, respektíve na základe miery interakcie medzi človekom a strojom. Druhý prístup zahŕňa definície, ktoré sú založené na sofistikovanosti rozhodovacích schopností stroja.

2.1 Autonómia ako vzťah medzi človekom a strojom

Prvý prístup vychádza z úlohy ľudského operátora v súvislosti s konečným rozhodnutím o výbere cieľa a útoku na cieľ. Daný prístup reprezentuje napríklad definícia ministerstva obrany Spojených štátov, ktorá opisuje autonómny zbraňový systém ako „zbraň, ktorá po aktivácii dokáže vybrať a zasiahnuť ciele bez ďalšieho zásahu ľudskej obsluhy“ (U.S. DoD 2023, 21).[2] Tento prístup zahŕňa aj definíciu navrhovanú organizáciou Human Rights Watch (HRW 2012, 1), ktorá vedie kampaň za zákaz „robotov zabijakov“, ako aj definíciu Medzinárodného výboru Červeného kríža (ICRC 2021, 5).

Autonómiu tu možno chápať ako schopnosť systému vykonať úlohu bez zásahu človeka. Autonómia však zároveň opisuje nejakú vzťahovú vlastnosť. V tomto zmysle daný prístup opisuje autonómiu systému vo vzťahu k ľudským

[1] Obidvom prístupom sa venujú napríklad Horowitz (2016b); Amoroso – Tamburrini (2021); Seixas-Nunes (2022).

[2] *Smernica 3000.09* bola aktualizovaná 23. januára 2023. Znenie definície autonómnych zbraňových systémov zostalo nezmenené.

používateľom. Autonómia taktiež predstavuje isté spektrum a nie je to vlastnosť, ktorú systém buď má, alebo nemá. V tomto zmysle autonómia systému súvisí s mierou zásahu človeka do činnosti stroja. Miera zásahu ľudského operátora ovplyvňuje mieru autonómie systému. Existuje rozpätie od úplnej kontroly človeka nad určitou funkciou, až po úplnú kontrolu stroja. O autonómii tak možno uvažovať ako o relatívnej nezávislosti alebo miere slobody, ktorú má umelý aktér nad vykonaním svojej úlohy vo vzťahu k ľudskému operátorovi. Na základe určenia úrovne autonómie stroja alebo stanovenia miery ľudského zásahu možno rozlišovať tri typy autonómnych zbraní.

Prvým typom sú polo-autonómne (*semi-autonomous*) zbraňové systémy, označované aj ako „človek v slučke" (*human-in-the-loop*). Slučka referuje na cyklus „pozorovanie – zorientovanie – rozhodovanie – konanie", známy ako *OODA loop*, ktorý sa používa v procese bojových operácii.[3] V systémoch typu *in-the-loop* sa ľudskí operátori aktívne podieľajú na výbere cieľa a rozhodnutí o použití sily. Buď priamo vyberajú cieľ alebo miesto útoku, alebo aktívne potvrdzujú cieľ vybraný programom. Zbraňový systém čaká na súhlas človeka, ktorý musí byť vyjadrený aktívnym konaním, a po každej úlohe stroj zastaví svoju činnosť (Ilachinski 2017, 147).

Druhým typom sú autonómne zbraňové systémy pod dohľadom (*supervised-autonomous*), označované aj ako „človek na slučke" (*human-on-the-loop*). V systémoch *on-the-loop* je úloha ľudského operátora obmedzenejšia. Operátori monitorujú činnosť systému a v prípade potreby môžu zasiahnuť, ale v konečnom dôsledku len reagujú (respektíve nereagujú) na ciele navrhnuté programom (Bode – Huelss 2022, 27). Systém sám aktívne vyberá a napáda ciele a ľudskí operátori môžu toto rozhodnutie zrušiť alebo zastaviť činnosť celého systému. Systém nečaká na aktívne potvrdenie vybraného cieľa. Ak operátor v určitom čase nezmení výber, stroj zaútočí bez aktívneho zásahu človeka. Rozhodnutie systému možno zmeniť len aktívnou činnosťou (stlačením „stop") zo strany operátora (Ilachinski 2017, 148 – 150).

Tretím typom sú plne autonómne (*fully autonomous*) alebo len autonómne zbraňové systémy, označované aj ako „človek mimo slučku" (*human-out-of-loop*). Po aktivácii môže takýto zbraňový systém vybrať a zasiahnuť ciele bez ďalšieho zásahu operátora. Človek nedohliada na činnosť stroja ani nemá možnosť zasiahnuť v prípade poruchy systému (Ilachinski 2017, 151).

[3] Strana, ktorá v boji dokáže uskutočniť proces pozorovania – zorientovania – rozhodovania – konania rýchlejšie, získava nad protivníkom zásadnú výhodu.

Rozdiel medzi autonómiou pod dohľadom a plnou autonómiou je však pomerne otázny. Oba typy zbraní dokážu samostatne – bez zásahu operátora – vybrať a zaútočiť na cieľ. Preto podľa stanoviska amerického ministerstva obrany (U.S. DoD 2023, 21) definícia autonómnych zbraňových systémov zahŕňa aj autonómne zbrane pod dohľadom, respektíve *on-the-loop* autonómne zbrane. Na jednej strane je rozlišovanie medzi dohliadanou a plnou autonómiou v istom zmysle významné. Je zjavne zásadný rozdiel medzi situáciou, keď človek nemusí, ale môže niečo urobiť v súvislosti s prevádzkou nebezpečnej technológie a situáciou, keď človek nemusí a nemôže nič urobiť. Na druhej strane obidva systémy vo svojej kritickej funkcii fungujú samostatne. Okrem toho, ak je ľudská obsluha zbrane s autonómiou pod dohľadom zranená alebo usmrtená, raz aktivovaný zbraňový systém sa v podstate stáva plne autonómnym.

Uvedený prístup k definícii autonómnych zbraní má dve výhody. Poskytuje spoločný jazyk na diskusiu, ktorý je prístupný širokému okruhu vlád a odbornej verejnosti bez ohľadu na úroveň ich technických znalostí. Jasne vymedzuje autonómne zbrane ako zbrane, ktoré sú schopné vyberať a útočiť na ciele bez potreby ľudskej interakcie.

Takéto chápanie autonómie sa však zároveň zameriava len na jeden aspekt komplexného celku. Autonómia chápaná ako vzťahový pojem sa zrejme netýka len vzťahu medzi umelým aktérom a jeho operátorom. Umelý aktér sa musí nachádzať v nejakom prostredí, aby bol rozpoznateľný ako autonómny aktér. Robot, ktorý má svoje senzory nastavené len na viditeľné spektrum svetla, v prostredí bez svetla prestáva byť v určitom aspekte autonómnym aktérom (Franklin – Graesser 1997, 26). Podobne komplexné prostredia, napríklad mestská výstavba alebo pohyb na dvojrozmernom povrchu vo všeobecnosti, predstavujú iné výzvy a obmedzenia voľnosti pohybu ako monotónna krajina alebo let v relatívne prázdnom trojrozmernom priestore. Autonómnosť systému taktiež závisí od časového trvania, v ktorom môže systém vykonávať úlohy. Čím dlhšie môže systém pracovať, tým je autonómnejší z hľadiska fungovania v čase. Zároveň dlhšia časová škála fungovania systému, môže mať vplyv na väčší priestorový dosah. V neposlednom rade je dôležitým faktorom zložitosť úloh a rozhodnutí, ktoré má systém vykonať. Inteligentný termostat a autonómne vozidlo môžu byť z hľadiska prevádzky autonómne. Autonómne rozhodovanie o tom, kedy zapnúť kúrenie, si však zvyčajne vyžaduje oveľa jednoduchšie schopnosti ako autonómne rozhodovanie o spôsobe jazdy na rušnej mestskej ceste (Christen et al. 2017, 40 – 41).

2.2 Autonómia ako zložitosť rozhodovacích schopností stroja

Druhý prístup k definovaniu autonómnych zbraňových systémov je založený na sofistikovanosti rozhodovacích schopností systému vo vzťahu k výberu cieľa a útoku na cieľ. Týka sa toho, ako prebieha rozhodovanie v rámci systému, respektíve rozsahu kognitívnych schopnosti systému vykonávať kontrolu nad vlastným správaním (Vries 2023, 45). Môžeme rozlíšiť širokú škálu komplexnosti strojov a rôzne stupne sofistikovanosti rozhodovacích schopností systému (Scharre, Horowitz 2015, 6). Skôr ako uvediem možné formulácie takto chápanej definície, pozrime sa na rozlíšenie, ktoré sa zdá pre tento prístup zásadné.

Podľa mnohých odborníkov a štátov (napríklad Taddeo – Blanchard 2022, 15; Seixas-Nunes 2022, 90; UK MoD 2018, 13; Crootof 2015, 1367; Vries 2023, 46) je ústrednou otázkou pri definovaní autonómnych zbraňových systémov rozdiel medzi „automatickými", „automatizovanými" a „autonómnymi" systémami. Domnievajú sa, že medzi nimi existuje jasne identifikovateľná hranica, a že v podstate ide o odlišné kategórie systémov. A to napriek tomu, že všetky tri uvedené pojmy sa vzťahujú na procesy, ktoré môžu od začiatku do konca prebiehať bez akéhokoľvek ľudského zásahu.

Označenie „automatický" sa zvyčajne používa pre systémy, ktoré majú veľmi jednoduché mechanické reakcie na podnety z prostredia a ich fungovanie nedokáže zohľadňovať neistoty v prevádzkovom prostredí. Automatickí aktéri jednoducho reagujú na vonkajší podnet a v podstate u nich nedochádza k rozhodovaniu (Willliams 2015, 38). Do tejto kategórie zbraní bude patriť pozemná mína, pretože jej činnosť je kauzálne podmienená konkrétnym spúšťačom, napríklad tým, že na ňu niekto stúpi alebo otrasom pôdy. Autonómne ani automatizované zbrane do tejto kategórie nepatria, pretože ich správanie nie je len reakciou spôsobenou prostredím (Taddeo – Blanchard 2022, 16). Byť autonómnym aktérom, ktorý vníma a koná v určitom prostredí, si vyžaduje istý odstup od podnetov z prostredia. Konanie aktéra nesmie byť úplne determinované aktuálnymi silami prostredia, ako je tomu v prípade biliardovej gule a mechanických síl. Biliardová guľa nevníma prostredie, pretože vnímanie predpokladá systém založený na informáciách (Castelfranchi – Falcone 2003, 106).

Stroje, ktoré na rozdiel od automatických zbraní dokážu kontrolovať svoju činnosť, možno označiť ako automatizované alebo autonómne. Otázkou teda je, čím sa odlišuje automatizovaný systém od autonómneho systému. Na túto otázku existuje niekoľko odpovedí.

Viacerí autori a zainteresované strany sa domnievajú, že automatizovaný systém je systém, ktorý je naprogramovaný tak, aby sa riadil vopred definovaným a pevným súborom pravidiel (najčastejšie v opakujúcom sa vzorci), s cieľom dosiahnuť želaný výsledok. Na základe svojho programu dokáže previesť istý súbor vstupov zo senzorov na deterministický a obmedzený súbor výstupov. Výstupy sú predvídateľné, ak je známy súbor pravidiel, v rámci ktorých pracuje (Williams 2015, 32 a 38). Systém zároveň dokáže pracovať len v štruktúrovanom prostredí a reagovať len na to, čo predvída človek.

Na rozdiel od týchto charakteristík sa za autonómne považujú tie systémy, ktoré sa riadia skôr širokými pravidlami a majú vysoký stupeň samosprávy riadenia a rozhodovania pri dosahovaní cieľov. Všeobecne sa uznáva, že autonómne systémy určitým spôsobom presahujú automatizované systémy, avšak presné kritériá sa značne líšia. Ministerstvo obrany Spojených štátov v dokumente *Integrovaný plán bezpilotných systémov na roky 2017 – 2042* (2017, 17) uvádza, že autonómia je (na rozdiel od automatizácie) definovaná ako „schopnosť subjektu nezávisle rozvíjať a vyberať si z rôznych spôsobov konania na dosiahnutie cieľov na základe vedomostí subjektu a poznania sveta, seba samého a situácie".

Definíciu autonómnych zbraní, ktorá môže byť výsledkom takéhoto chápania autonómie, uvádza ministerstvo obrany Veľkej Británie (UK MoD 2018, 13): „Autonómny systém je schopný pochopiť zámery a smerovanie na vyššej úrovni. Na základe tohto chápania a vnímania svojho prostredia je systém schopný prijať vhodné opatrenia na dosiahnutie požadovaného stavu. Je schopný rozhodnúť o spôsobe konania z viacerých alternatív bez toho, aby bol závislý od ľudského dohľadu a kontroly, hoci tie môžu byť stále prítomné. Hoci celková činnosť autonómneho bezpilotného lietadla bude predvídateľná, jednotlivé akcie nemusia byť predvídateľné."

Takáto definícia kladie na zbraňový systém veľmi náročné kognitívne požiadavky a skôr marginalizuje význam priameho ľudského dohľadu. Byť schopný porozumieť úmyslom na vyššej úrovni a podnikať kroky na ich dosiahnutie je nesmierne vysoká hranica toho, čo možno považovať za autonómne.[4] Žiadne existujúce a vyvíjané zbraňové systémy nechápu význam cie-

[4] Článok predpokladá dve odlišné koncepcie autonómie. Na jednej strane autonómiu ako koncept, ktorý je v západnej filozofickej tradícii neoddeliteľne spojený s človekom ako morálnou a racionálnou bytosťou. Na druhej strane autonómiu ako koncept používaný v počítačovej vede, robotike, umelej inteligencii a ďalších oblastiach pre označenie merateľnej a pozorovateľnej schopnosti programov, umelých aktérov, robotov alebo strojov.

ľov a zámerov na vyššej úrovni. V súčasnosti neexistuje ani kvalifikovaný odhad časového harmonogramu ich vývoja ani technológia na ich vytvorenie (UNIDIR 2017, 29). Autori tejto definície si to však uvedomujú a uvádzajú, že „Spojené kráľovstvo nedisponuje plne autonómnymi zbraňovými systémami a nemá v úmysle ich vyvíjať. Takéto systémy zatiaľ neexistujú a pravdepodobne nebudú existovať ešte mnoho rokov, ak vôbec" (UK MoD 2018, 14). Zároveň dodávajú, že bez ohľadu na to, či výrobcovia zbraní označia systémy ako autonómne, britská armáda ich bude považovať za automatické.

Definície majú svoju úlohu. Vzhľadom na prebiehajúcu kampaň zakázať alebo obmedziť používanie autonómnych zbraní možno neprekvapuje, že niektoré vlády pristupujú k diskusii o týchto zbraniach so zámerom prísne rozlišovať medzi autonómnymi a automatickými systémami. Definície autonómnych zbraňových systémov tak často vykazujú vysoké kognitívne a technologické požiadavky, aby sa vyhli regulácii alebo zákazu. Podobne použitie výrazu „automatický" namiesto „autonómny" má naznačovať vyššiu úroveň ľudskej kontroly (Bode – Huelss 2022, 19).

V nadväznosti na tieto okolnosti niektorí autori zdôrazňujú potrebu pochopiť dôležitosť rozlišovania medzi hypotetickými, budúcimi, autonómnymi zbraňami – systémami, ktoré sú ešte ďaleko od realizácie – a dosiahnuteľnými autonómnymi zbraňami, ktoré v súčasnosti nevyvolávajú ontologickú otázku týkajúcu sa ich statusu. V rámci diskutovaného prístupu k definícii autonómnych zbraní tak existujú aj iné, menej zaťažené definície, ktoré kladú na autonómiu odlišné nároky.

Michael Horowitz (2016a, 27) definuje autonómne zbraňové systémy ako „…zbraňové systémy, ktoré sú po aktivácii navrhnuté tak, aby vybrali a zasiahli ciele, ktoré predtým neurčil človek". Spojenie „ktoré predtým neurčil človek" má zohľadniť prípady, keď ľudia vyberú skupinu cieľov útoku bez toho, aby presne vedeli, ktorá munícia zasiahne presne ten ktorý cieľ. Bojové strety zahŕňajú salvy a ľudia nemajú pod kontrolou muníciu od miesta odpálenia po miesto dopadu. Podľa definície Rebbecy Crootof (2015, 1854), autonómny zbraňový systém je schopný samostatne vyberať a zasahovať ciele na základe záverov odvodených zo zhromaždených informácií a vopred naprogramovaných obmedzení. Autonómny systém musí byť schopný aspoň minimálnej úrovne nezávislej analýzy a konať na základe zhromaždených a analyzovaných informácií.

Definície, ktoré majú poukázať na kategóriu existujúcich alebo technologicky pravdepodobných autonómnych systémov prinášajú aj ďalší odborníci a štáty. Robert Sparrow (2016, 95) sa domnieva, že ide o „zbraň, ktorá je

schopná identifikovať možné ciele a vybrať si, na ktoré zaútočí bez ľudského dohľadu, a ktorá je dostatočne zložitá, takže aj keď funguje dokonale, zostáva určitá neistota, na ktoré objekty a/alebo osoby zaútočí a prečo". Čínska delegácia na stretnutí expertov CCW/GGE uviedla definíciu autonómnych zbraní ako nerozlišujúcich smrtiacich systémov, ktoré nemajú ľudský dohľad a kontrolu počas procesu vykonávania úlohy a ich činnosť nemôže byť po aktivácii ukončená. Zároveň takéto zbraňové systémy budú schopné vyvíjať a učiť sa prostredníctvom interakcie s prostredím, a to tak, že rozšíria svoje funkcie a schopnosti spôsobom, ktorý presahuje ľudské očakávania (PRC 2022).

Mariarosaria Taddeo a Alexander Blanchard (2022, 15) uvádzajú, že chcú poskytnúť hodnotovo neutrálnu definíciu autonómie a autonómny zbraňový systém možno podľa nich definovať ako „umelého aktéra, ktorý je prinajmenšom schopný meniť svoje vlastné vnútorné stavy na dosiahnutie daného cieľa (…) môže byť tiež vybavený určitými schopnosťami meniť svoje vlastné prechodové pravidlá (…) s cieľom vyvinúť kinetickú silu proti fyzickej entite a na tento účel je schopný identifikovať, vybrať alebo zaútočiť na cieľ bez zásahu iného aktéra".

Andrew P. Williams (2015, 56 – 57) ponúka definíciu, autonómnych zbraní podľa ktorej autonómia referuje na „…schopnosť systému, platformy alebo softvéru splniť úlohu bez zásahu človeka prostredníctvom správania vyplývajúceho z interakcie počítačového systému s externým prostredím". Systém môže plniť úlohy pomocou „…rôznych spôsobov správania, ktoré môžu zahŕňať rozmýšľanie, riešenie problémov, adaptáciu na neočakávané situácie, samostatné riadenie a učenie sa." Alfonso Seixas-Nunes uvádza, že najlepší spôsob, ako definovať autonómny zbraňový systém je, že ide o zbraňový systém navrhnutý tak, aby bol adaptívny a identifikoval, vyberal a zasahoval vojenské ciele bez ľudského zásahu (Seixas-Nunes 2022, 82).

Mary L. Cummings (2021, 277) v tomto kontexte píše, že autonómne systémy predstavujú výrazný skok v zložitosti oproti automatizovaným systémom, a to najmä z dôvodu úlohy pravdepodobnostného usudzovania v takýchto systémoch. Zatiaľ čo automatizované systémy sa riadia jasne definovanými, deterministickými pravidlami „ak – tak – inak" (*if – then – else*), autonómne systémy uvažujú pravdepodobnostne. To znamená, že odhadujú najlepšie možné spôsoby konania vzhľadom na vstupné údaje zo senzorov. To robí výstupy automatizovaných systémov predvídateľnými, zatiaľ čo autonómne systémy nebudú produkovať konzistentne rovnaké správanie (Cummings 2017, 3).

Iní odborníci však vidia rozdiel medzi automatizovanými a autonómnymi systémami len v miere či stupni samosprávy systému. Autonómne systémy považujú za zložitejšie formy automatizovaných systémov. Autonómia je len sofistikovanou formou automatizácie a úplná autonómia je na druhom konci (nepretržitej) škály zvyšujúcej sa automatizácie. To, čo odlišuje autonómny systém od automatizovaného systému, je vnímanie zložitosti funkcií, ktoré sú však v princípe automatické. Neexistuje jasná hranica medzi tým, čo sa vníma ako automatická funkcia a autonómny systém (Hagström 2016, 23). Podobne Alex Leveringhaus (2021, 177) tvrdí, že z technologického a koncepčného hľadiska sa autonómne zbrane prekrývajú s automatickými a hranice medzi nimi sú skôr plynulé ako pevné. Autonómne zbrane sú automatizované na vyššej úrovni, a preto sú schopné vykonávať zložitejšie úlohy ako ich automatizované príbuzné. Relevantné rozdiely sú skôr vecou stupňa než druhu.

3. Problémy definícií autonómnych zbraňových systémov

Ponúknuť definíciu je vždy zložitá úloha. O to viac keď hrozí, že zadefinovaná vec môže byť zakázaná. Na účely diskusie o regulácii môže byť užitočné mať definíciu autonómnych zbraní, ktorá vymedzí systémy, ktoré možno považovať za autonómne a stanoví, čím sa líšia od predchádzajúcich alebo podobných technológií.

Autonómia ako vzťah medzi človekom a strojom vymedzuje autonómne systémy tým, že ide o systémy, ktoré sú schopné pracovať bez ľudského zásahu. Čo v prípade autonómnych zbraní znamená vybrať a zaútočiť na cieľ. Hoci táto definícia hovorí, že autonómny zbraňový systéme vyberá a útočí na cieľ bez ďalšej interakcie s operátorom, nedokáže rozlíšiť, respektíve ponecháva otvorenú otázku, či ide o niektorú z existujúcich automatizovaných zbraní alebo zbraň s kognitívnymi schopnosťami blízkymi človeku. Definície prvého prístupu potenciálne zahŕňajú aj viaceré v súčasnosti existujúce zbrane. Zbraňový systém blízkej obrany ako napríklad Phalanx v jednom zo svojich štyroch režimov fungovania môže vyberať a útočiť na ciele bez zásahu operátora a podľa uvedenej definície ide o autonómny zbraňový systém. Kritéria definície budú spĺňať aj niektoré ďalšie zbraňové systémy ako napríklad americký Patriot, britský Brimstone a izraelský Iron Dome, a Harpy.

To je však v rozpore s argumentmi kampane na zastavenie vraždiacich robotov, ako aj mnohých vlád a autorov, ktorí tvrdia, že autonómne zbraňové

systémy dnes ešte neexistujú.[5] Ako však poznamenáva Horowitz (2016, 92), je zrejme značne nepravdepodobné, že by sa armády zbavili účinných systémov, ktoré dlhodobo a bez vážnejších kontroverzií používajú, a zdá sa, že niektoré mimovládne organizácie si tento problém uvedomujú. Taktiež snaha zakázať všetky zbrane, ktoré spadajú do takto koncipovanej definície by mohla skončiť zákazom súčasnej a budúcej generácie presnej munície, ktorá zvyšuje presnosť použitia sily a znižuje pravdepodobnosť zabitia civilistov (Horowitz 2016, 97). Možno tak namietať, že uvedená definícia je príliš široká a samotná absencia ľudskej interakcie neumožňuje správne definovať autonómne zbraňové systémy (Vries 2023, 39). Podľa kritikov uvedený prístup k definícii autonómnych zbraní v podstate nerozlišuje medzi skutočne autonómnymi systémami a systémami, ktoré by sa mali považovať len za automatizované.

Autonómia ako zložitosť rozhodovacích schopností stroja vymedzuje autonómne zbraňové systémy na základe toho, ako prebieha rozhodovanie v rámci systému. Autonómia je tu chápaná ako akýsi „kognitívny motor", ktorý poháňa takéto systémy. Bez autonómie sú roboty len prázdnymi nádobami, ktoré sú plne závisle od ľudskej kontroly (Scharre 2019, 25). Autonómia signalizuje, že kognitívne schopnosti stroja sú na takej úrovni, že umožňujú robiť rozhodnutia a konať na bojisku (Seixas-Nunes 2022, 109). Takéto uvažovanie sa snaží odlíšiť autonómne zbrane od minulých a existujúcich zbraní, ktoré sú síce schopné vykonávať ich kritickú funkciu bez zásahu človeka, avšak nemajú ďalšiu kvalitu, ktorá by ich robila autonómnymi.

Definícia autonómnych zbraní Veľkej Británie sa snaží vymedziť ostrú hranicu voči existujúcim automatickým systémom dôrazom na robustné kognitívne schopnosti stroja. Definícia hovorí, že autonómia sa vzťahuje na schopnosť rozumieť a že autonómny systém môže sám rozhodovať o rôznych spôsoboch konania bez ľudského dohľadu alebo kontroly. V súčasnosti však zrejme nemáme spôsob, ako testovať, či zbraňový systém skutočne „rozumie" a „chápe zámer veliteľa" alebo či dokáže pochopiť, „prečo" mu človek nariadil vykonať určitú úlohu (UNIDIR 2017, 29). Tieto komplexné kognitívne schopnosti by vyžadovali autonómnu zbraň, ktorá má schopnosť rozumieť pojmom a jazyku tak, ako mu rozumejú ľudia. Na dosiahnutie tejto úrovne

[5] Viaceré štáty, napríklad USA, Francúzsko, Veľká Británia, mimovládne organizácie, napríklad HRW (2012) a odborníci, napríklad Sparrow (2007), odmietajú, že autonómne zbrane v súčasnosti existujú. Niektorí autori a mimovládne organizácie sa naopak domnievajú, že autonómne zbrane v súčasnosti existujú, napríklad Crootof (2015), Horowitz (2015), Scharre (2019), ICRC (2021).

schopnosti by systémy museli mať kapacity, ktoré sú v súčasnosti mimo dosahu výskumu umelej inteligencie, strojového učenia a robotiky (Seixas-Nunes 2022, 79). Definícia tak hovorí o systémoch, ktoré neexistujú a je otázne, či vôbec niekedy budú existovať.

Istým riešením sa zdá byť skombinovať obidve spektrá autonómie v jednej zloženej alebo sekvenčnej definícii a znížiť nároky kladené na autonómny systém. Takéto definície ponúka mnoho odborníkov, štátov a medzinárodných organizácií. Domnievam sa, že ani v tomto prípade však nejde o bezproblémový počin. Na jednej strane to síce vymedzuje autonómne zbrane ako zbrane fungujúce bez zásahu človeka, na druhej strane však stále zostáva nejasné, čo robí autonómne zbrane autonómnymi v zmysle sofistikovanosti rozhodovania stroja. Hoci existuje určitý konsenzus v tom, že autonómne zbrane nemusia mať kognitívne schopnosti blízke človeku, spornou zostáva najmä otázka rozdielov medzi automatizovanými a autonómnymi systémami.

Kritici poukazujú na to, že v praxi je hranica medzi automatickými, automatizovanými a autonómnymi systémami veľmi tenká a nejasná (Scharre 2019, 44). Je ťažké odmerať a určiť, do ktorej kategórie stroj patrí. Stroje navyše často pracujú v niekoľkých režimoch, ktoré kombinujú rôzne typy správania. Scharre (2019, 45) poznamenáva, že nový stroj považujeme za „autonómny", pretože ešte nerozumieme, ako funguje. Až vďaka skúsenosti dokážeme porozumieť logike jeho správania. Tak dochádzame k tomu, že stroj je vlastne len „automatizovaný".

V skutočnosti existuje viacero rôznych pohľadov na to, čo vlastne charakterizuje autonómny zbraňový systém v zmysle komplexnosti a sofistikovanosti stroja a ako ho to odlišuje od podobných technológií. Niektorí autori používajú termín „autonómia" na označenie toho, že bola dosiahnutá určitá minimálna prahová úroveň zložitosti systému, napríklad schopnosť systému spracovať informácie a vyvodiť závery pre reakciu (Crootof 2015, 1855). Ďalší zdôrazňujú schopnosť vybrať cieľ, ktorý nebol vopred vybraný operátorom (Horowitz 2016a, 27).

Mnohí autori vyhradzujú termín „autonómny zbraňový systém" pre systémy, ktoré sú schopné určitej formy strojového učenia, adaptability alebo emergentného správania, ktoré nie je priamo predvídateľné (Williams 2015, 56 – 57; Taddeo – Blanchard 2022, 17; Seixas-Nunes 2023, 82; Vries 2023, 47). Autonómne systémy, na rozdiel od automatizovaných, sa preto považujú za nie úplne predvídateľné. Nepredvídateľnosť vyplýva z toho, že autonómny systém je schopný učiť sa, prispôsobovať sa a fungovať v neštruktúrovanom a dynamickom prostredí. V inom zmysle tento pohľad zdôrazňuje rozdiel

medzi používaním deterministických a nedeterministických algoritmov. Nedeterministické algoritmy prinášajú určitú nepredvídateľnosť, pretože operátor nie vždy presne predvída, ako zbraňový systém zareaguje na konkrétne okolnosti. Zásadnú úlohu tu zohráva schopnosť systému nezávisle vybrať postup z viacerých alternatív a konať spôsobom, ktorý považuje za najlepší na splnenie pridelenej úlohy, a to bez zásahu človeka (Vries 2023, 48).

Iní autori skôr naznačujú, že autonómne zbraňové systémy je v konečnom dôsledku potrebné charakterizovať ako stroje, ktoré fungujú nielen nezávisle od ľudského operátora, ale aj bez kontroly zo strany ich dizajnérov a programátorov (Sparrow 2007, 69 – 70; Sparrow 2016, 108; PRC 2022).

4. Záver

Súčasné snahy o definovanie autonómnych zbraní riešia otázku, či by sa ich definícia mala zakladať na vzťahu medzi človekom a strojom alebo na zložitosti rozhodovania a správania stroja.

Definície autonómnych zbraní, založené na miere interakcie človeka so strojom, umožňujú vyčleniť množinu zbraní, v prípade ktorých stroj vyberá a útočí na ciele bez zásahu človeka v reálnom čase. Autonómia zbraňových systémov je v zásade o tom, ako operátor interaguje so systémom v otázke kritickej funkcie zbrane. Daný prístup poskytuje základný rámec pre diskusiu a všeobecne sa používa na charakterizovanie autonómnych zbraní. Definície tohto prístupu však zároveň stanovujú prah autonómie tak široko, že do jedného celku spájajú existujúce automatizované zbrane, zbrane vybavené umelou inteligenciou a strojovým učením, ako aj potenciálne vysoko sofistikované zbrane budúcnosti. Zdá sa, že takýmto definíciám tak uniká podstata toho, čo je nové na vznikajúcich technológiách, a primerane nezohľadňujú zložitosť a dôsledky technológií, o ktorých uvažujú. Zároveň definície tohto prístupu zrejme zahŕňajú toľko súčasných zbraňových systémov, že kategória autonómnych zbraňových systémov by už nebola užitočnou kategóriou na účely regulácie.

Definície autonómnych zbraní založené na zložitosti rozhodovania a správania stroja síce v princípe vylučujú množstvo súčasných zbraní, ktoré by v prípade prvého prístupu spadali medzi autonómne zbrane, ale zároveň prinášajú iné výzvy. Problematické je hlavne to, že hranice medzi stupňami zložitosti, od automatického cez automatizovaný, až po autonómny sú veľmi tenké a každé vytýčenie hraníc medzi týmito systémami sa javí skôr ako pokus nakresliť čiaru v piesku (Crootof 2015, 1855). Je otázne či autonómne systémy predstavujú len systémy s vysokou úrovňou automatizácie alebo ide o odlišnú kategóriu zbraní. Práve nejasnosť hraníc medzi automatizovanými

a autonómnymi zbraňami je ústredným dôvodom, prečo sa vedú spory o tom, či autonómne zbrane v súčasnosti existujú alebo nie.

Pri pohľade na jednotlivé definície druhého prístupu vidíme, že autori označujú za autonómne zbrane pomerne širokú škálu zbraňových systémov. Termín „autonómny" je často vyhradený pre stroje, ktoré sa samy riadia, učia alebo sa správajú tak, že ich správanie nie je priamo predvídateľné. Iné definície označujú ako autonómne tie systémy, ktoré majú kognitívne schopnosti na úrovni človeka. Ďalšie sa sústredia na rôzne aspekty, ako je nezávislá analýza informácií, sloboda výberu cieľa útoku alebo naznačujú, že ide o zbrane bez ľudskej kontroly.

Definície môžu mať nezamýšľané dôsledky a môžu vytvárať svojvoľné a neužitočné hranice, rovnako ako užitočné hranice ignorovať. Niektoré charakteristiky nemusia byť samy osebe dostatočné na to, aby definovali autonómne zbraňové systémy, najmä vzhľadom na rýchly vývoj technológií. Definície by však mali byť primerane odolné voči rýchlemu vývoju a zmenám. Nemali by sa zakladať na porozumení, ktoré je na druhý deň zastaralé a vlastnostiach, ktoré sú prekonané skôr, ako uschne atrament. Diskusia by sa preto mala sústrediť skôr na ďalšie charakterizovanie, než definície. Snaha o definíciu je pochopiteľná a žiadúca, avšak obidva prístupy v súčasnom diskurze neposkytujú definície, ktoré by sa mohli stať základom pre reguláciu autonómnych zbraňových systémov.

Literatúra

AMOROSO, D. – TAMBURRINI, G. (2021): Toward a Normative Model of Meaningful Human Control over Weapons Systems. *Ethics & International Affairs*, 35 (2), 245 – 272. DOI: https://doi.org/10.1017/S0892679421000241

ARKIN, R. C. (2009): *Governing Lethal Behavior in Autonomous Robots*. Boca Raton: CRC Press.

BODE, I. – HUELSS, H. (2022): *Autonomous Weapons Systems and International Norms*. Montreal: McGill-Queen's University Press.

CASTELFRANCHI, C. – FALCONE, R. (2003): From Automaticity to Autonomy: The Frontier of Artificial Agents. In: Hexmoor, H. – Castelfranchi, C. – Falcone, R. (eds.): *Agent Autonomy, Multiagent Systems, Artificial Societies, and Simulated Organizations*. Dordrecht: Kluwer, 103 – 136. DOI: https://doi.org/10.1007/978-1-4419-9198-0_6

CHRISTEN, M. et al. (2017): *An Evaluation Schema for the Ethical Use of Autonomous Robotic Systems in Security Applications*. University of Zurich Digital Society Initiative White Paper Series, No. 1. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3063617

CROOTOF, R. (2015): The Killer Robots Are Here: Legal and Policy Implications. *Cardozo Law Review*, 36 (5), 1837 – 1915.

CUMMINGS, M. L. (2017): *Artificial Intelligence and the Future of Warfare*. Research Paper, January 2017. London: Chatham House, The Royal Institute of International Affairs, 1 – 17.

Available at: https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf

CUMMINGS, M. L. (2021): The Human Role in Autonomous Weapon Design and Deployment. In: Galliot, J. – MacIntosh, D. – Ohlin, J. D. (eds.): *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare*. Oxford: Oxford University Press, 273 – 287.

FRANKLIN, S. – GRAESSER, A. (1997): Is it an Agent or Just a Program? A Taxonomy for Autonomous Agents. In: Muller, J. P. – Wooldridge, M. J. – Jennings, N. R. (eds.): *Intelligent Agents III. Agent Theories, Architectures, and Languages*. Vol. 1193, Berlin: Springer, 21 – 35.

HAGSTRÖM, M.: Characteristics of Autonomous Weapon System. In: ICRC Expert Meeting. *Autonomous Weapon Systems. Implications of Increasing Autonomy in the Critical Functions of Weapons*. Versoix, Switzerland, 15. – 16. March 2016, 23 – 25. Available at: https://icrcndresourcecentre.org/wp-content/uploads/2017/11/4283_002_Autonomus-Weapon-Systems_WEB.pdf

HOROWITZ, M. C. (2016a): The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons. *Daedalus*, 145 (4), 25 – 36. DOI: https://doi.org/10.1162/DAED_a_00409

HOROWITZ, M. C. (2016b) Why Words Matter: The Real World Consequences of Defining Autonomous Weapons Systems. *Temple International and Comparative Law Journal*, 30 (1), 85 – 98.

Human Rights Watch (HRW) (2012): *Losing Humanity: The Case against Killer Robots*. Report.

International Committee of the Red Cross (ICRC) (2014): *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*. Expert Meeting. Geneva, Switzerland. Available at: https://www.icrc.org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014

International Committee of the Red Cross (ICRC) (2016): *Autonomous Weapon Systems. Implications of Increasing Autonomy in the Critical Functions of Weapons*. Expert Meeting. Versoix, Switzerland.

International Committee of the Red Cross (ICRC) (2021): *ICRC Position on Autonomous Weapon Systems*. ICRC Position and Background Paper. Available at: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems

ILACHINSKI, A. (2017): *AI, Robots, and Swarms: Issues, Questions, and Recommended Studies*. Center for Naval Analysis, CNA Corporation. Available at: https://www.cna.org/archive/CNA_Files/pdf/drm-2017-u-014796-final.pdf

LEVERINGHAUS, A. (2021): Autonomous Weapons and the Future of Armed Conflict. In: Galliot, J. – MacIntosh, D. – Ohlin, J. D. (eds.): *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare*. Oxford: Oxford University Press, 175 – 188.

People's Republic of China (PRC) (2022): *Working paper on Lethal Autonomous Weapon Systems*. Submitted by the People's Republic of China. CCW Group of Governmental Experts on Lethal Autonomous Weapon System. Available at: https://documents.unoda.org/wp-content/uploads/2022/07/Working-Paper-of-the-Peoples-Republic-of-China-on-Lethal-Autonomous-Weapons-Systems%EF%BC%88English%EF%BC%89.pdf

SEIXAS-NUNES, A. (2022): *The Legality and Accountability of Autonomous Weapon Systems*. Cambridge: Cambridge University Press.

SCHARRE, P. (2019): *Armáda strojov: Autonómne zbrane a budúcnosť vojny*. Bratislava: Ikar.

SCHARRE, P. – HOROWITZ, M. C. (2015): *An Introduction to Autonomy in Weapon Systems*. Working Paper. Center for New American Security.
SPARROW, R. (2007): Killer Robots. *Journal of Applied Philosophy*, 24 (1), 62 – 77. DOI: https://doi.org/10.1111/j.1468-5930.2007.00346.x
SPARROW, R. (2016): Robots and Respect: Assessing the Case against Autonomous Weapon Systems. *Ethics and International Affairs*, 30 (1), 93 – 116. DOI: https://doi.org/10.1017/S0892679415000647
TADDEO, M. – BLANCHARD, A. (2022): A Comparative Analysis of the Definitions of Autonomous Weapons. *Science and Engineering Ethics*, 28 (5), 1 – 22. DOI: https://doi.org/10.1007/s11948-022-00392-3
TOTSCHNIG, W. (2020): Fully Autonomous AI. *Science and Engineering Ethics*, 26 (5), 2473 – 2485. DOI: https://doi.org/10.1007/s11948-020-00243-z
United Nations Institute for Disarmament Research (UNIDIR). (2017): *The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics, and Definitional Approaches*. UNIDIR Resources, No. 6. Available at: https://unidir.org/files/publication/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf
United Kingdom Ministry of Defence (UK MoD). (2018): *Joint Concept Note 1/18: Human-Machine Teaming*. Shrivenham: The Development, Concepts and Doctrine Centre. Available at: https://assets.publishing.service.gov.uk/media/5b02f398e5274a0d7fa9a7c0/20180517-concepts_uk_human_machine_teaming_jcn_1_18.pdf
United States Department of Defense (U.S. DoD) (2023): *Directive 3000.09 on Autonomy in Weapon Systems*. Available at: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf
United States Department of Defense (U.S. DOD) (2017): *Unmanned Systems Integrated Roadmap: FY2017–2042*. Defense Technical and Information Center. Available at: https://apps.dtic.mil/sti/citations/AD1059546
VRIES, de B. (2023): *Individual Criminal Responsibility for Autonomous Weapons Systems in International Criminal Law*. Leiden – Boston: Brill Nijhoff.
WILLIAMS, A. P. (2015): Defining Autonomy in Systems: Challenges and Solutions. In: Williams, A. P. – Scharre, P. D. (eds.): *Autonomous Systems: Issues for Defence Policymakers*. NATO Communications and Information Agency, 27 – 62.

Ivan Koniar
Katolícka univerzita v Ružomberku
Filozofická fakulta
Katedra filozofie
Hrabovská cesta 1B
034 01 Ružomberok
Slovenská republika
e-mail: ivan.koniar@ku.sk
ORCID ID: https://orcid.org/0009-0008-6003-9392