Etika letálních autonomních vojenských robotů

David Černý

Centrum Karla Čapka pro studium hodnot ve vědě a technice
Ústav informatiky AV ČR, v. v. i.
Pod Vodárenskou věží 2, 18207 Praha, Česká republika
Email: david.cerny@ilaw.cas.cz

Abstrakt

V tomto příspěvku se zaměřuji na stručnou etickou reflexi důvodů pro a proti nasazování autonomních letálních vojenských robotů v reálných podmínkách válečných konfliktů. Tyto důvody jsou primárně konsekvencialistické, tj. věnují se benefitům a rizikům spojeným s touto moderní vojenskou technologií.

1 Úvod

Možná prvním vojenským robotem v naší západní kulturní imaginaci byl měděný obr Tálos, kterého na Diův příkaz vytvořil bůh Héfaistos, tvůrce mnoha mechanických divů, které dnes můžeme považovat za roboty (Mayor, 2018). Od doby řeckých mýtů však uplynulo mnoho času a možnost vytvořit plně autonomního vojenského robota, který dokáže plnit všechny bojové úlohy bez zásahu člověka, je již realitou. Mnozí lidé to z morálního hlediska považují za chybný krok, který bychom měli napravit úplným mezinárodním zákazem jejich vývoje a nasazování v reálných bojových podmínkách. Takové zákazy jsou však nerealistické; těžko můžeme čekat, že tak strategicky významné technologie by se státy skutečně vzdaly. Je proto velmi důležité snažit se o její etickou reflexi a nastavení podmínek jejich využívání, které budou v souladu s moderní doktrínou spravedlivé války.

V tomto příspěvku se nemohu věnovat celé problematice etické reflexe autonomních robotů ve službě armády. Představím proto jen některé, primárně konsekvencialistické, důvody pro jejich nasazování a některé důvody v jejich neprospěch. Komplexní uchopení celé problematiky by si vyžádalo zpracování formou monografie.

2 Letální vojenští roboti

Ani mezi odborníky neexistuje jednoznačná definice robota. Jak říká americký robotik Illah Nourbakhsh (Nourbakhsh, 2013):

> Nikdy se robotika neptejte, co je to robot. Odpověď se mění příliš rychle. V okamžiku, kdy vědci ukončí nejnovější diskusi o tom, co je a co není robot, zrodí se zcela nová interakční technologie a hranice se posune dál.

K určité pracovní definici robotů se nejlépe dopracujeme prostřednictvím pojmu agenta. Budeme ho definovat následujícím způsobem:

- **Agent**. Agentem rozumíme systém, jenž vnímá své okolí prostřednictvím senzorů a interaguje se světem pomocí aktuátorů.

Nejjednoduššího agenta si můžeme představit jako systém, který je vybaven senzory, aktuátory, a specifickým programem (*agent program*). Senzory zprostředkovávají kognitivní kontakt s prostředím formou perceptů, jejichž úplná historie se označuje jako sekvence perceptů. Agentský program je konkrétní implementací agentské funkce, jež zobrazuje percepty na množinu jednání. Agent může disponovat velmi jednoduchou agentskou funkcí ve formě „jestliže… pak“, např. „jestliže je před tebou překážka, zahni vpravo“., její podoba však může být výrazně sofistikovanější.

V zájmu jasnosti je třeba rozlišovat mezi vojenskými roboty a vojenskými systémy bez posádky (*unmmaned systems*).; nadále je budeme označovat jako US. Některé US mohou být chápány jako vojenští roboti, jiné zase ne. Dron, který je po celou dobu letu řízený nějakým operátorem, není robotem; pokud však v některých částech své mise jedná na operátorovi nezávisle, lze ho v nich považovat za robota.

Některé roboty lze chápat jako US (nemají-li posádku), jiné zase ne. A některé US jsou roboty (alespoň v některé části svého pracovního cyklu), jiné zase ne. Obě třídy – robotů a US – mají neprázdný průnik, vojenské roboty však nelze definovat jako systémy umělé inteligence bez posádky. Předběžná definice vojenských robotů může vypadat následovně:

- **Vojenský robot.** Vojenský robot je robot, tj. ve fyzickém světě existující agent, který
  1. vnímá své prostředí;
  2. zpracovává své percepty a vytváří si plány jednání;
  3. využívá své aktuátory k aktualizaci vybraného jednání;

4. je alespoň v určité části svého operačního cyklu autonomní;
5. plní vojenské cíle.

Podle toho, jaké cíle vojenští roboti plní lze rozlišit celou řadu robotů, nás zde ale budou zajímat plně autonomní letální vojenští roboti (AVR), tj. takoví roboti, kteří dokáží splnit svůj úkol, včetně výběru a likvidace cíle, bez zásahu člověka.

Definovat autonomii robotů není snadný úkol. K bližší definici autonomie musíme rozlišit mezi systémem S, úkolem t, systémem S^*, vůči němuž je S v nějaké vztahu závislosti či nezávislosti, a konečně prostředím e, v němž má S splnit svůj úkol t (Tamburini, 2016). Smyslem zavedení systému S^* je rozlišit situace, kdy S závisí na různých typech systémů, např. na jiném systému umělé inteligence, či na lidském operátoru. Autonomie vojenských robotů je tedy čtyřčlenou strukturou $< S, t, S^*, e >$. Například robotický systém SGR-A1 společnosti Samsung může nahradit stráže na hranicích mezi Severní a Jižní Koreou a operovat v částečně či plně autonomním režimu. Pokud by operoval v plně autonomním režimu, dokázal by identifikovat cíle, sám se rozhodnout je zneškodnit a posléze tak učinit. Neexistoval by tedy žádný systém kontroly S^* (lidský či jiný), jenž by na systém S (SGR-1A) během jeho činnosti jakkoli působil. Tato autonomie však závisí na prostředí: pokud tyto stroje budou nasazeny na střežení hranic, kam není nikomu dovolený přístup, není identifikace cílů obtížná; každý objekt určený jako člověk může být zneškodněný. V jiném prostředí by však plná autonomie možná nebyla; systém by jednoduše nedokázal dobře rozlišovat mezi legitimními a nelegitimními cíli. Byla by proto nezbytná existence systému S^* (lidské kontroly), který by v nějaké fázi operace systému SGR-1A intervenoval a rozhodoval o tom, zda smí či nesmí provést určitou činnost. Plně autonomní vojenští roboti nevylučují dohled člověka, ten však nemusí být strategicky vhodný a zajišťující dostatečně rychlou reakci na měnící se podmínky na bojišti.

3 Důvody pro zavádění vojenských robotů

Existuje celá řada dobrých důvodů, proč by vojenští roboti měli postupně nahrazovat lidské kombatanty (Arkin, 2009; Galliot, 2015) První a zřejmě nejevidentnější spočívá v tom, že když namísto lidí budou bojovat roboti, budou války méně krvavé; ušetříme mnoho životů a zdraví. I když je vedení moderních válek stále sofistikovanější, vojáci zůstávají těmi, kdo riskují své životy a zdraví na bitevních polích. Nehrozí jim smrt pouze z přímé konfrontace s živým nepřítelem. Bombardování, využívání dronů, miny, útoky raketami apod. umožňují efektivní zabíjení a mrzačení vojáků. Čím bude vojáků na bitevním poli méně, tím více lidských životů může být ušetřeno. A to je nepochybně důležitým pozitivním faktorem.

Životy a zdraví vojáků však nemusí být ušetřeny pouze tím, že se jejich přímé zapojení do válečné vřavy bude snižovat. Bojoví roboti by měli být schopni reagovat bez emocí a zasahovat mnohem přesněji než lidé. Vzhledem k tomu, že usmrcení nepřátel není nutným cílem k naplnění cílů spravedlivé války, mohli by roboti namísto zabíjení zraňovat. Zřejmě by se muselo jednat o vážnější zranění, aby vojáci skutečně nepokračovali v boji a minimalizovala se šance, že se do války zapojí po svém uzdravení, zranění je stále lepší než smrt.

Vojáci nejsou vystaveni pouze riziku smrti a fyzického zranění. Mnozí si odnášejí psychické problémy, obtížně se vracejí do starého života, nemohou spát, mají těžké sny a trpí postraumatickou stresovou poruchou. Nelze se tomu divit, válka je děsivá zkušenost, a i když vojáci přežijí ve zdraví, jen těžko se vyrovnávají s tím, že museli zabíjet, že jejich přátelé a spolubojovníci umírali, často děsivými a bolestivými způsoby. Nemálo vojáků spáchá po návratu do své vlasti sebevraždu. Veteráni válek často potřebují odbornou pomoc a někteří autoři odhadují, že kdyby jim byla skutečně poskytnuta, finančně by to ohrozilo systémy veřejného zdravotnictví. Nasazování vojenských robotů by mohlo omezit tyto negativní důsledky a tím i ve válce vzniklou, nicméně mnohá léta po válce existující újmu.

Využívání vojenských robotů by však mohlo přispět ještě jinak k minimalizaci ztrát na životech a újmě na zdraví nejen kombatantů, ale i civilistů. Vojáci pocházejí z různých společenských skupin, liší se svou výchovou, vzděláním, ale také motivací, proč slouží v armádě (je-li služba vojenská služba dobrovolná). Mohou mít rozdílné názory na morálku a jejich pohled na nepřátelské kombatanty a civilisty může být zatížený předsudky, rasismem či jinými negativními postoji. Válka je navíc velmi stresující událost, která může výraznou měrou ovlivňovat jejich úsudek a jednání. Citujeme zprávu Surgeon General Office (USA), která zkoumala duševní zdraví a etiku na bojišti u vojáků účastnících se vojenské operace v Iráku. Výzkum zjistil, že (Galliot, 2015)

1. 10 % vojáků a příslušníků americké Námořní pěchoty uvedlo, že se nekorektně chovalo k civilistům a jejich majetku (bití, kopání, ničení obydlí apod.).
2. Pouze 47 % vojáků a 38 % příslušníků americké Námořní pěchoty souhlasilo s tím, že k civilistům je třeba se chovat s úctou a respektem.
3. Více než třetina vojáků a příslušníků americké Námořní pěchoty souhlasila s mučením, pokud by mohlo zachránit životy či umožnilo získat důležité informace.
4. 17 % vojáků a příslušníků americké Námořní

pěchoty se domnívalo, že s civilisty by se mělo jednat jako s povstalci.

5. Třetina příslušníků americké Námořní pěchoty a více než čtvrtina vojáků se domnívá že jim jejich velitelé nedali jasný pokyn, aby se k civilistům chovali korektně.

6. Třebaže všichni vojáci absolvovali kurzy etiky, 28 % vojáků a 31 % příslušníků americké Námořní pěchoty uvedlo, že se ocitli v situaci, kdy nevěděli, jakým způsobem ji eticky řešit.

Tyto výsledky jsou alarmující, byť nejsou úplně překvapující. Smyslem etiky spravedlivé války je minimalizovat újmu na lidských životech a zdraví. To samozřejmě předpokládá, že jsou její principy skutečně aplikovány v praxi. Ze strany armádního velení tedy musíme očekávat, že jejich strategické plány a konkrétní rozkazy vždy berou principy spravedlivé války v potaz, zatímco u vojáků požadujeme, aby se jimi ve svém činnosti řídili. K tomu je však třeba, aby byli dostatečně trénováni, aby zvládali své emoce, aby dokázali poněkud abstraktní principy aplikovat v konkrétních situacích, aby se nenechávali strhávat svými předsudky, negativními postoji, žalem, strachem a dalšími faktory ovlivňujícími jejich rozhodování a konání. Je pochopitelné, nicméně z hlediska spravedlivé války nepřípustné, pokud se jim to nedaří dokonale. Důvodů tohoto nezdaru je hned několik:

1. Ztratili ve válce přátelé a ovládla je touha po pomstě.

2. Dehumanizují si své nepřátele, například tím, že je soustavně označují urážlivými termíny.

3. Jejich trénink nebyl dostatečný a/či nemají dostatek bojových zkušeností.

4. Propadají depresím a rostoucímu pocitu frustrace.

5. Mají potěšení z pocitu nadvlády a ze zabíjení.

Lze předpokládat, že v případě využívání vojenských robotů se podaří tyto negativní faktory ovlivňující jednání vojáků odstranit. Roboti nebudou mít emoce, nebudou se chtít mstít, nebudou plakat nad ztrátou svých „druhů“, nebudou dehumanizovat nepřítele, budou důsledně rozlišovat mezi kombatanty a civilisty, mezi legitimními cíli a těmi nelegitimními, neovládne je deprese či frustrace, nebudou mít potěšení ze zabíjení. V důsledku toho bude válka humánnější v tom smyslu, že ubudou zbytečné útoky a zabíjení, nebude docházet k zabíjení a mučení civilistů, ženy nebudou znásilňovány, nikdo nebude ničit majetek civilistů, zranění vojáci budou v bezpečí před případnou potřebou pomsty či vybitím si frustrace. To vše představuje další důležitý pozitivní faktor.

Mezi další pozitivní faktory se uvádí nižší finanční náročnost vojenských robotů v porovnání s vojenskými stroji s posádkou na palubě či nižší dopady na životní prostředí (stroje bez posádky nemusí mít tolik bezpečnostních prvků a konzumují méně pohonných hmot). Za nejdůležitější faktory hovořící ve prospěch využívání vojenských robotů, včetně AVR, však považuji ty spojené s minimalizací újmy, ať již na životech, zdraví a majetku, a umožňující důslednější aplikaci principů *ius in bello*.

4 Důvody proti zavádění vojenských robotů

Kritici zavádění a využívání vojenských robotů ve válečných konfliktech nejčastěji argumentují tak, že domnělé výhody se velmi snadno mohou změnit v nevýhody; z pozitivních faktorů na negativní. Kromě toho se mohou objevit i další problémy, které hovoří v neprospěch těchto moderních vojenských technologií. Uvedu zde jen některé z nich.

Moderní umělá inteligence, zvláště rozpoznávání obrazu ze senzorů, řeči a obecně nějakých vzorů a vztahů mezi nimi, využívá neuronové sítě a celou řadu algoritmů. Tyto sítě se dokáží naučit činnost, kterou od nich člověk očekává, kvalita jejich výkonu však závisí na množství a variabilitě dat. Jestliže se například neuronové sítě autonomních aut budou učit rozpoznávat lidi na vzorku, který bude obsahovat málo zástupců jiné barvy pleti, může mít později problémy rozpoznat černocha či Hispánce jako člověka, což samozřejmě povede k diskriminačnímu jednání. Podobné problémy se mohou objevit i u vojenských robotů, zvláště jejich plně autonomní varianty. Jedním z klíčových požadavků na vojenské roboty musí být, aby dokázali bezchybně rozlišovat legitimní a nelegitimní cíle, tedy minimálně kombatanty své armády od těch nepřátelské strany, a samozřejmě civilisty od kombatantů. Opomenutí v tréninkové sadě dat může vést k tomu, že se vojenští roboti budou cíle plést; chybně identifikují civilistu jako kombatanta či vlastního vojáka jako nepřátelského. Tyto chyby mohou mít vážné dopady a vojenští roboti, místo aby přispívali k naplňování podmínek i*us in bello*, mohou přispívat k nespravedlnosti a zvyšovat utrpení.

Další problémy s využíváním vojenských robotů spočívají v tom, že hluboké neuronové sítě jsou náchylné k tzv. *adversarial attack*. Některé studie ukazují, že na obrázku postačí změnit pouhý jeden pixel a neuronová síť bude psa klasifikovat jako kočku. Kdyby se nepřátelskému vojsku podařilo proniknout do kamerových systémů robotů a pozměnit percepty (vizuální obrazy přicházející z kamer ke zpracování), mohlo by to mít nedozírné následky. Roboti by mohli zcela chybně vyhodnocovat percepty pocházející z kamer a vést útoky na vlastní vojáky či dokonce civilisty.

Útok na vojenské roboty by mohl být globálnější

a pokusit se je ovládnout zcela (hacknout). Ti by se obrátili proti vlastním vojákům a způsobili by nemalé škody; nacházeli by se mezi „svými“, blízko potenciálním cílům, navíc by mohli útočit zcela nečekaně a zastihnout své oběti zcela nepřipravené. Teroristické organizace či některé státy by dokonce mohly využít hacknuté bojové roboty k útokům na civilisty, vznášet mezi ně zmatek, paniku a podkopat tak bojovou morálku celého národa a ochotu se i nadále angažovat ve spravedlivém válečném konfliktu.

Nepodceňujeme tyto výhrady proti využívání vojenských robotů, zvláště AVR, nezdá se nám však, že převažují nad pozitivními faktory. Důvodů je několik, hlavní však spočívá v tom, že se vlastně nejedná o negativní faktory jako spíše o rizika. Rizika jsou spojená s každou moderní technologií. Například u autonomních vozidel hrozí, že se do jejich systému někdo naboural a pokusí se je využít jako prostředek zabíjení. Pozitiva autonomní dopravy však převažují negativa + rizika. Kromě toho, identifikace rizik je důležitým nástrojem jejich předcházení. Víme-li, že nedostatečné sady trénovacích dat mohou vést k chybné identifikaci legitimních cílů, nemusíme to chápat jako důvod zákazu autonomních vojenských robotů, ale jako podmínku, která stanoví způsob trénování neuronových sítí a klade podmínky na jejich využívání v praxi.

Jednou z důležitých námitek proti vývoji AVR je riziko, že povede k novým závodům ve zbrojení (Asaro, 2019).[1] Současně by mohlo být snadnější válečné konflikty zahájit (Sharkey, 2008). Pokud ve válkách nebudou umírat lidé, nebo jich bude umírat výrazně méně, mohl by poklesnout případný odpor veřejnosti vůči válkám a tlak na vlády, aby se jim snažily vyhnout a využívaly jiné prostředky řešení mezinárodních sporů. Tato námitka je v podstatě konsekvencialistická a snaží se upozornit na riziko, že vyvíjení AVR povede k neblahým důsledkům. Těmi může být buď to, že posílí závody ve zbrojení a poroste napětí mezi státy, nebo rostoucí počet válečných konfliktů. I kdyby potom platilo, že AVR mají potenciál ušetřit lidské životy, tento pozitivní faktor by byl převážený počtem těchto konfliktů. V konečném důsledku by mohlo umírat stejné, nebo dokonce i větší množství lidí.

Připouštíme, že tato námitka je poměrně vážná. Pokud by skutečně vývoj a využívání AVR vedly k masivním závodům ve zbrojení a jisté řekněme lehkovážnosti v zahajování válečných konfliktů, potom by negativní důsledky využívání této moderní technologie skutečně mohly převážit pozitivní faktory. Bylo by však nerealistické předpokládat, že tyto obavy zastaví vývoj a posléze využívání AVR ve válečných konfliktech (Scholz, 2016). Některé země, jako je např. Rusko, vyvíjí AVR zcela otevřeně a všem státům je jasné, že zastavit jejich vývoj by v situaci nedůvěry a soupeření mezi státy nebylo moudré.

[1] Závody ve zbrojení zde rozumím soupeření mezi dvěma či více státy ve vývoji vojenských technologií. Cílem závodu ve zbrojení je dosáhnout kvalitativní (typy a kvalita zbraní) či kvantitativní (množství zbraní) výhody oproti soupeřícím státům.

5 Závěr

Je zjevné, že existují poměrně dobré (pro někoho přesvědčivé) důvody k využívání AVR v moderním válčení. Neznamená to, že s touto technologií nejsou spojena vážná rizika, ale jak jsme již uvedli, rizika nemusíme chápat jako překážky, ale spíše jako vodítka, která dokáží nasměřovat naši pozornost správným směrem a umožnit nám přijmout opatření k jejich minimalizaci.

Poděkování

Tento příspěvek vznikl za podpory programu Strategie AV21 „Filozofie a umělá inteligence“.

Literatúra

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

Asaro, P. (2019). What Is an Artificial Intelligence Arms Race Anyway? *A Journal of Law and Policy for the Information Society*, (15):45–64.

Galliot, J. (2015). *Military Robots. Mapping the Moral Landscape*. Routledge, London.

Lucas, G. (2023). *Law, Ethics and Emerging Technologies. Confronting Military Technologies*. Routledge, New York.

Mayor, A. (2018). *Gods and Robots. Myths, Machines, and Ancient Dreams of Technology*. Princeton University Press, Princeton.

Nourbakhsh, I. R. (2013). *Robot Futures*. The MIT Press, Cambridge, Mass.

Scholz, J. a kol. (2016). Ethical Weapons. A Case for AI in Weapons. V *Moral Responsibility in Twenty-First Century Warfare. Just War Theory and the Ethical Challenges of Autonomous Weapons Systems*, str. 181–213. SUNY Press, Albany.

Sharkey, N. (2008). Cassandra or the False Prophet of Doom: AI Robots and War. *IEEE Intelligent Systems}*, (23):14–17.

Tamburini, G. (2016). On Banning Autonomous Weapons Systems. V *Autonomous Weapons Systems. Law, Ethics, Policy*, str. 122–141. Cambridge University Press, Cambridge.