Etika a dynamika využívania systémov AI v súčasných ozbrojených konfliktoch

Abstrakt:

Súčasné ozbrojené konflikty vytvárajú priestor pre prvé skutočné a plošné nasadenie systémov umelej inteligencie (AI) vo vojenskej oblasti a zároveň predstavujú veľmi dynamické a realistické "laboratórium" nielen pre ich ďalší vývoj a modifikácie, ale i pre inovatívne spôsoby ich hybridného využitia a priameho použitia v boji.

Etické a morálne aspekty nasadenia AI v týchto konfliktoch sú veľkou výzvou nielen pre hodnotenie doterajších etických rámcov a odporúčaní, ale i pre zásadnú revíziu spôsobu a hľadanie mechanizmov, ako docieliť ich implementáciu a dodržiavanie v budúcich konfliktoch. Osobitný aspekt tejto výzvy predstavujú budúce systémy AI a stratégie vojenského nasadenia.

Kľúčové slová:

umelá inteligencia, smrtiace autonómne zbraňové systémy, dohľad, etika AI

Autor:

ThLic. Ing. Peter Šantavý, PhD.
Univerzita Komenského v Bratislave
Rímskokatolícka cyrilometodská bohoslovecká fakulta
E-mail: peter.santavy@uniba.sk

1. Úvod

Idea inteligentných strojov sprevádza ľudské pokolenie už niekoľko storočí. S rozvojom výpočtovej techniky a postupným rozmachom 3. priemyselnej revolúcie táto idea dostáva reálne kontúry.

I keď pri zrode fenoménu *artificial intelligence* (AI)[1] boli definované všeobecné témy, ktoré sú aktuálne i dnes (strojové spracovanie ľudskej reči, neurónové siete, strojové učenie, abstraktné koncepty a myslenie, kreativita,...), veľmi skoro vybadali potenciál AI aj predstavitelia spravodajských služieb a armády. V budúce schopnosti, resp. možnosti využitia systémov umelej inteligencie verili až do tej miery, že boli schopní financovať jej rozvoj aj v tzv. zimných obdobiach AI, v ktorých bol celý tento koncept v útlme.[2]

Vzhľadom na predikované možnosti umelej inteligencie už od čias Dartmouthského seminára stál rozvoj vojenskej techniky a armádne nasadenie ako jeden z prvých čakateľov v rade na reálne využitie. A akonáhle ostatná jar AI prerástla do prakticky kontinuálneho rozvoja inteligentných strojov, umelá inteligencia sa stala jednou z primárnych technológií nielen moderných zbraňových systémov, ale vojenského využitia v celej svojej šírke a komplexnosti.[3]

V ostatných dvoch dekádach sa na poli výskumu a vývoja systémov umelej inteligencie udialo veľmi veľa. Nové poznatky, algoritmy, kvantum dostupných dát, výkonné systémy, nástup generatívnych systémov – to všetko prispelo a prispieva k nebývalému rozvoju systémov AI a nepreberným možnostiam ich civilného využitia.

Nie menšiu priazeň v tomto období zažíva AI aj v spravodajských službách a v armáde. Osobitne s rozvojom moderných analytických, dohľadových a zbraňových systémov sa technológie AI stali nepostrádateľnými a kopírujúc rozmach v civilnom sektore, dostali zelenú v celej šírke armádneho a spravodajského spektra. Naviac badať určité stieranie hraníc, resp. prienik množín medzi armádnym a spravodajským využitím prostriedkov AI – niektoré oblasti nasadenia sú rovnaké, líšia sa len stupňom intenzity konfliktu (studená vojna, resp. jej mierový ekvivalent hybridnej vojny alebo skutočný horúci vojenský konflikt).

V súčasnosti armáda a spravodajské služby už dávno s umelou inteligenciou nekoketujú, ani sa nesnažia o neisté spolužitie na „múčnu knižku“[4], ale vytvárajú regulárny „manželský“ vzťah, v ktorom spojenie „pokiaľ nás smrť nerozdelí“ dostáva úplne nový obsah.

---

1 Za skutočný zrod umelej inteligencie sa však považuje Dartmouthský seminár – dvojmesačný výskumný projekt zaoberajúci sa umelou inteligenciou, ktorý sa uskutočnil v lete 1956 v Dartmouth college v USA.
V podstate išlo o pracovný seminár, ktorý organizoval mladý matematik John McCarthy v spolupráci s Marvinom Minskym (bývalým kolegom zo štúdií, ktorý s ním zdieľal fascináciu inteligentnými počítačmi a neskôr sa stal zakladateľom laboratória umelej inteligencie na MIT), Claudem Shannonom (kolegom z Bell Labs/IBM a vynálezcom teórie informácií) a Nathanielom Rochesterom (pionierom v oblasti elektroniky a architektom viacerých počítačov IBM).
ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán?*, 34.

2 V tzv. zimných časoch umelej inteligencie, t.j. v obdobiach veľkého útlmu na poli AI, bol vývoj udržiavaný v zásade len v rámci základného výskumu viacerých univerzitných centier a pokroku v príbuzných oblastiach (napr. robotika a kybernetika, výpočtová technika, dátová a počítačová veda, v rámci priemyslu a vojenského vývoja). Veľkú roľu v tejto oblasti dlhodobo zohrávala a zohráva DARPA (Defense Advanced Research Projects Agency) – agentúra ministerstva obrany USA zodpovedná za výskum a vývoj nových vojenských technológií.
ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán?*, 114.

3 Ibid.

4 V 50-tych rokoch dvadsiateho storočia sa tak v bývalom Československu nazývali páry, ktoré spolu žili, no neboli zosobášené.

2. Uvedenie do problematiky AI

Armáda už dlho pozná a využíva mnohoraké zbraňové systémy, ktoré sa vyznačujú určitou mierou automatizácie, pri ktorej však ide len o automatizované vykonávanie činností bez, resp. s minimálnym zásahom človeka.[5]

Skutočné systémy umelej inteligencie musia spĺňať dve základné vlastnosti – *autonómnosť a adaptívnosť*. Autonómne systémy sú schopné samostatne konať. Ide o schopnosť systému vykonávať úlohy v komplexnom prostredí bez neustáleho vedenia používateľom. Adaptívne systémy vykazujú schopnosť sa prispôsobovať, t.j. schopnosť zlepšovať svoj výkon (a schopnosti) učením sa zo skúseností.[6]

Komplexná ľudská inteligencia zahŕňa celé spektrum prejavov a vlastností, ktoré funkcionalistická umelá inteligencia nemá. I tak však o strojovej inteligencii môžeme hovoriť v kategóriách kontinuity (jeden objekt je inteligentnejší ako druhý) a viacerých dimenzií (inteligencia v rôznych oblastiach). V tejto optike rozlišujeme umelú inteligenciu *úzku a všeobecnú*.[7]

Úzko špecializované systémy umelé inteligencie (ANI – Artificial Narrow Intelligence) sú adaptívne a autonómne len v určitej oblasti, t.j. sú schopné riešiť určité úlohy „inteligentným spôsobom“, pričom v ostatných oblastiach zlyhávajú. Ide teda o vysoko špecializované systémy, ktoré sú optimalizované na zvládnutie konkrétnej úlohy, resp. množiny úloh.

Všeobecné systémy umelej inteligencie (AGI – Artificial General Intelligence) dokážu zvládnuť akúkoľvek intelektuálnu úlohu. V podstate ide o umelú inteligenciu, ktorá je na úrovni človeka.

Analogicky kategorizujeme umelú inteligenciu ako *silnú a slabú* na základe rozlišovania medzi skutočne inteligentnými systémami a systémami, ktoré síce inteligentne konajú, avšak inteligenciu len simulujú (napodobňujú).[8]

Silná umelá inteligencia (strong AI) je skutočne inteligentná (a potencionálne i „uvedomelá“), t.j. skutočne aj rozumie tomu, čo rieši a vykonáva. Skutočne inteligentná AI je súčasne aj všeobecnou, pretože má schopnosť generalizovať, t.j. zovšeobecňovať a prenášať, či adaptovať naučené schopnosti na iné úlohy (čo mimochodom patrí k základom ľudského myslenia).

Slabá umelá inteligencia (weak AI) vykazuje inteligentné správanie na základe modelov a aplikovaných metód i dát, na ktorých sa učí (je natrénovaná). Ide teda o systémy, ktoré sú zamerané na riešenie konkrétnych úloh a sú závislé na ľudskom vstupe a konfigurácii.[9] Slabá AI reprezentuje systémy, ktoré aktuálne máme, t.j. počítačové systémy, ktoré vykazujú inteligentné správanie.

---

5 Pokročilou automatizáciou sa vyznačujú napr. v lodných systémoch staré, radarom navádzané systémy CIWS Phalanx, ktoré sa používajú od 70. rokov 20. storočia, pri tankových napr. ruský systém Arena, izraelský Trophy a nemecký AMAP-ADS.
ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán?*, 122.

6 ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán?*, 38.

7 Cf. Ibid.

8 Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán?*, 39.

9 Vytvorenie výborne fungujúceho systému AI vyžaduje skutočných odborníkov, schopných aplikovať správne metódy i správne nakonfigurovať a parametrizovať daný systém. Pre neznalých to môže vyzerať ako mágia či voodoo:-)
Cf. MITCHELL, *Artificial Intelligence*, 98.

V súčasnosti skratkou ANI vyjadrujeme úzku a slabú umelú inteligenciu (narrow and weak AI) a pod skratkou AGI zvykneme rozumieť všeobecnú a silnú umelú inteligenciu (general and strong AI).[10]

Všetky metódy a systémy umelej inteligencie, ktoré dnes používame, spadajú do kategórie špecializovaných systémov AI (ANI), pričom súčasný rozvoj v tejto oblasti napreduje míľovými krokmi, takže viacerí odborníci hovoria o najlepších súčasných systémoch ako o BAGI (treba dodať, že približne rovnako veľká skupina odborníkov s týmto názorom nesúhlasí).[11]

Vzhľadom na našu nedostatočnosť chápania inteligencie a z nej prameniacu neschopnosť uchopiť vytvorenie skutočnej umelej inteligencie, sa miesto jasnej cesty rozvoja AI vývoj a následný pokrok dosahuje v takej rozmanitosti a spletitosti metód AI, že môžeme hovoriť o anarchii.[12]

I napriek uvedenej anarchii však môžeme aspoň principiálne (filozoficky) rozdeliť túto širokú množinu metód do dvoch za nimi stojacich prístupov – symbolická AI a subsymbolická AI.[13]

Symbolická umelá inteligencia ide cestou vytvárania umelej inteligencie na báze ľudského myslenia, t.j. pojmov, slov, fráz (= symboly) a vzťahov medzi nimi. Symbolické systémy na základe definovaných pravidiel a postupov („ak niečo, tak potom toto“) môžu jednotlivé symboly spracovávať a vykonávať priradené úlohy. Pre symbolickú AI sa význam jednotlivých symbolov odvíja od spôsobu ich kombinácie, vzájomných vzťahov a operácií, ktoré môžu byť nad nimi vykonávané.[14] Keďže logika je nutnou podmienkou nášho chápania všeobecnej inteligencie, symbolické systémy sa snažia logickými operáciami riešiť rôznorodé úlohy vyžadujúce nejaký stupeň inteligencie. Striktná formálnosť logického uvažovania umožňuje algoritmizovať ho a previesť do strojovej formy.

Subsymbolická umelá inteligencia a jej rozvoj boli inšpirované pokrokom v neurovede. Subsymbolický prístup k AI sa snaží uchopiť naše myšlienkové procesy, ktoré by sme mohli nazvať niekedy nevedomými či automatickými, a ktoré sú základom tzv. rýchleho vnímania (fast perception), čo využívame napr. pri rozpoznávaní tvárí alebo identifikácii hovorených slov. Subsymbolické programy umelej inteligencie tak neobsahujú súbor presných softvérových postupov na úrovni logického myslenia, ale sú tvorené len stohom rovníc, pre nezainteresovaného len húštinou ťažko interpretovateľných operácií s číslami. Subsymbolické systému sú navrhnuté tak, aby sa učili vykonávať úlohy na základe dát.[15]

Symbolický prístup – využívajúci prísnu logiku a stavajúci na jasne definovaných charakteristikách prostredia a vzťahov medzi objektami činnosti systémov AI – dokázal byť úspešný len pri riešení

10 Naviac – v kontexte aktuálnych pokrokov v oblasti generatívnych systémov AI – sa stále viac hovorí o rôznych stupňoch AGI s viac menej otvorenou snahou radiť tie najlepšie jazykové, resp. multimodálne modely na pomedzie medzi ANI a AGI. Delenie na ANI a AGI nestačí, a preto sa zvykne AGI deliť do viacerých stupňov, napríklad: BAGI (below-human), HAGI/HLAGI (human-level AGI), MAGI (moderately-superhuman AGI), SAGI/ASI (superintelligent AGI).
Cf. GABE. *Four Phases of AGI*.

11 Na platforme X je veľmi známe štipľavé, no zároveň odborné odmietanie radenia súčasných generatívnych systémov do kategórie BAGI od známeho profesora Gary Marcusa (https://x.com/garymarcus), kognitívneho vedca a jedného z autorov pokročilých testov inteligenčných schopností systémov AI.

12 Cf. LEHMAN, CLUNE, RISI, *An Anarchy of Methods: Current Trends*, 56-62.

13 Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán?*, 40-44.

14 Cf. MITCHELL, *Artificial Intelligence*, 21-23.

15 Cf. MITCHELL, *Artificial Intelligence*, 24.

špecifických typov problémov založených na konkrétnych charakteristikách prostredia, vybraných interakciách systémov AI s týmto prostredím a zadanej konkrétnej množine cieľov. Všetko ostatné, čo by mohlo a malo byť predmetom činnosti akejkoľvek inteligencie, však bolo pre tieto systémy tabu – symbolické systémy zlyhávali a zlyhávajú pri riešení problémov, ktoré sa nedajú exaktne opísať, a v reálnych prostrediach, ktoré nie je možné deterministicky uchopiť a sú plné nejasných informácií. Preto väčšina moderných implementácií systémov umelej inteligencie (medzi ktoré patria algoritmy strojového učenia vo všeobecnosti a osobitne neurónové siete) vychádza zo subsymbolického prístupu, ktorý sa snaží tieto problémy uchopiť a v určitej miere aj úspešne riešiť – či už klasickými metódami, ako sú napríklad regresné a štatistické metódy, alebo emuláciou činnosti mozgu na úrovni neurónov prostredníctvom tzv. neurónových sietí.[16]

30. novembra 2022 spoločnosť OpenAI predstavila ChatGPT – veľký jazykový model, ktorý na základe textových vstupov generoval požadované odpovede. Technologicky vychádzal z 3. generácie modelu GPT (Generative Pre-trained Transformer), ktorý bol predstaviteľom sľubne sa rozvíjajúcej tzv. *generatívnej umelej inteligencie* (genAI). Ako už názov napovedá, algoritmy generatívnych systémov AI sú schopné generovať rôznorodý obsah – text, obrázky, audio a video a pod.[17]

Jadrom veľkých jazykových modelov sú tzv. GPT moduly, ktoré fungujú na základe analýzy vstupnej sekvencie a použitia zložitej matematiky na predpovedanie najpravdepodobnejšieho výstupu. GPT modely, ako aplikácia technológie hlbokého učenia umelej inteligencie, dokážu spracovať výzvy v prirodzenom jazyku a generovať relevantné textové odpovede podobné ľudským. Rozsiahle súbory trénovacích dát umožňujú GPT napodobniť schopnosť porozumieť ľudskému jazyku.[18]

Schopnosti modelov GPT pramenia z dvoch kľúčových aspektov. Ide o generatívne predtrénovanie, ktoré učí model rozpoznávať vzory v neoznačených údajoch a následne tieto vzory aplikovať na nové vstupy, a architektúru transformerov, ktorá umožňuje modelu paralelne spracovať všetky časti vstupnej sekvencie a generovať výstupy.

Generatívne systémy, či už ide o modely jazykové, grafické alebo generujúce video, vedia byť úžasné v spracúvaní obsahu zadania, generovaní odpovedí a výstupov, pričom častokrát dokážu rýchlo ponúknuť lepšie výsledky než ľudia. Keďže vývoj genAI neustáva, stretávame sa s rôznymi pokročilými aplikáciami a modifikáciami týchto systémov. Osobitne môžeme spomenúť tzv. *uvažujúce modely,* ktoré sú schopné pri generovaní odpovedí napodobňovať pokročilé spôsoby uvažovania a *agentov AI,* ktorí sú schopní na základe zadaných príkazov samostatne vykonávať komplexné úlohy pre dosiahnutie zadaného cieľa. Využitie je skutočne širokospektrálne a môže byť úspešné pri chápaní rizík a dôslednom aplikovaní zásad správneho použitia a kontroly.

16 ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán?,* 43-44.

17 ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 79.

18 Na základe štatistickej analýzy obrovského množstva textov sa modely učia syntaktické vzory a vzťahy, ktoré sú zachytené v neurónových váhach (nie v explicitných gramatických pravidlách).
Emergentným dôsledkom sú tzv. sémantické reprezentácie, ktoré nie sú formálnou sémantickou analýzou tradičnej lingvistiky, ale naučením sa významov a vzťahov medzi slovami z kontextu.
Veľké jazykové modely, resp. genAI vo všeobecnosti, nevykonávajú syntaktickú a sémantickú analýzu – ony ju „vnímajú“, resp. realizujú prostredníctvom naučených vzorov. Preto „porozumenie“ jazyka nie je výsledkom formálneho spracovania gramatiky a významu, ale je ovocím pravdepodobnostného učenia na veľkom množstve trénovacích dát.
ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 80.

3. Rizikové faktory systémov AI a ich dôsledky

Využitie systémov AI sa stáva tak samozrejmým, že ich bežné používanie si ani neuvedomujeme, avšak ich akceptujeme a pomaly – zvykajúc si na benefity ich nasadenia – ich až vyžadujeme, keďže v mnohom nám prinášajú pridanú hodnotu, ktorej sa nechcem vzdať. Vo všeobecnosti platí, že zameranie sa na komfort a benefity informačných technológií je mnohokrát silnejšie, než opatrnosť a bezpečné správanie sa v rámci ochrany pred kybernetickými hrozbami, únikom a zneužitím osobných údajov a pod. Podobná ľahkovážnosť sprevádza spoločnosť aj pri využívaní systémov umelej inteligencie. Súčasná miera akceptácie jednoduchých systémov AI a spôsob ich využívania dáva tušiť, že používanie sofistikovaných systémov AI môže obnášať riziká, ktorých dosah si ani nevieme predstaviť.

V tejto kapitole sa aspoň okrajovo venujeme problémom, o ktorých vieme a treba s nimi rátať pri návrhu, tvorbe a využívaní systémov umelej inteligencie. Ide ponajprv o *technologické zlyhania, limity a riziká systémov AI*. Ich logickým dôsledkom sú *riziká spojené s využívaním systémov AI v reálnom nasadení*, medzi ktoré patrí nielen zneužitie systémov AI človekom, ale i komplexné dopady a vplyvy využívania systémov AI na človeka a spoločnosť.[19]

Stále sa pritom pohybujeme v oblasti ANI, teda úzko špecializovaných systémov umelej inteligencie (narrow AI), ktoré sú optimalizované na zvládnutie konkrétnej úlohy, resp. množiny úloh. Ide súčasne o systémy slabej umelej inteligencie (weak AI), ktoré vykazujú inteligentné správanie na základe modelov a aplikovaných metód i tréningových dát. I pri najnovších sofistikovaných technológiách AI, medzi ktoré nesporne patria pokročilé uvažujúce modely a agenti genAI, hovoríme stále o systémoch, ktoré sú zamerané na riešenie konkrétnych (množín) úloh a sú závislé na ľudskom vstupe a konfigurácii.[20]

Jedno zo základných rizík vyplýva zo skutočnosti, že *prakticky vôbec alebo len veľmi málo chápeme, na základe čoho robia hlboké neurónové siete svoje rozhodnutia*. Vieme, ako nadizajnovať neurónovú sieť pre konkrétnu oblasť použitia. Vieme, ako ju natrénovať a v rámci možností aj otestovať. Keďže však neurónová sieť neobsahuje súbor presných softvérových postupov na úrovni logického myslenia, ale je tvorená len stohom rovníc, len húštinou ťažko interpretovateľných operácií s číslami, ktoré fungujú na základe správneho nastavenia váh, konštánt a prahových hodnôt, v zásade nevieme, čo presne sa neurónová sieť naučila a ako spoľahlivo to dokáže aplikovať nielen v bežnej prevádzke, ale osobitne v hraničných situáciách za extrémnych podmienok na vstupe či pri činnosti systému. Táto miera nevedomosti rastie s mierou komplexnosti neurónovej siete, resp. celého sofistikovaného systému AI. Problematike *čiernej skrinky* je pomerne komplexná, zahŕňa aj také fenomény ako je „alchýmia“ hyper-parametrov (základný dizajn a vyladenie systému), „pohyby“ v modeloch (generovanie rozdielnych výstupov na základe nuáns v zadaní), skrývanie dôvodov uvažovania (neschopnosť prinútiť uvažujúce modely uvádzať podstatné kroky svojho uvažovania), atď...[21]

19 ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 139.

20 Keďže v súčasnosti neexistujú systémy AGI, t.j. systémy silnej a všeobecnej umelej inteligencie, zameranie na ANI je samozrejmé. Naviac – vzhľadom na aktuálnu absolútnu preferenciu neurónových sietí a strojového učenia – sa v prevažnej miere sústredíme na ich súčasnú prezentáciu hlbokými neurónovými sieťami (deep neural networks) a hlbokým učením (deep learning).
Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 138-139.

21 Sofistikovaný systém AI sa javí ako black box – čierna skrinka, ktorá niečo vykonáva, ale ako a prečo tak robí, nie je úplne jasné.

Nech už hovoríme o čiernej skrinke, alchýmii, mnohorozmernom pohybe alebo skrývaní, súčasné systémy AI sprevádzajú oprávnené a vážne obavy z toho, že *ak nechápeme ako tieto systémy pracujú, nemôžeme im reálne dôverovať a ťažko dokážeme predpovedať okolnosti, za ktorých tieto systémy zlyhajú*.

Odpoveďou na problematiku čiernej skrinky sú tzv. *vysvetliteľné a interpretovateľné systémy AI*. Ide o snahu vyvinúť systémy, ktoré sú vysvetliteľné (vieme popísať, ako pracujú) a interpretovateľné (vieme uviesť, čo výstupy znamenajú). Napriek tomu, že ide o veľmi dynamicky sa rozvíjajúcu oblasť vývoja technológií AI, treba uznať, že stále platí – čím sofistikovanejší systém AI, tým menšia, resp. problematickejšia je vysvetliteľnosť a interpretovateľnosť.[22]

V rámci niekoľkých dekád vývoja neurónových sietí a systémov strojového učenia boli postupne identifikované mnohé *rizikové faktory a zraniteľnosti systémov AI*. Interdisciplinárna skupina FutureTech na MIT spravuje priebežne aktualizovanú komplexnú databázu rizík AI, kategorizovaných podľa príčiny a oblasti rizika. Vychádza zo 43 rôznych AI rámcov ktoré pochádzajú od výskumných, vládnych a priemyselných organizácií. Databáza AI Risk Repository je skutočné „živá“, v roku 2025 mesačne pribúdalo cca. 100 nových rizík a v súčasnosti (marec 2026) ich katalogizuje viac než 1700![23]

Uveďme aspoň tie podstatné (názorné a vyskytujúce sa naprieč širokým spektrom systémov).[24]

Malá množina trénovacích dát (training dataset). Úspešnosť väčšiny súčasných systémov umelej inteligencie je extrémne závislá na rozsiahlych a kvalitných súboroch správne označených trénovacích dát, resp. tréningových iterácií. Bez nich sa súčasné systémy strojového učenia nedajú vytrénovať a ich nedostatok vedie v lepšom prípade k nekvalitným, v tom horšom k nesprávnym výsledkom a fatálnym zlyhaniam. Naviac v kontexte ďalších typoch zraniteľností a rizikových faktorov systémov AI je problematika trénovacích dát oveľa širšia...

Nesprávne zvolená či nekvalitná množina trénovacích dát a predsudky (biases). Na základe nesprávne zvolenej alebo nekvalitnej množiny trénovacích dát sa systém AI naučí robiť chybné uzávery alebo podávať výsledky „s predsudkami“. V mnohých prípadoch hrozí, že systémy AI trénované na zaujatých dátach môžu tieto predsudky násobiť a spôsobiť reálne škody. Ďalším aspektom „zaujatých“ systémov AI je znížená úspešnosť ich činnosti na základe predsudkov.

Nadmerné prispôsobovanie sa trénovacím údajom (overfitting to training data). Ide o nežiadúce správanie sa systému strojového učenia, ku ktorému dochádza, keď model strojového učenia poskytuje presné odpovede pre údaje, na ktorých bol natrénovaný, ale nie pre akékoľvek iné vstupy. Nadmerne prispôsobený model môže poskytovať nepresné predpovede alebo sa naučí z trénovacích dát rozlišovať niečo iné, než to, čo sa mal naučiť.

Efekt dlhého chvosta (long-tail effect). Týmto termínom sa v oblasti umelej inteligencie rozumie veľký rozsah možných neočakávaných situácií, s ktorými by sa systém AI mohol stretnúť.

Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 141-144.

22 Vysvetliteľnosť a interpretovateľnosť smeruje k vyššej transparentnosti a systémov AI. Zaužívaný je termín eXplainable AI (XAI), ktorý zahŕňa celé spektrum metód podieľajúcich sa na zvýšení dôveryhodnosti modelov AI. Cf. ALI, ABUHMED, EL-SAPPAGH, et al., *Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence*.

23 *AI Risk Repository.*

24 Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 147-159.

V reálnom svete jednoducho nedokážeme všetko popísať a predložiť systémom strojového učenia na vytrénovanie.

Klamanie hlbokých sietí a ich zraniteľnosti (fooling deep neural networks and vulnerability to hacking). Neurónové siete sú náchylné na zlyhanie pri záškodníckych dátach (adversarial examples). Dôsledkom je celé spektrum veľmi jednoduchých spôsobov ako oklamať hlboké neurónové siete. Žiaľ, mnohé z možných útokov sú prekvapivo robustné, pričom dokážu účinne oklamať rôzne a diametrálne odlišné sofistikované systémy strojového učenia.

Povery (superstition). Poverou zvykneme nazývať mylnú vieru, že určitá akcia či úkon môžu pomôcť zapríčiniť dobrý alebo zlý výsledok. V oblasti umelej inteligencie ide o problém prevažne v rámci algoritmov učenia formou odmeňovania (reinforcement learning), pri ktorých sa trénovanie modelu (agenta) realizuje prostredníctvom interakcie s prostredím metódou pokus-omyl. V rámci trénovania systému AI vzniká povera vtedy, ak sa daný systém chybne naučí vykonávať nejaký nepotrebný, ba možno až nebezpečný úkon, resp. činnosť pre dosiahnutie požadovaného cieľa.

I napriek neodškriepiteľnému úspechu vývoja a nasadenia súčasných systémov umelej inteligencie v širokom spektre akademického i reálneho prostredia *musíme mať neustále na pamäti, že tieto systémy môžu zlyhávať najrozličnejšími a často neočakávanými spôsobmi* v dôsledku nemožnosti pripraviť dostatočne veľkú množinu trénovacích dát, prípadne ich nesprávnej voľby bez dostatočnej kvality alebo s predsudkami, nadmernému prispôsobovaniu sa tréningovým údajom, efektu dlhého chvosta, rizikám plynúcim z oklamania hlbokých sietí, ich zraniteľností a povier, pod čo sa podpisuje i nedostatok odbornej erudovanosti potrebnej pre dizajn a ladenie hyperparametrov pri príprave funkčného a úspešného riešenia. Pri hlbšom pohľade na uvedené problémy nás môžu dobiehať aj ich ďalšie dôsledky – nielen riziká priameho zlyhania, ale aj realita výsledkov, ktoré môže byť ťažké správne interpretovať (čo sa vlastne sieť naučila, čo výstup z daných dát na vstupe vlastne znamená) a neschopnosť predvídať, kedy sa jednotlivé zlyhania prejavia (za akých podmienok, pri akej súhre okolností, s dôsledku akej dynamiky vnútorného vývoja, resp. činnosti systému AI).[25]

Generatívne systémy AI prinášajú aj ďalšie potencionálne i reálne problémy, napr. halucinovanie (vymýšľanie si odpovedí, ktoré systém predkladá ako relevantné a správne), predsudky a neobjektívne výstupy (riziko poloprávd a nesprávnych odpovedí), nejasný spôsob narábania s údajmi (dôsledkom môže byť únik dôverných dát, prehrešky voči ochrane osobných údajov i problémy s autorskými právami), problémy so stratou kompetencií v dôsledku nahradenia niektorých pracovných pozícií, dôsledky pre spoločnosť a demokraciu (propaganda a manipulácia), zneužitie ako nástroj kybernetickej kriminality, strata kritického myslenia a zmyslu pre realitu, riziko digitálnej demencie a mozgovej hniloby, digitálne rozdelenie, narušenie psychického vývoja, erózia identity, deepfake, algoritmického modelovania histórie a pod...[26]

I keď sa javí, že systémy genAI „rozmýšľajú“ podobne ako ľudia, musíme si uvedomiť, že ide o štatistické modely a dynamické systémy. *Modus operandi týchto systémov je výrazne odlišný od spôsobu práce ľudského mozgu*.[27]

25 ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 160.
26 Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 190-210.
27 Napríklad generatívne systémy zo svojej podstaty nevedia, či je odpoveď správna a či nie, keďže ponúkajú len štatisticky najpravdepodobnejšie výstupy.

Pri generatívnych systémov AI preto upozorňujeme, že *ich nemôžeme vnímať ako faktografické, spoľahlivé a etické zdroje, a že potrebujú človeka, ktorý ich vie správne používať a ich výsledky kontrolovať*.

Doteraz uvedené rizikové faktory implikujú závažné dôsledky:[28]

Systémy AI *„rozmýšľajú“ úplne inak než ľudia*. Ich zlyhania sú odlišné od ľudských, ťažko predvídateľné, niekedy ľahko vykonateľné a mnohokrát prekvapivo robustné.

Systémy AI v skutočnosti nerozmýšľajú – *inteligenciu len napodobňujú*. Systémy AI nechápu zmysel, skutočnú inteligenciu nemajú, len ju simulujú.

Činnosť systémov AI je spojená s vážnymi etickými dôsledkami. *Nie sú schopné rozlišovať morálne dobré a zlé! Nechápu zmysel a dôsledky!*

4. Všeobecné etické princípy v oblasti AI[29]

Základným princípom pre akýkoľvek systém umelej inteligencie je zameranie na dobro človeka, teda známy a všeobecne prijímaný princíp *human-centered and beneficial artificial intelligence.*

Tento základný princíp by mal byť chápaný v duchu kresťanskej antropológie, stavať na antropológii biologickej a kultúrnej, zachovávať ľudskú dôstojnosť a podporovať integrálny rozvoj ľudskej osoby a spoločnosti, zahŕňať každú ľudskú bytosť a nikoho nediskriminovať, mať na zreteli dobro ľudstva a spoločnosti, chrániac pri tom a rešpektujúc dobro každej ľudskej bytosti a vyznačovať sa starostlivosťou o náš „spoločný a zdieľaný domov“, teda o celý stvorený svet.

Principiálny postoj orientácie na dobro človeka sa tak stáva ekvivalentným problematike *dôveryhodnosti umelej inteligencie*, pričom je treba stanoviť podmienky, bez splnenia ktorých by nasadenie systémov AI do reálneho sveta, v ktorom interagujú s človekom a vplývajú na spoločnosť, nemalo byť umožnené.

Podľa Etického usmernenia pre dôveryhodnú umelú inteligenciu Skupiny expertov na umelú inteligenciu pri EÚ i novších nariadení, resp. usmernení EÚ a UNESCO[30] formulujeme *základné požiadavky na dôveryhodné systémy AI*, ktoré musia byť:

– *funkčné a užitočné* (functional and useful) – navrhnuté a realizované tak, aby vykonávali požadovanú činnosť.

– *legálne* (lawful) – vyhovujúce požadovaným normám, zákonom i reguláciám a spĺňajúce všetky platné zákony a predpisy.

– *etické* (ethical) – rešpektujúce etické zásady a hodnoty.

28 Cf. ŠANTAVÝ, KUBICOVÁ, *Systémy umelej inteligencie skutočnú inteligenciu nemajú, len ju simulujú. Nerozlišujú morálne dobré a zlé*.

29 Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 280-295.

30 *Rome Call for AI Ethics.*
Ethics guidelines for trustworthy AI.
The global landscape of AI ethics guidelines.
Artificial Intelligence Act.
UNESCO Recommendation on the Ethics of AI.

– *odolné, resp. robustné* (resilient and robust) – dosahujúce potrebné štandardy bezpečnosti a spoľahlivosti nielen z technologického hľadiska, ale zohľadňujúce aj sociálne prostredie a dopady na spoločnosť.[31]

Etické požiadavky, navrhnuté autorom na základe odporúčaní EÚ, IEEE a UNESCO, sú tieto:

1. Pri vývoji, výrobe, nasadení, poskytovaní a používaní systémov umelej inteligencie musí byť zaručená ochrana slobody, dôstojnosti a bezpečia každej ľudskej osoby i celej spoločnosti.

2. Technológie umelej inteligencie musia byť plne pod ľudskou kontrolou a ovládateľné človekom.

3. Algoritmy i výsledky činnosti systémov AI musia byť človekom pochopiteľné a revidovateľné.

4. Akékoľvek nasadenie technológií AI musí byť prospešné pre človeka a spoločnosť.[32]

5. Systémy umelej inteligencie nesmú byť nástrojom digitálneho rozdelenia.

6. Technológie umelej inteligencie nesmú škodiť nášmu spoločnému domu a mali by prispievať k spoločenskému a environmentálnemu blahobytu.

Uvedené požiadavky ešte viac akcentujú vyššie uvedený rámec dôveryhodných systémov AI, nakoľko bez splnenia kritérií funkčnosti, legálnosti a odolnosti ich nie je možné v plnej miere uskutočniť.

5. Špecifické etické princípy a odporúčania[33]

V oblasti armádneho využitia, spravodajských služieb a algokracie[34] okrem doteraz predstavených všeobecných princípov treba akcentovať i ďalšie odporúčania a nutné podmienky dôveryhodného využívania systémov AI.

Vzhľadom na špecifikum a dosah nasadenia technológií AI *v oblasti pokročilého riadenia štátu, spravodajských služieb a plošného dohľadu* pre ľudské práva, ochranu demokracie a slobôd si myslíme, že táto oblasť by okrem technologického rámca mala byť principiálne pokrytá už základnými legislatívnymi mechanizmami a verejným dohľadom demokratickej spoločnosti.

Bez dôsledne implementovaných kontrolných mechanizmov na úrovni zákonov a ústavy pravdepodobne nebude možné efektívne a dlhodobo zabezpečiť implementáciu kritérií pre dôveryhodné systémy AI. Naviac ak uvážime, že v rámci súčasnej legislatívy bývajú armádne a spravodajské systémy AI vyňaté a majú výnimku.

Navrhujeme tiež, aby oblasť exportu produktov a technológií umelej inteligencie, ktoré môžu byť zneužité v oblasti pokročilého riadenia štátu, spravodajstva a plošného dohľadu, bola predmetom medzinárodnej regulácie s cieľom zamedziť ich vývoz do rizikových krajín. Sankcie by však nemali

31 Hovoríme o bezpečnosti technologickej (security) a spoločenskej (safety).
32 Musia minimalizovať toxické psychologické a spoločenské dôsledky.
33 Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 296-305.
34 Ide o využívanie systémov AI v oblasti pokročilého riadenia štátu. Zahŕňa celú paletu oblastí a stupňov využitia od algoritmickej podpory činnosti úradov, súdov, rozhodovacích procesov až po vládu podľa algoritmov, algoritmický právny poriadok, algoritmické spravovanie spoločnosti a pod.

byť nástrojom (geo) politických zápasov, ale skutočnej zodpovednej angažovanosti na poli etickéhovyužívania systémov AI.

Oblasť armádneho vývoja, nasadenia a využívania je zložitejšia.

Implementácia systémov AI v tejto oblasti neprichádza do právneho vákua, ale do existujúcich právnych rámcov, v rámci ktorých *treba zachovávať základné princípy*: princíp právnej regulácie vojny, princíp humanity, princíp rozlišovania (vojenské/civilné ciele), princíp proporcionality (primerane dosahovanému cieľu) a nevyhnutnosti.

Tiež treba rozlišovať právne rámce mimo (napr. ius ad bellum – právo začať vojnu) a počas vojnových konfliktov (ius in bello – právo platné počas vojny).

V čase mieru by mnohí štátni aktéri chceli aplikovať latinské si *vis pacem, para bellum* i na oblasť armádneho nasadenia technológií AI v zbraňových systémoch a ich potenciálom upevňovať svoje postavenie. Pre ďalších – pozerajúc k horizontu možností autonómnych zbraňových systémov – by sa ich zavádzanie do výzbroje mohlo podobať odstrašujúcemu potenciálu jadrových zbraní. Ak však pozeráme za horizont súčasných možností vojenských systémov AI, vidíme technológie, ktorých rizikový potenciál môže dokonca prekračovať nebezpečenstvo prameniace z terajších arzenálov jadrových zbraní.

Oblasť armádneho nasadenia je však pod veľkým tlakom potenciálnej technologickej výhody, resp. nevýhody.

Schopnosti umelou inteligenciou poháňaných autonómnych zbraňových systémov a kybernetických zbraní i napriek všetkým rizikám vedú k zvyšujúcim sa tlakom na financovanie a zavádzanie útočných kybernetických zbraní. Jednotlivé krajiny sa nedokážu vzdať tak lákavej technologickej výhody a sú pevne rozhodnuté technológie umelej inteligencie implementovať v celej šírke možného zmysluplného využitia. Problematika obmedzenia týchto útočných systémov je skomplikovaná aj reálnym stieraním hraníc medzi obranným a útočným nasadením takmer vo všetkých oblastiach vojenského využitia technológií umelej inteligencie.[35]

Keďže akékoľvek obmedzovanie technológií umelej inteligencie vo vojenskej oblasti môže byť chápané ako bezpečnostné riziko a zníženie bojaschopnosti modernej armády, jednostranne prijaté regulácie nemusia byť účinné – nielen pre to, že sa jednou stranou ťažko prijímajú (aj keď pre hodnotovo orientovanú spoločnosť by to malo byť povinnosťou), ale i pre malú šancu na ich extra teritoriálne rozšírenie a akceptovanie.

Okrem toho zbraňové systémy poháňané umelou inteligenciou môžu otvoriť nové otázky, s ktorými súčasné právne rámce nerátali. Vždy by však mala platiť minimálne Martensová klauzula humanitárneho práva.[36]

---

35 Nie je problém s humanitárnym, zdravotníckym, komunikačným a vo väčšine prípadov aj obranným nasadením systémov AI v armáde.

36 Martensová klauzula slúži ako „bezpečnostná poistka“ a stanovuje, že v prípadoch, ktoré nie sú vyslovene upravené platnými medzinárodnými dohovormi, zostávajú civilisti a bojovníci pod ochranou zásad medzinárodného práva, ktoré vyplývajú zo zvyklostí zavedených medzi civilizovanými národmi, z humanitárnych zásad a z požiadaviek verejného svedomia.
Cf. GEFFERT, *Martensova klauzula v medzinárodnom práve vojnových konfliktov.*

Mnohé vojenské systémy využívajúce technológie AI môžu pracovať autonómne. V zásade rozlišujeme tri rozmery autonómie vzhľadom na vzťah medzi človekom a strojom (human-in-the-loop weapons, human-on-the-loop, human out-of-the-loop weapons), komplexnosť stroja (automatický, automatizovaný, autonómny) a typ automatizovaného rozhodnutia (aká úloha má byť plnená, schopnosť ju splniť a ako ju splniť).

Vzhľadom na mieru autonómie v rozhodovaní bojových systémov *sú z pohľadu etiky najväčšou hrozbou plne automatizované smrtiace zbraňové systémy (LAWs)*. V ďalšom texte si však ukážeme, že *môžu byť smrteľne nebezpečné aj iné systémy, pri ktorých z rôznych dôvodov predávame kontrolnú a rozhodujúcu právomoc strojom*.

Vo všeobecnosti sa možno stotožniť s princípmi dokumentu *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense z Defense Innovation Board USA*[37] s podstatnými závermi, ku ktorým patrí zodpovednosť (rozhodovacia právomoc), opatrnosť pri príprave testovacích dát a návrhu systému, dosledovateľnosť, spoľahlivosť a ovládateľnosť.

V prípade akýchkoľvek autonómnych zbraňových systémov kontrolovateľných človekom musia platiť nasledovné zásady:

– nutnou podmienkou prevádzky ľubovoľného systému AI, ktorý môže predstavovať riziko pre akúkoľvek ľudskú osobu, je schopnosť a možnosť človeka prebrať kedykoľvek kontrolu nad týmto systémom, resp. právo a možnosť verifikovať a prehodnotiť výsledky jeho činnosti.

– limity, regulácia a obmedzenia LAWs by mali predstavovať etický rámec stanovený na základe morálnych hodnôt ľudskej spoločnosti, nie na základe relativistickej tzv. „následnej regulácie“.

V prípade plne automatizovaných smrtiacich zbraňových systémov (LAWs) treba zaujať jasné stanovisko: *Technológiami umelej inteligencie poháňané automatické smrtiace zbraňové systémy, systémy automatického zameriavania a vyberania cieľov, automatické systémy schopné bez zásahu človeka rozhodnúť o smrtiacej reakcii akéhokoľvek druhu (od útoku dronu až po rozpútanie jadrového konfliktu) musia byť zakázané*.

V optike prebiehajúcich vojenských konfliktov, pri ktorých sú využívané rôzne systémy AI, je vhodné poukázať na spomínanú relativistickú „následnú reguláciu“. Ide o postoj určitých armádnych a politických kruhov, ktoré odmietajú limity, regulácie a obmedzenia LAWs na základe morálnych hodnôt ľudskej spoločnosti. Pri „následnej regulácii“ by sa mal uplatňovať vyčkávací prístup a regulácia by sa mala odvíjať od toho, ako sa objavujú nové pokroky vo vývoji a možnostiach nasadenia LAWs.

V zásade ide o špekuláciu ohľadom vývoja etiky a hodnotového rámca, prispôsobiac sa tak stavu vývoja zbraňových systémov na báze umelej inteligencie. Právni vedci, ako napríklad Kenneth Anderson a Matthew Waxman, ktorí tento prístup obhajujú, tvrdia, že regulácia bude musieť vznikať priebežne spolu s technológiou a domnievajú sa, že etika a mantinely morálnej obhájiteľnosti sa budú vyvíjať spolu s technologickým rozvojom.[38] Podporovatelia „následnej

37 Cf. *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense.*

38 Cf. ANDERSON, WAXMAN, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*.

regulácie“ sa tak ľahko môžu dostať do oblasti hodnotového a morálneho relativizmu, snažiac sa prehodnotiť argumenty o nenahraditeľnosti ľudského svedomia a morálneho úsudku.

I napriek zníženiu konkurenčnej schopnosti a malej šanci na akceptovanie inými štátmi, by jednostranne prijaté eticko-právne regulácie mali byť pre hodnotovo orientovanú spoločnosť povinnosťou. Pozývame k tomuto kroku z viacerých dôvodov:

– žiadny štát, pokiaľ sa zriekne etických princípov a morálnych zásad, nemá právo obhájiť svoju účasť na vojnovom konflikte a zvíťaziť.

– pri nasadení moderných zbraňových systémov s celoplošnými účinkami a technológií, zasahujúcich v hybridných vojnách a vojnách 4. generácie prakticky celé populácie štátov, sa koncept spravodlivej vojny stáva neprijateľný.

– len na základe konkrétne prijatých záväzkov sa môže celosvetová diskusia mocností, tlak verejnosti, angažovanosť jednotlivých častí spoločnosti v rôznych regiónoch sveta a úsilie zodpovedných strán pretaviť v postupné prijatie celosvetových pravidiel i záväzkov pre oblasť vývoja a nasadenia rizikových vojenských systémov vybavených technológiami umelej inteligencie.

Záverom tejto kapitoly sumarizujme osobitné etické požiadavky vojenského nasadenia technológií AI, ktoré s*ú postavené na základnom rámci dôveryhodných systémov AI*.

Dôležitá je zásada *ius in bello*, ktorá zahŕňa proporcionalitu, rozlišovanie, pripočítateľnosť a zodpovednosť i zákaz zbytočného utrpenia.

Podmienky vojenského nasadenia AI – základná požiadavka mať systémy AI pod kontrolou:

– zodpovednosť (rozhodovacia právomoc) pri nasadení;

– opatrnosť pri príprave trénovacích dát (unintended bias);

– dosledovateľnosť každého kroku činnosti systému;

– spoľahlivosť pri nasadení;

– ovládateľnosť počas každého kroku činnosti.

Obmedzenia vojenského nasadenia AI:

– nutná podmienka prevádzky systému AI, ktorý môže predstavovať riziko pre akúkoľvek ľudskú osobu: schopnosť a možnosť človeka prebrať kedykoľvek kontrolu nad týmto systémom, resp. právo a možnosť verifikovať a prehodnotiť výsledky jeho činnosti.

– principiálny postoj v oblasti LAWs: technológiami AI poháňané automatické smrtiace zbraňové systémy, systémy automatického zameriavania a vyberania cieľov, automatické systémy schopné bez zásahu človeka rozhodnúť o smrtiacej reakcii akéhokoľvek druhu musia byť zakázané.

– etický rámec pre limity, regulácie a obmedzenia LAWs: musí byť postavený na základe morálnych hodnôt ľudskej spoločnosti, nie na základe relativistickej tzv. „následnej regulácie“.

6. Armádne využitie AI vo všeobecnosti[39]

Využívanie technológií AI v armáde nie je zviazané len s krajinami, u ktorých by sme to vzhľadom na vedecko-technologický či vojenský potenciál predpokladali. S dostupnosťou moderných systémov založených na umelej inteligencii a ich komercializáciou sú tieto systémy lákadlom prakticky pre kohokoľvek.

Implementácia a využívanie systémov AI vo vojenskej oblasti napreduje predovšetkým v týchto krajinách:

– veľmoci (USA, Čína, Rusko,...), ktoré majú i dostatočný vedecko-technologický potenciál i rozsiahle armádne celky;

– vysoko technologicky rozvinuté štáty (Izrael, Japonsko, niektoré štáty EÚ,...), ktorých technologické portfólio takmer prirodzene zahŕňa aj armádne využitie;

– vysoko militarizované krajiny (India, Turecko,...), ktoré budujú moderné armádne celky a investujú do najmodernejších technológií v tejto oblasti;

– z pohľadu NATO problematické štáty (Irán, Severná Kórea,...), ktoré pre svoje obranné, ideologické a politické ciele kladú dôraz na rozvoj vojenského potenciálu, no v rámci svojich možností sa zameriavajú na pre nich dostupné a zároveň efektívne technológie, ku ktorým patria aj systémy AI.

Uvedené krajiny sa snažia uchopiť využívanie umelej inteligencie vo vojenskej oblasti pomerne komplexne. V súčasnosti však nielen ony, ale takmer každá krajina, ktorá sa snaží rozvíjať, resp. budovať moderné armádne zložky, využíva systémy AI aspoň v niektorej z oblastí, medzi ktoré patrí:

– vojenské spravodajstvo,

– modelovanie technológií, konfliktov a operácií,

– podpora pre velenie,

– trenažéry, simulátory a výcvik,

– autonómne zbraňové systémy,

– skupinové riadenie bojových prostriedkov a autonómnych systémov,

– vedenie vojny v kybernetickom priestore,

– vylepšovanie živej sily (skin-in a skin-out),

– ochrana a záchrana živej sily,

– obmedzovanie škôd a zabezpečovanie základných životných potrieb.

39 Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, 205-235.

Z uvedeného zoznamu vyplýva, že využívanie technológií umelej inteligencie v armáde sa neviaže len na zbraňové systémy a riadenie bojových operácií. *Nasadenie systémov AI môže mať pozitívny rámec* v ochrane civilistov a vojenskej živej sily, v záchranný operáciách, pri obmedzovaní škôd a zabezpečovaní základných životných potrieb, náhrade nebezpečných výcvikových priestorov trenažérmi a simulátormi, atď.

Viaceré aktuálne konflikty, ich jednotlivé fázy a priebeh môžeme zaradiť k vojnám štvrtej generácie, ktoré sú popisované stieraním hraníc medzi vojnou a politikou[40], medzi armádou a civilným obyvateľstvom, decentralizovaným vedením vojny, guerilovou taktikou a prvkami terorizmu, dezinformačným pôsobením a propagandou, útokom na kultúru a psychologickými metódami na oslabenie protivníka.[41]

Vojny štvrtej generácie sa vyznačujú masívnym nástupom využívania prostriedkov umelej inteligencie, takže mnohé vojenské systémy poháňané AI (osobitne z kategórie vedenia vojny v kybernetickom priestore) môžeme zaradiť k tzv. prostriedkom vojen štvrtej generácie.

7. AI v súčasných ozbrojených konfliktoch

Ostatné roky na medzinárodnej scéne boli a sú veľmi turbulentné. Viaceré vážne spory prerástli do otvorených vojenských konfliktov, ktoré presahujú „bežný“ rámec menších lokálnych konfliktov. Osobitne ide o vojnu na Ukrajine, izraelsko-palestínsky konflikt v Gaze a vojnový útok Izraela a USA na Irán. Aktérmi všetkých troch konfliktov sú armády, ktoré majú v portfóliu moderné autonómne zbraňové systémy a technológie AI. Tieto konflikty, ktoré sú tragédiou ľudskosti a zdrojom veľkého utrpenia, sa súčasne stali aj veľkým laboratóriom vojenských technológií a stratégií. Veľký priestor v tomto laboratóriu dostáva umelá inteligencia.

7.1. Drony a robotické systémy na frontovej línii v konflikte na Ukrajine

Vojna na Ukrajine priniesla stret technológií armád NATO a Ruska. Tento technologický stret nie je statický, ale extrémne dynamický, konkrétne stratégie nasadenia techniky a operačné postupy sa menia na mesačnej, niekedy až na týždennej báze. Menia sa nielen postupy, ale aj samotné technológie.

Jedným z podstatných fenoménov vojenských aktivít je masívne využívanie dronov. Armády Ruska i Ukrajiny majú svoje špecializované jednotky (napr. ruský Rubikon), ktoré sa viac podobajú vysoko špecializovaným technologickým tímom, než jednotkám bojového nasadenia. Okrem tejto špecializácie sa menia operačné postupy a doktríny, takže prieskumné i bojové využívanie dronov sa dostáva i do bežných jednotiek.

Účinné nasadenie dronov je štandardne viazané na činnosť operátorov, ktorí drony diaľkovo ovládajú. Diaľkové ovládanie sa tak stáva Achillovou pätou ich úspešného operačného nasadenia.

40 Tiež treba podotknúť, že aktérmi nemusia byť len štáty, resp. štátne zoskupenia. Môže ísť nielen o zástupné organizácie niektorých štátov, ale aj o akékoľvek iné mimovládne činitele.

41 Zaujímavým konkrétnym príkladom komplexnosti štvrtej generácie je štúdia prestížneho amerického think-tanku RAND Corporation o možnostiach destabilizácie a ekonomického vyčerpania Ruska.
Cf. DOBBINS, COHEN, CHANDLER et al., *Overextending and Unbalancing Russia: Assessing the Impact of Cost-Imposing Options*.

Preto sa v priebehu konfliktu extrémne rýchlo rozvíjajú nielen spôsoby ochrany techniky a ničenia útočiacich dronov, ale ešte viac prostriedky rádio-elektronického boja (REB), ktoré sú schopné účinne rušiť diaľkové ovládanie týchto strojov. Rádiovo ovládané drony sa tak pre niektoré oblasti frontovej línie stávajú nepoužiteľné.

Odpoveďou na komplexný a kvalitný REB je vývoj alternatívnych foriem ovládania. Veľký úspech napríklad dosiahli ruské drony ovládané prostredníctvom špeciálneho optického vlákna. Ďalšou technológiou je vývoj riadiacich jednotiek poháňaných umelou inteligenciou, vďaka ktorým v oblasti tzv. poslednej míle (frontová línia, silné REB) dokážu drony pracovať autonómne.

Väčšina dronov operujúcich na frontovej línii spadá do kategórie tzv. FPV dronov, ktoré sa zameriavajú na živé ciele. Vzhľadom na súčasný stav rozvoja systémov AI a rozmerové i hmotnostné obmedzenia je implementácia riadiacich jednotiek AI v týchto zariadeniach problematická. *AI v týchto dronoch nie je schopná bližšie analyzovať potencionálny cieľ, t.j. určiť, či ide o civilistu alebo legitímny vojenský cieľ*. Ak je to vojak, či útočí alebo sa vzdáva sa, prípadne či je ranený.

Ďalším problémom je analytické využívanie systémov AI v kombinácii s pokročilým satelitným snímkovaním a monitoringom.[42] Blízky a hlboký bojový priestor sa tak stáva takmer transparentným. Preto sa stratégia zameriava na vyhľadávanie nepriateľských síl a zároveň na oklamanie ich pozorovacích systémov. Výsledkom je, že frontová línia medzi oboma silami, približne do vzdialenosti 40 km na oboch stranách, je teraz veľmi smrtiacou zónou, cez ktorú sa ťažko prebojovať k víťazstvu.[43]

I z tohoto dôvodu Ukrajina a Rusko postupne aktualizujú súčasné poloautonómne letecké, pozemné a námorné systémy pomocou umelej inteligencie. Vďaka tomu budú tieto robotické systémy oveľa menej zraniteľné voči rušeniu elektronického boja (lokálna automatizácia algoritmami AI) a budú môcť autonómne rozpoznať nepriateľský cieľ a (samostatne) zaútočiť. Problematika vyššie uvedených rizík a nedostatkov riadiacich jednotiek AI dronov sa vzťahuje i na tieto systémy.

7.2. Lavender: vojenská analytika v smrtiacom nasadení v Gaze[44]

Izraelská armáda (IDF) sa snaží implementovať a využívať systémy AI v zbraňových systémoch už dlhší čas. IDF svoju 11-dňovú kampaň v Gaze v máji 2021 dokonca nazvala „prvou vojnou umelej inteligencie“. Pri súčasnom útoku v Gaze Izrael využíva prostriedky AI v troch kategóriách:

– systémy smrtiacich autonómnych zbraní (LAWS) a poloautonómne zbrane (semi-LAWS), napr. drony, robotické strelecké veže, a pod.;

– systémy rozpoznávania tváre a biometrický dohľad nad obyvateľmi Gazy s cieľom vytvoriť rozsiahlu databázu biometrických dát obyvateľov;

42 Na strane Ukrajiny/NATO sa využívajú napr. systémy Palantir Maven, technológie Palantiru integrované s veľkými jazykovými modelmi Anthropixu a OpenAI, GIS Arta a pod.

43 Cf. LAYTON, *Artificial intelligence at war.*

44 Cf. FATAFTA, LEUFER, *Artificial Genocidal Intelligence: how Israel is automating human rights abuses and war crimes*.
Cf. ABRAHAM, *‘Lavender’: The AI machine directing Israel’s bombing spree in Gaza*.
Cf. LAYTON, *Artificial intelligence at war.*

– automatizované systémy generovania cieľov: Lavender, Where is Daddy? a Gospel.

Izraelské spravodajské služby už pred konfliktom v Gaze využívali analytický systém Lavender, ktorý pomocou umelej inteligencie vyhodnocoval dáta z palestínskeho prostredia a dokázal identifikovať vedúcich predstaviteľoch Hamasu a generovať potencionálne individuálne ľudské ciele. Gospel generuje infraštruktúrne ciele a Where is Daddy? sa zameriava na sledovanie a zameriavanie podozrivých militantov, keď sú doma so svojimi rodinami.

V pôvodnom režime boli výsledky systému Lavender predmetom ľudskej supervízie, t.j. kontroly a validácie potencionálnych cieľov. Systém sa zároveň zameriaval špeciálne na vedúcich predstaviteľov Hamasu.

Od vypuknutia konfliktu v Gaze sa využívanie systému Lavender zmenilo – systém je využívaný na identifikáciu akéhokoľvek člena Hamasu a zároveň sa zmenil proces kontroly generovaných výsledkov: miesto dôkladného preverenia výstupov sa postupne skracoval čas supervízie na niekoľko desiatok až jednotiek sekúnd! *De facto sa ľudské schvaľovanie potvrdzovania cieľov delegovalo na stroje*. Miesto ľudskej kontroly a rozhodovania prichádza k rozhodovaniu systému AI o živote a smrti.[45]

Okrem delegovania rozhodovania na stroje sú so systémom Lavender spojené i ďalšie problémy:

– *dáta, na ktorých bol systém trénovaný, boli chybné*, lebo obsahovali aj informácie o nevojenských zamestnancoch vlády Hamasu v Gaze, čo viedlo k tomu, že systém Lavender omylom označil za ciele osoby s komunikačnými alebo behaviorálnymi vzorcami podobnými tým, ktoré majú známi militanti Hamasu. Medzi nimi boli policajti a pracovníci civilnej obrany, príbuzní militantov a dokonca aj osoby, ktoré mali len rovnaké meno ako členovia Hamasu.

– napriek tomu, že *systém Lavender má cca. 10% chybovosť* pri identifikácii príslušnosti jednotlivca k Hamasu, IDF získala všeobecné schválenie na automatické prijatie jeho zoznamov osôb určených na likvidáciu „akoby išlo o ľudské rozhodnutie“. Vojaci nemali povinnosť dôkladne alebo nezávisle overovať správnosť výstupov systému Lavender ani jeho zdrojov spravodajských údajov; jedinou povinnou kontrolou pred schválením bombardovania bolo uistiť sa, že označený cieľ je muž, čo trvalo približne „20 sekúnd“.

Neexistuje žiadny etický ani humánny spôsob, ako používať systémy ako Lavender alebo Where is Daddy?, pretože sú založené na zásadnej dehumanizácii ľudí. Keďže využívanie týchto systémov sa započalo ešte pred vypuknutím otvoreného konfliktu, celá „mierová“ infraštruktúra sledovania, biometrické databázy a ďalšie nástroje, ktoré umožňujú nasadenie takýchto systémov vo vojnových zónach, treba eliminovať aj v mierových časoch.

7.3. Generatívne systémy AI použité pri útokoch na Irán

Podľa dostupných informácií zohrávajú analytické a generatívne systémy umelej inteligencie dôležitú úlohu vo viacerých fázach vojenských operácií USA a Izraela voči Iránu:

45 Tzv. „kill chain“, t.j. reťazec od zistenia cieľa až po samotný útok, sa výrazne skracuje.

– analytický AI systém Mosaic spoločnosti Palantir sa pravdepodobne podieľal na nesprávnom vyhodnotení rizika vytvorenia iránskych jadrových zbraní v horizonte niekoľkých týždňov.[46]

– model Claude od spoločnosti Anthropic bol použitý v rámci spravodajských operácií a plánovania útoku na Irán. Model mal pomáhať pri spracovaní spravodajských dát, podpore výberu cieľov a simuláciách možných scenárov. Pri tomto nasadení systém nerozhoduje o konkrétnom cieli a dopade bomby. Ide o analytický nástroj, ktorý dokáže v priebehu sekúnd spracovať obrovské množstvo dát, satelitné snímky, zachytenú komunikáciu, logistické pohyby, a z toho vytvoriť prehľad rizík či odporúčaní.[47] Zaujímavým je fakt, že toto použitie nastalo až po federálnom zákaze využívania systémov AI spoločnosti Anthropic v štátnych agentúrach USA. Dôvodom zákazu je oficiálne bezpečnosť, no podľa spoločnosti Anthropic išlo o nedovolené použitie, nakoľko jej modely nemajú slúžiť priamo na vojenské zásahy a operácie. Pentagon však argumentuje opačne: ak je AI kritická pre národnú bezpečnosť, technologická firma by podľa neho nemala určovať limity jej použitia.

– Irán využíva čínske technológie umelej inteligencie na presné zameriavanie amerických základní na Blízkom východe. Ide o komerčné riešenie čínskej firmy MizarVision, ktoré na základe rozpoznávania objektov, dátovej analýzy a dlhodobej digitálnej stopy dokáže označiť lietadlá, radary, sklady paliva či koncentrácie vojsk a pripraviť podklady pre raketové útoky a útoky dronov. Samotný plán konkrétneho útoku je tiež predmetom plánovania generatívneho systému AI. To, čo bolo predtým predmetom zdĺhavej manuálnej analýzy, je dnes záležitosťou niekoľkých minút. Pomocou technológií AI sa Irán snaží asymetricky čeliť technologicky silnejšiemu protivníkovi.[48]

Z uvedeného vyplýva niekoľko dôležitých skutočností.

Potenciál generatívnych systémov môže byť veľkým prínosom pre analytické spracovanie obrovského množstva rôznorodých dát, vytváranie prehľadov rizík a odporúčaní, modelovanie priebehu konfliktov a operácií, plánovanie konkrétnych útokov i pre komplexnú podporu velenia. Hranica medzi civilným a vojenským využitím technológií sa stiera, dáta sa stávajú rovnako dôležité ako zbrane a nastáva boj nielen medzi „klasickými“ zbraňami, ale aj medzi algoritmami. Odvrátenou stranou technológií genAI je realita ich rizík a limitov. *Bez aplikácie komplexných rámcov riadenia rizík je spoľahlivé nasadenie generatívnych systémov diskutabilné a v klasifikovaných prostrediach so smrtiacimi dôsledkami a esenciálnym rozhodovacím faktorom priam nebezpečné.* Využívanie technológií genAI v najnovších konfliktoch akcentuje etický problém sofistikovaných systémov AI, ktoré principiálne spadajú do kategórie neútočných smrtiacich autonómnych zbraňových systémov. Realita ukazuje, že *sú schopné pracovať a byť nasadené aj v režime, ktorý by mohol byť eticky neprijateľný*.

Argument Pentagónu „ak je AI kritická pre národnú bezpečnosť, technologická firma by podľa neho nemala určovať limity jej použitia“ je veľmi blízko filozofii „následnej regulácie“, ktorú sme už spomínali v časti Špecifické etické princípy a odporúčania. Nerešpektujúc etické obmedzenia sa

46 SANTOS, *Tulsi Gabbard now says Iran could produce nuclear weapon 'within weeks'*.
47 Cf. STOJKOVSKI, *US forces used Claude in Iran strikes for intelligence, targeting even after Trump's ban*.
48 Cf. SERVAES, *Iran Uses Chinese AI Satellite Imagery to Target U.S. Military Bases and Equipment in Middle East*.
Cf. SINHA, *Iran using AI-enhanced satellite images from China to hit US bases in Middle East*.

ľahko môžeme dostať do oblasti hodnotového a morálneho relativizmu, snažiac sa prehodnotiť argumenty o nenahraditeľnosti ľudského svedomia a morálneho úsudku.[49]

V roku 2023 som napísal: „na margo eticko-právneho diskurzu, z ktorého veľká časť prebieha v armádnych kruhoch USA, treba povedať, že súčasné nastavenie je priaznivé realizácii jasného etického rámca a pravidiel. Vládny *Defense Innovation Board* v roku 2019 v dokumente *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense* navrhol princípy využívania systémov umelej inteligencie v americkej armáde, pričom je pre tieto princípy podstatné, že rozhodovacia právomoc zostáva na človeku, osobitne ak ide i misie s použitím smrtiacej sily.“[50] Žiaľ, stačila relatívne krátka skúsenosť niekoľkých horúcich konfliktov a realita aplikácie etického rámca vyzerá úplne inak.

7.4. Vízia blízkej budúcnosti

Príklady z konfliktov na Ukrajine, v Gaze i Iráne poukazujú na nesmiernu dynamiku vývoja a nasadenia systémov AI v súčasných vojnových konfliktoch (zmeny prebiehajú prakticky na mesačnej báze). Nemenej dynamická je i procesná stránka veci, ktorá je poplatná aktuálnym strategickým a politickým cieľom. *V tejto dynamike prakticky chýba akékoľvek etické zhodnotenie a regulačný dohľad. Sme tak svedkami nekontrolovanej implementácie technológií AI v súčasných konfliktoch.*

Tento nelichotivý stav nás privádza k zamysleniu, aká môže byť vízia využívania systémov AI v blízkej budúcnosti. Určite k tejto vízii patria aj nasledovné skutočnosti:[51]

– synergia technológií, t.j. stupňujúci sa prechod od samostatných systémov ku komplexnému ekosystému zahŕňajúcemu komplexné systémy AI pre riadenie (analytický „neobmedzený“ Palantir, dohľadový Flock, genAI,...) a skupinové riadenie bojových prostriedkov a autonómnych systémov (napr. „materské lode“);

– intenzívne nasadenie v hybridných operáciách, pri ktorých sa stiera rozdiel medzi útokom a obranou, aktívnymi (napr. hacking) a pasívnymi (napr. dohľad) operáciami, intenzívne sa využívajú moderné formy propagandy a manipuluje sa s postojom verejnosti (napr. predikcia a modelovanie nálad...);

– využívanie najnovších technológií AI (napr. genAI) bude rásť i napriek nedostatočnému etickému a regulačnému pokrytiu;

– aplikácia tzv. agentic AI a emergentný potenciál multiagentových systémov pri dosahovaní cieľov môže znamenať nové a komplexné riziká, osobitne v oblasti delegovanie rozhodovania na stroje, nemožnosti kontrolovať a dosledovať podstatné kroky procesov prebiehajúcich v týchto systémoch a slepého rastu dôvery vo výsledky činnosti týchto agentov;

– úsilie o dosiahnutie AGI (všeobecná a silná umelá inteligencia) bude narastať. Príchod AGI by znamenal extrémny posun v schopnostiach systémov AI. Paradigmatická zmena, ktorú by príchod

49 Cf. SHANKLIN, *Pentagon's pressure campaign*.
50 ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán?*, 138-139.
51 Cf. LAYTON, *Artificial intelligence at war.*
Cf. ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*.

AGI znamenal, nie je spoločnosť schopná v súčasnosti absorbovať a tieto systémy by sa vymykali kontrole až do tej miery, že by sme ich v súčasnosti nevedeli bezpečne používať!

V tomto kontexte sa vedú debaty, ako by bolo možné zvíťaziť vo vojne, ktorú riadi a ovplyvňuje umelá inteligencia. Takáto vojna by sa vyznačovala:

– neustálou snahou využívať sofistikovanejšie a lepšie vytrénované systémy AI;

– schopnosťou správne nasadiť a kombinovať systémy AI s klasickými bojovými prvkami;

– schopnosťou dlhodobo prevádzkovať systémy AI na vysokej úrovni účinnosti;

– schopnosťou ich operatívne zakomponovať do operačných a taktických plánov;

– snahou rýchlo a účinne nasadiť novinky, ktoré ešte protistrana nemá, resp. nevie na ne účinne odpovedať;

– aplikovaním metódy „nájdi a oblbni“ (klamať systémami AI a vďaka AI nedať sa prekabátiť).

V optike uvedeného prichádza aj k redefinícii pojmu *hypervojna*.[52] V očakávaní budúcich konfliktov, ktoré riadi a ovplyvňuje umelá inteligencia, to znamená:

– konflikt poháňaný umelou inteligenciou a riadený strojmi;

– bezkonkurenčná rýchlosť, ktorú umožňuje automatizácia rozhodovania a súbežnosť akcií;

– ľudské rozhodovanie takmer úplne absentuje v slučke pozorovanie - orientácia - rozhodovanie - konanie (OODA).

Rýchlosť činnosti systémov AI, s ktorou sa ľudské schopnosti nedajú porovnať, výrazná tendencia dôverovať výsledkom činnosti umelej inteligencie a extrémna snaha odovzdať rozhodovacie právomoci strojom sú veľkým nebezpečenstvom pre ľudstvo a ozbrojené konflikty, ktoré našu civilizáciu budú ešte dlho sprevádzať.

Napriek tomu, že ide o technológie nedokonalé, chybujúce a inteligenciu len simulujúce, dôsledky na človeka nezvládnuté a systémy vo vojnovom nasadení vysoko problematické, nevieme ich nástup zastaviť. To však neznamená, že by sme sa mali zrieknuť úsilia o nápravu. Zápas o etické princípy a regulačné rámce využívania technológií AI v armádnom prostredí a vojnových konfliktoch nekončí a možno v skutočnosti ešte len začína...

52 Cf. ALLEN, HUSAIN, *On Hyperwar.*

Zoznam použitej literatúry

Alain SERVAES, *Iran Uses Chinese AI Satellite Imagery to Target U.S. Military Bases and Equipment in Middle East*, Army Recognition Group, 2026. https://www.armyrecognition.com/news/army-news/2026/iran-uses-chinese-ai-satellite-imagery-to-target-u-s-bases-in-middle-east (8. apríl 2026)

Bojan STOJKOVSKI, *US forces used Claude in Iran strikes for intelligence, targeting even after Trump's ban, Interesting Engineering*, 2026. https://interestingengineering.com/military/us-forces-used-claude-iran-strikes (11. marec 2026).

James DOBBINS, Raphael S. COHEN, Nathan CHANDLER et al., *Overextending and Unbalancing Russia: Assessing the Impact of Cost-Imposing Options,* Santa Monica, CA: RAND Corporation, 2019.

John R. ALLEN, Amir HUSAIN, *On Hyperwar,* in: Proceedings 143 (2017). U.S. Naval Institute.

Joel LEHMAN, Jeff CLUNE, Sebastian RISI: *An Anarchy of Methods: Current Trends*, in: *How Intelligence Is Abstracted in AI,* IEEE Intelligent Systems 29 (2014). s. 56-62.

Kenneth ANDERSON, Matthew WAXMAN, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*. Stanford University: Hoover Institution Press, 2013.

Melanie MITCHELL. *Artificial Intelligence*. Farrar, Straus and Giroux, 2019. ISBN: 978-0-374-71523-6.

Peter LAYTON, *Artificial intelligence at war,* Australian Strategic Policy Institute, 2024. https://www.aspistrategist.org.au/artificial-intelligence-at-war/ (31. marec 2026).

Peter ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán?*, Bratislava: RKCMBF UK, 2023. ISBN 978-80-88696-91-9.

Peter ŠANTAVÝ, *Umelá inteligencia – dobrý sluha a zlý pán? Druhé, rozšírené vydanie*, Bratislava: RKCMBF UK, 2026. ISBN 978-80-88696-96-4.

Peter ŠANTAVÝ, Júlia KUBICOVÁ, *Systémy umelej inteligencie skutočnú inteligenciu nemajú, len ju simulujú. Nerozlišujú morálne dobré a zlé,* Bratislava: Nové mesto, 2023. ISSN: 2729-9597.

Sajid ALI, Tamer ABUHMED, Shaker EL-SAPPAGH, et al., *Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence*, Information Fusion 99 (2023). ISSN: 1566-2535. https://doi.org/10.1016/j.inffus.2023.101805 (12. októbra 2025).

Slavomír GEFFERT, *Martensova klauzula v medzinárodnom práve vojnových konfliktov,* in: *Vybrané otázky medzinárodného práva súkromného a verejného,* BRATISLAVA: PF UK, 2021. ISBN: 978-80-7160-584-3

Sujita SINHA, *Iran using AI-enhanced satellite images from China to hit US bases in Middle East*, Interesting Engineering, 2026. https://interestingengineering.com/military/iran-china-satellite-images-target-us-bases (8. apríl 2026)

Will SHANKLIN, *Pentagon's pressure campaign*, Engadget, 2026. https://www.engadget.com/ai/anthropic-weakens-its-safety-pledge-in-the-wake-of-the-pentagons-pressure-campaign-183436413.html (31. marec 2026).

Informácie z internetu

Artificial Intelligence Act. https://www.consilium.europa.eu/en/policies/artificial-intelligence-act/ (30. marec 2026).

AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense. https://admin.govexec.com/media/dib_ai_principles_-_supporting_document_-_embargoed_copy_(oct_2019).pdf (30. marec 2026).

AI Risk Repository. https://airisk.mit.edu/ (20. marec 2026).

Anna JOBIN, Marcello IENCA, Effy VAYENA, *The global landscape of AI ethics guidelines.* https://doi.org/10.1038/s42256-019-0088-2 (30. marec 2026).

Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (30. marec 2026).

Marwa FATAFTA, Daniel LEUFER, *Artificial Genocidal Intelligence: how Israel is automating human rights abuses and war crimes*. https://www.accessnow.org/publication/artificial-genocidal-intelligence-israel-gaza/ (31. marec 2026).

Michael GABE, *Four Phases of AGI.* https://www.lesswrong.com/posts/qeJomTN2yp5tQG4rL/four-phases-of-agi (22. december 2025).

Rome Call for AI Ethics. https://www.romecall.org/wp-content/uploads/2022/03/RomeCall_Paper_web.pdf (30. marec 2026).

UNESCO Recommendation on the Ethics of AI. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics (30. marec 2026).

Yuval ABRAHAM, *‘Lavender’: The AI machine directing Israel’s bombing spree in Gaza*. https://www.972mag.com/lavender-ai-israeli-army-gaza/ (31. marec 2026).