Etika a dynamika využívania systémov AI v súčasných ozbrojených konfliktoch

Peter Šantavý

APVV-23-0509

Na čo treba pamätať – pripomenutie...

- (~~automatické/automatizované~~) autonómne a adaptívne systémy
- neurónové siete sú ako čierna skrinka + technologické riziká a limity
- „rozmýšľajú“ úplne inak než ľudia
  Zlyhania sú odlišné od ľudských, ťažko predvídateľné, niekedy ľahko vykonateľné a mnohokrát prekvapivo robustné...
- v skutočnosti nerozmýšľajú – inteligenciu len napodobňujú
  Systémy AI skutočnú inteligenciu nemajú, len ju simulujú.
- s vážnymi etickými dôsledkami
  Nie sú schopné rozlišovať morálne dobré a zlé! Nechápu zmysel a dôsledky!
- „neľudské“ a nepredvídateľné zlyhania, halucinovanie, predsudky a neobjektívne výstupy, nejasný spôsob narábania s údajmi, dohľad bez kontroly, manipulácia, relativizácia hodnôt a pravdy, digitálna demencia, mozgová hniloba, kognitívna kapitulácia, strata zmyslu pre realitu,...

Armádne využitie AI vo všeobecnosti

- 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

Nielen AKO, ale i KTO?

QR kód: ***Umelá inteligencia – dobrý sluha a zlý pán?*** 1. a 2. vydanie, kap. 2.7., resp. 2.8. a 4.4.2.

Vojny štvrtej generácie – smerovanie k AI

- supermoderné zbraňové systémy
- kybernetické vedenie vojny
- stieranie hraníc: vojna – politika, armáda – civilné obyvateľstvo
- decentralizované vedenie vojny, guerilová taktika a prvky terorizmu
- dezinformačné pôsobenie a propaganda, útoky na kultúru a psychologické metódy na oslabenie protivníka
- aktérmi nemusia byť len veľmoci, resp. štátne zoskupenia; môžu to byť i malé štáty, zástupné organizácie niektorých štátov, ba i akékoľvek iné mimovládne činitele
- + plánovanie v reálnom čase, automatizácia rozhodovania a súbežnosť akcií

Vyznačené sú oblasti **s masívnym nástupom využívania prostriedkov AI**.

Všeobecné etické princípy v oblasti AI

UMELÁ INTELIGENCIA ZAMERANÁ NA DOBRO ČLOVEKA
(human-centered and beneficial artificial intelligence)

= dôveryhodné systémy AI, ktoré musia byť:

- funkčné a užitočné
- legálne
- etické
- odolné, resp. robustné
  [spoľahlivé a bezpečné (technologicky – security a nasadením – safety]

Dôveryhodné systémy AI sú rozoberané v 4. kapitole odkazovanej knihy (QR kód)...

Etika vojenského nasadenia AI

- základné princípy sú stále platné
  – princíp právnej regulácie vojny, princíp humanity, princíp rozlišovania (vojenské/civilné ciele), princíp proporcionality (primerane dosahovanému cieľu) a nevyhnutnosti,...

- zásady ius in bello – právo platné počas vojny (...ius ad bellum – právo začať vojnu)
  – proporcionalita, rozlišovanie, pripočítateľnosť a zodpovednosť, zákaz zbytočného utrpenia

- nové etické výzvy
  – limity a riziká AI, morálna zodpovednosť pri činnosti inteligentných strojov, ľudské práva
  – mimo právnych rámcov minimálne platí Martensova klauzula humanitárneho práva

- základný etický rámec pre reguláciu a obmedzenia
  – postavený na základe morálnych hodnôt ľudskej spoločnosti alebo na základe relativistickej tzv. „následnej regulácie“

Pri následnej regulácii sa regulácie *prispôsobujú* pokrokom vo vývoji a využívaní AI. Ide o hodnotový a morálny relativizmus, snažiaci sa prehodnotiť argumenty o nenahraditeľnosti ľudského svedomia a morálneho úsudku.

Požiadavky vojenského nasadenia AI

- 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

Potreba mať systémy pod kontrolou...

Recommendations on the Ethical Use of Artificial Intelligence by the US Department of Defense.

Obmedzenia vojenského nasadenia AI

- nutná podmienka prevádzky systému AI, ktorý môže predstavovať riziko pre akúkoľvek ľudskú osobu:
  schopnosť a možnosť človeka kedykoľvek **prebrať 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“.

Súčasné nasadenie – Ukrajina

- drony na frontovej línii
  – riešenie extrémneho REB pomocou AI
  – masívna aktualizácia súčasných polo-autonómnych leteckých, pozemných a námorných systémov pomocou jednoduchých AI (väčšia odolnosť voči REB, samostatná činnosť)

- transparentný hlboký bojový priestor
  – analytické využívanie v kombinácii s pokročilým satelitným snímkovaním a monitoringom
  – vzniká široká smrtiaca zóna

- etické dôsledky
  – neschopnosť rozlišovať medzi civilistami a legitímnymi cieľmi
  – smrtiaca zóna prakticky mimo pravidiel vedenia vojny

Uvedené sú len vybrané príklady prezentujúce problematické nasadenie AI.

Súčasné nasadenie – Gaza

- aktuálne využívanie systémov AI
  – LAWS a poloautonómne semi-LAWS, napr. drony, robotické strelecké veže,...
  – systémy rozpoznávania tváre a biometrický dohľad nad obyvateľmi Gazy
  – automatizované systémy generovania a vyhľadávania cieľov:
  Lavender, Where is Daddy? a Gospel

- Lavender: vojenská analytika v smrtiacom nasadení v Gaze
  – od strategických informácií o vedení Hamasu k operačnému nasadeniu pre elimináciu bežných členov
  – od ľudského potvrdzovania **k automatizácii schvaľovania ľudských cieľov**
  – od ľudskej supervízie **k prenechaniu rozhodovania na stroje**

- problémy systému Lavender
  – dáta, na ktorých bol systém trénovaný, boli chybné
  – systém Lavender má cca. 10% chybovosť

Uvedené sú len vybrané príklady prezentujúce problematické nasadenie AI.

Súčasné nasadenie – Irán

- nástup generatívnych systémov AI (genAI)
  – využívanie genAI v USA pri vyhodnocovaní strategických rizík Iránu (Palantir)
  – spracovanie spravodajských dát a plánovanie útoku na Irán (Claude)
  – komerčne dostupné čínske technológie využívané na zameriavanie amerických základní (MizarVision + nešpecifické genAI systémy)

- postrehy
  – obrovský potenciál systémov genAI v rýchlom spracovaní spravodajských informácií, plánovaní a riadení vojenských operácií
  – precedens v spôsobe využívania systémov genAI
  [„následná regulácia“ v USA, Iránske využitie rôznych dostupných technológií]

Uvedené sú len vybrané príklady prezentujúce problematické nasadenie AI.

Problematické súčasné nasadenie

- kognitívna kapitulácia
  – prenechanie rozhodovania na stroje
  – dôvera a automatizácia schvaľovania výsledkov činnosti systémov AI

- obrovská dynamika vývoja a využívania systémov AI
  (zmeny prakticky na mesačnej báze)
  – systémy AI sú pre takéto využitie ešte „nezrelé“ (chyby, riziká, limity)
  – nekontrolovaný vývoj a nasadenie v boji
  – bez etického zhodnotenia a regulačného dohľadu

Pokud není AI dost chytrá, aby mohla zbraně sama používat, nemůžete jí v boji moc důvěřovat.
A pokud je tak chytrá, aby dokázala zbraně používat sama, nemůžete jí důvěřovat nikdy. *J. Campbell*

Vízia blízkej budúcnosti I.

- synergia technológií
  – komplexné systémy AI pre riadenie (analytický „neobmedzený“ Palantir, dohľadový Flock,...)
  – skupinové riadenie bojových prostriedkov a autonómnych systémov (napr. „materské lode“)
- intenzívne nasadenie v hybridných operáciách
  – útok aj obrana; aktívne, napr. hacking, i pasívne, napr. dohľad; propaganda a predikcia nálad...
- využívanie najnovších technológií AI (napr. genAI, agentic AI)
  – pridaná hodnota, osobitne v kombinácii s inými systémami (napr. dáta z Palantiru,...)
  – potencionálne veľké riziká vyplývajúce z limitov a problémov genAI
  – aplikácia tzv. agentic AI a emergentný potenciál multiagentových systémov pri dosahovaní cieľov môže znamenať nové a komplexné riziká
- úsilie o dosiahnutie AGI (všeobecná a silná umelá inteligencia)
  – extrémny skok v schopnostiach systémov AI
  – paradigmatická zmena systémov AI, ktoré by sme v súčasnosti nevedeli bezpečne používať!

Vízia blízkej budúcnosti II.

- ako zvíťaziť vo vojne, ktorú riadi a ovplyvňuje umelá inteligencia
  – využívať sofistikovanejšie a lepšie vytrénované systémy AI
  – vedieť ich správne nasadiť a kombinovať s klasickými bojovými prvkami
  – vedieť ich dlhodobo prevádzkovať na vysokej úrovni účinnosti
  – vedieť ich operatívne zakomponovať do operačných a taktických plánov
  – vedieť rýchlo a účinne nasadiť novinky, ktoré ešte protistrana nemá, resp. nevie na ne účinne odpovedať
  – aplikovať metódu „nájdi a oblbni“ (klamať systémami AI a vďaka AI nedať sa prekabátiť)
- redefinovanie pojmu hypervojna
  – 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)

Ako riešiť etický rámec a regulácie?

Podnety na diskusiu...

Zápas o eticko-právne regulácie

- celosvetové obmedzovanie zavádzania rizikových zbraňových systémov AI sa v súčasnosti javí ako nereálne
  – ako sa vysporiadať s protivníkom, ktorý systémy AI bude využívať bez akýchkoľvek pravidiel,...
- eticko-právne regulácie vojenského nasadenia AI by mali byť pre hodnotovo orientovanú spoločnosť povinnosťou
  – 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
  – základ pre celosvetovú diskusiu mocností, tlak verejnosti, angažovanosť jednotlivých častí spoločnosti v rôznych regiónoch sveta a úsilie zodpovedných strán pretaviť v postupné prijatie celosvetových pravidiel

Ako riešiť etický rámec a regulácie? Podnety na diskusiu...

Podklady pre prezentáciu

Etika a dynamika využívania systémov AI v súčasných ozbrojených konfliktoch
[pdf, SK]

Ethics and Dynamics of AI System Use in Contemporary Armed Conflicts
[pdf, EN]

Preprint článku do zborníka z konferencie Legitimizovanie a deeskalácia vojen a ozbrojených konfliktov.

Ďakujem za pozornosť

APVV-23-0509

ThLic. Ing. Peter Šantavý, PhD., UK v Bratislave
peter.santavy@uniba.sk

Kredit: rawpixel.com on Freepik