Sławomir Nikiel[1]
Uniwersytet Zielonogórski,

e-mail: s.nikiel@wez.uz.zgora.pl
ORCID: 0000-0003-3648-6359
https://doi.org/10.34768/8aj8-3h79

Nr: 4 (2022)
ISSN (on line) 2658-154X

"Speed is the essence of war" (Szybkość jest esencją wojny)
Sun Tzu "The Art of War"

Sztuczna Inteligencja na wojnie – Autonomiczne Systemy Broni

AI at war- Autonomous Weapon Systems

Abstrakt

W ostatnich latach/miesiącach zauważono gwałtowny rozwój Sztucznej Inteligencji. Zastosowanie w działaniach militarnych wymaga podejmowania decyzji przez SI w zakresie identyfikacji i eliminacji celów w kontekście typowych możliwości dostarczanych przez uczenie maszynowe. W proponowanej prezentacji/artykule dokonano identyfikacji obecnie dostępnych rozwiązań w zakresie rozwiązań militarnych SI ze szczególnym podkreśleniem roli polityki zbrojeniowej, złożoności przetwarzania SI i zakresu wykorzystywanych technologii informatycznych.

[1]Sławomir Nikiel, dr hab. inż., prof. uczelniany w Katedrze Logistyki i Systemów Informacyjnych, Wydział Ekonomii i Zarządzania Uniwersytetu Zielonogórskiego. Działalność naukowa dotyczy głównie zagadnień systemów wirtualnej rzeczywistości VR, rozszerzonej rzeczywistości XR oraz systemów sztucznej inteligencji AI.W latach 2009-2013 uczestniczył z ramienia Sztabu Generalnego Wojska Polskiego, Zarządu Planowania Systemów Dowodzenia i Łączności w pracach NATO-wskiej komisji Link16 I&IWG (Implementation and Interoperability Working Group)w grupie roboczej dedykowanej przez MIRB (MIDS International Review Board) oraz był członkiem International Data Link Society.

Przedstawiono dyskusję „bojowego-SI” oraz zasugerowano potencjalne podejścia do osiągnięcia przewagi w bezpośredniej konfrontacji człowiek-SI na arenie działań wojennych.

Abstract

In recent years/months a tremendous development of Artificial Intelligence has been noticed. Its application in military operations requires AI driven decision-making mostly in the field of identifying and eliminating targets in the context of typical cases ‘trained’ by Machine Learning. The proposed presentation/paper identifies currently available solutions in the field of the AI warfare, with particular emphasis on the role of warfare policy, AI processing complexity and the underlying ICT technologies. The "combat-AI" discussion is forwarded alongside with some potential approaches to gain advantage in direct human-AI confrontation on the battleground.

Słowa kluczowe: Sztuczna Inteligencja, Uczenie Maszynowe, Autonomiczne Systemy Broni, Czynnik Ludzki, Osobliwość

Keywords: Artificial Intelligence, Machine Learning, Autonomous Weapon Systems, Human Factors, Singularity

Wprowadzenie

Najczęściej pojawiającym się argumentem za wprowadzeniem Sztucznej Inteligencji i w pełni autonomicznych systemów jest „dostrzegalna” korzyść płynąca z ich używania w szczególności do zapobiegania urazom lub nawet śmierci, tak jak w przypadku zaawansowanych systemów bezpieczeństwa stosowanych w samochodach. Jeśli chodzi o śmiercionośne systemy zaprojektowane do celów wojskowych LAWS (ang. Lethal Autonomous Weapon System) wykorzystuje się podobną argumentację: ograniczenie lub wręcz brak strat „własnych” żołnierzy, ogromna precyzja uderzenia i zapobieganie strat wśród cywilów (ang. collateral damage). Te, brzmiące nadzwyczaj racjonalnie, korzyści z zastosowania LAWS na polu walki mogą być jednak przeważone przez długoterminowe konsekwencje[2] . Przykładowo rozwiązania LAWS nie tylko ograniczają ryzyko odniesienia ran przez własnego żołnierza ale ich wykorzystanie może również skutkować tempem walki, które przekracza ludzką reakcję i refleksyjne zdolności decyzyjne dowódców. Roje (ang. swarms) małych uzbrojonych dronów mogą wręcz zamienić pola bitewne w strefy całkowicie zabójcze dla każdego człowieka a tempo działań wojennych może nasilać się poza znaczącą ludzką kontrolą (24/7/365)[3]. Dzisiaj nawet laicy wiedzą, że „roboty” i „oprogramowanie” nie

[2]Vardi M., *On Lethal Autonomous Weapons*, Communications of the ACM, Vo. 58, No.12, 2015

[3]Scharre P., *Army of None: Autonomous Weapon Systems and the future of War*, N.Y. Norton&Company. 2018

potrzebują czasu na regenerację… Pojawia się więc pytanie, czy jest jeszcze miejsce na działania człowieka oraz czy jest nadzieja na zwycięską konfrontację w zderzeniu człowiek-SI?

1. Złożoność wojny - C5ISR

Dzisiaj praktycznie wszystkie „rozwinięte" państwa postrzegają zaawansowaną technologicznie broń jako niezbędną do ich przetrwania. Obecnie USA i inne państwa nie mają żadnych wdrożonych systemów uzbrojenia które mogą działać w pełni autonomicznie. Jednak już sama nazwa systemów współczesnego pola walki C5ISR (ang. Command, Control, Communications, Computers, Cyber, Intelligence, Surveillance and Reconnaissance) pokazuje, jak złożony jest problem na styku człowiek-technologia. Osiągnięcie zamierzonego efektu operacyjnego EBO (ang. Effect Based Operations) zależy w dużym stopniu od tych usług które będą zintegrowane w środowiska sieciocentryczne np. dla NATO jest to NII/NNEC (ang. NATO Information Infrastructure/ NATO Network Enabled Capability) oraz gotowe do użycia na poziomie taktycznym, w celu wsparcia dowódcy i stratega ale także żołnierza na froncie. Obecnie istnieje wiele wojskowych systemów uzbrojenia działających półautonomicznie, wymagając pewnej kontroli człowieka lub danych wejściowych aby np. wybrać cele albo zaprogramować algorytm do wykonania uderzenia. Efekt końcowy jest wciąż wynikiem wielu działań i decyzji, niektórych podejmowanych przez ludzi a także niektórych podejmowanych/wykonanych przez maszyny. Definiowanie przyczynowości przy omawianiu działań przez bardzo złożone systemy jest prawie niemożliwe. Niektórzy przywódcy wojskowi, jak i sami żołnierze są jednak zaniepokojeni tym, że służba wojskowa zostanie pozbawiona jakichkolwiek „cnót" i postaw etycznych, na polu walki zostanie jedynie techniczna skuteczność czyli osiągniecie „efektu operacyjnego" (EBO)…

2. Polityka zbrojna a Autonomiczne Systemy Broni

Niektórzy stratedzy uważają, że skutki robotyzacji wojny całkowicie zmienią dotychczasową naturę wojny[4]. „Myślę, że to całkiem jasne, że przewaga militarna w XXI wieku będzie w dużej mierze polegać na umiejętnym mieszaniu ludzi i inteligentnych maszyn" twierdził John Arquilla, profesor i przewodniczący Departamentu Analiz Obronnych

[4]Danet D., *Digitization and Robotization of the Battlefield: Evolution or Robolution?,*w: *Robots on the Batllefield, Contemporary Issues and Implications for the Future*, red. Doare R. US Army Combined Arms Center, Fort Leavenwort 2014

w Szkole Podyplomowej Marynarki Wojennej Stanów Zjednoczonych w Monterey, Kalifornia[5].

Dyrektywa DoD Directive 3000.09 z 2012 roku zakładała, że autonomiczne i półautonomiczne systemy uzbrojenia są projektowane w sposób umożliwiający dowódcy i operatorom narzędzia w celu wyegzekwowania odpowiedniego poziomu osądu ludzkiego nad użyciem takiej broni (tzw. kill switch czyli możliwość „manualnego" wyłączenia systemu w każdym momencie)[6]. Od tego czasu postęp technologiczny, jaki dokonał się w uczeniu maszynowym ML (ang. Machine Learning) i Sztucznej Inteligencji SI/AI (ang. Artificial Intelligence) spowodował rozważenie bardziej niejednoznacznych sytuacji, w których człowiek ceduje pełną decyzyjność na autonomiczny system. Założenia przyjęte w CODE (ang. Collaborative Operations in Denied Environments) oraz TRACE (ang. Target Recognition and Adaptation in Contested Environments) biorą pod uwagę sytuacje, w której z braku kontaktu z ośrodkiem dowodzenia (niezależnie czy przyczyną jest brak łączności, zakłócenie czy przejęcie „inicjatywy") system LAWS samodzielnie podejmuje decyzję o ataku dopasowując się do aktualnych okoliczności a celem nadrzędnym jest dokończenie „misji".

Jak wspomniano wcześniej, istnieje dzisiaj wiele wojskowych systemów uzbrojenia działających półautonomicznie. W arsenale aktualnie stosowanych rozwiązań możemy znaleźć systemy takie jak: pociski kierowane/naprowadzane PGM (ang. Precision Guided Munitions) takie jak Brimstone, Javeline, Excalibur…, systemy przeciwokrętowe LRASM (ang. Long Range Anti-Ship Missile) np. Harpoon ale też systemy obrony morskiej Aegis, drony bojowe UCAV(ang. Unmanned Combat Aerial Vehicle) do których możemy zaliczyć MQ Reaper, Predatory czy też tureckie Bayraktary, drony-samobójcy w postaci amunicji „krążącej" (ang. Loitering Munition) np. Switchblade których system sam dokonuje wyboru celu i sposobu podejścia do jego zniszczenia wykorzystując algorytmu ATR (ang. Automatic Target Recognition) oraz BDA (ang. Battle Damage Assessment). Odrębną kategorię stanowią "roje" (ang. Swarms) broni, w której dziesiątki/setki/tysiące uzbrojonych dronów realizują zadanie eleminacji wroga wykorzystując szereg specyficznych dla tego rozwiązania taktyk, poczynając od bezpośredniego ataku „wszyscy na jednego" tzw. „Greedy Shooter" poprzez zachowania stadne typowe dla stad ptaków/rojów owadów (ang. flocks) gdzie kilka

[5]Kirkpatrick K., *Can We Trust Autonomous Weapons?*, Society, Comminications of the ACM, Dec. 2016, Vol.59/No. 12, pp. 27-29.

[6]DoD Directive 3000.09, 2012

jednostek „dowodzi" a reszta stada naśladuje przywódcę, kończąc na kolektywnym współdziałaniu FLA (ang.Fast Lightweight Autonomy) w którym role pętli decyzyjnej OODA (ang. Observe, Orient, Decide, Act) są współdzielone przez wszystkie drony w stadzie…

W powyższych rozwiązaniach mowa jest o oprogramowaniu kontrolującym „inteligentną" broń, ale jak wspomniano wcześniej rosnące tempo działań wojennych może skutkować wprowadzeniem autonomiczności także na poziomie planowania i dowodzenia operacją zbrojną. Powołanie przez Pentagon biura CDAO (ang. Chief Digital and AI Office) ma za zadanie przyśpieszyć adaptację SI i uczenia maszynowego ML (ang. Machine Learning) w systemach wojskowych i doprowadzenie w najbliższym czasie do pełnej zdolności operacyjnej. Biuro te ma wspierać działanie Departamentu Obrony USA w opracowaniu strategii użycia ww technologii w armii USA. Najciekawszym obszarem w CDAO jest chyba broń algorytmiczna (CDAO for Algorithmic Warfare), która jest związana bezpośrednio z rozwiązaniami SI. 30 czerwca 2022 NATO ogłosiło, że tworzy fundusz innowacji o wartości 1 miliarda dolarów, który będzie inwestował w start-upy na wczesnym etapie rozwoju i fundusze venture capital rozwijające „priorytetowe" technologie, takie jak sztuczna inteligencja, przetwarzanie dużych zbiorów danych i automatyzacja[7].Według raportu Georgetown Center for Security and Emerging Technologies, chińska armia prawdopodobnie wydaje co najmniej 1,6 miliarda dolarów rocznie na sztuczną inteligencję. Wojna na Ukrainie jeszcze bardziej zwiększyła naglącą potrzebę wprowadzenia na pole bitwy większej liczby technologii LAWS.

3. Uczenie Maszynowe i Sztuczna Inteligencja

Sztuczna Inteligencja SI/ AI (ang. Artificial Intelligence) nawet dzisiaj nie jest dobrze zdefiniowana- najczęściej pojawiająca się definicja sztucznej inteligencji to „inteligencja demonstrowana przez maszyny, w przeciwieństwie do naturalnej inteligencji przejawianej przez ludzi i inne zwierzęta". Aby wyjaśnić różnice należy zacząć od tego, że generalnie istnieją dwa rodzaje SI: „silne" SI i „słabe" SI. Silna Sztuczna Inteligencja jest tym, czym może być większość ludzi myślą o tym, kiedy słyszą o SI, a słabe SI to wysoce wyspecjalizowane algorytmy zaprojektowane w celu odpowiedzi na konkretne, przydatne pytania w wąskim zakresie zdefiniowane dziedziny problemowe. O ile silne SI jeszcze nie

[7]https://www.nato.int/cps/en/natohq/news_197494.htm

powstało, to słabe SI gwałtownie się rozwija w ostatnich latach co jest skutkiem zarówno rozwoju technologii informatycznych jak też badań naukowych nad uczeniem maszynowym ML (ang. Machine Learning) i inżynierią danych (ang. Data Science). Uczenie maszynowe to szczególny sposób tworzenia inteligencji maszynowej. Podejście do uczenia maszynowego jest już dobrze znane, zamiast próbować napisać program komputerowy oparty o model matematyczny, buduje się system danych, który potrafi wymyślić własny zestaw zinternalizowanych modeli-klasyfikatorów po pokazaniu wielu przykładów. Olbrzymia ilość danych zastępuje model matematyczny. Przykładowo, kilka kilobajtów kodu we współczesnym języku programowania wystarczy do opracowania prostych działań arytmetycznych (dodawanie, odejmowanie, mnożenie i dzielenia), model samouczącego się systemu ML będzie potrzebował kilkadziesiąt miliardów parametrów/danych aby znaleźć poprawne rozwiązanie. O ile arytmetyka jest dość prosta do modelowania matematycznego, to rozpoznawanie twarzy, głosu, emocji itp. znacznie wykracza poza algebrę, analizę i geometrię (np. rozpoznawanie mowy wymaga ponad 500 miliardów parametrów). Wtedy modele ML okazują swoją przewagę. Rodzaj modelu, którego się używa ma ogromne znaczenie. Określa on, w jaki sposób Sztuczna Inteligencja się uczy, jakiego rodzaju są dane, z których może się uczyć i jakie pytania możesz zadać SI. Wyzwaniem uczenia maszynowego jest zatem tworzenie i wybór odpowiednich modeli dla właściwych problemów (czym zajmuje się inżynieria danych). Potrzebne są modele, który są wystarczająco wyrafinowane, aby uchwycić skomplikowane relacje i struktury, ale są na tyle prosty, że można je zaprogramować w ML.

Sieci neuronowe NN (ang. Neural Networks) to rodzaj modelu uczenia maszynowego, który wykorzystuje strukturę podobną do neuronów w mózgu do wykonywania obliczeń i przewidywania. Neurony w sieciach neuronowych są zorganizowane w warstwy: każda warstwa wykonuje zestaw prostych obliczeń i przekazuje odpowiedź do następnej. Układanie warstw umożliwia wykonywanie bardziej złożonych obliczeń. Prosta sieć z kilkoma warstwami neuronów wystarczy, aby odtworzyć proste funkcje -klasyfikatory. Głębokie sieci neuronowe DNN (ang. Deep Neural Networks) to sieci z mnóstwem warstw — dziesiątkami, a nawet setkami które są w stanie tworzyć niesamowicie potężne modele. Są one w stanie nauczyć się wszelkiego rodzaju zawiłości bez konieczność posiadania badacza-człowieka, który określi zasady, co pozwoliło nam stworzyć algorytmy do rozwiązywania wszelkiego rodzaju problemów. Jest jeszcze jeden aspekt sieci neuronowych, który przyczynił się do ich sukcesu: trening. „Pamięć” modelu to zestaw parametrów numerycznych, które

decydują o tym, w jaki sposób generuje on odpowiedzi na zadawane pytania. Trening modelu oznacza więc dostrojenie tych parametrów, aby model dał najlepsze odpowiedzi, jakie może. Zbiór danych, który jest wykorzystywany do trenowania sieci neuronowej przeszkolony na może trafić na miliony przykładów, dobrych i złych. Nie ma jednoetapowego rozwiązania analitycznego. Ale można zacząć od złej sieci neuronowej, a potem ulepszyć ją, wprowadzając stopniowe poprawki. Odbywa się to m.in. poprzez wprowadzenie oceny klasyfikatorów i tzw „punktacji". Ogromna różnorodność modeli uczenia maszynowego opiera się na tej idei. Najtrudniejszą rzeczą jest wymyślenie „punktacji" używanej do treningu sieci. Najlepszym sposobem do szkolenia sieci NN jest użycie innej sieci neuronowej. Technika ta nazywa się generatywnymi sieciami adwersarialnymi lub GANN (ang. Generative Adversarial Neural Networks). Mamy zwykle dwie sieci neuronowe, które działają przeciwko sobie: sieć, która próbuje generować dane (losowo) i inna sieć, która jest wyszkolona, aby spróbować odróżnić prawdziwe przykłady od fałszywych, używając kilku przykładów „realnych" danych. Te dwie sieci są następnie przeciwstawiane w konkurencyjnej grze. W tym miejscu wkracza „Adversarial" część nazwy. Sieć generatywna próbuje zrobić przekonujące podróbki, a sieć dyskryminacyjna próbuje dowiedzieć się, co jest prawdziwe, a co nie. Ta technologia zyskała na popularności w ostatnich latach i jest wykorzystywana szeroko w GPT3 (Open AI), ALPHA GO i DEEP MIND[8]. Ta sam technologia miała być rozwijana dla Pentagonu w ramach projektu Maven w celu zbudowania systemów rozpoznawania obrazu dla poprawy zdolności bojowych dronów ale po protestach i oburzeniu pracowników, w 2018 roku Google wycofał się z projektu. Epizod wywołał gorącą debatę na temat praw człowieka i moralności opracowywania Sztucznej Inteligencji dla broni autonomicznej. Ponieważ kontrowersje wokół Mavena wydają się dziś przeszłością, głosy wzywające do zwiększenia AI w obronie stały się coraz głośniejsze w ciągu ostatnich kilku lat. Aby złagodzić obawy, departament Obrony USA opracował wytyczne dotyczące „odpowiedzialnej sztucznej inteligencji" dla twórców sztucznej inteligencji i ma własne wytyczne etyczne dotyczące korzystania z SI. NATO posiada strategię SI, która określa dobrowolne wytyczne etyczne dla swoich państw członkowskich. Wszystkie te wytyczne wzywają wojsko do korzystania ze sztucznej inteligencji w sposób zgodny z prawem,

[8]https://arstechnica.com/features/2019/04/from-ml-to-gan-to-hal-a-peak-behind-the-modern-artificial-intelligence-curtain/

odpowiedzialny, niezawodny i identyfikowalny oraz ma na celu łagodzenie uprzedzeń wbudowanych w algorytmy.

4. Glitch - czy da się „oszukać" AI?

Firmy zajmujące się technologiami Sztucznej Inteligencji proponują wojsku rozwiązania które według nich mogą pomóc we wszystkim, od logistyki do działań bojowych, od przesiewania życiorysów kandydatów do służby do przetwarzanie multispektralnych danych z satelitów lub dronów aby pomóc żołnierzom w podejmowaniu szybszych decyzji na polu bitwy. Wsparte przez SI oprogramowanie do rozpoznawania obrazów może pomóc w identyfikacji celów a w pełni autonomiczne drony mogą być wykorzystywane do nadzoru lub ataków na lądzie, w powietrzu lub na wodzie. Technologie LAWS są jednak wciąż w fazie testów na prawdziwym polu bitwy, czasami bez większych sukcesów. Strefy walki są prawdopodobnie jednymi z najbardziej wymagających technicznie obszarów, w których próbuje się wdrażać Sztuczną Inteligencję, ponieważ istnieje niewiele odpowiednich danych szkoleniowych. Może to spowodować awarię systemów autonomicznych w „złożony i nieprzewidywalny sposób".

Błędy powodowane przez maszynę mają różny zakres i skalę, podobnie jak ich konsekwencje. Wypadki drogowe spowodowane przez „autopilota" Tesli, przypadkowe zwolnienia pracowników firmy po tym, jak system komputerowy HR nieumyślnie rozwiązał umowy o pracę lub przypadek chatbotów Facebooka, gdzie firma stworzyła sztuczną inteligencję, która mogła ze sobą rozmawiać, ale wkrótce programiści zdali sobie sprawę, że boty opracowały własny „tajny"(niezrozumiały) język, którego używają do komunikowania się między sobą. Nowe technologie niosą ze sobą nowe błędy i awarie, których nikt się nie spodziewał. Błędy z SI często zaczynają się na poziomie kodowania i opracowania modeli danych. Jedna z teorii walki z niepowodzeniami sztucznej inteligencji mówi, że pozwalamy maszynom popełniać błędy i uczyć się na nich, tak jak robią to ludzie (wspomniane wcześniej GANN). Może to też być celowe działanie. Yannic Kilcher wyszkolił Sztuczną Inteligencję, używając 3 miliony wątków z niesławnie toksycznej politycznie niepoprawnej tablicy 4chan /pol/. Następnie wypuścił bota z powrotem na 4chan z przewidywalnymi wynikami – Sztuczna Inteligencja była „perfekcyjnie" zła jak posty na których była szkolona, wypowiadając rasistowskie obelgi i angażując się w wątki antysemickie[9].Systemy robotów,

[9]https://www.vice.com/en/article/7k8zwx/ai-trained-on-4chan-becomes-hate-speech-machine?utm_source=reddit.com 2012.

sterowane przez duże zbiory danych i modele GANN, które zawierają „klasyfikatory" ludzi, są również obarczone dużym ryzykiem ogólnego wzmacnianie złośliwych stereotypów[10]..Sztuczna Inteligencja zupełnie przy tym nie wykazuje zrozumienia relacji człowiek-człowiek. Praktycznie wszystkie badania nad SI pomijają zastosowanie „modelu ludzkiego", aby zrozumieć zachowania SI jako podmiotu społecznego. Wyniki nielicznych badań pokazały, że ludzie są bardziej skłonni do przyjęcia celowej postawy po interakcji z bardziej dostępnym społecznie i podobnym do człowieka robotem lub awatarem, podczas gdy nie pojawiło się przyjęcie postawy intencjonalnej w kierunku robota mechanicznego.

Pojawia się pytanie czy nie będąc konstruktorem SI jest możliwe celowe wpłynięcie na jego działanie. Przykładem takich działań są tzw. ataki adwersarialne na modele uczenia maszynowego, które wykorzystują błędy w sieciach neuronowych. Zyskały one rosnące zainteresowanie w ostatnich latach. Dokonując tylko subtelnych zmian na wejściu sieci neuronowej, wyjście sieci może dać zupełnie inny wynik. Pierwsze ataki odbywały się przez nieznacznie zmienione wartości pikseli obrazu wejściowego, aby oszukać klasyfikator. Inne podejścia wykorzystywały „łatki" (ang. patches), które można było zastosować do obiektu aby oszukać detektory i klasyfikatory. Wykazano, że ataki te są możliwe do zrealizowania w świecie rzeczywistym, tj. poprzez modyfikację obiektu[11].Z powodzeniem wykorzystali tę technikę studenci podczas protestów w Hongkongu w 2019 roku, więc możliwe jest opracowanie specjalnego „kamuflażu" chroniącego przed LAWS. Innym sposobem jest wyjście poza schemat zachowań i „ukrycie" się w nietypowym środowisku i grupie ludzi…

5. Konkluzja i „mroczna przyszłość"

Potencjał Sztucznej Inteligencji i robotyki został dostrzeżony przez polityków, wojskowych i przemysł zbrojeniowy co spowodowało rozważenie użycia tej technologii w zastosowaniach militarnych. Technologia jest w niewystarczającym stopniu dojrzała a skuteczność jej wykorzystania wciąż zależy w bardzo dużym stopniu od ludzi. Istnieje globalna kampania o nazwie Stop Killer Robots, która ma na celu zakazanie śmiercionośnej broni autonomicznej, takiej jak roje dronów. Aktywiści, urzędnicy wysokiego szczebla, m.in. szef ONZ António Guterres oraz rządy takie jak Nowa Zelandia twierdzą, że broń autonomiczna jest głęboko nieetyczna, ponieważ daje maszynom kontrolę nad decyzjami dotyczącymi życia i śmierci

[10]Birhane A. , Prabhu V., Kahembwe E. 2021. Multimodal datasets: misogyny, pornography, and malignant stereotypes. ArXiv abs/2110.01963 (2021)

[11]Thys S. at Al., Fooling automated surveillance cameras: adversarial patches to attack person detection, arXiv:1904.08653v. 2019

i może nieproporcjonalnie zaszkodzić zmarginalizowanym społecznościom poprzez błędy algorytmiczne. Niemniej jednak powszechnie uważa się, że w przeciwieństwie do broni ABC (Atomowej Biologicznej Chemicznej), z definicji skierowanej do zabijania ludzi w skali masowej broń inteligentna jest bronią konwencjonalną, w której decyzję o „likwidacji" w de facto indywidualnego przeciwnika podejmuje wciąż człowiek lub w wersji pełni autonomicznej Sztuczna Inteligencja. Biorąc pod uwagę ostatnie doniesienia o eksperymencie, w którym Sztuczna Inteligencja wykorzystywana do poszukiwania bardziej skutecznych leków potrzebowała mniej niż sześć godzin, aby wynaleźć 40 000 potencjalnie śmiertelnych molekuł trudno oprzeć się wrażeniu, że znajdujemy się w przełomowym momencie kreacji „Puszki Pandory". Naukowcy zaprogramowali SI aby pokazać, jak łatwo można ją nadużywać np. do opracowania nowej broni biologicznej. Jedyne, co musieli zrobić, to zmienić swoją metodologię, aby szukać zamiast wyeliminować toksyczność generowanych związków biochemicznych. SI wymyśliła dziesiątki tysięcy nowych substancji, z których niektóre okazały się podobne do VX, który jest inhibitorem tak zwanej acetylocholinoesterazy. Sposób, w jaki VX jest zabójczy polega na tym, że faktycznie powstrzymuje przeponę i mięśnie płuc, przed poruszaniem się przez co płuca zostają sparaliżowane. Największym zaskoczeniem naukowców było to, że wiele nowych unikalnych wygenerowanych związków okazało się potencjalnie dużo bardziej toksycznych niż VX[12]…

1. **Bibliografia:**

2. Birhane A. , Prabhu V., Kahembwe E. 2021. *Multimodal datasets: misogyny, pornography, and malignant stereotypes*. ArXiv abs/2110.01963 (2021)
3. Calma J.,*AI suggested 40,000 new possible chemical weapons in just six hours*, 2022,https://www.theverge.com/2022/3/17/22983197/ai-new-possible-chemical-weapons-generative-models-vx
4. Carras T. et Al., *A Style-Based Generator Architecture for Generative Adversarial Network*, Neural and Evolutionary Computing, 2019.
5. Danet D., *Digitization and Robotization of the Battlefield: Evolution or Robolution?*, w: *Robots on the Batllefield, Contemporary Issues and Implications for the Future*, red. Doare R.US Army Combined Arms Center, Fort Leavenwort 2014
6. Demarest C., *Pentagon's AI, data office fully operational as leadership posts filled*, Pentagon 2022
7. Department of Defence Directive, NUMBER 3000.09, 2012

[12]Calma J., AI suggested 40,000 new possible chemical weapons in just six hours, 2022, https://www.theverge.com/2022/3/17/22983197/ai-new-possible-chemical-weapons-generative-models-vx

.

8. Dorier J., *This AI Learned the Design of a Million Algorithms to Help Build New AIs Faster* ,SingularityHub, 2022, https://singularityhub.com/2022/01/31/this-ai-learned-the-design-of-a-million-algorithms-to-help-build-new-ais-faster/
9. Gault M., *AI Trained on 4Chan Becomes "Hate Speech Machine",* Motherboard, Tech by vice, 2022
10. Huang H., *The basics of modern AI—how does it work and will it destroy society this year?,* Ars Technica 2019
11. https://arstechnica.com/features/2019/04/from-ml-to-gan-to-hal-a-peak-behind-the-modern-artificial-intelligence-curtain/
12. Knight W., *The Pentagon Inches Toward Letting AI Control Weapons* ,Wired 2022
13. Kirkpatrick K., *Can We Trust Autonomous Weapons*?, *Society, Comminications of the ACM*, Dec. 2016, Vol.59/No. 12, pp. 27-29
14. Kopeć R., *Dyplomacja dronów*, „Kultura i Polityka", 2015, nr.17, s. 65-83
15. Kopeć R., *Kodeksy etyczne robotów bojowych*, w:*Wojna/Pokój – Humanistyka wobec wyzwań współczesności*, red. R.Sapeńko/ P.Pochyły, Zielona Góra 2017
16. Krishan A., *Killer Robots. Legality and Ethicality of Autonomous Weapons*, Ashate, Farnham 2009
17. McDonald G., Slaughterbots' Video Depicts a Dystopian Future of Autonomous Killer Drones | Space, 2017https://www.space.com/38820-slaughterbots-video-depicts-a-dystopian-future-of-autonomous-killer-drones.html
18. Moosavi-Dezfooli M. et Al., *Deep fool: a simple and accurate method to fool deep neural networks*. w: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2574–2582, 2016
19. Rose J., *Image-Generating AI Keeps Doing Weird Stuff We Don't Understand*, Motherboard, Tech by vice,2022
20. Scharre P., *Robotics on the Battlefield*, Washington 2014
21. Scharre P., *Army of None: Autonomous Weapon Systems and the future of War,* N.Y. Norton&Company. 2018
22. Thys S. at Al., *Fooling automated surveillance cameras: adversarial patches to attack person detection*, arXiv:1904.08653v. 2019
23. Vincent J., *DeepMind's AI agents conquer human pros at StarCraft II* - The Verge, 2022
24. Vardi M., *On Lethal Autonomous Weapons*, Communications of the ACM, Vo. 58, No.12, 2015
25. Wei J. et Al., *Emergent Abilities of Large Language Models*, Computer Science, Computation and Language, arXiv:2206.07682. 2022
26. Windeck A., *Preface*, w: *Robots on the Battlefield, Contemporary Issues and Implications for the Future*, red. R. Doare et Al.,Fort Leavenworth 2013

Summary

The proposed paper identifies currently available solutions in the field of the AI warfare, discusses complexity of war and role of Autonomous Weapon Systems, AI/ML processing complexity and the underlying ICT technologies. Finally the concept of 'glitch'is exploited as a chnce for mankind to survive direct confrontation with combat-AI.