SPRAWY MIĘDZYNARODOWE

2022, t. 75, nr 2
ISSN 0038-853X
e-ISSN 2720-0361
DOI 10.35757/SM.2022.75.2.06

Dezinformacja i operacje wpływu

KAROL KUMALSKI
Uniwersytet Marii Curie-Skłodowskiej
ORCID: 0000-0003-1753-4714
k.kumalski@onet.pl

Sztuczna inteligencja jako instrument intensyfikacji zagrożeń hybrydowych w domenie informacyjnej

Artificial intelligence as an instrument of intensification hybrid threats in the information domain

Słowa kluczowe:
sztuczna inteligencja, AI, zagrożenia hybrydowe, dezinformacja, ingerencja, legitymizacja, manipulacja, przestrzeń informacyjna

Keywords:
artificial intelligence, AI, hybrid threats, disinformation, interference, legitimization, manipulation, infosphere

Sztuczna inteligencja jako instrument intensyfikacji zagrożeń hybrydowych w domenie informacyjnej

Ze względu na swój polimorficzny i wielowymiarowy charakter zagrożenia hybrydowe stały się użytecznym narzędziem implementowanym w szarej strefie interakcji między państwami autorytarnymi i demokratycznymi. Jednocześnie spektrum tych zagrożeń może być dynamicznie ulepszane i poszerzane dzięki rozwojowi technologicznemu. Sztuczna inteligencja jest jednym z czynników, które mogą zintensyfikować siłę i zasięg zagrożeń hybrydowych. Wykorzystując model koncepcyjny zaproponowany przez Europejskie Centrum Doskonałości w zakresie Przeciwdziałania Zagrożeniom Hybrydowym (Hybrid CoE), można określić potencjał sztucznej inteligencji w zakresie oddziaływania na domenę informacyjną. Właśnie tam możliwe jest wykorzystanie tej technologii do manipulowania przestrzenią informacyjną, głównie poprzez dezinformację i ingerencje. Dla realizacji tego celu może ona pełnić funkcje analityczno-decyzyjną, narracyjną i aksjologiczną, uderzając w podstawy demokracji oraz zniekształcając i podważając relacje między państwem a społeczeństwem. Warto również zauważyć, że o ile do tej pory głównymi operatorami operacji hybrydowych były państwa, o tyle rozwój tej technologii może zwiększyć znaczenie aktorów niepaństwowych, ponieważ stanie się ona dla nich bardziej dostępna.

Artificial intelligence as an instrument of intensification hybrid threats in the information domain

Due to polymorphic and multidimension character, hybrid threats became an useful tool implemented in grey zone of interaction between authoritarian and democratic states. Simultaneously, their spectrum might be dynamically upgraded and expanded, thanks to technological development. Artificial intelligence is one of these tools, which can intensify power and range of hybrid threats. Using a conceptual model proposed by the European Centre of Excellence in Countering Hybrid Threats (Hybrid CoE), it is possible to determine the potential of artificial intelligence in influencing an information domain. There, it could be used to manipulate an infosphere, mainly through disinformation and interference. For this purpose AI can play analytical-decision making, narrative and axiological functions by striking the foundations of democracy and distorting and undermining relations between state and society. It is also worth to note, that while so far states have been the main operators of hybrid operations, the development of this technology may increase the non-state actors importance as this technology will become more accessible for them.

Wstęp

Celem niniejszego artykułu jest próba określenia potencjału sztucznej inteligencji (AI) jako czynnika mogącego zintensyfikować zagrożenia hybrydowe w domenie informacyjnej. Wybór tej technologii jako obiektu badań podyktowany jest jej unikalną istotą i specyfiką, a także coraz większą obecnością w codziennym życiu. Znajduje się ona w grupie czynników określanych w literaturze anglosaskiej terminem *emerging threats* (zagrożenia wschodzące). Natura sztucznej inteligencji wydaje się też zgodna z polimorficzną i wielowymiarową specyfiką zagrożeń hybrydowych, w szczególności w domenie informacyjnej. Autor przyjął tym samym hipotezę, że AI może stać się instrumentem intensyfikacji zagrożeń hybrydowych, na co wskazują jej specyfika, obszary wykorzystywania, wzrost dostępności czy też fakt, że zaliczana jest do technologii podwójnego zastosowania. Ponadto dzięki algorytmom sztucznej inteligencji tradycyjne zagrożenia hybrydowe mogą się nasilać oraz uzyskiwać nowe formy i sposoby oddziaływania. Do weryfikacji przyjętej hipotezy wykorzystana zostanie synteza takich metod i technik badawczych jak:

– analiza czynnikowa, mająca wykazać wpływ wykorzystania sztucznej inteligencji w działaniach hybrydowych na bezpieczeństwo międzynarodowe;

– analiza stanu literatury poświęconej tematyce AI w stosunkach międzynarodowych, możliwościom jej zastosowania oraz refleksjom na temat jej potencjału w kontekście specyfiki zagrożeń hybrydowych;

– analiza przypadków, mająca udokumentować dotychczas odnotowane zastosowania AI.

Z uwagi na fakt, że zarówno sztuczna inteligencja, jak i zagrożenia hybrydowe to relatywnie nowe zjawiska w stosunkach międzynarodowych, istnieją naturalne ograniczenia dostępu do źródeł pierwotnych. Z pomocą przychodzą tu jednak źródła wtórne, publikowane głównie w Stanach Zjednoczonych i Europie Zachodniej.

Artykuł podzielony został na cztery części. Pierwsze dwie mają charakter teoretyczny, a poświęcone są istocie i specyfice sztucznej inteligencji oraz zagrożeniom hybrydowym. Przedstawiono w nich również model koncepcyjny, wykorzystany w późniejszej analizie. Trzecią część stanowi próba dokonania prognozy wpływu AI na zagrożenia hybrydowe i zlokalizowania jej w domenie informacyjnej. Czwarta służy wreszcie omówieniu ról, w których sztuczna inteligencja może służyć intensyfikacji tych zagrożeń.

Sztuczna inteligencja – istota i specyfika

Na wstępie należy zaznaczyć, że jak dotąd nie wypracowano jednolitej powszechnie przyjętej definicji terminu *sztuczna inteligencja*. Taki stan utrzymuje się od 1956 r., gdy po raz pierwszy pojawił się on w nauce. Rozwój AI przypomina sinusoidę, na amplitudy której przypadają zarówno okresy rozkwitu, jak i momenty, w których badania były praktycznie zawieszone[1]. Tym samym w ciągu ostatnich sześćdziesięciu lat rozumienie pojęcia ulegało ewolucji, na którą wpływał postęp technologiczny w informatyce. Przykładem jest tu definicja przyjęta przez Ziemowita Jacka Pietrasia pod koniec lat osiemdziesiątych XX w. na fali popularności systemów eksperckich i ich pionierskiego omówienia na gruncie polskiej politologii. Analizując różne modele definicyjne, badacz ten przyjął, że istotą sztucznej inteligencji jest tworzenie komputerowych systemów niealgorytmicznego przetwarzania symboli, w pewnym zakresie zdolnych do adaptacji. Utożsamiał zatem sztuczną inteligencję systemu z jego samodzielną zdolnością do adaptacji[2].

Fiasko rozwoju AI w oparciu o systemy eksperckie sprawiło, że współcześnie jest ona najczęściej utożsamiana z architekturą głębokich sieci neuronowych[3]. Mają one co najmniej kilka warstw neuronów, które dzięki algorytmom propagacji wstecznej[4] stopniowo przekazują sygnał do wyższych poziomów, ucząc się przy tym. Proces ten zakłada automatyczne ulepszanie komputerowego algorytmu poprzez doświadczenie i nazywany jest uczeniem maszynowym[5]. System potrzebuje milionów przykładów,

1 N. Bostrom, *Superinteligencja. Scenariusze, strategie, zagrożenia*, Helion, Gliwice 2016, s. 23–31.

2 Z. J. Pietraś, *Sztuczna inteligencja w politologii. Heurystyczne modelowanie procesów adaptacji politycznej*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 1990, s. 104.

3 Obok głębokich sieci neuronowych w pracach nad rozwojem AI wykorzystuje się algorytmy ewolucyjne i metody logiki rozmytej. W przeszłości spore nadzieje wiązano też m.in. z systemami eksperckimi.

4 Metoda uczenia sieci wielowarstwowej, w której błąd ostatniej warstwy jest przesyłany wstecz i wykorzystywany do zmiany przez system parametru wartości sygnałów wejściowych (wag połączeń) w poprzednich warstwach. Parametr ten jest zmieniany poziom po poziomie wstecz, dzięki czemu system się uczy.

5 Uczenie maszynowe uważane jest obecnie za najprostsze rozwiązanie prowadzące do zbudowania sztucznej inteligencji. W tematyce tej często pojawia się

aby z sukcesem zakończyć proces uczenia się. Pełni funkcję zadaniową, a więc może być wykorzystywany wyłącznie w określonym celu. Jego zaletami są szybkość przetwarzania informacji oraz bardzo duża skuteczność – w niektórych zastosowaniach przerastająca możliwości człowieka. Michael Horowitz określa w związku z tym sztuczną inteligencję jako użycie komputera w sposób symulujący inteligentne zachowanie człowieka[6]. Missy Cummings uważa, że jest to zdolność systemów komputerowych do wykonywania zadań na co dzień wymagających ludzkiej inteligencji, takich jak przetwarzanie obrazu, rozpoznawanie mowy czy podejmowanie decyzji[7]. Inne, bardzo szerokie spojrzenie na sztuczną inteligencję definiuje ją jako zaprogramowanie urządzenia sterowanego przez komputer w taki sposób, aby umożliwić mu dostrzeganie, rozumowanie i działanie lub zapewnić automatyzację inteligentnego zachowania[8]. Bardzo prostą definicję zaproponował natomiast Max Tegmark, dla którego jest to po prostu inteligencja niebiologiczna[9].

W powyższych definicjach sztuczna inteligencja postrzegana jest przez pryzmat antropocentryczny. Charakteryzuje się to poszukiwaniem podobieństw i porównań do inteligencji ludzkiej. Taki kierunek myślenia poddany został zdecydowanej krytyce przez profesora filozofii na Uniwersytecie Kalifornijskim w Berkeley Johna Searle'a. Jego zdaniem komputery nie mogą myśleć tak jak ludzie, ponieważ symulują myślenie, ale nie duplikują go jako procesu[10]. Rozumowanie odległe od podejścia antropocentrycznego

także pojęcie *głębokie uczenie* (*deep learning*). Jest ono jednym ze sposobów uczenia maszynowego obok takich rozwiązań jak: uczenie drzew decyzyjnych, uczenie Bayesowskie, uczenie przez wzmacnianie. Szerzej zob. T. M. Mitchell, *Machine learning*, McGraw-Hill, Ohio 1997.

6 M. Horowitz, *Artificial intelligence. International competition and balance of power*, „Texas National Security Review" 2018, vol. 1, issue 3, s. 40.

7 M. L. Cummings, *Artificial intelligence and the future of warfare*, [w:] *Artificial intelligence and international affairs. Disruption anticipated*, ed. M. L. Cummings i in., The Royal Institute of Foreign Affairs, London 2018, s. 7.

8 S. de Spiegeleire, M. Maas, T. Sweijs, *Artificial intelligence and the future of defence*, The Hague Centre for Strategic Studies, The Hague 2017, s. 27.

9 M. Tegmark, *Życie 3.0. Człowiek w erze sztucznej inteligencji*, Prószyński i S-ka, Warszawa 2019, s. 58.

10 J. Kaplan, *Sztuczna inteligencja. Co każdy powinien wiedzieć*, PWN, Warszawa 2019, s. 95–98.

zaprezentowali również eksperci z waszyngtońskiego Centre for Strategic and International Studies. Uważają oni, że AI to skonkretyzowany algorytm stworzony pod kątem osiągnięcia precyzyjnego celu lub rozwiązania określonego problemu. Jest to zatem oprogramowanie, którego kod zawiera algorytmy podejmujące decyzje na podstawie przetworzonych danych w celu implementacji świadomego wykonywania zadań przez maszyny. Oznacza to, że sztuczna inteligencja nie musi myśleć w sposób odpowiadający ludzkiej inteligencji bądź świadomości, a realizacja postawionych przed nią zadań może nie mieć związku z tym, co powszechnie uważane jest za logiczne myślenie[11]. Podejście to rozwija Nick Bolstrom, twierdząc, że sztuczna inteligencja nie musi przypominać działania ludzkiego mózgu. Zaawansowane systemy AI będą bowiem prawdopodobnie skrajnie różne, a pod względem architektury poznawczej mogą być całkiem odmienne niż biologiczne istoty rozumne[12]. Zdaniem autora takie ujęcie omawianej problematyki wydaje się bardziej trafne i bliżej oddające jej istotę, a także mające potwierdzenie w praktyce. Wykorzystanie sieci neuronowych, które implementują proces głębokiego uczenia, sprawiło, że decyzje podejmowane przez algorytmy wykraczają często poza zakres wiedzy człowieka i możliwości klasycznej interpretacji. Przetwarzanie olbrzymich ilości danych przez tego typu algorytmy pozwala bowiem na dostrzeżenie wyników lub zależności nieodkrytych wcześniej przez człowieka. Słuszność takiego rozumowania potwierdza zachowanie AI w tradycyjnych grach planszowych, gdzie stosowane przez nią strategie, które powszechnie uznaje się za przegrywające, prowadziły algorytm do zwycięstwa.

Reasumując rozważania dotyczące prawidłowego zdefiniowania sztucznej inteligencji, autor niniejszego artykułu proponuje następujące sformułowanie: sztuczna inteligencja to uprzednio zaprogramowany przez człowieka algorytm mający zdolność autonomicznego podejmowania decyzji na podstawie zautomatyzowanego przetwarzania olbrzymiej ilości danych i pozbawiony bufora hierarchii wartości lub wpływu czynnika emocjonalnego.

Rozwój AI doprowadził do wyodrębnienia kilku poziomów tej technologii. Zasadniczo rozróżnia się trzy rodzaje sztucznej inteligencji. Jej obecnym

11 L. Sheppard i in., *Artificial intelligence and national security. The importance of the AI ecosystem*, CSIS, Washington 2018, s. 6.

12 N. Bostrom, *Superinteligencja…*, s. 55.

etapem rozwoju są systemy określane mianem wąskiej lub słabej sztucznej inteligencji (*artificial narrow intelligence*), przez co rozumie się automatyczne systemy działające na poziomie przewyższającym możliwości obliczeniowe człowieka, wyspecjalizowane w realizacji konkretnego zadania. Ich zaprogramowana inteligencja jest równa lub wyższa od ludzkiej, ale wyłącznie w realizowanym zadaniu, np. grze w szachy czy go, filtrach sieciowych, translatorach internetowych, systemach transakcyjnych wykorzystywanych na giełdach papierów wartościowych. Za drugi, wyższy poziom sztucznej inteligencji uważa się tzw. ogólną (silną) sztuczną inteligencję (*artificial general/strong intelligence* – AGI). Mają to być systemy zaprogramowane w sposób odpowiadający ludzkiemu działaniu w każdym zadaniu lub celu. Za trzeci, najwyższy poziom uważa się sztuczną superinteligencję (*artificial superintelligence*), która przewyższy ludzki punkt odniesienia w każdym wymiarze[13].

W literaturze anglojęzycznej sztuczna inteligencja zaliczana jest do grupy zagrożeń wschodzących (*emerging threats*)[14]. Definiuje się je jako czynniki zaistniałe w rezultacie rozwoju technologicznego i mogące się przyczynić do powstania istotnych zagrożeń dla funkcjonowania płaszczyzn aktywności uczestników stosunków międzynarodowych. Iwan Danilin zastrzega, że chodzi tu o technologie, które wciąż są bardziej kwestią przyszłości niż teraźniejszości, a analiza ich wpływu na stosunki i bezpieczeństwo międzynarodowe koncentruje się raczej na oddziaływaniu prognozowanym niż rzeczywistym. Jak jednak słusznie zauważa, przy obecnym tempie zmian wyraźne rozgraniczenie faktycznego i prognozowanego wpływu technologii na stosunki międzynarodowe staje się coraz mniej klarowne[15].

W ramach badań nad zagrożeniami wschodzącymi określono kilka parametrów charakteryzujących ten obszar bezpieczeństwa międzynarodowego. Po pierwsze uznano, że zagrożenia te nie są powszechnie obecne w dyskursie naukowym lub eksperckim. Widać to na przykładzie sztucznej inteligencji, której poświęca się sporo miejsca w prasie codziennej,

13 Szerzej zob.: tamże, s. 46–94; S. de Spiegeleire, M. Maas, T. Sweijs, *Artificial...*, s. 30.

14 I. Danilin, *Emerging technologies and their impact on international relations and global security*, „Hoover Institution" [online], 3 X 2018 [dostęp: 17 XII 2022]: <https://www.hoover.org/research/emerging-technologies-and-their-impact-international-relations-and-global-security>.

15 Tamże.

branżowej i mediach popularnonaukowych, podczas gdy liczba poruszających tę tematykę materiałów *stricte* naukowych jest relatywnie uboga w porównaniu z podnoszeniem innych czynników bezpieczeństwa międzynarodowego. Po drugie, część z tych zagrożeń jest albo niedostrzegalna, albo całkowicie niedojrzała, jak skutki eksploracji przestrzeni pozaziemskiej dla bezpieczeństwa międzynarodowego. Niektóre z nich, np. zmiany klimatyczne, mają charakter utajony, mimo że od kilku lat postrzega się je jako wyzwania dla bezpieczeństwa międzynarodowego. W praktyce oznacza to, że stały się już zagrożeniem, ale skutki ich oddziaływania są rozłożone w czasie i mogą być odczuwalne dopiero za kilka, a nawet kilkadziesiąt lat. Zagrożenia tradycyjne ulegają z kolei różnym mutacjom związanym z postępem technologicznym, który zmienia ich istotę i specyfikę. Niektóre z nich pojawiają się w sposób nagły i całościowy. Przykładem z przeszłości jest stworzenie bomby atomowej w 1945 r., obecnie zaś istnieje co najmniej kilka technologii, które mogą generować zagrożenia wschodzące[16]. Choć niektórzy badacze skłonni są przyporządkowywać AI do kategorii *disruptive technology*[17], to w rzeczywistości jest to podejście uproszczone, pomijające m.in. skomplikowaną naturę samej technologii, a także odpowiedzi na pytania, czym właściwie ma ona być i gdzie są jej granice. Trafnie bowiem zauważył Frank Sauer, że AI jest technologią jednocześnie przeszacowaną i niedoszacowaną przez decydentów oraz opinię publiczną[18].

Zdaniem wielu ekspertów[19] w najbliższej przyszłości niemożliwe jest pojawienie się sztucznej inteligencji, która dorównywałaby ludzkim

16 Por. G. Herd, D. Puhl, S. Costigan, *Emerging security challenges: framing the policy context*, „GCSP Policy Paper" 2013/5: <https://www.gcsp.ch/publications/emerging-security-challenges-framing-policy-content> [dostęp: 21 XI 2022].

17 Pojęcie *disruptive technology* stosuje się wobec tych rozwiązań, które zamieniają i wypierają te już funkcjonujące w przestrzeni społecznej.

18 F. Sauer, *Military applications of artificial intelligence: nuclear risk redux*, [w:] *The impact of artificial intelligence on strategic stability and nuclear risk*, vol. 1: *Euro-Atlantic perspective*, ed. V. Boulanin, Stockholm International Peace Research Institute, Solna 2019, s. 85.

19 Zakładają oni bezpieczną perspektywę czasową przekraczającą 30–50 lat. Szerzej zob.: V. Muller, N. Bostrom, *Future progress in artificial intelligence: a survey of expert opinions*, [w:] *Fundamental issues of artificial intelligence*, ed. V. Muller, Berlin 2014; S. Baum, B. Goertzel, T. Goertzel, *How long until human-level AI? Results from an expert assessment*, „Technological Forecasting & Social Change" 2011, vol. 78, issue 1, s. 185–195.

zdolnościom intelektualnym. Należy przez to rozumieć niewielkie szanse na stworzenie AGI. Nie oznacza to jednak, że słaba forma AI nie będzie wpływać na stosunki międzynarodowe, zmiany technologiczne oddziałują bowiem na stabilność systemu międzynarodowego, zwłaszcza w wymiarach militarnym i ekonomicznym, o ile zostają spełnione określone warunki. Można do nich zaliczyć m.in. zdolność podmiotów do wykorzystania danej technologii, tempo jej rozprzestrzeniania się oraz jej rzeczywisty wpływ na daną płaszczyznę stosunków międzynarodowych, w tym sposób prowadzenia wojen. Koreluje to z oceną ekspertów z Geneva Centre for Security Policy, którzy zauważają, że wobec gwałtownie zachodzących zmian technologicznych decydenci działają reaktywnie i mimo ostrzeżeń nie są na nie przygotowani[20].

Mimo obserwowanego w ostatnich latach postępu badania nad sztuczną inteligencją wciąż są niewystarczająco zaawansowane. Kierunek rozwoju tej technologii zależy od intelektualnych zdolności pracujących nad nią osób, zasobów finansowych, dostępności danych, nakładanych barier i ograniczeń oraz granic rozbudowy możliwości obliczeniowych obecnych komputerów. Niemniej zainteresowanie tą technologią ze strony państw, a także jej specyfika pozwalają widzieć w niej wielowymiarowy czynnik zagrożenia dla bezpieczeństwa międzynarodowego, w tym niebezpieczny instrument intensyfikacji zagrożeń hybrydowych.

Geneza i natura zagrożeń hybrydowych

Hybrydowość zagrożeń bezpieczeństwa międzynarodowego nie jest zjawiskiem nowym, choć w ostatnim czasie zyskało ono na popularności i doczekało się dopracowania metodologicznego. Sam termin wprowadzony został do nomenklatury politologicznej i spopularyzowany w 2007 r. przez Franka Hoffmana. Analizując ewolucję sposobu prowadzenia konfliktów zbrojnych, zauważył on możliwość ich konwergencji w kierunku multimodalności lub – jak to określił – wojen hybrydowych. Te ostatnie są mieszanką zabójczego konfliktu państwowego z fanatyczną i długotrwałą żarliwością działań formacji nieregularnych. Organizacyjnie mogą być zarówno podporządkowane hierarchicznym strukturom politycznym, jak i sprzężone

20 G. Herd, D. Puhl, S. Costigan, *Emerging...*

z komórkami zdecentralizowanymi[21]. Mimo że początkowo hybrydowość konfliktów utożsamiana była z zagrożeniami asymetrycznymi, takimi jak terroryzm czy przestępczość zorganizowana, to w 2014 r. sposób przeprowadzenia przez Federację Rosyjską aneksji należącego do Ukrainy Krymu zmienił sposób percepcji tego typu zagrożeń. Zmiana ta wyewoluowała na kanwie wygłoszonej kilka miesięcy wcześniej przez szefa sztabu rosyjskiej armii gen. Walerija Gierasimowa opinii nt. sposobu prowadzenia konfliktów zbrojnych – nie tylko środkami militarnymi, ale także instrumentami miękkiej siły. Choć za sprawą brytyjskiego politologia Marka Galeottiego wypowiedź ta utrwaliła się jako doktryna Gierasimowa, to jak zauważa Eugene Rumer, stanowiła raczej próbę twórczego rozwinięcia i uzupełnienia doktryny Primakowa, która funkcjonuje w polityce zagranicznej Rosji od 1996 r.[22] Jakiś czas później sam Galeotti przeprosił zresztą na łamach „Foreign Policy" za to nieprecyzyjne określenie[23].

W literaturze przedmiotu znajdujemy wiele definicji zagrożeń hybrydowych. Sojusz Północnoatlantycki określa je jako działania militarne i niemilitarne oraz jawne i niejawne środki obejmujące dezinformację, ataki cybernetyczne, presję ekonomiczną, użycie nieregularnych grup zbrojnych i wojsk regularnych mające na celu rozmywanie granicy między wojną i pokojem oraz dezorientowanie społeczeństw[24]. Dla Unii Europejskiej stanowią one połączenie działań konwencjonalnych i niekonwencjonalnych (militarnych i niemilitarnych) stosowanych w skoordynowany sposób przez państwowych i niepaństwowych aktorów ukierunkowanych na osiągnięcie celów politycznych. Charakteryzują się one wielowymiarowością oraz wykorzystaniem środków przymusu i dywersji, są trudne do wykrycia i przypisania sprawstwa, dezorientujące i hamujące procesy decyzyjne, co podważa

21 Por. F. Hoffman, *Conflict in the 21st century: the rise of hybrid wars*, Potomac Institute for Policy Studies, Arlington VA 2007.

22 Por. E. Rumer, *The Primakov (not Gerasimov) doctrine in action*, Carnegie Endowment for International Peace, Washington 2019: <https://carnegieendowment.org/2019/06/05/primakov-not-gerasimov-doctrine-in-action-pub-79254> [dostęp: 9 I 2022].

23 M. Galeotti, *I'm sorry for creating the „Gerasimov Doctrine"*, „Foreign Policy" [online], 5 III 2018 [dostęp: 13 XI 2022]: <https://foreignpolicy.com/2018/03/05/im-sorry-for-creating-the-gerasimov-doctrine/>.

24 *NATO's response to hybrid threats*, „North Atlantic Treaty Organization" [online], 16 III 2021 [dostęp: 10 I 2022]: <https://www.nato.int/cps/en/natohq/topics_156338.htm>.

zaufanie obywateli do instytucji rządowych i pogłębia podziały społeczne[25]. Niklas Nilsson ze sztokholmskiego Institute for Security and Development Policy zauważa ponadto, że strona używająca instrumentów hybrydowych chce osiągnąć założony rezultat bez wojny, drogą podważania i uszkadzania systemu politycznego atakowanego obiektu poprzez przemoc, kontrolę, manipulację, działalność wywrotową oraz rozpowszechnianie dezinformacji. Celem ataku nie są więc siły zbrojne, ale społeczeństwo[26]. Na polskim gruncie istotę zagrożeń hybrydowych starali się pogłębić m.in. Krzysztof Liedel[27] i Robert Kupiecki[28].

Na potrzeby niniejszego artykułu przyjęty został kompleksowy model koncepcyjny analizy zagrożeń hybrydowych, który opracowało Europejskie Centrum Doskonalenia w zakresie Przeciwdziałania Zagrożeniom Hybrydowym (Hybrid CoE). Zgodnie z nim stanowią one mnożniki siły i taktyki przymusu stosowanej w celu wsparcia polityki lub strategii, która nie przynosi oczekiwanych rezultatów. W zaproponowanym modelu aktorzy państwowi i niepaństwowi poprzez szereg narzędzi oddziałują w czterech różnych fazach na wybrane z trzynastu domen funkcjonowania atakowanego państwa[29]. Charakterystycznymi cechami zagrożeń hybrydowych w tym modelu są: tworzenie dwuznaczności, ukrywanie rzeczywistego celu, rzucanie wyzwań państwom demokratycznym przez autorytarnych adwersarzy, a także kreatywne i odpowiednio skrojone narzędzia oddziaływania.

25 *A Europe that protects: countering hybrid threats. Factsheet*, „European Union External Action" [online], 13 VI 2018 [dostęp: 10 I 2022]: <https://eeas.europa.eu/topics/economic-relations-connectivity-innovation/46393/europe-protects-countering-hybrid-threats_en>.

26 N. Nilsson i in., *Security challenges in the grey zone. Hybrid threats and hybrid warfare*, [w:] *Hybrid warfare*, ed. M. Weissman i in., London 2021, s. 2–3.

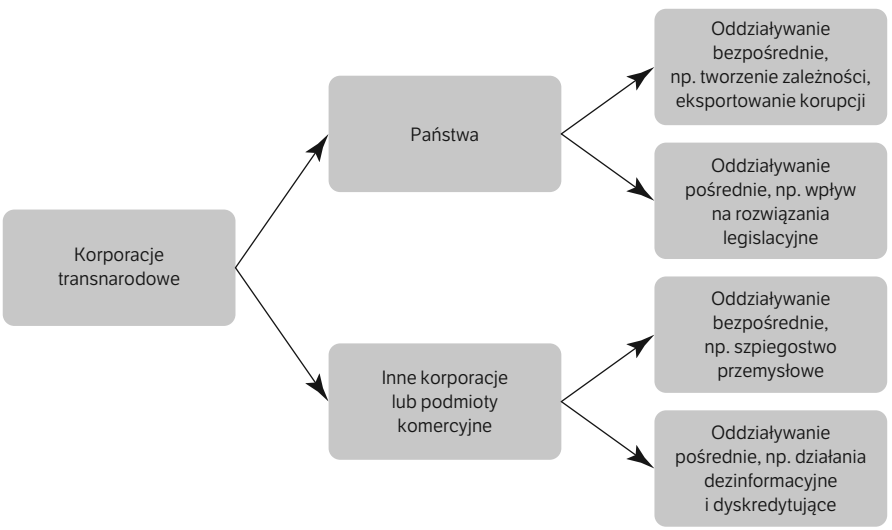
27 Por. K. Liedel, *Zagrożenia hybrydowe. Jak zmienia się środowisko bezpieczeństwa RP?*, „Przegląd Bezpieczeństwa Wewnętrznego" 2015, wydanie specjalne: *Wojna hybrydowa*, s. 51–58.

28 R. Kupiecki, *Dezinformacja w stosunkach międzypaństwowych. Geneza, cele, aktorzy, metody – zarys problemu*, [w:] *Platforma przeciwdziałania dezinformacji – budowanie odporności społecznej*, red. R. Kupiecki i in., Warszawa 2021, s. 15–30.

29 Por. G. Giannopoulos, H. Smith, M. Theocharidou, *The landscape of hybrid threats: a conceptual model*, European Commission, Luxembourg 2021: <https://www.hybridcoe.fi/wp-content/uploads/2021/02/conceptual_framework-reference-version-shortened-good_cover_-_publication_office.pdf> [dostęp: 11 I 2022].

Prognoza wpływu sztucznej inteligencji na intensyfikację zagrożeń hybrydowych

Sztuczna inteligencja, podobnie jak wcześniej rozwój technologii internetowych, może wprowadzić do zagrożeń hybrydowych nową jakość. Postęp w tej technologii, badania nad nią oraz jej implementacja będą prowadzić przede wszystkim do zwiększenia liczby podmiotów generujących zagrożenia hybrydowe. Tradycyjnie były one domeną państw o ustroju autorytarnym, działających samodzielnie lub poprzez *proxy*. Upowszechnienie AI powinno zatem doprowadzić do sytuacji, w której na znaczeniu zyskają podmioty inwestujące w sztuczną inteligencję i zdolne do jej zoperacjonalizowania w wymiarze praktycznym. Oznacza to prawdopodobny wzrost znaczenia podmiotów komercyjnych, takich jak korporacje transnarodowe mogące stać się nie tylko źródłem nowych zagrożeń hybrydowych, ale także podmiotami, które nimi operują (wykres 1)[30].



Wykres 1. Korporacje transnarodowe jako potencjalne źródło zagrożeń hybrydowych (oprac. własne)

30 Tamże, s. 22.

Wdrożenie, a następnie komercjalizacja rozwiązań opartych na AI będą też sprzyjać innym aktorom niepaństwowym, dla których sprzęt i oprogramowanie służące do generowania zagrożeń hybrydowych staną się bardziej dostępne. Niewykluczone zatem, że państwa utracą monopol na tworzenie i wykorzystywanie tego typu zagrożeń. Z drugiej strony zmiany obejmą również same państwa. Na znaczeniu stracą te, które zbyt mało inwestują w przedmiotową technologię lub nie mają sił i środków przeznaczonych na ten cel. Jednocześnie państwa, które obecnie znajdują się na drugim biegunie, o ile będą w stanie właściwie wykorzystać AI, zyskają nowe możliwości zarówno generowania zagrożeń hybrydowych, jak i ochrony przed nimi. Oprócz powyższego można się spodziewać zmian przedmiotowych. Przede wszystkim jakościowej zmianie ulegną obecnie istniejące rodzaje zagrożeń hybrydowych oraz instrumentarium ich stosowania. Dzięki umiejętności autonomicznego przetwarzania ogromnej ilości danych, szybkości działania czy ciągłemu doskonaleniu się w oparciu o algorytmy uczenia maszynowego AI powinna zmultiplikować zakres i charakter stosowanych narzędzi i technik. Co więcej, należy się liczyć z możliwością powstania nowych, obecnie nieznanych zagrożeń hybrydowych oraz płaszczyzn i narzędzi ich wykorzystywania. W takich warunkach sztuczna inteligencja może otwierać zupełnie nowe możliwości oddziaływania poprzez manipulowanie informacją w przestrzeni publicznej. Możliwości oferowane przez tę technologię powinny również objąć takie domeny jak gospodarcza, militarna, społeczna, polityczna czy cyberprzestrzeń. Niemniej jednak eksperci z Hybrid CoE wskazują domenę informacyjną jako prawdopodobnie najbardziej znaczący obszar podatny na zagrożenia hybrydowe. Właśnie tam może być ona bowiem wykorzystywana do podważania percepcji bezpieczeństwa obywateli poprzez agresywne wzmacnianie podziałów politycznych, społecznych i kulturowych. Powstające w ten sposób zamęt i nieporządek wzmacniają w nich poczucie braku bezpieczeństwa. Dla operatora operacji hybrydowej domena informacyjna wiąże się też ze stosunkowo niskim ryzykiem wykrycia, toleruje pewien margines błędu, a nawet może być przedmiotem *outsourcingu*[31]. Co więcej, dzięki wykorzystaniu sztucznej inteligencji paradoksalnie znacznie łatwiej będzie destabilizować społeczeństwa, niż je kontrolować.

31 Tamże, s. 32.

Wydaje się zatem, że istota i specyfika AI dobrze komponują się z uwarunkowaniami prowadzenia działań hybrydowych w domenie informacyjnej. Potwierdzeniem takiego założenia mogą być niektóre narzędzia oddziaływania zdefiniowane w modelu koncepcyjnym Hybrid CoE[32]. W ocenie autora spośród około czterdziestu przykładów na uwagę zasługują następujące: a) wykorzystywanie rozłamów społeczno-kulturalnych; b) promowanie niepokojów społecznych; c) tworzenie sprzecznych narracji; d) dyskredytowanie przywódców, członków rządów czy potencjalnych kandydatów; e) wspieranie określonych grup politycznych; f) oddziaływanie na media; g) kampanie dezinformacyjne i propagandowe.

Wymienione powyżej narzędzia korespondują z prognozami wskazującymi na ryzyko wykorzystania AI w wojnie hybrydowej. W zbiorczej publikacji ekspertów m.in. z Future of Humanity Institute, Centre for a New American Security czy Centre for the Study of Existential Risk wymieniono potencjalne zagrożenia związane z niewłaściwym wykorzystaniem sztucznej inteligencji. Niektóre z nich wpisują się w charakterystykę działań hybrydowych, np.: a) doniesienia typu *fake news* wzmocnione realistycznie sfabrykowanym przekazem w formacie audio i wideo (*deep fake*); b) zautomatyzowane, wysoce spersonalizowane kampanie dezinformacyjne; c) zautomatyzowane kampanie oddziaływania; d) prowadzenie ataków typu *denial of information* polegających na zalaniu kanałów informacyjnych fałszywą lub rozpraszającą uwagę zawartością i utrudnianiu w ten sposób pozyskiwania informacji prawdziwej; e) manipulowanie dostępnością informacji[33].

W podobny sposób zagrożenia generowane przez sztuczną inteligencję definiuje Katarina Kertysova, współpracowniczka The Hague Centre for Strategic Studies. Zwraca uwagę, że proliferacja tej technologii wśród państw autorytarnych zwiększa długoterminowe ryzyko dla wartości demokratycznych, zwłaszcza że znajduje ona zastosowanie nie tylko w wymiarze politycznym, ale także w manipulowaniu informacją. Jako zagrożenia tego rodzaju autorka wymienia: a) profilowanie użytkowników i ich

32 Tamże, s. 33–35.

33 M. Brundage i in., *The malicious use of artificial intelligence: forecasting, prevention, and mitigation*, Future of Humanity Institute i in., 2018, s. 29: <https://arxiv.org/abs/1802.07228> [dostęp: 23 VII 2020].

segmentację; b) zhiperpersonalizowane targetowanie; c) *deep fakes*; d) społeczną świadomość utraty kontroli nad systemami AI – zjawisko *out of the loop*[34]. Na zagrożenie, jakie sztuczna inteligencja może stanowić dla domeny informacyjnej, wskazują też eksperci z Centre for New American Security. Stwierdzają oni słusznie, że technologia ta dostarcza mechanizmy tworzenia bardzo dobrze przygotowanej propagandy dopasowanej do odbiorcy będącego obiektem oddziaływania. Jednocześnie ułatwia jej rozpowszechnianie, zwiększając zasięg i efektywność w dużej skali. Z drugiej strony dzięki AI pojawią się także nowe instrumenty obrony przed tego typu narzędziami działań hybrydowych[35].

Na potrzeby niniejszego artykułu dokonano próby usystematyzowania zastosowań sztucznej inteligencji do intensyfikacji zagrożeń hybrydowych w domenie informacyjnej. Punktem wyjścia dla dalszej analizy jest założenie o dążeniu do kontrolowania i modelowania narracji w przestrzeni publicznej przez operatora działań hybrydowych. Zakotwiczenie się w tym obszarze otwiera przed nim nowe możliwości oddziaływania na proces wyborczy oraz kształt dyskusji politycznej, a następnie na świadomość opinii publicznej. Zgodnie z zaproponowaną przez Hybrid CoE klasyfikacją należy zakładać, że zasadniczo technologia ta może być stosowana w dwóch fazach wdrażanych działań hybrydowych, tj. w przygotowawczej (*priming*) oraz destabilizacyjnej (*destabilization*). W obu przypadkach AI może mieć przypisane określone funkcje i realizować konkretne zadania obliczone na zaburzenie wewnętrznej stabilności obiektu ataku, a przez to zwiększenie nad nim kontroli przez operatora. Analizując dostępną literaturę, zdecydowano się podzielić potencjalne zastosowania sztucznej inteligencji w oparciu o kryterium przedmiotowe. Zgodnie z nim wyodrębniono trzy funkcje, jakie może ona pełnić w działaniach hybrydowych w domenie informacyjnej: analityczno-decyzyjną, narracyjną oraz aksjologiczną. Nie są one sztywno przypisane do konkretnej fazy, co oznacza, że mogą być wykonywane na każdej z nich. Niemniej zarówno w fazie przygotowawczej,

34 K. Kertysova, *Artificial intelligence and disinformation: how AI changes the way disinformation is produced, disseminated, and can be countered*, „Security and Human Rights" 2018, vol. 29, s. 64.

35 M. Horowitz i in., *Artificial intelligence and international security*, Washington 2018, s. 4–5: <https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security> [dostęp: 6 I 2022].

jak i destabilizacyjnej ich efektywność będzie najwyższa. Ponadto w przypadku każdej z funkcji możliwe będzie kontrolowanie i modelowanie narracji w przestrzeni publicznej poprzez narzędzia realizacji działań hybrydowych wykorzystujące AI.

Funkcja analityczno-decyzyjna

Wykorzystanie AI w przypadku tej funkcji wydaje się najbardziej skuteczne w fazie przygotowawczej. Ze zgromadzonych danych wynika, że sztuczna inteligencja może być zastosowana w: 1. analizie sytuacji wewnątrzpolitycznej danego państwa, m.in. do prognozowania wyników wyborów w czasie rzeczywistym i identyfikacji potencjalnych obszarów zapalnych; 2. identyfikacji preferencji wyborczych mieszkańców danego państwa i wygenerowaniu strategii mikrotargetingu; 3. wspieraniu procesów decyzyjnych w samym ośrodku decyzyjnym operatora poprzez analizę *big data*.

Wszechstronna analiza sytuacji wewnątrzpolitycznej danego państwa stała się bardziej dostępna i efektywna dzięki sieciom społecznościowym, w których użytkownicy afiszują się ze swoimi poglądami politycznymi, przekonaniami światopoglądowymi czy komentarzami na temat otaczającej ich rzeczywistości. Odpowiednio skonstruowane oprogramowanie wspierane przez algorytmy sztucznej inteligencji może zostać wykorzystane do oceny sytuacji w danym państwie, a także stanowić narzędzie do weryfikowania informacji pozyskiwanych innymi kanałami. Swoistym studium przypadku co do możliwości oferowanych przez AI stała się sytuacja zaobserwowana w Argentynie. Zauważono, że technologia ta ma zdolność przewidywania nie tylko preferencji wyborczych, ale także samych wyników wyborów. Zhenkun Zhou i Hernan Makse w opracowaniu poświęconym roli AI w prognozowaniu wyników wyborów prezydenckich w Argentynie w 2019 r. wykorzystali algorytmy oparte na uczeniu maszynowym, analizie *big data* i teorii sieci celem przeanalizowania milionów wiadomości umieszczonych w sieci społecznościowej Twitter. Na tej podstawie przygotowali prognozy dotyczące wyników głosowania, dostrzegając przy tym, że przeanalizowane dane w dużym stopniu pokrywają się z rezultatem obowiązkowych prawyborów. Nie tylko otrzymali więc prognozy bardzo zbliżone do rzeczywistych wyników, ale też doszli do wniosku, że na podstawie zgromadzonych i przenalizowanych danych możliwe jest określenie trendów światopoglądowych w społeczeństwie, np. odnośnie do postawy wobec

zmian klimatycznych, politycznych czy edukacyjnych[36]. Zhou i Makse zgromadzili 45 milionów wpisów z Twittera wygenerowanych od marca do października 2019 r. przez 2 miliony użytkowników platformy. Odrzucili następnie wpisy zamieszczone przez automaty, czyli tzw. boty, a pozostałe podzielili według *hashtagów*. W kolejnym kroku AI dokonała błyskawicznych obliczeń, przyporządkowując dzięki *hashtagom* wpisy umieszczone w Twitterze do grup zwolenników trzech najważniejszych kandydatów[37]. Okazało się, że na tej podstawie możliwe było przede wszystkim określenie wyników prawyborów w czasie rzeczywistym. Była to o tyle istotna informacja, że według przeanalizowanych i zinterpretowanych przez AI danych prawyborcze zwycięstwo kandydata opozycji Alberto Fernándeza nad urzędującym prezydentem Mauricio Macrim miało się dokonać różnicą około 16 proc. głosów, czego nie wychwycił żaden z tradycyjnych ośrodków badania opinii publicznej. Co więcej, te ostatnie wskazywały raczej nieznaczne zwycięstwo drugiego z nich[38]. Przedstawione w badaniu prognozy co do rezultatu właściwych wyborów nie były już tak dokładne, jeśli chodzi o różnicę w wynikach obu kandydatów – AI oceniła ją na 16 proc., podczas gdy w rzeczywistości wyniosła prawie 8 proc. Niemniej w granicach błędu określono poparcie dla Fernándeza na poziomie 47,5 proc., ostatecznie wygrał bowiem z wynikiem 48,24 proc.[39] Rezultaty badania przeprowadzonego przez Zhou i Maksego dowodzą, jak duży potencjał tkwi w wykorzystaniu sztucznej inteligencji do analizy sytuacji politycznej danego państwa, i stanowią interesujący punkt wyjścia do pogłębionych badań w tym obszarze. Nie można

36 Z. Zhou, H. Makse, *Artificial intelligence for elections: the case of 2019 Argentina primary and presidential election*, 24 X 2019 [dostęp: 7 VI 2020]: <https://arxiv.org/pdf/1910.11227.pdf>.

37 Tamże.

38 Tamże.

39 Tradycyjne ośrodki badania opinii publicznej i tym razem podawały znacznie bardziej nietrafione dane. Zwycięstwo Fernándeza przy pesymistycznym scenariuszu miało zostać odniesione głosami 50,3 proc. wyborców, a inne ośrodki wskazywały na jeszcze wyższe poparcie. Z kolei różnica pomiędzy oboma kandydatami została oszacowana w przedziale 18–21 proc. Por. M. Lammertyn, *Argentine Peronist challenger seen winning presidency outright on October 27: polls*, „Reuters" [online], 18 X 2019 [dostęp: 7 VI 2020]: <https://www.reuters.com/article/us-argentina-election-polls/argentine-peronist-challenger-seen-winning-presidency-outright-on-october-27-polls-idUSKBN1WX23K>.

bowiem zapominać, że sztuczna inteligencja jest technologią wciąż udoskonalaną, a jej możliwości będą stale rosnąć. Jednocześnie w kontekście zagrożeń hybrydowych zastosowanie AI do przewidywania na przykład wyników wyborów otwiera przed operatorem realizowanych działań dodatkowe możliwości. Znając prawdopodobny rezultat głosowania, ma on możliwość ingerencji w jego przebieg poprzez nasilenie operacji inspirujących i dezinformujących na ostatnim etapie kampanii wyborczej, a także uzyskuje szansę wyprzedzającego przygotowania swojej polityki oraz komunikacji do nadchodzących zdarzeń.

AI zdaje się także znajdować zastosowanie w identyfikowaniu preferencji wyborczych i generowaniu indywidualnie dopasowanych strategii mikrotargetingu, co także może zostać wykorzystane w działaniach hybrydowych. Aleksandra Przegalińska definiuje mikrotargeting jako celowe spersonalizowane komunikowanie określonych treści do wytypowanej osoby w internecie celem wpływania na postawy wyborców czy konsumentów[40]. Jednym z pionierów mikrotargetingu był polski psycholog Michał Kosiński, który opracował metodę tworzenia profili psychologicznych użytkowników na podstawie ich zachowania w sieci społecznościowej Facebook[41]. W rezultacie badań okazało się, że na podstawie aktywności behawioralnej użytkownika z dużym prawdopodobieństwem można określić jego kolor skóry, płeć, preferencje seksualne, a nawet wykonywany zawód. O ile intencją Kosińskiego było wykorzystanie tego narzędzia w celach naukowych, o tyle znalazły one szereg innych zastosowań, zwłaszcza w biznesie. Z rezultatów jego badań skorzystała m.in. firma Cambridge Analytica, stanowiąca swoiste studium przypadku ingerencji w proces wyborczy. W 2014 r., wykorzystując platformę Facebook, pozyskała dane od ok. 250 tys. użytkowników, którzy wyrazili zgodę na analizę nie tylko prywatnych danych umieszczonych w ich profilach, ale również danych ich znajomych. W rezultacie pozyskano dane od ok. 87 mln użytkowników, z których 70,6 mln mieszkało w Stanach Zjednoczonych. Właśnie

40 P. Oksanowicz, A. Przegalińska, *Sztuczna inteligencja. Nieludzka, arcyludzka*, Wydawnictwo „Znak", Kraków 2020, s. 267.

41 M. Kosiński, D. Stillwell, T. Graepel, *Private traits and attributes are predictable from digital records of human behaviour*, „Proceedings of the National Academy of Sciences of the United States of America" 2013, vol. 110, No. 15, s. 5802–5805: <https://doi.org/10.1073/pnas.1218772110> [dostęp: 26 V 2020].

część z tych danych[42] miała trafić do firmy Cambridge Analytica, która wykorzystała je do profilowania preferencji wyborczych poprzez badanie i segregowanie danych, do tworzenia na ich podstawie grup odbiorców i wyszukiwania występujących pomiędzy nimi połączeń, które bazują na potencjalnych zainteresowaniach lub sympatiach politycznych. Obiekt oddziaływania znajdował się zatem pod wpływem określonego sposobu dostarczania informacji (w tym przypadku spersonalizowanych postów), przy czym jego preferencje były przedmiotem uprzedniej kompleksowej analizy zgromadzonych danych. Ujawnienie wykorzystywania tej metody w walce wyborczej stało się przyczyną licznych kontrowersji, choć jak zauważa Louise Amoore, spektrum codziennych zastosowań tych algorytmów jest bardzo szerokie, mimo że są one kojarzone głównie ze zwycięstwami zwolenników wyboru Donalda Trumpa na prezydenta USA czy wynikiem referendum o wyjściu Wielkiej Brytanii z Unii Europejskiej. Sprawczość narzędzi Cambridge Analytica nie byłaby jednak tak duża, gdyby nie zastosowanie sztucznej inteligencji[43]. Dzięki niej możliwe było poszukiwanie optymalnej formy spersonalizowanej wiadomości, m.in. poprzez uprzednie testowanie tysięcy jej wariantów[44]. Przygotowywane treści były tym samym w pełni zoptymalizowane, a w rezultacie bardzo skuteczne. Wydaje się, że profilowanie użytkowników sieci społecznościowych i wykorzystywanie do tego celu algorytmów AI znajduje swoją kontynuację w tym, że m.in. sieć Facebook – w imię dochodu generowanego z tytułu wyświetleń – promuje np. mowę nienawiści. Choć niedawno problem ten ujawniono w mediach, to *de facto* dostrzeżony został już w 2015 r.

42 W zależności od źródeł chodziło o pozyskanie danych od 30 mln do 87 mln użytkowników sieci Facebook. Dostępne są liczne opracowania poświęcone temu studium przypadku, jednak różnią się detalami opisującymi historię Cambridge Analytica. Z uwagi na szczegóły techniczne w niniejszym artykule wykorzystano m.in. tekst: H. Kanakia, G. Shenoy, J. Shah, *Cambridge Analytica – a case study*, „Indian Journal of Science and Technology" 2019, vol. 12 (29): <https://indjst.org/articles/cambridge-analytica-a-case-study> [dostęp: 7 I 2022].

43 K. Woznicki, *The politics of artificial intelligence. An interview with Louise Amoore*, „Eurozine" [online], 23 X 2018 [dostęp: 18 V 2020]: <https://www.eurozine.com/politics-artificial-intelligence/>.

44 J. Delcker, *AI: decoded: how Cambridge Analytica used AI – No, Google didn't call for a ban on face recognition – Restricting AI exports*, „Politico" [online], 28 I 2020, [dostęp: 27 V 2020]: <https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-how-cambridge-analytica-used-ai-no-google-didnt-call-for-a-ban-on-face-recognition-restricting-ai-exports/>.

przez Anata Ben-Davida i Ariadne Matamoros-Fernández. Dokonali oni monitoringu profili użytkowników Facebooka związanych z ekstremalnie prawicowymi partiami politycznymi w Hiszpanii pod kątem mowy nienawiści i ukrytej dyskryminacji. Badania wykazały, że Facebook toleruje rosnącą liczbę ukrytych praktyk dyskryminacyjnych, które nie tylko rozpowszechniają dane i treści, ale także promują mowę nienawiści przy pomocy zwolenników danych partii. Autorzy ci zastrzegli jednak, że tego typu zjawisko nie dotyczy wyłącznie Hiszpanii, ale ma związek ze specyfiką portalu, który zakłada promowanie i dzielenie się zawartością mogącą przynosić koncernowi dochód[45]. Przedstawiony przez nich przykład znalazł swój epilog w 2019 r., gdy do hiszpańskiego parlamentu weszła skrajnie prawicowa partia Vox, zdobywszy 12,9 proc. głosów wobec popularności nieprzekraczającej 0,5 proc. w okresie prowadzonych badań. Może to zatem oznaczać, że specyfika algorytmów Facebooka była jednym z czynników, które dały temu ugrupowaniu sukces wyborczy.

AI może być również nieodzowna we wspieraniu procesów decyzyjnych w samym ośrodku decyzyjnym operatora poprzez analizę *big data*. Może wówczas dostarczyć pakiet przetworzonych informacji, które zawierają teoretycznie obiektywny i pozbawiony zniekształceń obraz, np. sytuacji politycznej[46]. Co więcej, dzięki swoim specyficznym właściwościom może odkrywać zależności i trendy niedostrzegalne przy stosowaniu tradycyjnych narzędzi analitycznych. W ten sposób z punktu widzenia metodologii działań hybrydowych ich operator może uzyskać istotną przewagę już na wstępie.

Funkcja narracyjna

Dzięki algorytmom uczenia maszynowego, zdolnościom do rozpoznawania i przetwarzania języka naturalnego oraz umiejętności przetwarzania olbrzymiej ilości danych sztuczna inteligencja umożliwia podmiotowi

45 A. Ben-David, A. Matamoros-Fernández, *Hate speech and covert discrimination on social media: monitoring the Facebook pages of extreme-right political parties in Spain*, „International Journal of Communication" 2016, vol. 10, s. 1167–1193: <https://ijoc.org/index.php/ijoc/article/download/3697/1585> [dostęp: 7 I 2022].

46 Kwestia dostarczania przez AI obiektywnego obrazu danej sytuacji jest dyskusyjna. W literaturze sporo miejsca poświęca się m.in. problematyce uprzedzeń poznawczych (*bias*), które mogą rzutować na wiarygodność rezultatów tej technologii.

zaangażowanemu w realizację operacji hybrydowej pozyskanie nowych narzędzi, które zwiększają efektywność jego działań. Zdolności te mogą zostać wykorzystane w szczególności do manipulowania zawartością przestrzeni medialnej. Dzięki AI możliwe będzie zarówno wzmocnienie efektu stosowania tradycyjnych narzędzi do modelowania narracji w domenie informacyjnej, jak i używanie zupełnie nowych rozwiązań. Obecnie, z uwagi na dużą popularność problematyki zagrożeń hybrydowych i dynamikę rozwoju technologicznego, w literaturze przedmiotu panuje chaos pojęciowy i nadinterpretowanie określenia *dezinformacja*.

Dezinformacja nie jest zjawiskiem nowym, a jako technika wprowadzania w błąd przeciwnika znana była już w starożytności. Samo pojęcie stworzone zostało w połowie XIX w. w Rosji, a na przestrzeni kolejnych dekad twórczo rozwinięte przez sowieckie organy bezpieczeństwa. Co więcej, w toczonych dyskusjach zupełnie zatarło się jego pierwotne znaczenie. Jak słusznie zauważa Tomasz Kacała, istotą dezinformacji jest przekazanie odbiorcy wiedzy pozornej, bezużytecznej lub wręcz szkodliwej, która następnie skłoni go do podejmowania błędnych decyzji, korzystnych z punktu widzenia podmiotu dezinformującego[47]. Można ją zatem zdefiniować jako świadome działanie lub szereg działań polegających na rozpowszechnianiu pozornie wiarygodnych, lecz nieprawdziwych i trudno weryfikowalnych informacji w celu osiągnięcia określonego skutku politycznego, gospodarczego lub społecznego. Takie rozumienie dezinformacji uległo jednak współcześnie zatarciu na skutek przypisywania jej niemal wszystkich działań tradycyjnych i hybrydowych, które obliczone są na oszukanie drugiej strony lub wywarcie na nią określonego wpływu. Choć dezinformacja stanowi istotny element manipulacji przestrzenią informacyjną, to jednak co do zasady ma charakter defensywny, a do pewnego stopnia także statyczny. Elementami dezinformacji mogą być: 1. bierne wywieranie wpływu na proces decyzyjny w ośrodkach decyzyjnych adwersarza (w swojej istocie jest to najbliższe definicji Kacały); 2. *fake news*, w tym *deep fakes*; 3. kamuflowanie. Działania te zmierzają do celowego i świadomego wprowadzenia w błąd jednostek, grup społecznych i ośrodków decyzyjnych celem osiągnięcia wcześniej określonych zadań. W odróżnieniu od propagandy, której istotą jest konsolidacja środowiska wewnętrznego i ewentualne wprowadzenie w błąd

47 T. Kacała, *Dezinformacja i propaganda w kontekście zagrożeń dla bezpieczeństwa państwa*, „Przegląd Prawa Konstytucyjnego" 2015, nr 2 (24), s. 51.

obserwatora zewnętrznego, dezinformacja skierowana jest do zewnątrz, a jej oddziaływanie ma charakter bierny. Stąd też z punktu widzenia dokładności w analizowaniu narzędzi oddziaływania w domenie informacyjnej celowe wydaje się wyodrębnienie instrumentu wpływu, który należałoby określić terminem *ingerencja*. Ma ona charakter aktywny, ofensywny i dynamiczny, przy czym to w szczególności ona jest beneficjentem postępu technologicznego i oferowanych przez niego rozwiązań (diagram 1).

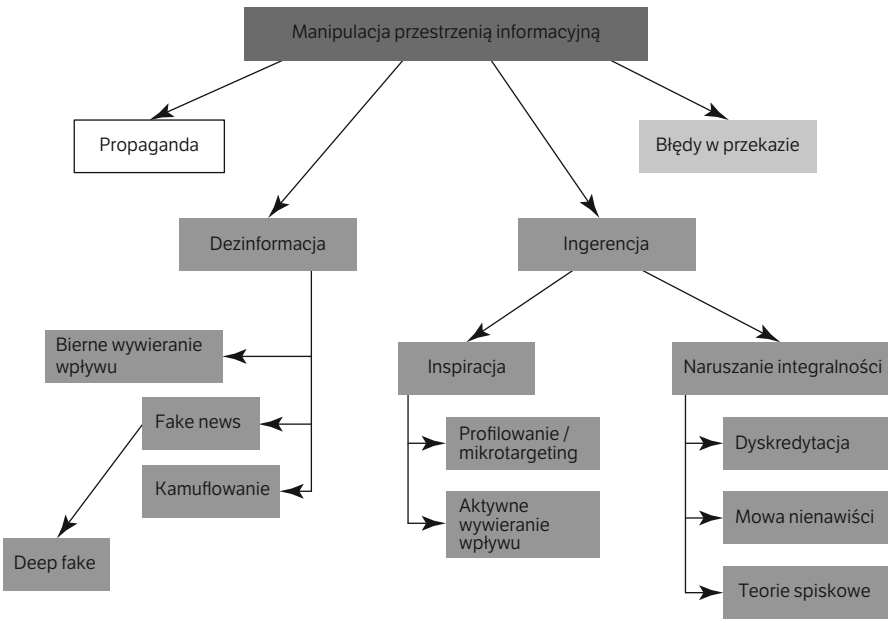


Diagram 1. Rodzaje manipulowania przestrzenią informacyjną (oprac. własne)

Ingerencję można zdefiniować jako aktywne i dynamiczne działania ofensywne obliczone na wywarcie wpływu w społeczeństwie i ośrodku decyzyjnym atakowanego podmiotu, tworzenie w ich przestrzeni rozdźwięków, podziałów, napięć i kontrowersji, a także inspirowanie ich do określonych postaw i aktywności. Współcześnie działania te prowadzone są najczęściej w cyberprzestrzeni, z wykorzystaniem świadomych bądź nieświadomych mieszkańców atakowanego podmiotu, przy stosunkowo niskich kosztach i ryzyku zdemaskowania. W rezultacie oddziaływania m.in. poprzez przestrzeń informacyjną i sieci społecznościowe atakowanemu podmiotowi mogą grozić niepokoje społeczne, a nawet może się on stać obiektem destabilizacji lub dezintegracji wewnętrznej.

Zarówno dezinformacja, jak i ingerencja mogą znacznie zwiększyć swoją siłę oddziaływania, jeżeli zostaną wsparte przez algorytmy sztucznej inteligencji. W tym przypadku technologia może również stworzyć nowe, nieznane wcześniej narzędzia do prowadzenia wojny hybrydowej, czego dowodem jest *deep fake*, tj. zjawisko, które w dyskusji pojawiło się po raz pierwszy dopiero w 2017 r.[48] Algorytmy sztucznej inteligencji są w stanie generować realistyczne efekty dźwiękowe i wizualne, tworzyć realistyczne obrazy wyłącznie na podstawie tekstowych opisów czy przygotowywać informacje. AI może być wykorzystywana m.in. do tworzenia *fake news*, profilowania użytkowników mediów społecznościowych, naruszania integralności przekazu czy aktywnego wywierania wpływu. O tym, jak duży potencjał ma wykorzystanie sztucznej inteligencji do generowania *fake news*, świadczy stworzony w instytucie badawczym Open AI model tej technologii o nazwie GPT-2. Rezultaty jego funkcjonowania były tak dobre, że twórcy w obawie przed konsekwencjami nie zdecydowali się na ujawienie całego kodu źródłowego[49]. Opracowany algorytm po przeanalizowaniu 40 GB zawartości stron internetowych doskonale radził sobie w przygotowywaniu krótkich nieprawdziwych informacji. Był w stanie stworzyć dowolny *fake news*, a także komentować inne posty w sposób nieodróżnialny od ludzi. Powyższy przykład, ale także wspomniane wcześniej profilowanie użytkowników sieci społecznościowych i wywieranie wpływu na ich aktywność poprzez sztucznie generowaną narrację zgodną z interesami operatora, potwierdzają możliwości wykorzystania sztucznej inteligencji do operacji hybrydowych w domenie informacyjnej.

48 Po raz pierwszy technologia ta zastosowana została w pornografii, poprzez wykorzystanie opartego na licencji publicznej oprogramowania do zmiany twarzy. Z czasem za jego pomocą m.in. generowano realistyczne zdjęcia ludzkich twarzy używanych w sieciach społecznościowych do tworzenia fikcyjnych kont czy realistycznych nagrań wideo z udziałem znanych osób. Por. M. Somers, *Deepfakes, explained*, „MIT Sloan School of Management", 21 VII 2020 [dostęp: 8 I 2022]: <https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained>.

49 J. G. Botha, H. Pieterse, *Fake news and deepfakes: a dangerous threat for 21st century information security*, [w:] *Proceedings of the 15th International Conference on Cyber Warfare and Security, Old Dominion University, Norfolk, Virginia, USA, 12–13 March 2020*, ed. B. K. Payne, H. Wu, ACPI, 2020: <http://hdl.handle.net/10204/11946> [dostęp: 12 I 2022].

Funkcja aksjologiczna

Rozwój sztucznej inteligencji i wykorzystywanie jej przez państwo w działaniach hybrydowych może nasilać napięcia pomiędzy państwem i społeczeństwem oraz podważać legitymizację instytucji publicznych, m.in. poprzez eksponowanie różnic aksjologicznych i światopoglądowych. Za jedno z zagrożeń należy uznać nadmierne wykorzystywanie AI przez aparat państwowy, co może prowadzić do utracenia przez niego legitymizacji społecznej. Współcześnie wielu państwom zarzuca się zarówno nadmierny technokratyzm, jak i pogłębiające się oderwanie rządzących elit od rzeczywistości. Pokusa wykorzystania narzędzi i możliwości oferowanych przez sztuczną inteligencję sprawia, że tendencje te mogą ulec tylko wzmocnieniu. Jak zauważają eksperci z Oxford Insights, sztuczną inteligencję opartą na algorytmach głębokiego uczenia charakteryzuje ultratechnokratyzm, który wzbudza w ludziach strach przed kształtowaniem przez nią zasad i mechanizmów codziennego funkcjonowania społeczeństwa. To z kolei może prowadzić do kwestionowania przez społeczeństwo legitymizacji władzy, zwłaszcza jeśli AI otrzyma zbyt dużą samodzielność w podejmowaniu decyzji lub będzie pełnić rolę pośrednika w kontaktach między państwem a obywatelem[50]. W tej sytuacji napięcia na linii władza–społeczeństwo, które wynikają na kanwie spadającej legitymizacji, mogą stać się źródłem zagrożeń zarówno dla bezpieczeństwa danego państwa, jak i w szerszym wymiarze – dla bezpieczeństwa międzynarodowego. W systemie, jakim jest państwo, może zostać wykreowana przestrzeń charakteryzująca się niestabilnością wewnętrzną, podatna na oddziaływanie podmiotu stosującego metody i narzędzia wojny hybrydowej. W przypadku państw demokratycznych zagrożenie to staje się o tyle istotniejsze, że narzędziem społecznej weryfikacji skuteczności i popularności aparatu państwowego jest proces wyborczy.

Innym zagrożeniem związanym z AI i dotyczącym głównie państw demokratycznych będzie eksponowanie przez ich autorytarnych adwersarzy naruszania praw człowieka i swobód obywatelskich. Jest to o tyle istotna kwestia, że oprócz wywoływania kolejnych podziałów na linii

50 W. Pasquarelli, *Finding legitimacy in the age of AI: challenges & opportunities*, „Centre for Public Impact" [online], 18 X 2018 [dostęp: 24 IV 2020]: <https://www.centreforpublicimpact.org/finding-legitimacy-age-ai-challenges-opportunities/>.

państwo–społeczeństwo i oddziaływania na warstwę aksjologiczną fałszywa narracja i podejmowane działania będą utrudniać państwom demokratycznym budowanie odporności i przeciwdziałanie złośliwemu wykorzystaniu tej technologii m.in. poprzez samą AI. Głębokość ingerencji państwa w swobody i prywatność obywateli z wykorzystaniem sztucznej inteligencji pozostaje nieznana. Wydaje się natomiast pewne, że możliwości i narzędzia oferowane przez tę technologię mogą znacznie wykraczać ponad standardy kontroli i monitoringu społeczeństwa, które znamy obecnie. O ile można przewidywać, że państwa autorytarne będą z nich korzystać w pełni, o tyle nie do końca wiadomo, jak zachowają się państwa demokratyczne. W tym kontekście należy odnotować zrealizowane w 2019 r. badania Carnegie Endowment for International Peace, zgodnie z którymi co najmniej 75 ze 176 uwzględnionych państw wykorzystywało sztuczną inteligencję do celów inwigilacyjnych[51]. Zostały one podzielone według ustroju politycznego na: 1. demokracje liberalne; 2. demokracje nieliberalne, ale spełniające kryteria wolności słowa, zgromadzeń czy transparentności procesu wyborczego; 3. państwa autorytarne z systemem wyborczym niezapewniającym standardów demokratycznych; 4. państwa całkowicie autorytarne, nieprzeprowadzające wyborów lub robiące to w warunkach braku konkurencji. Wykazano, że w 2019 r. z inwigilacji opartej na algorytmach AI korzystało 51 proc. państw sklasyfikowanych jako demokracje liberalne (np. Stany Zjednoczone, Wielka Brytania, Holandia, Malta), 41 proc. będących demokracjami nieliberalnymi (np. Singapur, RPA, Meksyk, Izrael), 41 proc. należących do trzeciej grupy (np. Rosja, Egipt, Pakistan, Turcja) oraz 31 proc. państw całkowicie autorytarnych (np. Chiny, Katar, Arabia Saudyjska)[52].

Sztandarowym przykładem wykorzystania AI do celów inwigilacyjnych jest chiński system oceniający zachowania obywateli, znany pod nazwą Bystre Oko. Opierając się na algorytmach AI, przyznaje on punkty społeczne za pozytywne i negatywne zachowania. W rezultacie osoby

51 Do technologii wykorzystywanych w inwigilacji zaliczono łącznie: systemy wykrywania i identyfikacji twarzy, platformy typu inteligentne/bezpieczne miasto oraz systemy wspomagające działania policyjne poprzez gromadzenie danych lokalizacyjnych, rodzajów popełnianych przestępstw, dane biometryczne, sieci społecznościowe itp. Por. S. Feldstein, *The global expansion of AI surveillance*, Carnegie Endowment for International Peace, Washington 2019.

52 Tamże, s. 25–28.

konformistyczne będą nagradzane przez państwo ułatwieniami w codziennym życiu, a sytuacja prezentujących postawy przeciwne będzie wyglądać dokładnie odwrotnie[53]. System ten oparty jest na bazie miliardów zdjęć twarzy, milionach kamer rozmieszczonych w całym kraju i algorytmach AI, które dzięki zgromadzonym danym błyskawicznie określają sposób i prawidłowość zachowania danego obywatela, odnosząc to do uprzednio zdefiniowanych kryteriów. Choć przykład ten wychodzi dalece poza demokratyczne standardy, to próba zastosowania AI do pomocy w zapewnianiu bezpieczeństwa w państwach demokratycznych może wzbudzać kontrowersje i obawy przed powieleniem rozwiązań z Państwa Środka. To z kolei może przenieść spór ideologiczny wokół praw człowieka i swobód obywatelskich w nowy, nieznany dotąd wymiar, stanowiąc dobrą płaszczyznę dla operacji hybrydowych oraz dyskontowania napięć społecznych w danym państwie przez jego rywali na arenie międzynarodowej. Warto jednak zauważyć, że instytucje państwowe dostrzegają napięcia związane z gromadzeniem danych, które potem mogą być wykorzystywane przez algorytmy sztucznej inteligencji. Szczególnie surowe podejście do tej kwestii prezentuje Unia Europejska, która wprowadziła rozporządzenie z 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych, znane w Polsce pod skrótem RODO[54]. Nakłada ono na gromadzenie i przetwarzanie danych na terenie UE szereg ograniczeń. Pod tym względem Wspólnota wyprzedza USA czy Chiny, które wciąż mają trudności z prawidłowym uregulowaniem tej sfery. W szerszej perspektywie pozwala to natomiast postawić hipotezę, że państwa demokratyczne, pomimo oferowanych przez AI narzędzi, mogą być mniej zainteresowane bezprawnym wykorzystywaniem tej przewagi. Taki kierunek może potwierdzać projekt *Rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie*

53 G. Lindenberg, *Ludzkość poprawiona*, Wydawnictwo „Otwarte", Kraków 2020, s. 134–135.

54 Por. *Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych)*, „Dziennik Urzędowy Unii Europejskiej", 4 V 2016, L 119: <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN> [dostęp: 28 VII 2020].

sztucznej inteligencji) zmieniającego niektóre akty ustawodawcze Unii, w którym m.in. zdefiniowano zakazane praktyki w zakresie AI, które niosą niedopuszczalne bądź wysokie ryzyko sprzeczne z wartościami UE. Zaliczono do nich m.in. działania wykorzystujące sztuczną inteligencję do manipulowania ludźmi czy oceny zachowań społecznych[55].

Uwagi końcowe

Specyficzna natura zagrożeń hybrydowych wydaje się właściwym środowiskiem do wykorzystania sztucznej inteligencji. Dotyczy to w szczególności domeny informacyjnej, gdzie możliwości oferowane przez tę technologię wydają się mieć szereg zastosowań pozwalających na relatywnie łatwe dotarcie do społeczeństwa przy jednocześnie niskim ryzyku ujawnienia źródła dezinformacji i ingerencji. Istnieją także przesłanki wskazujące, że intensyfikowanie działań hybrydowych poprzez AI będzie łatwiejsze w przypadku państw autorytarnych, które w złej wierze mogą wykorzystywać fundamenty państw demokratycznych, jakimi są wolne wybory, wolność słowa czy niezależność mediów. Z drugiej strony przewaga technologiczna oraz systematycznie budowana świadomość społeczna powinny pozwolić państwom demokratycznym na tworzenie skutecznych mechanizmów obrony i profilaktyki przed tego typu zagrożeniami, także przy wykorzystaniu sztucznej inteligencji. Warto też mieć na uwadze, że sama technologia, mimo widocznego postępu, wciąż jest niedojrzała, zwłaszcza jeśli spojrzymy na nią przez pryzmat możliwych etapów rozwoju oraz fakt, że nadal wykonuje działania, których kontekstu i znaczenia nie rozumie. Niemniej nawet przy tych ograniczeniach jej dotychczasowe możliwości powodują, że może stać się instrumentem zagrożeń hybrydowych, o ile nie staną jej na przeszkodzie ograniczenia technologiczne, intelektualne lub surowcowe.

55 Por. *Wniosek Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) zmieniające niektóre akty ustawodawcze Unii*, „Eur-Lex" [online, dostęp: 13 XI 2022]: <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:52021PC0206&from=PL>.

Bibliografia

Artificial intelligence and international affairs. Disruption anticipated, ed. M. L. Cummings, H. M. Roff, K. Cukier, J. Parakilas, H. Bryce, The Royal Institute of Foreign Affairs, London 2018.

Baum S., Goertzel B., Goertzel T., *How long until human-level AI? Results from an expert assessment*, „Technological Forecasting & Social Change" 2011, vol. 78, issue 1.

Ben-David A., Matamoros-Fernández A., *Hate speech and covert discrimination on social media: monitoring the Facebook pages of extreme-right political parties in Spain*, „International Journal of Communication" 2016, vol. 10: <https://ijoc.org/index.php/ijoc/article/download/3697/1585> [dostęp: 7 I 2022].

Bostrom N., *Superinteligencja. Scenariusze, strategie, zagrożenia*, Helion, Gliwice 2016.

Botha J. G., Pieterse H., *Fake news and deepfakes: a dangerous threat for 21st century information security*, [w:] *Proceedings of the 15th International Conference on Cyber Warfare and Security, Old Dominion University, Norfolk, Virginia, USA, 12–13 March 2020*, ed. B. K. Payne, H. Wu, ACPI, 2020: <http://hdl.handle.net/10204/11946> [dostęp: 12 I 2022].

Brundage M., Avin Sh., Clark J., Toner H., Eckersley P., Garfinkel B., Dafoe A., Scharre P., Zeitzoff Th., Filar B., Anderson H., Roff H., Allen G. C., Steinhardt J., Flynn C., hÉigeartaigh S. Ó, Beard S., Belfield H., Farquhar S., Lyle C., Crootof R., Evans O., Page M., Bryson J., Yampolskiy R., Amodei D., *The malicious use of artificial intelligence: forecasting, prevention, and mitigation*, Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation, Open AI, 2018: <https://arxiv.org/abs/1802.07228> [dostęp: 23 VII 2020].

Cummings M. L., *Artificial intelligence and the future of warfare*, [w:] *Artificial intelligence and international affairs. Disruption anticipated*, ed. M. L. Cummings, H. M. Roff, K. Cukier, J. Parakilas, H. Bryce, The Royal Institute of Foreign Affairs, London 2018.

Danilin I., *Emerging technologies and their impact on international relations and global security*, „Hoover Institution" [online], 3 X 2018 [dostęp: 21 XI 2022]: <https://www.hoover.org/research/emerging-technologies-and-their-impact-international-relations-and-global-security>.

Delcker J., *AI: decoded: how Cambridge Analytica used AI – No, Google didn't call for a ban on face recognition – Restricting AI exports*, „Politico" [online], 28 I 2020, [dostęp: 27 V 2020]: <https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-how-cambridge-analytica-used-ai-no-google-didnt-call-for-a-ban-on-face-recognition-restricting-ai-exports/>.

A Europe that protects: countering hybrid threats. Factsheet, „European Union External Action" [online], 13 VI 2018 [dostęp: 10 I 2022]: <https://eeas.europa.eu/topics/economic-relations-connectivity-innovation/46393/europe-protects-countering-hybrid-threats_en>.

Feldstein S., *The global expansion of AI surveillance*, Carnegie Endowment for International Peace, Washington 2019.

Fundamental issues of artificial intelligence, ed. V. Muller, Berlin 2014.

Galeotti M., *I'm sorry for creating the „Gerasimov Doctrine"*, „Foreign Policy" [online], 5 III 2018 [dostęp: 13 XI 2022]: <https://foreignpolicy.com/2018/03/05/im-sorry-for-creating-the-gerasimov-doctrine/>.

Giannopoulos G., Smith H., Theocharidou M., *The landscape of hybrid threats: a conceptual model*, European Commission, Luxembourg 2021: <https://www.hybridcoe.fi/wp-content/uploads/2021/02/conceptual_framework-reference-version-shortened-good_cover_-_publication_office.pdf> [dostęp: 11 I 2022].

Herd G., Puhl D., Costigan S., *Emerging security challenges: framing the policy context*, „GCSP Policy Paper" 2013/5: <https://www.gcsp.ch/publications/emerging-security-challenges-framing-policy-content> [dostęp: 21 XI 2022].

Hoffman F., *Conflict in the 21st century: the rise of hybrid wars*, Potomac Institute for Policy Studies, Arlington VA 2007.

Horowitz M., *Artificial intelligence. International competition and balance of power*, „Texas National Security Review" 2018, vol. 1, issue 3.

Horowitz M., Allen G. C., Saravalle E., Cho A., Frederick K., Scharre P., *Artificial intelligence and international security*, Washington 2018: <https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security> [dostęp: 6 I 2022].

Hybrid warfare, ed. M. Weissman, N. Nilsson, B. Palmertz, P. Thunholm, London 2021.

The impact of artificial intelligence on strategic stability and nuclear risk, vol. 1: *Euro-Atlantic perspective*, ed. V. Boulanin, Stockholm International Peace Research Institute, Solna 2019.

Kacała T., *Dezinformacja i propaganda w kontekście zagrożeń dla bezpieczeństwa państwa*, „Przegląd Prawa Konstytucyjnego" 2015, nr 2 (24).

Kanakia H., Shenoy G., Shah J., *Cambridge Analytica – a case study*, „Indian Journal of Science and Technology" 2019, vol. 12 (29): <https://indjst.org/articles/cambridge-analytica-a-case-study> [dostęp: 7 I 2022].

Kaplan J., *Sztuczna inteligencja. Co każdy powinien wiedzieć*, PWN, Warszawa 2019.

Kertysova K., *Artificial intelligence and disinformation: how AI changes the way disinformation is produced, disseminated, and can be countered*, „Security and Human Rights" 2018, vol. 29.

Kosiński M., Stillwell D., Graepel T., *Private traits and attributes are predictable from digital records of human behaviour*, „Proceedings of the National Academy of Sciences of the United States of America" 2013, vol. 110, No. 15: <https://doi.org/10.1073/pnas.1218772110> [dostęp: 26 V 2020].

Kupiecki R., *Dezinformacja w stosunkach międzypaństwowych. Geneza, cele, aktorzy, metody – zarys problemu*, [w:] *Platforma przeciwdziałania dezinformacji – budowanie odporności społecznej*, red. R. Kupiecki, T. Chłoń, F. Bryjka, K. Kozłowski, J. Misiuna, J. Podemska, P. Podemski, Warszawa 2021.

Lammertyn M., *Argentine Peronist challenger seen winning presidency outright on October 27: polls*, „Reuters" [online], 18 X 2019 [dostęp: 7 VI 2020]: <https://www.reuters.com/article/us-argentina-election-polls/argentine-peronist-challenger-seen-winning-presidency-outright-on-october-27-polls-idUSKBN1WX23K>.

Liedel K., *Zagrożenia hybrydowe. Jak zmienia się środowisko bezpieczeństwa RP?*, „Przegląd Bezpieczeństwa Wewnętrznego" 2015, wydanie specjalne: *Wojna hybrydowa*.

Lindenberg G., *Ludzkość poprawiona*, Wydawnictwo „Otwarte", Kraków 2020.

Mitchell T. M., *Machine learning*, McGraw-Hill, Ohio 1997.

Muller V., Bostrom N., *Future progress in artificial intelligence: a survey of expert opinions*, [w:] *Fundamental issues of artificial intelligence*, ed. V. Muller, Berlin 2014.

NATO's response to hybrid threats, „North Atlantic Treaty Organization" [online], 16 III 2021 [dostęp: 10 I 2022]: <https://www.nato.int/cps/en/natohq/topics_156338.htm>.

Nilsson N., Weissmann M., Palmertz B., Thunholm P., Häggström H., *Security challenges in the grey zone. Hybrid threats and hybrid warfare*, [w:] *Hybrid warfare*, ed. M. Weissman, N. Nilsson, B. Palmertz, P. Thunholm, London 2021.

Oksanowicz P., Przegalińska A., *Sztuczna inteligencja. Nieludzka, arcyludzka*, Wydawnictwo „Znak", Kraków 2020.

Pasquarelli W., *Finding legitimacy in the age of AI: challenges & opportunities*, „Centre for Public Impact" [online], 18 X 2018 [dostęp: 24 IV 2020]: <https://www.centreforpublicimpact.org/finding-legitimacy-age-ai-challenges-opportunities/>.

Pietraś Z. J., *Sztuczna inteligencja w politologii. Heurystyczne modelowanie procesów adaptacji politycznej*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 1990.

Platforma przeciwdziałania dezinformacji – budowanie odporności społecznej, red. R. Kupiecki, T. Chłoń, F. Bryjka, K. Kozłowski, J. Misiuna, J. Podemska, P. Podemski, Warszawa 2021.

Proceedings of the 15th International Conference on Cyber Warfare and Security, Old Dominion University, Norfolk, Virginia, USA, 12–13 March 2020, ed. B. K. Payne, H. Wu, ACPI, 2020.

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych), „Dziennik Urzędowy Unii Europejskiej", 4 V 2016, L 119: <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN> [dostęp: 28 VII 2020].

Rumer E., *The Primakov (not Gerasimov) doctrine in action*, Carnegie Endowment for International Peace, Washington 2019: <https://carnegieendowment.org/2019/06/05/primakov-not-gerasimov-doctrine-in-action-pub-79254> [dostęp: 9 I 2022].

Sauer F., *Military applications of artificial intelligence: nuclear risk redux*, [w:] *The impact of artificial intelligence on strategic stability and nuclear risk*, vol. 1:

Euro-Atlantic perspective, ed. V. Boulanin, Stockholm International Peace Research Institute, Solna 2019.

Sheppard L., Karlén R., Hunter A. P., Balieiro L., *Artificial intelligence and national security. The importance of the AI ecosystem*, CSIS, Washington 2018.

Somers M., *Deepfakes, explained*, „MIT Sloan School of Management", 21 VII 2020 [dostęp: 8 I 2022]: <https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained>.

Spiegeleire S. de, Maas M., Sweijs T., *Artificial intelligence and the future of defence*, The Hague Centre for Strategic Studies, The Hague 2017.

Tegmark M., *Życie 3.0. Człowiek w erze sztucznej inteligencji*, Prószyński i S-ka, Warszawa 2019.

Wniosek Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) zmieniające niektóre akty ustawodawcze Unii, „Eur-Lex" [online, dostęp: 13 XI 2022]: <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:52021PC0206&from=PL>.

Woznicki K., *The politics of artificial intelligence. An interview with Louise Amoore*, „Eurozine" [online], 23 X 2018 [dostęp: 18 V 2020]: <https://www.eurozine.com/politics-artificial-intelligence/>.

Zhou Z., Makse H., *Artificial intelligence for elections: the case of 2019 Argentina primary and presidential election*, 24 X 2019 [dostęp: 7 VI 2020]: <https://arxiv.org/pdf/1910.11227.pdf>.