DOI: 10.52694/ThPSR.2024.3.14

Monika Szyłkowska

Sztuczna Inteligencja – strategiczne wyzwania i ryzyka w obszarze prewencji kryminalnej – zarys problemu

Artificial Intelligence – strategic challenges and risks in the area of crime prevention – an outline of the problem

Sztuczna inteligencja uznawana jest obecnie za jeden z najważniejszych czynników wpływających na przemiany gospodarcze i społeczne w skali globu, posiadającej w zasadzie nieograniczone możliwości zastosowań. Celem artykułu jest przedstawienie możliwości zastosowania sztucznej inteligencji w teraźniejszości i przyszłości w kontekście wyzwań oraz strategicznych ryzyk ze szczególnym odniesieniem do prewencji kryminalnej.

Słowa kluczowe: bezpieczeństwo, prewencja, strategiczne ryzyko, autonomia decyzyjna

Artificial intelligence is now recognized as one of the most important factors influencing economic and social change on a global scale, with essentially unlimited application possibilities. The purpose of the article is to present the possibilities of applying artificial intelligence in the present and future in the context of challenges and strategic risks with particular reference to crime prevention.

Key words: security, prevention, strategic risk, decision-making autonomy

Unijny akt w sprawie Sztucznej Inteligencji

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13 czerwca 2024 r. *w sprawie ustanowienia zharmonizowanych przepisów*

dotyczących sztucznej inteligencji (…) (akt w sprawie sztucznej inteligencji)1 w założeniu ma na celu ustanowienie jednolitych ram prawnych, „w szczególności w zakresie rozwoju, wprowadzania do obrotu, oddawania do użytku i wykorzystywania systemów sztucznej inteligencji (zwanych dalej „systemami AI") w Unii, zgodnie z wartościami Unii, w celu promowania upowszechniania zorientowanej na człowieka i godnej zaufania sztucznej inteligencji (AI)"[2] przy jednoczesnym zapewnieniu wysokiego poziomu ochrony: zdrowia, bezpieczeństwa, praw podstawowych zapisanych w Karcie praw podstawowych Unii Europejskiej[3]. W dokumencie podkreślono zarówno korzyści, jakie stwarza AI we wszystkich obszarach działalności – m.in. poprzez umożliwianie optymalizacji procesów czy lepszego prognozowania, które docelowo może zapewnić podmiotom kluczową przewagę konkurencyjną, a instytucjom – wzmacniać korzyści np. w obszarze zarządzania infrastrukturą, energetyką, usługami publicznymi, bezpieczeństwa czy wymiaru sprawiedliwości, ale także zagrożenia i potencjalne szkody, które może wyrządzić AI dla interesu publicznego i praw podstawowych chronionych przepisami prawa Unii, w zależności od okoliczności jej konkretnego zastosowania, wykorzystania oraz od poziomu rozwoju technologicznego. Szkody te mogą być zarówno materialne, jak i niematerialne, w tym również takie, jak: fizyczne, psychiczne, społeczne lub ekonomiczne[4].

Do niewłaściwego zastosowania AI zaliczono m.in. narzędzia do praktyk manipulacji, wyzyskiwania i kontroli społecznej (np. mające na celu nakłonienie osób do niepożądanych zachowań lub podejmowania decyzji pod wpływem błędu w sposób ograniczający autonomię i swobodę wyboru; stosujące techniki podprogowe – bodźce dźwiękowe, bodźce obrazowe).

Należy wskazać, że na potrzeby dokumentu przyjęto następujące definicje:

– system AI – „system maszynowy, który został zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu, a także który

[1] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13 czerwca 2024 r. *w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji).*

[2] Ibidem (1)

[3] Op.cit. (1).

[4] Op.cit. (5).

– na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na podstawie otrzymanych danych wejściowych wyniki, takie, jak: predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne;
– dane biometryczne – dane osobowe będące wynikiem specjalnego przetwarzania technicznego, które dotyczą cech fizycznych, fizjologicznych lub behawioralnych osoby fizycznej, takich jak wizerunek twarzy lub dane daktyloskopijne;
– dentyfikacja biometryczna – oznacza zautomatyzowane rozpoznawanie fizycznych, fizjologicznych, behawioralnych lub psychologicznych cech ludzkich w celu ustalenia tożsamości osoby fizycznej przez porównanie danych biometrycznych tej osoby z danymi biometrycznymi osób fizycznych przechowywanymi w bazie danych;
– weryfikacja biometryczna – oznacza zautomatyzowaną weryfikację typu jeden-do-jednego, w tym: uwierzytelnianie, tożsamości osób fizycznych przez porównanie ich danych biometrycznych z wcześniej przekazanymi danymi biometrycznymi”[5].

Wyłączeniem stosowania przepisów przedmiotowego rozporządzenia zostały objęte systemy AI do celów: wojskowych, obronnych lub celów bezpieczeństwa narodowego – niezależnie od tego, jaki podmiot wykonuje te działania – publiczny czy prywatny. Wyjątek od niestosowania przepisów rozporządzenia stanowią systemy wykorzystywane do celów wojskowych, obronnych lub celów bezpieczeństwa narodowego, które zostały tymczasowo lub na stałe wykorzystywane do innych celów, na przykład cywilnych lub humanitarnych, ścigania przestępstw lub bezpieczeństwa publicznego (system taki objęty zostanie zakresem stosowania rozporządzenia).

Pozostałe wyjątki objęte wyłączeniem stosowania przepisów stanowią:
– systemy i modele AI rozwinięte oraz oddane do użytku wyłącznie do celów badań naukowych i rozwojowych,
– systemy wykorzystywane do ścigania przestępstw, ale tylko w zakresie enumeratywnie wskazanych spraw w załączniku do rozporządzenia, w których wykorzystanie jest „bezwzględnie konieczne do realizacji istotnego interesu publicznego, którego waga przeważa nad ryzykiem. Sytuacje te obejmują poszukiwanie określonych ofiar przestępstw,

[5] Artykuł 3 Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13 czerwca 2024 r. *w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji).*

w tym osób zaginionych; zapobieganie niektórym zagrożeniom życia lub bezpieczeństwa fizycznego osób fizycznych lub atakowi terrorystycznemu; oraz lokalizowanie lub identyfikowanie sprawców przestępstw lub podejrzanych o popełnienie przestępstw”[6] wskazanych w załączniku (…),

– „w przypadku, gdy przestępstwa te podlegają w danym państwie członkowskim karze pozbawienia wolności lub środkowi polegającemu na pozbawieniu wolności przez okres, którego górna granica wynosi co najmniej cztery lata, zgodnie z ich definicją w prawie tego państwa członkowskiego,
– - możliwość przeprowadzania przez organy ścigania, organy kontroli granicznej, organy imigracyjne lub organy azylowe kontroli tożsamości w obecności danej osoby zgodnie
– z warunkami określonymi w prawie Unii i prawie krajowym (…) w celu zidentyfikowania osób, które podczas kontroli tożsamości odmawiają identyfikacji lub nie są w stanie podać
– lub dowieść swojej tożsamości – bez konieczności uzyskiwania uprzedniego zezwolenia
– na podstawie rozporządzenia (np. w stosunku do osoby, która nie chce lub – w wyniku wypadku lub z powodu stanu zdrowia – nie jest w stanie ujawnić swojej tożsamości organom ścigania)”[7].

Wskazane w przepisie *odpowiednie zezwolenie* dotyczy organu wymiaru sprawiedliwości lub niezależny organ administracyjny państwa członkowskiego, którego decyzje są wiążące. Zezwolenie powinno być uzyskane przed wykorzystaniem systemu AI, jednak zakłada się możliwość stosowania wyjątków w sytuacjach nadzwyczajnych z ograniczeniem do bezwzględnie niezbędnego minimum wraz z krajową procedurą warunków. Niezależnie od wyjątkowych zastosowań bez uprzedniego zezwolenia zastosowania systemu – organ ścigania powinien wystąpić o zezwolenie, podając powody, dla których nie był w stanie wystąpić o nie wcześniej (nie później niż w ciągu 24 godzin). W przypadku odmowy udzielenia zezwolenia wykorzystywanie systemów identyfikacji biometrycznej w czasie rzeczywistym powinno zostać wstrzymane ze skutkiem natychmiastowym,

[6] Ibidem, (33)

[7] Artykuł 3 Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13 czerwca 2024 r. *w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji).*

a wszystkie dane (wejściowe oraz uzyskane wyniki) związane z takim wykorzystaniem powinny zostać odrzucone i usunięte. Podkreślenia wymaga zapis punktu (35) wskazujący, iż: „W każdym przypadku żadnej decyzji wywołującej niepożądane skutki prawne dla osoby nie należy podejmować wyłącznie na podstawie wyników uzyskanych z systemu zdalnej identyfikacji biometrycznej”[8].

Wykaz przestępstw, o których mowa w rozporządzeniu, stanowi Załącznik II wskazujący *numerus clausus*, tj.:

– terroryzm,
– handel ludźmi,
– wykorzystywanie seksualne dzieci i pornografia dziecięca,
– nielegalny obrót środkami odurzającymi lub substancjami psychotropowymi,
– nielegalny handel bronią, amunicją lub materiałami wybuchowymi,
– zabójstwo, ciężkie uszkodzenie ciała,
– nielegalny obrót organami lub tkankami ludzkimi,
– nielegalny handel materiałami jądrowymi lub promieniotwórczymi,
– uprowadzenie, bezprawne przetrzymywanie lub wzięcie zakładników,
– przestępstwa podlegające jurysdykcji Międzynarodowego Trybunału Karnego,
– bezprawne zawładnięcie statkiem powietrznym lub statkiem,
– zgwałcenie,
– przestępstwo przeciw środowisku,
– rozbój w formie zorganizowanej lub rozbój przy użyciu broni,
– sabotaż,
– udział w organizacji przestępczej uczestniczącej w co najmniej jednym z wyżej wymienionych przestępstw”[9].

AI – przykłady zastosowań

Tytułem wprowadzenia w obszar istniejących przykładów zastosowań sztucznej inteligencji warto ponownie odwołać się do unijnego rozporządzenia, w kontekście ryzyka niepożądanego wpływu na prawa podstawowe

[8] Op.cit.

[9] Załącznik II do Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13 czerwca 2024 r. *w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji).*

zagwarantowane w Karcie oraz procesowych praw podstawowych, takich, jak: prawo do skutecznego środka prawnego i dostępu do bezstronnego sądu czy prawo do obrony i domniemania niewinności. W szczególności, jeżeli system AI nie spełnia odpowiednich wymogów pod względem skuteczności jego działania, dokładności lub solidności, lub nie został odpowiednio zaprojektowany i przetestowany przed wprowadzeniem do obrotu lub oddaniem do użytku (…). Techniczne „wady" systemów AI mogą z kolei prowadzić do nieobiektywnych wyników i wywoływać skutki w postaci dyskryminacji. Ponadto „bezpośredniość oddziaływania i ograniczone możliwości późniejszej kontroli lub korekty wykorzystania takich systemów działających w czasie rzeczywistym niosą ze sobą zwiększone ryzyko dla praw i wolności osób zainteresowanych w związku z działaniami organów ścigania lub na które działania te miały wpływ"[10]. Interesujący *kazus* w obszarze, chociaż nie opierający się na danych biometrycznych, a raczej na „poszlakach" przedstawia stosowane w USA narzędzie systemu AI w działaniach organów ścigania „ShotSpotter"[11]- inteligentny system wykrywania i lokalizacji wystrzałów.

System wykrywania i lokalizacji strzałów

Technologia oparta o czujniki, algorytmy i sztuczną inteligencję ma na celu wspierać służby w szybkiej reakcji na wystrzał z broni palnej. Zasada działania opiera się na sieci czujników dźwiękowych i mikrofonach rozmieszczonych w ramach infrastruktury miejskiej, identyfikująca rodzaj wystrzału oraz jego lokalizację a następnie przekazująca bezpośrednio informację do organów ścigania. Obecnie narzędzie jest używane w ponad 100 miastach USA. W celu eliminacji fałszywych alarmów producent posiada dział analityczny tzw. Centrum Weryfikacji Zdarzeń, w którym pracownicy sprawdzają przychodzące sygnały zanim zostaną przekazane policji. Krytycy rozwiązania wskazują na możliwość wpływu na program przez pracowników firmy, np. poprzez przypisanie dźwięku

[10] (32) Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13 czerwca 2024 r. *w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji*

[11] ShotSpotter Gunshot Location System® (GLS). Strona producenta SoundThinking: https://www.soundthinking.com/law-enforcement/leading-gunshot-detection-system/

zaklasyfikowanego przez program jako fałszywy alarm na wystrzał[12] oraz brak jednoznacznych badań wskazujących na skuteczność tego rodzaju rozwiązań. Jest to istotne w kontekście potencjalnych aresztowań tylko na podstawie wskazania programu, co m.in. miało miejsce w roku 2020 r. Sprawa Michaela Williamsa odbiła się szerokim echem w mediach po tym, jak został aresztowany na podstawie pliku dźwiękowego ShotSpotter i oskarżony o zabicie młodego mężczyzny z sąsiedztwa, który poprosił go o podwiezienie, po czym zginął od strzału napastnika znajdującego się w innym samochodzie. Prokuratorzy stwierdzili, że technologia oparta na tajnym algorytmie, który analizował odgłosy wykryte przez czujniki, wskazała, że Williams zastrzelił mężczyznę. Aresztowany spędził w więzieniu 11 miesięcy, zanim sędzia oddalił sprawę na wniosek prokuratorów, którzy stwierdzili, że nie mają wystarczających dowodów[13]. Dochodzenie Associated Press[14], oparte na przeglądzie tysięcy wewnętrznych dokumentów, e-maili, prezentacji i poufnych umów, wraz z wywiadami z dziesiątkami obrońców publicznych w społecznościach, w których ShotSpotter został wdrożony, zidentyfikowało szereg poważnych błędów w używaniu systemu jako wsparcia dowodowego dla prokuratorów"[15]. Jak podkreślają autorzy – Garance Burke, Martha Mendoza, Juliet Linderman i Michael Tarm – doświadczenie Williamsa podkreśla rzeczywisty wpływ rosnącej zależności społeczeństwa od algorytmów, które pomagają podejmować decyzje dotyczące wielu aspektów życia publicznego. Nigdzie nie jest to bardziej widoczne, niż w organach ścigania, które zwróciły się do firm technologicznych, takich, jak firma SoundThinking (producent ShotSpotter), w celu zwalczania przestępczości. Dowody ShotSpotter są coraz częściej dopuszczane w sprawach sądowych w całym kraju, których obecnie jest około 200[16]. Autorzy podkreślają, że stosowane przez firmę metody identyfikacji wystrzałów nie zawsze opierają się wyłącznie na technologii – pracownicy producenta mogą

[12] Zeznanie pracownika firmy – Paula Green`a podczas przesłuchania w sprawie strzelaniny w 2016 r. w Rochester w stanie Nowy Jork oraz w sprawie o morderstwo w roku 2017.

[13] Jak podano w publikacji AP:" ShotSpotter utrzymuje, że ostrzegł prokuratorów, aby nie polegali na tej technologii w wykrywaniu strzałów oddanych wewnątrz pojazdów lub budynków. Firma podała, że zastrzeżenie można znaleźć drobnym drukiem w umowie z policją w Chicago.

[14] G. Burke, M. Mendoza, J. Linderman, M. Tar, *How AI-powered tech landed man in jail with scant evidence,* https://apnews.com/article/artificial-intelligence-algorithm-technology-police-crime-7e3345485aa668c97606d4b54f9b6220 (dostęp 20.09.2024)

[15] Ibidem

[16] Jak podaje producent rośnie wykorzystanie nagrań z ShotSpotter`a jako dowodów kryminalistycznych na salach sądowych – od 2010 r. około 200 razy w 20 stanach, z czego 91 przypadków w ciągu ostatnich trzech lat.

i często zmieniają źródło dźwięków odbieranych przez czujniki po odsłuchaniu nagrań audio, wprowadzając możliwość ludzkiej stronniczości do algorytmu wykrywania wystrzałów. Pracownicy mogą i modyfikują lokalizację lub liczbę strzałów – ponadto w przeszłości dyspozytorzy miejscy lub sami policjanci również mogli wprowadzać niektóre z tych zmian. Wśród ogólnokrajowej debaty, obrońcy prywatności i praw obywatelskich twierdzą, że system ShotSpotter i inne technologie oparte na algorytmach stosowane do ustalania wszystkiego – od wyroków więzienia po zasady zawieszenia, nie mają przejrzystości i nadzoru oraz pokazują, dlaczego system wymiaru sprawiedliwości w sprawach karnych nie powinien zlecać niektórych z najważniejszych decyzji kodowi komputerowemu. Zapytany o potencjalne błędy algorytmu firmy, dyrektor generalny producenta – Ralph Clark, odmówił omówienia szczegółów dotyczących wykorzystania sztucznej inteligencji, mówiąc, że „nie jest to naprawdę istotne”[17]. Z kolei szefowie policji nazywają ShotSpotter przełomem, który pomaga w szybszym dotarciu na miejsca przestępstw.

Podkreślenia wymaga fakt, że powstały już badania dotyczące wpływu ShotSpottera w społecznościach, w których jest on używany. W jednym z opublikowanych wyników badań w kwietniu w czasopiśmie Journal of Urban Health przeanalizowano wpływ wdrożenia technologii ShotSpotter na przemoc z użyciem broni palnej i aresztowania przez organy ścigania, w 68. dużych hrabstwach metropolitalnych w latach 1999–2016. W wyniku analizy okazało się, że technologia ta nie zmniejszyła przemocy z użyciem broni palnej ani nie zwiększyła bezpieczeństwa społeczności[18]. Interesującym jest, że w wyniku przeprowadzonych badań stwierdzono również, że to tzw. przepisy dotyczące pozwolenia na broń (tzw. PTP)[19] wpływają na zabójstwa z użyciem broni palnej w stanach i na poziomie hrabstw. Kontrolując obecność ShotSpotter, stwierdzono, że hrabstwa w stanach z wspomnianymi wyżej przepisami miały o 15,7% mniejszą liczbę zabójstw z użyciem broni palnej w porównaniu z hrabstwami w stanach

[17] G. Burke, M. Mendoza, J. Linderman., M. Tar, *How AI-powered tech landed man in jail with scant evidence,* https://apnews.com/article/artificial-intelligence-algorithm-technology-police-crime-7e3345485aa668c97606d4b54f9b6220 (dostęp 20.09.2024)

[18] M. L Doucette, Ch. Green, J. Necci Dineen, D. Shapiro, K. M Raissian, *Impact of ShotSpotter Technology on Firearm Homicides and Arrests Among Large Metropolitan Counties: a Longitudinal Analysis, 1999-2016*, National Library of Medicine, https://pubmed.ncbi.nlm.nih.gov/33929640,DOI: 10.1007/s11524-021-00515-4 (dostęp 20.09.2024)

[19] Przepisy prawa zezwalającego na zakup broni – tzw. PTP (permit-to-purchase) nakładają na osoby fizyczne obowiązek osobistego ubiegania się o pozwolenie na broń palną przed zakupem, angażując w ten proces lokalne lub stanowe organy ścigania.

bez takich przepisów[20]. Warto również przywołać za autorami, że producent ShotSpotter nie ujawnia dokładnego zasięgu spisu ludności; chociaż pojawiły się wezwania do upublicznienia tych danych w celu naukowej oceny wpływu oraz skuteczności technologii w ograniczaniu przemocy z użyciem broni palnej w ośrodkach miejskich, dlatego przedmiotowe badania uległy ograniczeniu w tym zakresie. Natomiast ograniczeniem stricte technicznym jest jego niezdolność do wykrywania wystrzałów w pomieszczeniach i tym samym nieskuteczny w pomaganiu organom ścigania w przypadku części incydentów strzelanin, które mają miejsce na przykład w domach"[21]. W kontekście samych kosztów wskazano, że opłata konfiguracyjna wynosi 10 000 USD, natomiast opłaty roczne wahają się od 65 000 USD do 90 000 USD. Zwrócono uwagę, że w sytuacji, kiedy inwestycja nie zmniejsza liczby przestępstw lub zgonów związanych z bronią ani nie poprawia wskaźników aresztowań sprawców – wydatki związane z wdrażaniem i utrzymaniem systemów mogą raczej zwiększać koszty, ponieważ nie przynoszą spodziewanych zwrotów z inwestycji wynikających z wdrożenia – w szczególności w zakresie skuteczności w ograniczaniu przemocy. Autorzy AP konstatują, że narzędzia kryminalistyczne, takie jak DNA i dowody balistyczne wykorzystywane przez prokuratorów, były uprzednio szczegółowo badane przez dziesięciolecia, ale w przypadku producenta ShotSpotter nie ma takiej możliwości, ponieważ oprogramowanie jest zastrzeżone i nie udostępnia on algorytmu. Sytuacja powyższa stoi w oczywistej sprzeczności z podstawowymi prawami większości systemów prawnokarnych – w szczególności w kontekście możliwości odniesienia się do dowodów w sprawie – w tym możliwości ich podważenia. Jak wskazała mec. Katie Higgins: „Mamy konstytucyjne prawo do konfrontacji ze wszystkimi świadkami i dowodami przeciwko nam, ale w tym przypadku system ShotSpotter jest oskarżycielem i nie ma sposobu, aby ustalić, czy jest dokładny, monitorowany, skalibrowany lub czy ktoś coś dodał. Najpoważniejszą konsekwencją jest skazanie za przestępstwo, którego nie popełniłeś, używając tego jako dowodu"[22]. Z kolei mec. Tania Brief specjalizująca się w sprawach dotyczących uchylania niesłusznych wyroków skazujących stwierdziła: „Obawy związane z wykorzystywaniem

[20] Por. *AJPH Study Shows That Permit to Purchase Laws Are a Promising Avenue to Reduce Suicides in Young Adults*, https://www.apha.org/news-and-media/news-releases/ajph-news-releases/2024/ajph-august-permit-laws

[21] G. Burke, M. Mendoza, J. Linderman, M. Tar, *How AI-powered tech landed man in jail with scant evidence,* https://apnews.com/article/artificial-intelligence-algorithm-technology-police-crime-7e3345485aa668c97606d4b54f9b6220 (dostęp 22.09.2024)

[22] Ibidem

ShotSpotter jako bezpośredniego dowodu wynikają z faktu, że po prostu nie ma badań ustalających ważność i niezawodność tej technologii. Nic". Z kolei polityka prywatności firmy stanowi, że lokalizacje czujników nie są ujawniane departamentom policji, chociaż są one widoczne dla wszystkich przechodniów np. na latarniach ulicznych... Firma zabezpiecza wewnętrzne dane i zapisy ujawniające wewnętrzne funkcjonowanie systemu, pozostawiając brak możliwości analizy technologii w celu zrozumienia specyfiki jej działania. Na przestrzeni ostatnich pięciu lat część amerykańskich miast zrezygnowała z korzystania z oprogramowania, pozostali nadal je wykorzystują jako wsparcie policji w zakresie predykcyjnych działań policyjnych, które wykorzystują powiadomienia o wystrzałach do „identyfikacji obszarów ryzyka"[23].

Warto na koniec dodać, że 2 lipca 2024 r. firma SoundThinking, Inc., opublikowało „list poparcia"[24] dla technologii ShotSpotter podpisanego przez najważniejszych funkcjonariuszy organów ścigania z Massachusetts oraz dyrektora SoundThinking, który stanowi odpowiedź na list wskazujący na zagrożenia na temat technologii akustycznego wykrywania wystrzałów ShotSpotter z 14 maja 2024 r. od senatorów: Eda Markeya, Elizabeth Warren, Rona Wydena i Ayanny Pressley do Inspektora Generalnego Departamentu Bezpieczeństwa Wewnętrznego Stanów Zjednoczonych[25]: „ShotSpotter jest skuteczny, dokładny i mile widziany w naszych społecznościach, a także ma udowodnioną zdolność do ratowania życia i pomagania nam w skutecznym reagowaniu na przemoc z użyciem broni" (…). Z szacunkiem stwierdzamy, że najlepszymi sędziami skuteczności, uczciwości i wartości ShotSpotter nie są odlegli wybrani urzędnicy w Waszyngtonie, ale raczej ci z nas na miejscu, łącznie z naszymi obywatelami, którzy na co dzień żyją w obliczu zagrożenia przemocy z użyciem broni i którzy słusznie popierają stosowanie technologii, aby pomóc chronić swoje sąsiedztwo"[26].

[23] Op.cit. (dostęp 22.09.2024)

[24] List otwarty, *ShotSpotter Saves Lives: Local MA Chiefs Support Gunshot Detection for Public Safety* https://www.soundthinking.com/wp-content/uploads/2024/06/RCG0234-Regan-Law-Enforce-Letter-HERALD_D.NoCrops.pdf (dostęp 22.09.2024)

[25] Strona producenta oprogramowania: https://www.soundthinking.com/press-releases/ma-police-leaders-pen-letter-supporting-shotspotters-accuracy-effectiveness-and-value/

[26] https://www.soundthinking.com/wp-content/uploads/2024/06/RCG0234-Regan-Law-Enforce-Letter-HERALD_D.NoCrops.pdf (dostęp 22.09.2024)

Program typowania „przyszłych przestępców"

Drugi z przykładów zastosowania systemów AI, ale już w obszarze związanym z identyfikacją przestępców stanowi – stosowany od 2015 r. i wdrożony przez Szeryfa hrabstwa Pasco – kontrowersyjny program polegający na typowaniu rzekomych „przyszłych przestępców", który uznał go za: „normalne działania prewencyjne policji – choć realizowane w dość niecodzienny sposób"[27]. Osoby wytypowane przez algorytm zostają w dalszej kolejności „sprawdzane" przez funkcjonariuszy, którzy – zdaniem krytyków – mają na celu „przyłapanie ich na jakimkolwiek złamaniu prawa – do skutku" (np. mandaty za drobne naruszenia za brak numeru na budynku). Wytoczony przez mieszkańców pozew ujawnił, że program typował „przyszłych przestępców" w oparciu o historię kryminalną danej osoby, historie aresztowań, nieokreślone informacje wywiadowcze i arbitralne decyzje analityków policyjnych. „Wytypowane" osoby otrzymywały dodatkowe „punkty" za każdym razem, gdy byli podejrzani, zostali aresztowani, nawet jeśli zarzuty zostały wycofane. W trzymiesięcznych interwałach czasowych komputer generował wyniki i tworzył wstępną listę przestępców, którą następnie analizowali pracownicy i ustalali, które 100 osób powinno się na niej znaleźć. W podręczniku stanowiącym podstawę działań analitycznych wymieniono cechy, które program uznaje za tzw. „czynniki ryzyka kryminogennego"; są to czynniki, co do których uważa się, że prowadzą do przyszłej przestępczości. Cechy te obejmują np.: bycie ofiarą przestępstwa, nieokreśloną „niską inteligencję", „aspołecznych rodziców" i „brak zaspokojenia potrzeb społeczno-ekonomicznych". W podręczniku wyjaśniono, że celem programu jest identyfikacja młodych ludzi „skazanych na życie przestępcze"… W dalszych etapach, po wytypowaniu, działania stróżów prawa względem zidentyfikowanych osób określone zostały w pozwie jako „stalking i nękanie niewinnych obywateli". Sam mechanizm ich typowania przyrównano do „typowania przestępców po kształcie twarzy". Przez dłuższy czas Biuro Szeryfa broniło programu, twierdząc, że pomógł on w ograniczeniu przestępstw

[27] K. McGrory, N. Bedi, *Targeted*, Tamba Bay Times, https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/intelligence-led-policing/; Walt Buteau , *Pasco's Sheriff Nocco sued again over 'intelligence led policing'*, (dostęp 20.09.2024) https://www.wfla.com/8-on-your-side/pasco-s-sheriff-nocco-sued-again-over-intelligence-led-policing/(dostęp 20.09.2024) D. Sullivan, M. Cohen, *Pasco sheriff discontinues controversial intelligence program, court documents say* https://www.tampabay.com/news/pasco/2023/03/23/pasco-sheriff-discontinues-controversial-intelligence-program-court-documents-say/ (dostęp 20.09.2024)

przeciwko mieniu, jednak „The Times" odkrył, że inne pobliskie jurysdykcje policyjne odnotowały podobny spadek liczby przestępstw przeciwko mieniu bez stosowania tego rodzaju narzędzi. Pozew, który został złożony w marcu 2021 r., ma na celu zakończenie raz na zawsze „dystopijnego programu predykcyjnej policji". Program ponownie znalazł się na pierwszych stronach gazet po tym, jak ujawniono, że biuro wysłało dziwaczne listy do osób objętych programem, aby „pogratulować" im włączenia („Miło nam poinformować, że zostałeś wybrany do programu") informując jednocześnie, że w ramach programu zostaną poddani wzmożonej kontroli policyjnej. Prawnicy powodów stwierdzili, że sprawa dostarczyła „istotnego materiału faktycznego, który świadczy o ciągłej kampanii nękania" wobec powodów i innych osób. Z dokumentu wynika również, że w ciągu sześciu lat biuro szeryfa przeprowadziło ponad 13 000 „częstych kontroli wytypowanych przyszłych przestępców" bez nakazów ani dowodów działalności przestępczej. Według dokumentów sądowych Biuro zaprzestało kontrowersyjnego programu i stosowanych praktyk. Łącznie skierowano przeciwko Szeryfowi hrabstwa Pasco cztery pozwy, w tym część z nich związana jest z wykorzystywaniem programu wobec nieletnich. Jak wskazano w dokumentach – zgodnie z danymi dotyczącymi przestępczości w Pasco od czasu uruchomienia „programu" dziesięć lat temu wynika, że ogólny wskaźnik przestępczości spadł o niecałe 50%, jednak przestępstwa z użyciem przemocy wzrosły o około 16%, a przemoc domowa wzrosła o około 38% w tym okresie…

Halucynacje AI – strategiczne ryzyko i wyzwania

Halucynacjami określa się określone zjawiska generowania przez AI fałszywych lub absurdalnych odpowiedzi. AI jako uczenie maszynowe oparte na modelach generatywnych, które mają na celu tworzenie nowych treści w oparciu o wzorce i relacje – przy niekompletnym, niedokładnym, stronniczym lub niewystarczającym zbiorze/bazie danych nie posiada zwyczajnie innych możliwości, zwiększając tym samym ryzyko tworzenia „fałszywych" treści, czyli takich, które nie istnieją w ich „bazach wiedzy". Jak podkreślają eksperci, modele sztucznej inteligencji mogą również doznawać tzw. efektu halucynacji, gdy zostaną „zmuszone do działania" poza granicami swojej wiedzy – wówczas mogą sfabrykować wiarygodnie brzmiącą treść, aby wypełnić luki w swojej wiedzy lub uczą się na danych,

które są wadliwe, sprzeczne lub intencjonalnie zakłamane[28]. Wyniki niezgodne ze stanem faktycznym lub wprowadzające w błąd mogą również wynikać z nieprawidłowych założeń (wzorców), błędnych danych użytych do szkolenia modelu, złożoność modelu, tzw. inżynierii podpowiedzi (niejednoznaczne lub wiodące podpowiedzi mogą spowodować wygenerowanie nieprawidłowych informacji)[29].

Interesujący eksperyment w tym obszarze przeprowadzili badawcze z Rangaraya Medical College, tworząc chatboxa w ChatGPT, aby wskazał 50 nowych tematów badań medycznych, które mogą być wykonywane przez studentów medycyny: „Zaproponuj 50 nowatorskich tematów badań medycznych, które mogą być przeprowadzone przez studentów medycyny w Indiach. Tematy muszą być wykonalne, interesujące, nowatorskie, etyczne i istotne". Program wskazał 50 propozycji, po czym został poinstruowany, aby napisał rozbudowany protokół badawczy na każdy z nich z odpowiednim wprowadzeniem, celami, metodologią, implikacjami i odniesieniami oraz aby podał cyfrowy identyfikator obiektu (DOI) dla wszystkich odniesień. Wszystkie 178 referencji i ich DOI zostały zweryfikowane niezależnie przez pięciu badaczy poprzez przeszukiwanie Internetu w wyszukiwarkach takich, jak np. Scopus, Google i PubMed. Spośród wskazanych referencji cytowanych przez ChatGPT, 69 z nich nie miało DOI, jednak po ponownej analizie w sieci okazało się, że istnieje 41 z nich, jednak DOI dostarczony przez ChatGPT dla tego odniesienia nie istniał w rzeczywistości lub należał do innego artykułu, w związku z czym autorzy uznali powyższy wynik za efekt częściowej halucynacji AI. Pozostałe 28 publikacji nie pojawiło się w wyszukiwarce Google ani nie miało istniejącego DOI, dlatego wynik został uznany za pełen efekt halucynacji AI[30]. Jak podkreślają autorzy: „halucynacje AI stanowią problem, ponieważ ograniczają zaufanie użytkownika do systemu AI, negatywnie wpływają na podejmowanie decyzji i mogą powodować szereg problemów etycznych i prawnych. Ulepszenie danych wejściowych poprzez włączenie różnorodnych, dokładnych i kontekstowo istotnych zestawów danych

[28] P. Kawecki, prezes ITBoom dla „Rz", https://cyfrowa.rp.pl/technologie/art38764981-niebezpieczne-halucynacje-sztucznej-inteligencji-potrafi-klamac-jak-najeta (dostęp 24.09.2024)

[29] Sembot, *Halucynacje sztucznej inteligencji: czym są, jakie są przykłady i jak się przed nimi chronić?, https://pl.sembot.com/blog/halucynacje-sztucznej-inteligencji-czym-sa-jakie-sa-przyklady-i-jak-sie-przed-nimi--chronic/* (dostęp 20.09.2024).

[30] A. Sai Anirudh , M. Sandeep Varma , V S R K. Manoj Kesapragada, Y. Vinee , D.Tirth , R.Tulasi Siri Duddumpudi, *Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References DOI: 10.7759/cureus.37432* (dostęp 20.09.2024).

wraz z częstymi informacjami zwrotnymi od użytkowników i włączeniem ludzkich recenzentów do oceny wyników generowanych przez system AI to niektóre rozwiązania tego problemu halucynacji AI”[31].

Rozważając kwestie ryzyka i wyzwań, warto odnieść się do założeń Rezolucji Zgromadzenia Ogólnego ONZ numer 78/266 z 21 marca 2024 r. – „w sprawie zasad regulujących wykorzystanie technologii sztucznej inteligencji”, które podkreślają globalne zobowiązanie do wykorzystania sztucznej inteligencji dla korzyści zbiorowych, przy jednoczesnym zapewnieniu przestrzegania praw człowieka oraz „bezpiecznej i godnej zaufania” w domenie niewojskowej. Bezpieczeństwo powinno opierać się na: niezawodności, wytłumaczalności, etyce, inkluzywności, w pełnym poszanowaniu, promowaniu i ochronie praw człowieka i prawa międzynarodowego – chroniąc prywatność[32]. Tym samym niewłaściwe lub złośliwe projektowanie, opracowywanie, wdrażanie i wykorzystywanie systemów sztucznej inteligencji, m.in. bez odpowiednich zabezpieczeń lub w sposób niezgodny z prawem międzynarodowym, stwarza ryzyko, które może utrudnić postęp w kierunku osiągnięcia celów zrównoważonego rozwoju oraz podważyć zrównoważony rozwój w jego trzech wymiarach – gospodarczym, społecznym i środowiskowym; pogłębiać podziały cyfrowe między krajami i wewnątrz nich; prowadzić do dyskryminacji; podważać integralność informacji i dostęp do informacji; podważać ochronę, promowanie i korzystanie z praw człowieka i podstawowych wolności, w tym prawa do niepodlegania bezprawnej lub arbitralnej ingerencji w prywatność; oraz zwiększać potencjalne ryzyko wypadków i potęgować zagrożenia ze strony złośliwych podmiotów[33]. W dokumencie podkreślono, że prawa człowieka i podstawowe wolności muszą być przestrzegane, chronione i promowane przez cały cykl życia systemów sztucznej inteligencji, wzywając wszystkie państwa członkowskie do „powstrzymania się lub zaprzestania korzystania z systemów sztucznej inteligencji, które nie mogą działać zgodnie z międzynarodowym prawem dotyczącym praw człowieka lub które stwarzają nadmierne ryzyko dla korzystania z praw człowieka, zwłaszcza dla osób znajdujących się w trudnej sytuacji, oraz potwierdza, że te same prawa, które przysługują ludziom offline, muszą

[31] Ibidem.
[32] Rezolucja Zgromadzenia Ogólnego ONZ numer 78/266 z 21 marca 2024 r. – w sprawie zasad regulujących wykorzystanie technologii sztucznej inteligencji.
[33] Op.cit.

być również chronione online, w tym przez cały cykl życia systemów sztucznej inteligencji”[34].

Podsumowanie

Wskazane powyżej zarys problematyki związanej z potencjalnym zastosowaniem AI wraz z jej ułomnościami miał na celu ukazać jego złożoność oraz wyzwania, jakie obecnie implikuje. Zgodnie ze standardem Europejskiej Federacji Stowarzyszeń Zarządzania Ryzykiem FERMA*: ryzyko to kombinacja prawdopodobieństwa wystąpienia zdarzeń oraz jego skutków, przy czym obejmuje ono zarówno negatywne, jak i pozytywne skutki zdarzeń*. Ryzyko strategiczne można definiować zarówno jako jeden z kluczowych obszarów ryzyka – a samo „zrozumienie ryzyk”, stanowi w tym wypadku warunek konieczny opracowania właściwej strategii każdego podmiotu, jak i jako działania umożliwiające wdrożenie nowych, ryzykownych rozwiązań, które pozwolą odpowiedzieć na pojawiające się nowe wyzwania – w szczególności w kontekście stosowania i wykorzystywania systemów opartych o AI. W obszarach działania tzw. wysokiego ryzyka, do których niewątpliwie należy szeroko rozumiane bezpieczeństwo publiczne – „niedoskonałość” czy „omylność” systemów i narzędzi AI może stanowić nie tylko problem, ale również implikować konkretne zagrożenia – zarówno dla samych obywateli (predykcja kryminalna vs zasada domniemania niewinności), jak i dla decydentów. Poprawa danych treningowych dla modeli sztucznej inteligencji poprzez wykorzystanie dokładnych, zweryfikowanych zestawów danych, a nie tylko dużej ilości danych może nie mieć znaczenia dla programów w obszarze prewencji, jak i sądownictwa (brak możliwości wpisania zbioru danych „niepewności”). Jak wskazano w przywoływanym uprzednio Rozporządzeniu PEiR[35] systemy AI „przeznaczone do wykorzystania w kontekście ścigania przestępstw, w którym dokładność, wiarygodność i przejrzystość są szczególnie ważne dla uniknięcia niepożądanego wpływu, zachowania zaufania publicznego oraz zapewnienia odpowiedzialności i skutecznego dochodzenia roszczeń,

[34] Op.cit.

[35] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13 czerwca 2024 r. *w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji)* – (59)

należy klasyfikować jako systemy wysokiego ryzyka, o ile ich wykorzystanie jest dozwolone zgodnie z właściwymi przepisami prawa Unii i prawa krajowego".

Wskazane przykłady rozwiązań znajdujących już zastosowanie w obszarze szeroko lub szczególnie rozumianej prewencji powinny zwrócić uwagę na konieczność dochowania wyjątkowej staranności we wprowadzaniu standardów i wymogów dla takich rozwiązań, jeśli będą wykorzystywane przez organy publiczne. Technologie bezpieczeństwa publicznego z pewnością będą wprowadzane w szerszym zakresie w obszarze europejskim, stawiając nowe wyzwania nie tylko przed ustawodawcami, ale przede wszystkim – instytucjami i organami je stosującymi. „Bezpieczeństwo nie ma ceny" w kontekście nakładów finansowych, jednak w obszarze praw podstawowych może powodować dylematy decyzyjne. Biorąc pod uwagę zasady działania technologii opartej o modele uczenia – nawet najdoskonalszy algorytm nie zastąpi czynnika ludzkiego i nie pozwoli na przykład na stosowanie zasady domniemania niewinności.

Wybór pomiędzy wolnością a bezpieczeństwem wydaje się oczywisty, jednak przy braku szczególnej ostrożności w wyborze narzędzi AI – ceną walki z przestępczością mogą stać się prawa i wolności obywatelskie, co uznać można za strategicznie kluczowe wyzwanie w kontekście potencjału zastosowania AI.

Bibliografia

Anirudh A.S. , Sandeep V.M. , V S R Krishna Manoj K., Vineel Y. , Tirth D. , Duddumpudi R.T.S., *Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References, "* Cureus 15(4)", DOI: 10.7759/cureus.37432

Burke G., Mendoza M., Linderman J., Tar M., *How AI-powered tech landed man in jail with scant evidence,* https://apnews.com/article/artificial-intelligence-algorithm--technology-police-crime-7e3345485aa668c97606d4b54f9b6220

Doucette M.L., Green Ch., Necci Dineen J., Shapiro D., Raissian K.M., *Impact of ShotSpotter Technology on Firearm Homicides and Arrests Among Large Metropolitan Counties: a Longitudinal Analysis, 1999-2016*, "National Library of Medicine", https://pubmed.ncbi.nlm.nih.gov/33929640, DOI: 10.1007/s11524-021-00515-4

Makowiec P., *Rezolucja ONZ ws. sztucznej inteligencji. „Te sama prawa online, co offline"* https://cyberdefence24.pl/polityka-i-prawo/rezolucja-onz-ws-sztucznej-inteligencji-te-sama-prawa-online-co-offline

Kawecki P., opinia, Rzeczpospolita, *Niebezpieczne halucynacje sztucznej inteligencji. Potrafi kłamać jak najęta,* https://cyfrowa.rp.pl/technologie/art38764981-niebezpieczne-halucynacje-sztucznej-inteligencji-potrafi-klamac-jak-najeta

McGrory K., Bedi N., *Targeted*, *Pasco's sheriff created a futuristic program to stop crime before it happens.It monitors and harasses families across the county.* Tamba Bay Times,

https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/intelligence-led-policing/;

Buteau W., *Pasco's Sheriff Nocco sued again over 'intelligence led policing'***,**

https://www.wfla.com/8-on-your-side/pascos-sheriff-nocco-sued-again-over-intelligence-led-policing/

Sullivan D., Cohen M., *Pasco sheriff discontinues controversial intelligence program, court documents say;*https://www.tampabay.com/news/pasco/2023/03/23/pasco-sheriff-discontinues-controversial-intelligence-program-court-documents-say/

Regulacje:

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13 czerwca 2024 r. *w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji),* https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=OJ:L_202401689

Rezolucja Zgromadzenia Ogólnego ONZ numer 78/266 z 21 marca 2024 r. – w sprawie zasad regulujących wykorzystanie technologii sztucznej inteligencji.

Pozostałe publikacje:

Publikacja Parlamentu Europejskiego, *Jakie są korzyści z ułatwienia udostępniania danych w UE?* https://www.europarl.europa.eu/topics/pl/article/20220331STO26411/jakie-sa-korzysci-z-ulatwienia-udostepniania-danych-w-ue; *Akt ws. sztucznej inteligencji: pierwsze przepisy regulujące sztuczną inteligencję,* https://www.europarl.europa.eu/topics/pl/article/20230601STO93804/akt-ws-sztucznej-inteligencji-pierwsze-przepisy-regulujace-ai

Publikacja Komisji Europejskiej, *Europejskie podejście do sztucznej inteligencji,* https://digital-strategy.ec.europa.eu/pl/policies/european-approach-artificial-intelligence

Publikacja CyberPolicy NASK, *Rezolucja ONZ w sprawie sztucznej inteligencji,*

https://cyberpolicy.nask.pl/aktualnosci/rezolucja-onz-w-sprawie-sztucznej-inteligencji/

The GIP Digital Watch Observatory team, *UN General Assembly adopts first-ever resolution on AI,* https://dig.watch/updates/un-general-assembly-adopts-first-ever-resolution-on-ai;

The GIP Digital Watch Observatory team*, UN General Assembly resolution on AI – Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development,* https://dig.watch/resource/un-general-assembly-resolution-on-ai

Sembot, *Halucynacje sztucznej inteligencji: czym są, jakie są przykłady i jak się przed nimi chronić?, https://pl.sembot.com/blog/halucynacje-sztucznej-inteligencji-czym-sa-jakie-sa-przyklady-i-jak-sie-przed-nimi-chronic/*

Strona producenta SoundThinking ShotSpotter Gunshot Location System® (GLS):

https://www.soundthinking.com/law-enforcement/leading-gunshot-detection-system/

List otwarty, *ShotSpotter Saves Lives: Local MA Chiefs Support Gunshot Detection for Public Safety* https://www.soundthinking.com/wp-content/uploads/2024/06/RCG0234-Regan-Law-Enforce-Letter-HERALD_D.NoCrops.pdf